Quantum Proof Systems and Entanglement Theory

by

Salman Abolfathe Beikidezfuli

B. S., Sharif University of Technology, 2004

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

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2009

Author .....
Department of Mathematics
May 1, 2009

Certified by .....
Peter W. Shor
Morss Professor of Applied Mathematics
Thesis Supervisor

Accepted by .....
Michel X. Goemans
Chairman, Applied Mathematics Committee

Accepted by .....
David Jerison
Chairman, Department Committee on Graduate Students

Quantum Proof Systems and Entanglement Theory

by

Salman Abolfathe Beikidezfuli

Submitted to the Department of Mathematics
on May 1, 2009, in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY

Abstract

Quantum complexity theory is important from the point of view of not only theory of computation but also quantum information theory. In particular, quantum multi-prover interactive proof systems are defined based on complexity theory notions, while their characterization can be formulated using LOCC operations. On the other hand, the main resource in quantum information theory is entanglement, which can be considered as a monotonic decreasing quantity under LOCC maps. Indeed, any result in quantum proof systems can be translated to entanglement theory, and vice versa. In this thesis I mostly focus on quantum Merlin-Arthur games as a proof system in quantum complexity theory.

I present a new complete problem for the complexity class QMA. I also show that computing both the Holevo capacity and the minimum output entropy of quantum channels are NP-hard. Then I move to the multiple-Merlin-Arthur games and show that assuming some additivity conjecture for entanglement of formation, we can amplify the gap in QMA(2) protocols. Based on the same assumption, I show that the QMA(k)-hierarchy collapses to QMA(2). I also prove that $\mathsf{QMA}_{\log}(2)$, which is defined the same as QMA(2) except that the size of witnesses is logarithmic, with the gap $n^{-(3+\epsilon)}$ contains NP. Finally, motivated by the previous results, I show that the positive partial transpose test gives no bound on the trace distance of a given bipartite state from the set of separable states.

Thesis Supervisor: Peter W. Shor
Title: Morss Professor of Applied Mathematics

Acknowledgments

This thesis would have never been written without help, support, and inspiration of my mother and my advisor. I am grateful to both of them.

My mother learned me how to stand firm against any difficulty. During these years I was always confident that there is someone who understands me and does not leave me alone. I proudly dedicate this thesis to her.

I cannot imagine that I could have an advisor better than Peter Shor. He was always patient of my infantile questions and ideas, and gently guided me for deeper understandings. Our discussions were usually taking no more than fifteen minutes not because he did not have time for me, but because after a few words of advice I had a new idea to think of. I do not remember anytime that I came out of his office without being happy. I wish all the best and pray for him forever.

I wish to thank Mohsen Bahramgiri who introduced me to Peter Shor and the wonderful world of quantum computation. I wish to thank Isaac Chuang for his introductory and at the same time high-level quantum computation course that he (with Peter Shor) taught me. I wish to thank Madhu Sudan for his great course on complexity theory; I started working on this subject after his course. I wish to thank Scott Aaronson for his interesting problem that influenced my research for two years. I wish to thank Edward Farhi for his wonderful weekly group meetings; I learned a lot in those meetings and am sure that I will miss them. I wish to thank Daniel Stroock for his stochastic processes course, serving on my qualifying exam committee, and also for all enjoyable discussions that we had. I wish to thank George Lusztig for his Lie algebra course and also serving on my qualifying exam committee. I wish to thank Albert Meyer who learned me a lot about teaching when I was his TA.

I am also grateful to many professors in my undergraduate who helped me to increase my knowledge of mathematics. I wish to thank Yahya Tabesh who encouraged me to think of mathematics as a major. I wish to thank Saeed Akbari who learned me a lot about algebra in eight courses in my undergraduate. I wish to thank Ebadollah Mahmoodian who was not only a great teacher but also available whenever I needed a consultation. I wish to thank Siamak Yassemi because of his helpful advice on how to change myself from being a math student to a math researcher.

I wish to thank those who have collaborated with me directly on papers, some of which form much of the content of this thesis: Peter Shor, Mohsen Bahramgiri, Scott Aaronson, Bill Fefferman, Andrew Drucker, Isaac Chuang, Bei Zeng, and Markus Grassl.

I like to send my best wishes to all of my friends at both SUT and MIT who learned me a lot and made these years much more joyful: Omid Naghshineh, Kasra Alishahi, Javad Ebrahimi, Reza Rezazadegan, Farzad Fathizadeh, Majid Hadian, Hoda Bidkhori, Babak Modami, and Mohammad-Hadi Hedayetzadeh at the mathematics department of SUT; Amin Aminzadeh, Ali Babai, Ali Shourideh, Amin Jafarian, Ali Sharifi, Mohammad Asadollahi, Amir Zabet, Ali Masoumi, Omid Kokabi and Saeed Moradi at the "6-unit" dormitory of SUT; Mohsen Bahramgiri, Eaman Eftekhari, Mohsen Razavi, Sanaz Sigaroudi, Hamed Mamani, Mojtaba Bateni, Mohammad-Reza Alam, Reza Karimi, Ali Hosseini, Saeed Bagheri, Ali Parandeh-Gheibi, Mohammad Araghchini, Ali Farahanchi, Pouyan Ghaemi, Danial Lashkari, Asadollah Kalantarian, Davoud Ebrahimi, Abolhasan Vaezi, Reza Sharifi, Sarah Paydavosi, Leila Farhadi, Ramis Movasagh, Amrit Sinha, Sherif Akl and Kandarp Bhatt at MIT.

Finally, I am thankful of my grandmother, my brother, and my sister.

Contents

Chapter 1

Introduction

The theory of *quantum computation* was started by Richard Feynman who raised the following seminal point in 1982: *it seems there are essential difficulties in simulating quantum mechanical systems on classical computers*, and suggested that designing computers based on the principals of quantum mechanics would allow us to avoid these difficulties [96]. This idea was followed by David Deutsch in a series of papers [48, 49, 49] in which he proved the existence of a *universal quantum computer*. In 1994, Peter W. Shor gave a strong evidence on the validity of Feynman's observation. He showed that the problem of integer factorization, which is believed to be a hard problem on classical computers, can be solved efficiently on a quantum computer [109, 110]. After Shor's algorithm, there has been an intensive effort to understand the advantages of using quantum computers compared to the classical ones. This attempt was the beginning of the theory of quantum complexity.

Computational complexity theory is the theory of the classification of problems based on resources that are required to solve them. This classification is important because it tells us which problems are hard and which problems are easy to solve. For example, the RSA protocol [104] for public-key cryptography is known to be safe because we believe that integer factorization is a hard problem.

Computational complexity theory is naturally extended in the presence of quantum computers. Shor's factoring algorithm [109, 110], as the first striking result of quantum computation, shows that there are hard problems that can be solved easily on a quantum computer; therefore, quantum resources may define completely different classes of problems. In the past decade, this area has been active to understand these differences [118].

Quantum complexity theory is important from the point of view of not only theory of computation, but also quantum information theory. For instance, Communication complexity with shared entanglement [36, 41, 93], and quantum multi-prover interactive proof systems with shared entanglement [71, 72, 73] are two models of computation which can be characterized using the properties of entanglement and LOCC maps. Conversely, the power of entanglement as a resource in information theory can be formulated in these two models. Therefore, studying quantum complexity theory, and in particular quantum proof systems, will help us to understand quantum information theory.

In this thesis I mostly focus on the complexity class QMA (quantum Merlin-Arthur) which is important in three different points of view. First, QMA is the quantum analogue of NP [119, 6], so understanding this class provides us the quantum version of the deep theory of NP-completeness. Second, the local Hamiltonian, which is a physically motivated problem, is QMA-complete [77, 6], and techniques of analyzing this problem are applied

to recognize the power of adiabatic quantum computation [7]. Third, an extension of this complexity class, called quantum multiple-Merlin-Arthur $\mathsf{QMA}(k)$, is another formulation of the problem of the power of entanglement in complexity theory.

I have organized my results in four chapters and two appendices.

In Chapter 2, after reviewing basic properties of QMA, I introduce the quantum version of the clique problem and show that it is QMA-complete. This problem is basically the problem of estimating the zero-error capacity of quantum channels. Motivated by this problem, I raise the question of computing the Holevo capacity of quantum channels. I prove that this problem is NP-complete. These results are based on joint work with Peter W. Shor [17].

In Chapter 3, I define the complexity class $\mathsf{QMA}(k)$, and show that if the weak additivity conjecture in quantum information theory holds, we can amplify the gap in $\mathsf{QMA}(2)$. Under the same assumption, I also prove that $\mathsf{QMA}(k) = \mathsf{QMA}(2)$, for any $k \geq 2$. To understand the relation between QMA and $\mathsf{QMA}(2)$, I express the problem of the existence of a map that disentangles any quantum state. I prove that a perfect disentangler does not exist. These results are based on joint work with Scott Aaronson, Andrew Drucker, Bill Fefferman and Peter W. Shor [2].

In Chapter 4, I show that $\mathsf{QMA}_{\log}(2)$, which is defined the same as $\mathsf{QMA}(2)$ except that the size of the witnesses sent by Merlins are logarithmic, contains NP. I prove this containment when the gap in $\mathsf{QMA}_{\log}(2)$ is $n^{-(3+\epsilon)}$, for every $\epsilon > 0$. This result is based on [16].

In Chapter 5, I try to answer the problem of the complexity of separability problem with constant gap, which is related to the power of $\mathsf{QMA}_{\log}(2)$. I show that the positive partial transpose test [102, 62] gives no bound on the distance of a bipartite state from separable states. I argue that the same result holds for other well-known separability tests such as reduction criterion [60], majorization criterion [97], and symmetric extension criterion [46, 47]. These results are based on joint work with Peter W. Shor [18].

In Appendix A, I investigate the zero-error capacity of quantum channels in more details. In addition, I present a new upper bound on the zero-error capacity of graphs. These results are not published.

In Appendix B, I state a precise definition for $\mathsf{PostQMA}(k)$, and show some properties of this class. These results are also unpublished.

In the rest of the this chapter, I describe preliminary concepts of quantum physics, theory of computation, complexity theory, and also quantum computation. The reader who is familiar with any of these subjects, can simply skip the relevant section since I use the common notations of these theories.

1.1 Mathematical Framework of Quantum Physics

The *Schrödinger equation* in quantum physics plays the rule of *Newton's laws* in classical physics. The Schrödinger equation describes the behavior of a *quantum system* in time. In a general form, it can be written as

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle, \tag{1.1}$$

where $i = \sqrt{-1}$, and $\hbar$ is *Planck's constant*. $|\psi(t)\rangle$ is called the *wave function* and represents the state of a quantum system at time t. Also, H is a self-adjoint operator, called the

Hamiltonian, and represents the constraints of the system.

In this section, I do not want to express the physical interpretation of this equation, and instead, will describe the terms wave function, and Hamiltonian in a mathematical framework so that quantum physics can be formulated based on a few axioms. A more comprehensive discussion on this part can be found in Nielsen and Chuang's book [96], and in Preskill's lecture notes [103].

1.1.1 Some Notations in Linear Algebra

A *Hilbert space* $\mathcal{H}$ is a vector space over the complex numbers $\mathbb{C}$ equipped with an inner product, which is complete[1] under the induced norm by the inner product. In this thesis, we deal only with the finite dimensional Hilbert spaces, so we can forget about the completeness condition since any finite dimensional inner production space is automatically complete [66].

A vector in $\mathcal{H}$ is represented in the form $|\psi\rangle$. Also its dual with respect to the inner product is shown by $\langle\psi|$. We can think of $|\psi\rangle$ as a column-vector, and $\langle\psi|$ as a row-vector which is the complex conjugate of $|\psi\rangle$ ($|\psi\rangle^\dagger = \langle\psi|$). Therefore, the inner product of two vectors $|\psi\rangle, |\phi\rangle$ is $\langle\psi|\phi\rangle = \langle\phi|\psi\rangle^*$ (x^* is the conjugate of the number x). More generally, for any linear operator $M : \mathcal{H} \to \mathcal{K}$ between two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, we represent its dual or *adjoint*[2] by $M^\dagger : \mathcal{K} \to \mathcal{H}$, so for any $|\phi\rangle \in \mathcal{H}$ and $|\psi\rangle \in \mathcal{K}$ we have $\langle\psi|M|\phi\rangle^* = \langle\phi|M^\dagger|\psi\rangle$. A linear map $M : \mathcal{H} \to \mathcal{H}$ is called *self-adjoint* or *hermitian* if $M^\dagger = M$. It is well-known, and can be proved easily, that all eigenvalues of a hermitian operator are real. A hermitian operator M is called *positive semi-definite* if all of its eigenvalues are non-negative. If M is positive semi-definite we denote it by $M \geq 0$. We also say $M \geq N$ if $M - N$ is positive semi-definite. A linear operator M is called *unitary* if $MM^\dagger = I$, where I is the identity map over $\mathcal{H}$.

For $|\phi\rangle \in \mathcal{H}$ and $|\psi\rangle \in \mathcal{K}$ we can think of $|\psi\rangle\langle\phi|$ as a rank-one linear operator $|\psi\rangle\langle\phi| : \mathcal{H} \to \mathcal{K}$ such that $(|\psi\rangle\langle\phi|)\,|x\rangle = (\langle\phi|x\rangle)\,|\psi\rangle$. Since any linear map can be written as a sum of rank-one linear operators, the set $\{\sum_i |\psi_i\rangle\langle\phi_i|\}$ contains all linear maps from $\mathcal{H}$ to $\mathcal{K}$.

A vector $|\psi\rangle$ is called *normal* if it has unit length, i. e. $\langle\psi|\psi\rangle = 1$. Letting $\dim\mathcal{H} = d$, an *orthonormal basis* $\{|1\rangle, |2\rangle, \ldots |d\rangle\}$ of $\mathcal{H}$ is a basis such that all of its vectors have unit length and they are mutually orthogonal, i. e. $\langle i|j\rangle = \delta_{ij}$, where δ_{ij} is the Kronecker delta function. From the definition it is clear that any unitary operator sends an orthonormal basis to an orthonormal basis.

For two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ we represent their *tensor product* by $\mathcal{H} \otimes \mathcal{K}$. This new Hilbert space is spanned by vectors $|\phi\rangle \otimes |\psi\rangle$, where $|\phi\rangle \in \mathcal{H}$ and $|\psi\rangle \in \mathcal{K}$, and its inner product is defined by $\langle\phi| \otimes \langle\psi| \cdot |\phi'\rangle \otimes |\psi'\rangle = \langle\phi|\phi'\rangle\langle\psi|\psi'\rangle$ and is extended by linearity. We usually represent $|\phi\rangle \otimes |\psi\rangle$ by $|\phi\rangle|\psi\rangle$, and if there is no ambiguity by $|\phi, \psi\rangle$.

1.1.2 State Spaces

> *For any physical system there is a corresponding Hilbert space. Any vector in the Hilbert space describes a state of the system. Two vectors that are a scaler multiple of each other correspond to the same state. Therefore, states of the systems are in one-to-one correspondence with the unit vectors of the Hilbert space up to a phase factor (a norm-one complex number).*

[1] A metric space is called complete if every Cauchy sequence in the space converges.

[2] Note that, since $\mathcal{H}$ and $\mathcal{K}$ are inner product spaces, they are canonically isomorphic to their dual spaces.

The simplest example of a quantum system is the *spin* of an electron which corresponds to a 2-dimensional Hilbert space with the orthonormal basis $\{|\uparrow\rangle, |\downarrow\rangle\}$. $|\uparrow\rangle$ describes a spin-up electron and $|\downarrow\rangle$ a spin-down electron. In general, any state of the system can be written as $a|\uparrow\rangle + b|\downarrow\rangle$, where $|a|^2 + |b|^2 = 1$. $a|\uparrow\rangle + b|\downarrow\rangle$ is called a *superposition* of states $|\uparrow\rangle$ and $|\downarrow\rangle$ with *amplitudes* a and b, respectively.

A 2-dimensional quantum system is usually called a *qubit* (quantum bit), and is represented by a Hilbert space with the orthonormal basis $\{|0\rangle, |1\rangle\}$. In the classical case, a bit can be either 0 or 1, while in the quantum case, a qubit can be in a superposition of these two states as well. Of course, there is nothing special about the basis $\{|0\rangle, |1\rangle\}$. $\{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$ is also an orthonormal basis for a qubit, and we can think of any state of a qubit as a superposition of vectors $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.

1.1.3 Density Matrices

In the previous section, we said that states of a quantum system correspond to unit vectors of a Hilbert space up to a phase factor. To eliminate this phase and get to a one-to-one correspondence, we can represent a system in the state $|\psi\rangle$ by the linear operator $|\psi\rangle\langle\psi|$. Such a linear map is positive semi-definite, rank-one, and $\mathrm{Tr}|\psi\rangle\langle\psi| = \langle\psi|\psi\rangle = 1$[3]. Hence, there is a one-to-one correspondence between matrices with this properties and states of a quantum system.

The advantage of this notation is that it provides a convenient way of describing probability distributions over states of a quantum system. Assume that a quantum system is in the state $|\psi_i\rangle$ with probability p_i. This system is described by the matrix $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. ρ is also positive semi-definite, and $\mathrm{Tr}\rho = 1$; however, it is not rank-one anymore. A positive semi-definite matrix with trace 1 is called a *density matrix*. A density matrix is in the form $|\psi\rangle\langle\psi|$ iff it is rank-one, and in this case is called a *pure state*.

> *States of a quantum system are described by density matrices over the corresponding Hilbert space. A rank-one density matrix describes a completely known state, while a density matrix generally describes a probability distribution over the Hilbert space.*

I should remark on two points regarding this notation.

First, representing quantum states with unit vectors and density matrices are equivalent, and any notion about one of them can be translated to the other. In the literature, when we talk about a general quantum state we use density matrices, and whenever the state is completely known it is represented by a vector.

Second, one may object that representing quantum states with density matrices is not unique. For instance, if a qubit is in the state either $|0\rangle$ or $|1\rangle$ each with probability 1/2, the corresponding density matrix would be $\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}I$. On the other hand, the density matrix corresponding to a qubit which is in the states $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ or $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ each with probability 1/2, is $\frac{1}{2}I$ as well. The point is that although these two qubits are generated by different processes, they describe the same state. In fact, as we will see, there is no physical experiment to distinguish these two qubits, so it is reasonable to have the same representation for them.

[3] $\mathrm{Tr}(\cdot)$ denotes the trace function.

1.1.4 Composition of Quantum Systems

Assume that $\mathcal{H}$ and $\mathcal{K}$ are two Hilbert spaces describing two quantum systems. Then the Hilbert space corresponding to the composition of the two systems is $\mathcal{H} \otimes \mathcal{K}$.

For example, two qubits are described by the Hilbert space with the orthonormal basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. The vector $\frac{1}{\sqrt{2}}(|00\rangle + |01\rangle) = \frac{1}{\sqrt{2}}|0\rangle \otimes (|0\rangle + |1\rangle)$ represents two qubits where the first one is in the state $|0\rangle$, and the second one is in the state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Note that in this example we can write the state of the composite system in the form $|\psi\rangle \otimes |\psi\rangle$ and say that the first system is in the state $|\psi\rangle$ and the second one is in the state $|\psi\rangle$. However, in general, this separation does not exist. For instance, one cannot write the state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, called an EPR[4] pair, in the form $|\psi\rangle \otimes |\psi\rangle$. Such a state is called *entangled.* Entanglement does not have any classical correspondence.

Moving to the density matrices, every density matrix over $\mathcal{H} \otimes \mathcal{K}$ that can be written in the form $\sum_i p_i \rho_i \otimes \sigma_i$, where $\{p_i\}$ is a probability distribution, and ρ_i and σ_i are density matrices over $\mathcal{H}$ and $\mathcal{K}$, respectively, is called a separable state. Other density matrices are called entangled.

1.1.5 Subsystems of a Composite System

In the previous section, we explained, given the description of two individual systems, how to find the state of the composite system. In this section we explain the converse process: given the state of a composite system, how to find the description of the individual systems.

Let A and B be two quantum systems. If the state of the composite system is $|\psi\rangle \otimes |\phi\rangle$, then the state of the individual system A is $|\psi\rangle$. More generally, if the state of the composite system is $\sum_i p_i \rho_i \otimes \sigma_i$, it means that the whole system is in $\rho_i \otimes \sigma_i$ with probability p_i. Therefore, the subsystem A is in state ρ_i with probability p_i, or equivalently, in state $\sum_i p_i \rho_i$.

So far we can find the state of a subsystem if the whole state is separable. For a general state we simply extend this definition by linearity.[5]

Definition 1.1.1 *The partial trace, denoted $Tr_{\mathcal{K}}(\cdot)$, is a linear map from matrices over $\mathcal{H} \otimes \mathcal{K}$ to matrices over $\mathcal{H}$, defined by $Tr_{\mathcal{K}}(M \otimes N) = Tr(N)M$.*

If ρ^{AB} is the density matrix of the composite system A and B, then the individual system A is described by the density matrix $\rho^A = Tr_B \rho^{AB}$.

1.1.6 Evolution

The question of the evolution of an *isolated* quantum system in time, is answered by Schrödinger equation. In Eq. (1.1), assume that the Hamiltonian H is time-independent. Then we have

$$|\psi(t)\rangle = e^{-\frac{i}{\hbar}tH}|\psi(0)\rangle.$$

Since H is hermitian, $e^{-\frac{i}{\hbar}tH}$ is unitary. This observation holds even if the Hamiltonian is time-dependent.

[4]EPR stands for Einstein, Podolsky and Rosen. The state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is called an EPR pair after their famous paper about the incompleteness of quantum physics [52].

[5]Note that Schrödinger equation (1.1) is also linear in terms of the wave-function.

The evolution of an isolated quantum system is described by unitary operators. More precisely, the states of system at times t_0 and t_1 are related by a unitary operation U, $|\psi(t_1)\rangle = U|\psi(t_0)\rangle$. This expression in terms of density matrices is $\rho_{t_1} = U\rho_{t_0}U^\dagger$.

Note that, in a mathematical point of view, unitary operators are the only linear maps that send unit-vectors to unit-vectors and density matrices to density matrices.

1.1.7 Measurement

We learned how to describe quantum systems, and also how to relate the states of systems in different times, but we still do not know how to extract information. In classical physics, for instance, we can measure the angular momentum of a particle without disturbing the particle. However, in the quantum world such a measurement is impossible; when we measure the spin of an electron, we find it either spin-up or spin-down. It means that, after the measurement the spin *cannot* be in a superposition of spin-up and spin-down anymore, and we have changed the spin during the measurement.

A measurement of a quantum system with the corresponding Hilbert space $\mathcal{H}$ is described by projections $P_1, P_2, \dots P_m$ with $\sum_i P_i = I$, where I is the identity operator on $\mathcal{H}$ and such that $P_iP_j = \delta_{ij}I$. If we measure a state $|\psi\rangle \in \mathcal{H}$, we get the i-th outcome with probability $\langle\psi|P_i|\psi\rangle$. In this case, the state after the measure will be parallel to $P_i|\psi\rangle$.

What is explained here is usually called a *projective measurement*. This type of measurement is not the general notion of measurement in quantum physics. Using these measurements as a tool, we can remove the constraint that $P_iP_j = \delta_{ij}I$. The idea is to extend the Hilbert space by attaching another system, consider a projective measurement in the extended space, and then trace out the extra part of the space, (see [96] for details). In fact, the only important condition on measurement operators is what usually referred as the *completeness equation:*

$$\sum_i P_i^\dagger P_i = I.$$

In this general setting, the outcome of the measurement is i with probability $\langle\psi|P_i^\dagger P_i|\psi\rangle$, and the state after the measurement collapses to

$$\frac{P_i|\psi\rangle}{\langle\psi|P_i^\dagger P_i|\psi\rangle^{\frac{1}{2}}}.$$

There are examples of measurements in which we do not care about the state after the measurement, but only the probability distribution of the outcome. This probability distribution can be expressed in terms of *positive operator-value measure* (POVM) elements $M_i = P_i^\dagger P_i$.

A POVM consists of positive semi-definite matrices M_i, with the completeness equation $\sum_i M_i = I$. If we apply this measurement on the density matrix ρ, the outcome will be i with probability $Tr(M_i\rho)$.

In a mathematical point of view, any linear map that sends density matrices to probability distributions can be expressed in terms of a POVM.

1.1.8 Example

In the previous sections we learned the general concepts of quantum physics. Here, by giving an example, we bring them all together.

Let $\sigma = a|0\rangle\langle 0| + b|1\rangle\langle 1|$, where $a, b > 0$ and $a + b = 1$, be the state of a qubit. We can think of this qubit as the second qubit of the two-qubit state $|\psi\rangle = \sqrt{a}\,|00\rangle + \sqrt{b}\,|11\rangle$: $\mathrm{Tr}_1|\psi\rangle\langle\psi| = \sigma$.The state $|\psi\rangle$ is called a *purification* of σ.

It is easy to see that such a purification is always possible; for any mixed state σ^A, there exists a system B and a pure state $|\psi^{AB}\rangle$ over the composite system AB, such that $\sigma^A = \mathrm{Tr}_B|\psi^{AB}\rangle\langle\psi^{AB}|$.

Now, assume that we apply the *Hadamard gate* on the first qubit. The matrix representation of Hadamard gate in the standard basis $\{|0\rangle, |1\rangle\}$ is

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{1.2}$$

Therefore, we have

$$H_1|\psi\rangle = H \otimes I|\psi\rangle = \sqrt{\frac{a}{2}}\,|00\rangle + \sqrt{\frac{a}{2}}\,|10\rangle + \sqrt{\frac{b}{2}}\,|01\rangle - \sqrt{\frac{b}{2}}\,|11\rangle.$$

Next, we measure the first qubit in the basis $\{|0\rangle, |1\rangle\}$ (the projective measurement with projectors $P_1 = |0\rangle\langle 0|$ and $P_2 = |1\rangle\langle 1|$). We get each of the outcomes $|0\rangle$ and $|1\rangle$ with probability $1/2$. In the first case, the second qubit collapses to $|\phi_1\rangle = \sqrt{a}|0\rangle + \sqrt{b}|1\rangle$, and in the second case to $|\phi_1\rangle = \sqrt{a}|0\rangle - \sqrt{b}|1\rangle$. (Since the two qubits are entangled, measuring the first qubit will change the state of the second qubit as well). After the measurement, the second qubit is in the state

$$\frac{1}{2}|\phi_0\rangle\langle\phi_0| + \frac{1}{2}|\phi_1\rangle\langle\phi_1| = a|0\rangle\langle 0| + b|1\rangle\langle 1|,$$

which is the same state as what we started with.

Such an equality always holds; if we have a state on two systems, any unitary operation or measurement on the first system, *without knowing the outcome of the measurement*, does not change the reduced density matrix of the second system. This fact can be proved easily using linearity.

1.1.9 Heisenberg Uncertainty Principle

Heisenberg uncertainty principle is usually stated as

$$\Delta x \Delta h \geq \frac{\hbar}{2},$$

where x denotes the position operator and h the momentum operator, and Δx and Δh denote the accuracy of their measurement.

This inequality means that if we measure the position of a particle up to precision Δx, then the best measurement of momentum (without changing the particle) contains at least $\frac{\hbar}{2\Delta x}$ error. In other words, we cannot measure both position and momentum at the same time.

When we say that x is the position operator, it means that x is a hermitian matrix over the Hilbert space of the particle, each of whose eigenvectors correspond to a certain position

(and the same for h). Let $\{|\psi_s\rangle\}_s$ be the set of eigenvectors of x which is an orthonormal basis for the Hilbert space. The vector $|\psi_s\rangle$ describes the state of a particle in position s. To measure the position of a given particle we should consider the projective measurement with operators $\{P_s = |\psi_s\rangle\langle\psi_s|\}_s$. Thus, after measurement, by getting the position s the state of system collapses to $|\psi_s\rangle$. Now assume we measure momentum. Under the new measurement we change the state (it collapses to an eigenvector of h), except if $|\psi_s\rangle$ is an eigenvector of h as well. However, x and h do not commute, so they have different eigenvectors. As a result, it is impossible to measure both x and h at the same time.

1.2 Quantum Channels

A *quantum channel* is any physical process that given a quantum state outputs another quantum state. As an example, suppose given a state ρ we measure this state using the POVM $\{M_i\}$, and if the outcome of measurement is i we output the state σ_i. This means that, in the output we have σ_i with probability $\mathrm{Tr}(M_i\rho)$; thus, we can write the channel in the form[6]

$$\Psi(\rho) = \sum_i \mathrm{Tr}(M_i\rho)\sigma_i. \tag{1.3}$$

Characterization of quantum channels is very useful because of their wide applications in quantum information theory, and also designing and analyzing quantum algorithms. Here, we present two (equivalent) characterizations of quantum channels from different points of view.

1.2.1 Physical Characterization

Based on what we have learned in Section 1.1, a quantum process consists of adding and deleting subsystems, applying unitary operations, and measurement. Here, it is useful to mention an important fact called the *principle of deferred measurement.*

> *Any measurement performed in the middle of a quantum process can be replaced by another measurement performed at the end of the process. It holds even if other operations are conditionally performed depending on the outcome of the measurement.*

This fact is a standard consequence of the techniques of designing quantum circuits, and we refer the reader to [96, 103] for a proof.

Using this result, we may assume that all the measurements are performed at the end of a process. On the other hand, since, in general, we do not know the outcome of the measurement, the state of the remaining system after the measurement would be the same if we just ignore the measured system and trace-out it (see Section 1.1.8). Hence, we may replace measurements with a trace-out at the end of a process.

The remaining operations that we can perform are adding extra subsystems and also applying unitary operators. Not that without loss of generality, we may assume that we add all the extra subsystems once at the beginning, and also we may assume that the state of the extra subsystems is a pure state[7] which we call $|e\rangle$.

[6] Channels in the form of Eq. (1.3) are called *entanglement breaking.*

[7] Use the purification idea explained in Section 1.1.8.

Therefore, the whole process is the following: given the input state ρ we add the state $|e\rangle$ to it, apply the unitary operator U, and then trace-out a subsystem which we call E. We can write the output state as

$$\Psi(\rho) = \mathrm{Tr}_E(U\rho \otimes |e\rangle\langle e|U^\dagger). \tag{1.4}$$

Every quantum channel can be written in the above form.

1.2.2 Mathematical Characterization

In this section we define quantum channels as maps over density matrices that satisfy certain properties.

First, quantum physics is linear, so any quantum channel Ψ should be linear as well. More precisely, for a probability distribution $\{p_i\}$ and mixed states $\{\rho_i\}$,

$$\Psi(\sum_i p_i\rho_i) = \sum_i p_i\Psi(\rho_i). \tag{1.5}$$

Second, a quantum channel must send quantum states to quantum states. In means that $\Psi(\rho)$ should be normalized: $\mathrm{Tr}\,\Psi(\rho) = 1$. This property is referred as Ψ being *trace preserving*[8].

Third, Ψ should send mixed states, that are described by positive semi-definite matrices, to positive semi-definite matrices. This property is referred as Ψ being *positive*. However, something more than the positivity should hold. Assume that we apply the channel Ψ to one part of a bipartite state. The outcome must still be positive. In other words, for the identity channel Id, that is applied to density matrices of an arbitrary dimension, $Id \otimes \Psi$ must also be a positive map. This stronger property is referred as Ψ begin *completely positive*[9].

These three properties characterize quantum channels from a mathematical point of view. Such a map with these properties is called *completely positive trace-preserving* (CPTP). It is not hard to prove that any CPTP map on density matrices can be written in the form

$$\Psi(\rho) = \sum_i E_i\rho E_i^\dagger, \tag{1.6}$$

where $\{E_i\}$ are arbitrary matrices, called *Kraus operators*, that satisfy $\sum_i E_i^\dagger E_i = I$.

It can be shown easily that the two representations (1.4) and (1.6) are equivalent: any channel in the form of Eq. (1.4) can be written in the form of Eq. (1.6) and vice versa.

1.2.3 No-Cloning Theorem

Now we are ready to state one of the most exiting results of quantum physics: there is no physical process for making a copy of an unknown quantum state [122]. This statement is called the *no-cloning theorem.*

[8]The trace preserving assumption is not crucial, and can be replaced with $0 \leq \mathrm{Tr}\,\Psi(\rho) \leq 1$. Indeed, we may think of a quantum process in which we ignore certain outcomes of a measurement. It is the same as to replace the completeness property of a POVM $\{M_i\}$ with $\sum_i M_i \leq I$.

[9]Consider the map that sends a matrix to its transpose. This map is clearly linear, trace preserving and positive, but it is not completely positive. For example, if we apply the transpose map on one qubit of an EPR pair we do not get a positive semi-definite matrix.

A more precise description of no-cloning theorem is that there is no quantum channel that sends $|\psi\rangle$ to $|\psi\rangle|\psi\rangle$. Using the characterization of quantum channels in Eq. (1.4), we can express this theorem in the following form. There is no unitary operation U such that

$$U|\psi\rangle|e_0\rangle|e_1\rangle = |\psi\rangle|\psi\rangle|\psi_*\rangle,$$

where $|\psi\rangle$ is an arbitrary state, $|e_0\rangle$, $|e_1\rangle$ are fixed, and $|\psi_*\rangle$ is some state which depends on $|\psi\rangle$.

For a proof suppose such a U exists. Since U is unitary and preserves inner product, for every two states $|\psi\rangle$ and $|\psi'\rangle$ we have

$$|\langle\psi|\psi'\rangle| = |\langle\psi|\psi'\rangle\langle e_0|e_0\rangle\langle e_1|e_1\rangle| = |\langle\psi|\psi'\rangle^2\langle\psi_*|\psi'_*\rangle| \leq |\langle\psi|\psi'\rangle|^2,$$

which is a contradiction if we let $\langle\psi|\psi'\rangle = 1/2$.

1.3 Distance Measures

In this section I will address the question of quantifying the difference between quantum states. This problem is easy in the special case of pure states because the only parameter that describes the relative position of two unit vectors $|\phi\rangle$ and $|\psi\rangle$ is the angle between them. However, in general, quantum states may be mixed, and the problem is not as easy as for the pure states.

1.3.1 Trance Distance

The *trace distance* of two mixed states ρ and σ is defined by

$$D(\rho,\sigma) = \|\rho-\sigma\|_{\mathrm{Tr}} = \frac{1}{2}\mathrm{Tr}|\rho-\sigma|, \tag{1.7}$$

where $|X| = \sqrt{X^\dagger X}$. For instance, if $\rho = |\phi\rangle\langle\phi|$ and $\sigma = |\psi\rangle\langle\psi|$ are pure, then

$$D(|\phi\rangle,|\psi\rangle) = \sqrt{1-|\langle\phi|\psi\rangle|^2}\,. \tag{1.8}$$

It is not hard to show that

$$\|\rho-\sigma\|_{\mathrm{Tr}} = \max_P \mathrm{Tr}\big(P(\rho-\sigma)\big), \tag{1.9}$$

where the maximum is taken over all hermitian matrices P, such that $0 \leq P \leq I$. Using this characterization of trace distance, it is easy to show that trace distance satisfies *triangle inequality*. The other useful property of trace distance is the *monotonicity under CPTP maps*: for any quantum channel Ψ, and quantum states ρ and σ we have

$$\|\Psi(\rho)-\Psi(\sigma)\|_{\mathrm{Tr}} \leq \|\rho-\sigma\|_{\mathrm{Tr}}.$$

Note that, a special choice of the channel Ψ is the partial trace map; hence, for bipartite states ρ^{AB} and σ^{AB} we have

$$\|\mathrm{Tr}_B\,\rho^{AB} - \mathrm{Tr}_B\,\sigma^{AB}\|_{\mathrm{Tr}} = \|\rho^A-\sigma^B\|_{\mathrm{Tr}} \leq \|\rho^{AB}-\sigma^{AB}\|_{\mathrm{Tr}}. \tag{1.10}$$

1.3.2 Fidelity

Fidelity between density matrices is a generalization of inner product of pure states. This measure was defined in [69].

$$F(\rho,\sigma) = \mathrm{Tr}\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}. \tag{1.11}$$

In the special case where one of the states is pure we have

$$F(\rho,|\psi\rangle) = \sqrt{\langle\psi|\rho|\psi\rangle}\,.$$

This formula can be generalized in the following form.

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

where the maximum is taken over all purifications $|\phi\rangle$ of ρ, and $|\psi\rangle$ of σ. Using this formula, Eq. (1.8), and inequality (1.10) it is easy to show that

$$1 - F(\rho,\sigma) \leq \|\rho-\sigma\|_{\mathrm{Tr}} \leq \sqrt{1-F(\rho,\sigma)^2}\,. \tag{1.12}$$

1.3.3 Von Neumann Entropy

Suppose, given a mixed state ρ, we want to quantify how far this state is from being pure. Of course, if ρ is rank-one, it is pure, but the rank of a matrix is not a continuous quantity and cannot be considered as a good measure. Notice that, if ρ is pure, then one of its eigenvalues is 1 and the rest are zero, so purity of a state can be measured by the distance of the set of its eigenvalues from that of pure states. Indeed, we can think of the set of eigenvalues of ρ as a probability distribution, and use all known classical techniques to quantify its properties.

Shannon entropy is a key concept in information theory. For a probability distribution $\{p_i\}$, its entropy is defined by

$$H(\{p_i\}) = \sum_i -p_i \log p_i. \tag{1.13}$$

If $p_1 = 1$ and $p_2 = \cdots = p_n = 0$, then $H(\{p_i\}) = 0$; also, the maximum value of Shannon entropy is attained at the uniform distribution $H(1/n,\ldots,1/n) = \log n$. Therefore, we can think of $H(\{p_i\})$ as the *uncertainty* of the probability distribution $\{p_i\}$.

The quantum version of Shannon entropy for quantum states is called *von Neumann entropy*, and is defined as the Shannon entropy of the eigenvalues of a density matrix [56, 59, 67]. Equivalently, von Neumann entropy is

$$S(\rho) = -\mathrm{Tr}(\rho\log\rho). \tag{1.14}$$

Quantum relative entropy is also defined by

$$S(\rho\|\sigma) = \mathrm{Tr}(\rho\log\rho) - \mathrm{Tr}(\sigma\log\rho). \tag{1.15}$$

Klein's inequality [96] says that $S(\rho\|\sigma)$ is always non-negative. In fact, it is zero if and only if $\rho = \sigma$. Hence, we can consider quantum relative entropy as a distance measure between quantum states.

1.4 Entanglement

Entanglement is a phenomenon in quantum physics with no classical analogue. Indeed, entanglement is considered as the main resource in quantum computation and quantum information theory. For instance, by *superdense coding* (see [20, 96, 103]) we can use entanglement as a resource, and transmit two bits of information by sending only one qubit. The first step toward understanding entanglement is to measure it. Several entanglement measures have been defined, and we mention some of them here.

1.4.1 Entanglement of Pure States

Before going to the general case, let us first consider the entanglement of pure states. Suppose $|\psi^{AB}\rangle$ is a bipartite pure state. Then, $\rho^A = \mathrm{Tr}_B|\psi^{AB}\rangle\langle\psi^{AB}|$ and also $\rho^B = \mathrm{Tr}_A|\psi^{AB}\rangle\langle\psi^{AB}|$ are pure if and only if $|\psi^{AB}\rangle$ is separable. It means that the amount of entanglement of $|\psi^{AB}\rangle$ is related to the amount of uncertainty ρ^A and ρ^B. On the other hand, it is easy to see that for any pure state $|\psi^{AB}\rangle$, the sets of eigenvalues of ρ^A and ρ^B, regardless of zeros, are the same. Thus, we can define the entanglement of $|\psi^{AB}\rangle$ by

$$E(|\psi^{AB}\rangle) = S(\rho^A) = S(\rho^B). \tag{1.16}$$

This measure of entanglement is called *entropy of entanglement.*

For example, the entanglement of an EPR pair is

$$E\big(\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)\big) = S(\frac{1}{2}I) = 1,$$

which is the maximum amount of entanglement between two qubits. An EPR pair is considered as a *unit* of entanglement.

1.4.2 Entanglement of Formation

Let

$$\rho^{AB} = \sum_i p_i|\psi_i^{AB}\rangle\langle\psi_i^{AB}|$$

be a mixed state. It means that the system is in state $|\psi_i^{AB}$ with probability p_i. Hence, one idea toward generalizing the entanglement measure defined in Eq. (1.16), is to consider the average of the entanglement of states $|\psi_i^{AB}\rangle$. However, the problem is that writing a mixed state as an average of pure states, is not unique. For example,

$$\sigma^{AB} = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|$$

is a separable state, and then its entanglement must be zero. On the other hand, we can write

$$\sigma^{AB} = \frac{1}{2}|\psi\rangle\langle\psi| + \frac{1}{2}|\phi\rangle\langle\phi|,$$

where $|\psi\rangle = \frac{1}{\sqrt{2}}(\,|00\rangle + |11\rangle)$ and $|\phi\rangle = \frac{1}{\sqrt{2}}(\,|00\rangle - |11\rangle)$ are both maximally entangled.

To resolve this problem we can simply consider the minimum of the average of Eq. (1.16) over all possible ensembles.

$$E_F(\rho^{AB}) = \min \sum_i p_i S\big(\mathrm{Tr}_B\, |\psi_i^{AB}\rangle\langle\psi_i^{AB}|\big), \tag{1.17}$$

where the minimum is taken over all $\{p_i, |\psi_i^{AB}\rangle\}$ such that $\rho^{AB} = \sum_i p_i |\psi_i^{AB}\rangle\langle\psi_i^{AB}|$. This measure of entanglement is called *entanglement of formation* [22]. Notice that, entanglement of formation is equal to entropy of entanglement for pure states. Also, if ρ^{AB} is separable then $E_F(\rho^{AB}) = 0$.

1.4.3 Local Operations and Classical Communication

Entanglement is not a classical phenomenon at all. It means that two far apart parties cannot generate an entangled state between themselves only by classical communication. As an example, assume that Alice and Bob talk on the phone and agree on a set of states $\{\rho_i^A, \rho_i^B\}$ and a probability distribution $\{p_i\}$. Then, Alice flips some coins and picks i according to the distribution $\{p_i\}$, and tells Bob the result. Now, Alice and Bob generate the states ρ_i^A and ρ_i^B, respectively, in their own labs. Hence, they generate the state $\sum_i p_i \rho_i^A \otimes \rho_i^B$, which is separable.

The point is that to produce more entanglement, Alice and Bob should have quantum correlation, and by *local operations and classical communication* (LOCC) they cannot generate entanglement. It means that, for a reasonable measure of entanglement E we should have

$$E(\Psi^{AB}(\rho^{AB})) \leq E(\rho^{AB}),$$

where Ψ^{AB} is an LOCC map. In this case we say that E is *monotonic decreasing under LOCC operations.* It is easy to show that entanglement of formation is monotonic decreasing under LOCC operations.

1.4.4 Entanglement Cost and Entanglement of Distillation

In Section 1.4.1 we mentioned that an EPR pair is considered as the unit of entanglement, so loosely speaking, the entanglement of a state ρ^{AB} is α means that there are α EPR pairs *inside* ρ^{AB}. To express this statement more precisely we should interpret the word "inside."

One idea is to say that α EPR pairs can be extracted from ρ^{AB}: if two far apart parties Alice and Bob share the state ρ^{AB}, they can transform this state to $\mathrm{EPR}^{\otimes\alpha}$ pair by LOCC operations.

Based on this idea, *entanglement of distillation* is defined as the rate of the number of EPR pairs that can be generated from arbitrary many copies of the state [21]:

$$E_D(\rho) = \sup_{\alpha} \{\, \exists \Psi_n \in \mathrm{LOCC} : \lim_{n\to\infty} \|\Psi_n(\rho^{\otimes n}) - \mathrm{EPR}^{\otimes(\alpha n)}\|_{\mathrm{Tr}} = 0\}. \tag{1.18}$$

It is not hard to see that entanglement of distillation is equal to zero for separable states; also it is monotonic decreasing under LOCC operations. However, there are states that are not separable while their entanglement of distillation is zero [63].

By the same idea as in the entanglement of distillation, we can define *entanglement cost*: having arbitrary many EPR pairs, how many states ρ can be generated using LOCC maps.

$$E_C(\rho) = \inf_{\alpha} \{\, \exists \Psi_n \in \mathrm{LOCC} : \lim_{n\to\infty} \|\Psi_n(\mathrm{EPR}^{\otimes(\alpha n)}) - \rho^{\otimes n}\|_{\mathrm{Tr}} = 0\}. \tag{1.19}$$

By definition, entanglement cost is equal to zero for separable states, and also it is monotonic decreasing under LOCC operations. Unlike the entanglement of distillation, entanglement cost is faithful, meaning that it is non-zero for any entangled state [123]. Another interesting property of entanglement cost is that it is equal to the *regularized* entanglement of formation [64]

$$E_C(\rho) = \lim_{n\to\infty} \frac{1}{n} E_F(\rho^{\otimes n}).$$

Here, I should mention that entanglement of distillation and entanglement cost are equal for pure states, and agree with the entanglement of formation [19].

1.4.5 Relative Entropy of Entanglement

The more a state is entangled, the farther the state is from separable states, so to quantify the entanglement we can consider some distance measure and define an entanglement measure to be equal to the distance of the state from separable states. Among distance measures that we have, trace distance and fidelity are not wise choices because they are always between 0 and 1, and then do not really distinguish between, say, 10 EPR pairs and 1000 of them. However, quantum relative entropy is a useful choice; we can define the *relative entropy of entanglement* [115, 114] as follows.

$$E_R(\rho) = \min_{\sigma} S(\rho\|\sigma),$$

where the minimum is taken over all separable states σ.

Relative entropy of entanglement is faithful, and also monotonic decreasing under LOCC operations. It also agrees with entanglement of formation on pure states.

Since relative entropy of entanglement is not additive [116], its regularized version is also of interest.

$$E_R^{\infty}(\rho) = \lim_{n\to\infty} \frac{1}{n} E_R(\rho^{\otimes n}).$$

1.5 Theory of Computation

In this section I discuss what a quantum algorithm is. Before that, I explain the classical case, describe the circuit model, and then try to define its quantum analogue. In addition, I will define some well-known classical and quantum complexity classes.

A more comprehensive discussion about theory of computation and complexity theory can be found in [111] and [100].

1.5.1 Universal Set of Gates

Suppose we have two integers $m > n > 0$ and want to compute their greatest common divisor (gcd). The following algorithm, called the Euclidean algorithm, does this job. If n divides m, output n; if not, replace n with the remainder of the division of m by n, replace m by n, and repeat. What is given here is a *high level* description of Euclidean algorithm, but a PC does not understand the words "replace", "division", "remainder", and we should somehow translate them.

A PC stores information in sequences of bits, where each bit contains either a 0 or a 1, and while running the algorithm it updates them one by one. Of course, we cannot ask a PC to update the bits by dividing two numbers in only one step. We should break "division" into some simple rules and ask PC to follow those steps. In order to build a real PC these *updating rules* should satisfy two basic properties. First, there should be a few of them, and second, each one of them should be very simple.

We can think of an updating rule as a *Boolean function* $f : \{0,1\}^m \to \{0,1\}^n$. A Boolean function is simple if m and n are small. Thus, to have a PC we should fix a few simple Boolean functions, called *gates*, and design a computer that has the ability to apply those gates. The only point that we should be careful about is that those gates must be *universal*, meaning that they should generate any Boolean function.

The gates AND, OR, NOT and NAND are some well-known gates.

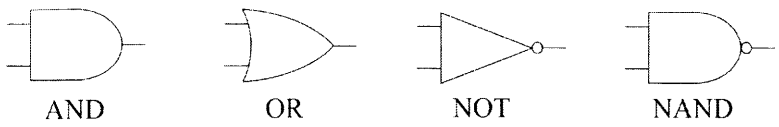

Figure 1-1: Distinctive shape of the gates AND, OR, NOT and NAND.

It is easy to show that the sets $\{\text{AND}, \text{NOT}\}$, $\{\text{OR}, \text{NOT}\}$, and $\{\text{NAND}\}$ are universal.

A sequence of gates is called a *circuit*. Figure 1-2 shows a circuit that computes the XOR of two bits using only NAND gate.

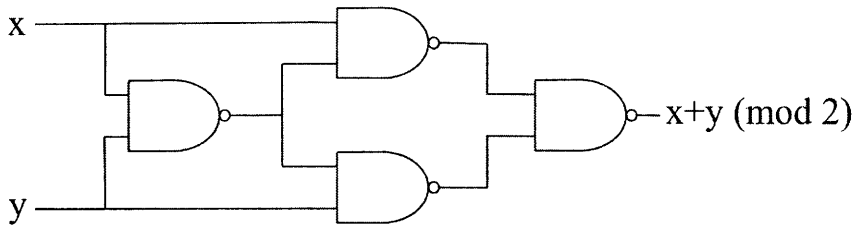

Figure 1-2: This circuit computes XOR of two bits using only NAND gate.

1.5.2 Languages

Before getting to a formal definition of an algorithm for a problem, we should define what we mean by a problem.

In the example of Euclidean algorithm, our problem has inputs m, n, and the output is $\gcd(m, n)$, which is a function from pairs of integers to natural numbers. However, for simplicity, in theory of computation, functions are usually considered with the range $\{0,1\}$[10]. Thus, a problem can be formulated as follows: it consists of a set of functions $\{f_n\}$

[10]Notice that, the gcd function can also be expressed using these restricted set of function; for example, for each k we can ask whether the k-th bit of the gcd is 0 or 1.

where $f_n : \{0,1\}^n \to \{0,1\}$. Each bit-string $x \in \{0,1\}^n$ is an input of the problem, and the answer to the problem x is $f_n(x)$.

There is a more efficient way of describing problems: we can define the *language* corresponding to the problem $\{f_n\}$ by

$$L = \bigcup_{n>0} \{x \in \{0,1\}^n : f_n(x) = 1\},$$

so the answer to the problem, on input x, is whether x is in the language or not.

An algorithm that solves such a problem is a set $\{C_n\}$, where C_n is a circuit with n-bit strings as input, and a single-bit as output, such that x of length n is in L if and only if the output of C_n on x is 1.

1.5.3 Complexity Theory

Any Boolean function can be computed using the set of gates $\{\text{AND}, \text{NOT}\}$, which is universal. However, a circuit that computes such a function may be very complicated, and it may take a decade for a PC to apply the circuit and finish the computation. Indeed, designing an algorithm, or equivalently a circuit, is not the main part of solving a problem because we know that it is always possible since we start with a universal set. The hard part is to design a circuit which is efficient based on our resources such as the time of computation.

Of course to solve a problem for a given input, the longer the length of the input is, the more resources we need. Hence, for instance, time as a resource should be specified as a function of the input-length.

If we assume that each gate in a circuit takes one time-step to be applied, the time of a computation is equal to the number of gates in the circuit. Thus, we say that a language has a $t(n)$-time algorithm, or a $t(n)$-size circuit, if for any input-length n, there exists a circuit C_n which contains $t(n)$ gates and solves the problem on inputs of length n. The *complexity class* P consists of languages for which there exists a *uniform* polynomial-size[11] set of circuits. Here, I do not define the word "uniform" precisely, but it basically excludes the case where circuits C_n, for different values of n, are completely independent.[12]

The set of rank-one matrices, connected graphs, and complete square integers are examples of languages in P. Another problem in this class is the *primality problem*; there is an efficient (polynomial-time) algorithm for deciding whether a given number is prime or not [5].

The same as P, EXP is defined to be the class of languages that have an exponential-time algorithm.[13] By definition, it is clear that $\mathsf{P} \subseteq \mathsf{EXP}$. The set of 3-colorable graphs is a language which is not known to be in P, but is insinde EXP.[14]

Although the language of 3-colorable graphs is not known to have a polynomial-time algorithm, it has an interesting property that makes it an easier problem than a typical one in EXP; if we are given a proposed 3-coloring of a graph, in polynomial-time we can

[11] A $p(n)$-size circuit for some polynomial $p(n)$. Here, all polynomial-time algorithms are considered in the same class because the running-time of an algorithm, depending on the choice of the model of computation (for example the universal set of gates), may change by a polynomial factor.

[12] Any algorithm that you can think of, is uniform.

[13] A $2^{p(n)}$-time algorithm for some polynomial $p(n)$.

[14] Here is an exponential-time algorithm: check all the possible colorings.

check whether it is a valid coloring or not. In other words, finding a valid coloring is a hard problem; however, checking whether a given coloring is eligible or not, is easy.

The class of languages that given a polynomial-size *hint* or *witness*, can be solved in polynomial-time, is called *non-deterministic polynomial-time* and is denoted by NP. More precisely, a language L is in NP if there is a polynomial-size circuit C such that for any $x \in L$ there exists w_x with $C(x, w_x) = 1$, and for any $x \notin L$ and any w, $C(x, w) = 0$.

By definition, $\mathsf{P} \subseteq \mathsf{NP}$. Although it is widely believed that this inclusion is strict, this is one of the most important open questions in mathematics that whether P is the same as NP or not [40].

The most well-known language in NP is the satisfiability problem, denoted by SAT [42]. This problem is that given Boolean variables $x_1, \ldots, x_n$ and clauses $c_1, \ldots, c_m$, where each clause is the "OR" of some of the variables, either x_i or $\bar{x}_i$,[15] decide whether there exists an assignment to the variables $x_1, \ldots, x_n$ such that the value of all clauses be 1. k-SAT is a special case of SAT in which each clause consists of k variables. Clearly, given an assignment we can efficiently (in polynomial-time) check whether it is a valid assignment or not; therefore, k-SAT is in NP. Indeed, k-SAT, for $k \geq 3$, is the hardest problem in NP, meaning that we can reduce any problem in NP to an instance of 3-SAT. In other words, if a model of computation can solve 3-SAT, then it can solve any problem in NP. Such a problem is called NP-hard. An NP-hard problem which is inside NP is called NP-complete. Both of SAT and 3-coloring are NP-complete. For a list of NP-complete problems see [53].

NEXP, non-deterministic exponential-time, is the exponential version of NP. It is the class of languages for which a *proof* can be verified in exponential-time.

Besides time, space is also a very important constraint in the classification of problems. By space we mean the amount of memory, or more precisely the number of Boolean variables that are required during computation. PSPACE is the class of languages that have polynomial-space algorithms.

Obviously, an algorithm that is run in polynomial-time cannot occupy more than polynomially many bits of memory; hence, $\mathsf{P} \subseteq \mathsf{PSPACE}$. NP is also in PSPACE because, for example in the case of 3-SAT, we can check the validity of one possible assignment, erase the memory, and then go for the next assignment. On the other hand, PSPACE is in EXP because on a computer that solves a problem in $p(n)$-space, for some polynomial $p(n)$, there are $p(n)$ Boolean variables that describe the configuration of the computer is each step; thus, the number of all configurations is at most $2^{p(n)}$, or equivalently the number of time-steps is at most $2^{p(n)}$.

1.5.4 Randomized Algorithms

Suppose we are given a sequence $x_1, \ldots, x_{2n}$ such that either all x_i's are 1, or at least half of them are 0, and want to distinguish between these two cases. To solve this problem we can check $x_1, \ldots x_{n+1}$; if all of them are 1, output the first case, and if there is a 0 between them, output the second case. In this algorithm we need $n + 1$ *query* from sequence $x_1, \ldots, x_{2n}$, and $n + 1$ is optimum in the sense that if we know only n bits of the sequence we cannot distinguish between the two cases with no *error*. However, if we allow some probability of error, only one query is enough; we can *randomly* choose a number i, $1 \leq i \leq 2n$, and check x_i; if it is 1, we output the first case, and if not, the second case. This algorithm gives the right answer with probability 1/2.

[15] $\bar{x}$ denotes the NOT of x.

This problem is an example of algorithms called *randomized algorithms*, that are the same as ordinary algorithms except that we have access to a source of randomness, say a fair coin, and we want to find the right answer with some probability of error. In the above example, the extra source of randomness and also relaxing our expectation in finding the right answer, helped us to solve the problem using fewer resources.

BPP, *bounded-error probabilistic polynomial-time* [55], is the class of languages for which there exists a polynomial-time randomized algorithm that outputs the right answer with probability at least 2/3.

Notice that, in the example the probability of success is 1/2; however, we can increase it to 2/3 by repeating the algorithm and checking, say, 10 random x_i's instead of only one of them. In general, by repeating a BPP algorithm polynomially many times, and taking the majority vote of the outputs, the probability of success can be increased to $1 - 2^{-p(n)}$, for any polynomial $p(n)$. Here, there is a gap between probability of error and probability of getting the right answer ($2/3 - 1/3 = 1/3$), and since the gap is greater than an inverse polynomial ($1/p(n)$ for some polynomial $p(n)$) we can use Chernoff bound.

By removing the assumption "bounded-error" we get to the complexity class PP *probabilistic polynomial* [55]. PP is the class of languages L that have a polynomial-time randomized algorithm that outputs "yes" with probability greater than 1/2 on inputs $x \in L$, and outputs "no" with probability at least 1/2 on inputs $x \notin L$. Note that here there is no gap between probability of error and probability of getting the right answer, so we cannot use Chernoff bound and *amplify the gap.*

By definitions $\mathsf{P} \subseteq \mathsf{BPP} \subseteq \mathsf{PP}$. Also by the same techniques as in the proof of $\mathsf{NP} \subseteq \mathsf{PSPACE}$ one can show $\mathsf{PP} \subseteq \mathsf{PSPACE}$.[16] Surprisingly. PP contains NP; here is an algorithm to solve 3-SAT in PP; pick a random assignment for variables $x_1, \ldots, x_n$ and check whether it is a satisfying assignment or not. If yes, output "yes", and if not output either "yes" or "no" each with probability 1/2.

1.5.5 Probabilistic Checkable Proofs

3-SAT is in NP: having an assignment for the Boolean variables, a polynomial-time verifier can check the validity of the assignment. Now assume that the verifier does not want to read all the variables, but a constant number of them. It is equivalent to say that the verifier checks the satisfiability of only a constant number of clauses, say only one of them. Thus, he picks one clause at random, reads the value of variables in that clause from the given assignment, and checks whether that clause is satisfiable or not. He accepts if it is satisfiable and rejects otherwise. This protocol shows that the verifier can solve NP-complete problems by reading only 3 bits from the given satisfying assignment. However, this algorithm has a large probability of error. For example, if all but one of clauses of an instance of 3-SAT can be satisfied simultaneously, the verifier picks that clause with probability $1/m$ (m is the number of clauses), so he gets to the right answer with probability $1/m$, which is very small.

The *probabilistic checkable proofs* (*PCP*) theorem [10, 11, 44] is one of the deepest results in complexity theory which says that we can solve NP-complete problems with a *constant* probability of error by reading only 3-bits of the proof.

The proof of PCP theorem consists of several steps for amplifying the gap from inverse polynomial to constant, without increasing the size of problem. Here, I do not give any detail about these steps and refer the reader to [44].

[16] The idea is to check all possible random resources.

1.5.6 Proof Systems

We can think of the complexity class NP as a *proof system* in which the *prover* sends a witness to the *verifier*, and the verifier checks whether it is a valid proof or not. There are several directions to generalize this setting.

We can simply consider the probabilistic version of NP: there is a prover, called *Merlin*, who sends a proof to a verifier, called *Arthur*, and Arthur in probabilistic polynomial-time decides to whether reject or accept. We assume that Arthur gets the right answer with probability 2/3. This complexity class is called MA (Merlin-Arthur) [14]. By definition, MA contains NP. The same argument as for proving NP $\subseteq$ PP (see Section 1.5.4) shows that MA $\subseteq$ PP.

Now consider the case where Arthur asks a question from Merlin, and after Merlin's answer he either accepts or rejects. Here, Arthur is again a probabilistic polynomial-time verifier. This complexity class is denoted AM (Arthur-Merlin) [14]. It is clear that MA $\subseteq$ AM.

We can go further and assume that the interaction (question and answer) between verifier and prover is more than one round, and they can interact as much as they want. However, the number of rounds has to be at most polynomial since the verifier is restricted to polynomial-time. This complexity class is denoted IP (interactive proof). Shamir in [106] has shown that IP = PSPACE.

Another generalization is to consider more than one prover. Two or more provers are useful because the verifier may ask one of them a question and check the answer with another one. For example, assume that the verifier has an exponential-size instance of 3-SAT. He cannot check a satisfying assignment of this problem in polynomial-time. Instead, he can use the advantage of two provers by the following protocol. He chooses one of the clauses randomly that contains variables, say, x, y and z. He sends this clause to one of the provers and asks for the value of x, y, z in the satisfying assignment. He also sends one of the variables x, y and z to the other prover. The first prover has to send a satisfying assignment for that particular clause. However, the other prover does not know which clause has been chosen. Therefore, if there is no satisfying assignment the verifier finds a disagreement between answers (with some probability of error) and rejects. The class of problems for which there exists a *multi-prover interactive proof*, is called MIP [24]. Notice that, in such a proof system there are polynomial many rounds, and also provers cannot communicate during the protocol. By the same idea in the above example, and also using the PCP theorem for reducing the probability of error, it can be shown that MIP = NEXP. At first this result had been shown by Babai, Fortnow, and Lund [15] without using PCP.

1.6 Quantum Computation

In this section, I explain the concepts of the previous section in the quantum setting. I discuss the quantum circuit model, quantum algorithms, and some well-known quantum complexity classes.

1.6.1 Quantum Gates

Any (classical) circuit consists of a sequence of gates which are chosen from a universal set of gates, where each one of these gates is basically a Boolean function over a constant number of bits. For a *quantum circuit* we should replace these Boolean functions with quantum

operations. Thus, a quantum circuit is a sequence of simple quantum operations which are either unitary operators or measurements[17]. However, by the principle of deferred measurement mentioned in Section 1.2.1, we may assume that there is only a single measurement at the end of circuit. Also, since the outcome of circuit is a single bit that shows "accept" or "reject", we may assume that it is the measurement of, say, the first qubit in the standard basis $\{|0\rangle, |1\rangle\}$; Outcomes $|0\rangle$ and $|1\rangle$ mean "reject" and "accept", respectively.

The remaining part is to explain what a universal set of unitary gates is. Unitary gates are unitary operation over a small number of qubits. Let us start by giving some examples. The (classical) NOT gate sends 0 to 1, and 1 to 0. We can extend this definition linearly and get to the following single-qubit unitary matrix

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{1.20}$$

The matrix X is not only unitary but also hermitian, and its eigenvalues are $+1, -1$. Another important unitary gate with these properties is the Z gate

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.21}$$

The group generated by X and Z is $\{\pm I, \pm iI, \pm X, \pm iX, \pm Y, \pm iY, \pm Z, \pm iZ\}$, where

$$Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{1.22}$$

The matrices X, Y and Z are called *Pauli matrices*, and the group generated by them is called *Pauli group*. Note that $X^2 = Y^2 = Z^2 = I$ and $XZ = -ZX = -iY$. Pauli matrices together with identity consist a basis for hermitian matrices over real numbers.

Hadamard gate is another important single-qubit gate which is defined in Eq. (1.2). We have $H^2 = I$ and $HXH^\dagger = Z$. It means that H is in the normalizer of the Pauli group.

The gate $S = Z^{\frac{1}{2}}$, called the *phase gate*, is also in the normalizer of Pauli group

$$S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}. \tag{1.23}$$

Now, let us move to two-qubit gates, and try to define the quantum analogue of XOR gate. Since a quantum gate must be unitary, the number of input and output qubits must be the same. To resolve this problem for defining quantum XOR gate, we can simply keep one of the qubits unchanged and store the XOR of the two qubits in another one: $|x\rangle|y\rangle \to |x\rangle|x + y\rangle$. This gate is called *controlled-NOT gate* since it flips the second qubit if the first one is $|1\rangle$.

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{1.24}$$

$CNOT$ is also in the normalizer of the Pauli group. Indeed, the normalizer of Pauli group, called the *Clifford group*, is generated by H, S and $CNOT$.[18]

[17] Notice that by the characterization of quantum channels in Section 1.2.1 we may assume our quantum operations are either unitary operations or measurements.

[18] Note that by definition the Clifford group contains the Pauli group.

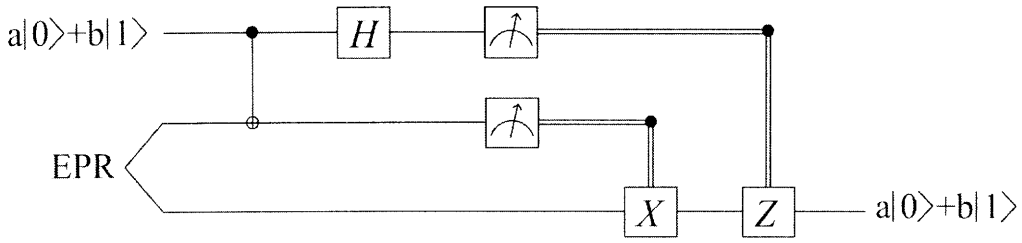

Figure 1-3: This is an example of a 3-qubit circuit. The first qubit is arbitrary, while the next two qubits are initialized in the EPR state. It first applies a $CNOT$ gate with the first qubit being the control qubit, next a Hadamard gate on the first qubit, and then measures the first and second qubits in the standard basis. It applies Z or X gates on the third qubit if the outcome of first or second measurements are 1. It is easy to show that the outcome of this circuit is always the same as the input state. This circuit shows that if two parties share an EPR pair, the first party by sending *two bits* can transmit *one qubit* to the second party. In other words, entanglement allows two parties to send quantum information to each other with only LOCC operation.

1.6.2 Universal Set of Gates

The same as in the classical case, a universal set of quantum gates is a set of unitary gates that can *approximate* any unitary operator. Note that, unlike the classical case, only an approximation of unitary operators is considered here because the set of all unitary matrices is a continuous group and cannot be generated by a finite set.

All the quantum gates that we have introduced are in the Clifford group. Note that, here by Pauli group and Clifford group, we mean the Pauli group over an arbitrary number of qubits and its normalizer. These groups are the group generated by the tensor products of Pauli matrices and its normalizer, which is again generated by the tensor products of $\{H, S, CNOT\}$. However, it is easy to see that, for any number n, the Clifford group over n qubits is finite.[19] Thus, Clifford operations cannot consist a universal set of gates.

It is not hard to prove that any unitary operator (over n qubits) can be written (without approximation) as a product of single-qubit gates and $CNOT$ [45], so if we find a universal set for single-qubit gates, by adding $CNOT$ to it we get a universal set of gates.

The following matrix, called $\pi/8$ gate, generates all the single-qubit gates (up to an approximation) together with H.

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}. \tag{1.25}$$

As a result, $\{H, T, CNOT\}$ is universal [31].

1.6.3 Bounded Error Quantum Polynomial-Time

Now it is clear that a quantum circuit is a sequence of quantum gates, from a universal set of gates (for instance $\{H, T, CNOT\}$), followed by the measurement of the first qubit in the standard basis. Also, an efficient (polynomial-time) quantum algorithm for a language L, is a uniform set of quantum circuits $\{C_n\}$ such that for an n-bit string x, if $x \in L$ ($x \notin L$), the outcome of circuit C_n on input $|x\rangle$ is $|1\rangle$ with probability at least 2/3 (at most 1/3).

[19] By definition Pauli group over n qubits is of size 4^{n+1}. Therefore, the size of Clifford group over n qubits, as the normalizer of Pauli group, is at most the number of permutations over 4^{n+1} objects.

The class of problems for which there exist efficient quantum algorithms is called *bounded error quantum polynomial-time* and denoted BQP [26].

Here, I should remark on a very important difference of quantum and classical circuits. In the circuit of Fig. 1-2 there are three places in which we copy a bit into two wires in order to input it into different gates. This operation is considered as a gate called FANOUT. However, by the no-cloning theorem (Section 1.2.3) FANOUT is not possible in the quantum world, so we should avoid it in the design of quantum circuits.

A quantum computer is as powerful as a classical computer. It means that any computation on a classical computer can be efficiently simulated on a quantum computer. To prove this statement, since every Boolean function can be generated using NAND and FANOUT gates, if we simulate these two gates on a quantum computer, then we can simulate any classical computation on a quantum computer. Notice that, here we should consider only the basis vectors because in a classical circuit we never get to a superposition of them.

NAND and FANOUT gates can be simulated using *Toffoli gate*, which is a 3-qubit gate defined by $|x\rangle|y\rangle|z\rangle \to |x\rangle|y\rangle|z + xy\rangle$ on the basis vectors.[20] If we let $z = 1$, then the third register of the output is the NAND of the first 2 registers. Also, if we let $y = 1$ and $z = 0$, then in the output we have two copies of x, which is a simulation of FANOUT. Therefore, any classical computation can be simulated by Toffoli gate acting on basis vectors. In other words, $\mathsf{BPP} \subseteq \mathsf{BQP}$.

BQP is in EXP because any quantum polynomial-time algorithm consists of polynomially many qubits, polynomially many simple unitary gates, and then a measurement, and all of these operations can be computed classically: the input of circuit is the state $|0\ldots\rangle$, which is an exponential size vector; then the state of the system after applying each gate can be updated by the multiplication of the matrix representation of the gate by the previous vector; at the end the probability of outcome $|1\rangle$ in the measurement can be estimated by computing the reduced density matrix of the first qubit. Since all of these computations can be done in exponential time, we get $\mathsf{BQP} \subseteq \mathsf{EXP}$. By a more careful analysis of this argument and also using some properties of PP, it is proved that $\mathsf{BQP} \subseteq \mathsf{PP}$ [4].

1.6.4 Shor's Factoring Algorithm

In this section, I present a high-level description of *Shor's factoring algorithm* [109, 110], which gives the prime factorization of a given number N, as an example of a quantum algorithm.

Assume that we have found a number x such that $x^2 \equiv 1 \mod N$, but $x \not\equiv \pm 1 \mod N$. Thus, N has non-trivial factors in $x-1$ and $x+1$, and we can break the problem into smaller ones. Therefore, our main problem is to find such a number x with the above properties.

Suppose $1 < a < N$ is such that a and N are coprime. Then, by Euler's theorem there exists r such that $a^r \equiv 1 \mod N$. Assume that r is the smallest number with this property, and also assume that r is even. Hence, if we let $x = a^{r/2}$, $x \not\equiv 1 \mod N$ and $x^2 \equiv 1 \mod N$, and then if x has the extra property that $x \not\equiv -1 \mod N$, we are done. Therefore, the whole algorithm is to choose a random a relatively prime to N, find r, if it is even compute x and then check $x \not\equiv -1 \mod N$, if it does not hold repeat by choosing another a. The point is that a random a with high probability satisfies these properties. Therefore, all steps of this algorithm can be done efficiently on a classical computer except finding r, so we use a quantum computer.

[20] Summation and multiplication are modulo 2.

Let $M = \phi(N)$ where ϕ is the Euler's function[21]. Consider two Hilbert spaces with orthonormal bases $\{|0\rangle, \ldots, |M-1\rangle\}$ and $\{|0\rangle, \ldots, |N-1\rangle\}$. Prepare the following state

$$\frac{1}{\sqrt{M}} \sum_{j=0}^{M-1} |j\rangle|0\rangle,$$

and apply the unitary operation U, which is defined by $U|j\rangle|t\rangle = |j\rangle|t + a^j \mod N\rangle$ on the basis vectors[22]. We get to

$$\frac{1}{\sqrt{M}} \sum_{j=0}^{M-1} |j\rangle|a^j \mod N\rangle = \frac{1}{\sqrt{M}} \sum_{s=0}^{r-1} \sum_{k=0}^{M/r-1} |kr+s\rangle|a^s \mod N\rangle.$$

Now apply the *Fourier transform* on the first registers. This unitary operator is defined by

$$F|j\rangle = \frac{1}{\sqrt{M}} \sum_{l=0}^{M-1} \omega^{jl} |l\rangle$$

on the basis vectors, where $\omega = e^{2\pi i/M}$. We get

$$\frac{1}{M} \sum_{s=0}^{r-1} \sum_{k=0}^{M/r-1} \sum_{l=0}^{M-1} \omega^{(kr+s)l} |l\rangle|a^s \mod N\rangle = \frac{1}{M} \sum_{s=0}^{r-1} \sum_{l=0}^{M-1} \omega^{sl} \Big(\sum_{k=0}^{M/r-1} \omega^{klr} \Big) |l\rangle|a^s \mod N\rangle.$$

Note that, if lr is not a multiple of M, $\sum_k \omega^{klr} = 0$, so we can write the state in the following form

$$\frac{1}{r} \sum_{s=0}^{r-1} \sum_{j=0}^{r-1} \omega^{sjM/r} |jM/r\rangle|a^s \mod N\rangle = \frac{1}{\sqrt{r}} \sum_{j=0}^{r-1} |jM/r\rangle \Big(\frac{1}{\sqrt{r}} \sum_{s=0}^{r-1} \omega^{sjM/r} |a^s \mod N\rangle \Big).$$

Therefore, if we measure the first register in the standard basis we get jM/r, for some $0 \leq j \leq r-1$, each with probability $1/r$. To find r, we can simply repeat this process polynomially many times, and extract r from numbers $j_1 M/r, j_2 M/r, \ldots$.

1.6.5 Quantum Proof Systems

As in the classical case we can define complexity classes based on *quantum proof systems.*

The quantum version of NP (or rather MA) is QMA, *quantum Merlin-Arthur*, defined by Watrous [119]. QMA is the class of problems that can be solved by a bounded-error quantum polynomial-time verifier (called Arthur), given a quantum witness (quantum state) by Merlin. Since any classical computation can be simulated on a quantum computer (see Section 1.6.3), QMA contains NP and MA. Also, PP is a classical upper-bound for QMA [77, 88]. This thesis is mostly about this complexity class and its variants.

[21] Of course we cannot compute M without knowing prime factors of N, but if we let $M = N^2$, all steps of algorithm work up to a small error.

[22] Although U is not a simple gate, it can be generated efficiently using a universal set of gates. It is the case for the next unitary operation called Fourier transform as well.

The quantum version of IP is QIP, *quantum interactive proof* [120]. Its definition is the same as IP except that messages sent by prover and verifier are quantum states, and the verifier is a quantum polynomial-time computer. Clearly, QIP contains IP = PSPACE. Kitaev and Watrous in [78] have proved that without loss of generality we may assume that number of messages passed between prover and verifier is at most 3. Using this result they have shown that QIP $\subseteq$ EXP.

We can go farther and define QMIP, *quantum multi-prover interactive proof system*, the same as MIP except that messages are quantum states, and the verifier is a quantum polynomial-time one [80]. Of course, the same as MIP provers should not send neither classical nor quantum messages to each other during the protocol; however, they may share an entangled state among themselves before that.

Sharing entanglement among provers makes the complexity class QMIP a very complicated one. As we have seen in Fig. 1-3, two far apart parties can apply non-trivial operations using entanglement, but we still do not know how powerful these operations are. We do not know that whether shared entanglement gives more power to provers to cheat the verifier, or whether it allows the verifier to ask harder questions, and then solve harder problems. That is why the power of QMIP is still unknown.

Clearly, if we restrict QMIP so that provers are not allowed to share entanglement, the resulting class contains NEXP = MIP. It was proved that the other direction also holds, and the resulting complexity class is actually equal to NEXP [80].

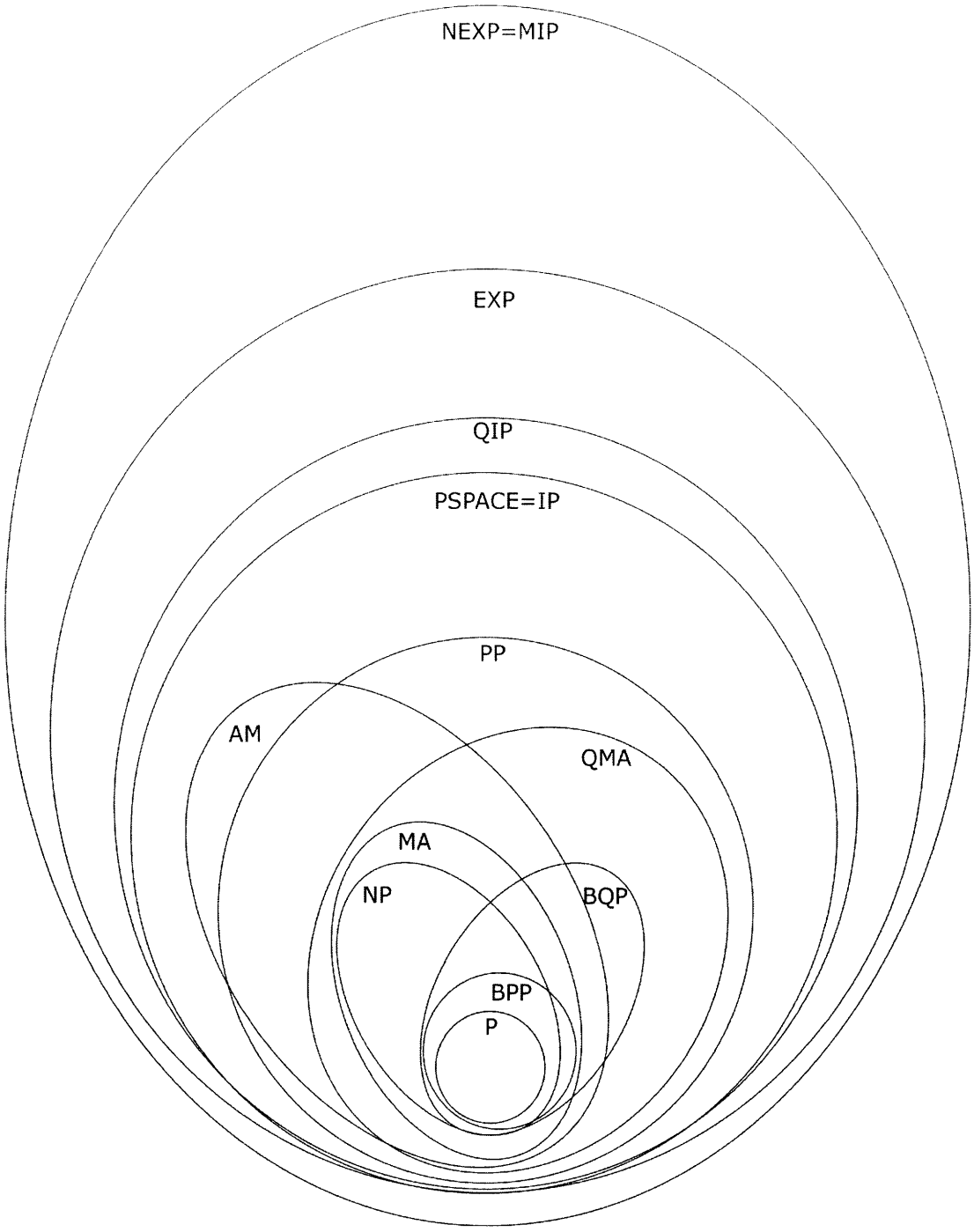

Figure 1-4: Complexity classes that we have learned and their inclusion relations.

Chapter 2

A New QMA-Complete Problem

One of the basic results of computational complexity theory is Cook-Levin theorem, which says that SAT, the problem of whether a Boolean formula has a satisfying assignment or not, is NP-complete. In fact, the Cook-Levin theorem was the beginning of the theory of NP-completeness. After SAT, a series of natural problems in graph theory and combinatorics were shown to be NP-complete as well (see [111]). Hamiltonian cycle, clique problem, graph coloring, subset sum, and vertex cover are examples of such problems. The rich theory of NP-completeness has been used in other parts of complexity theory as well. For instance, the basic ideas of such important results as IP = PSPACE [106], and the PCP theorem [10, 44], are from this theory.

In quantum complexity theory, the complexity class QMA has been defined by Watrous as the quantum analogue of NP [119]. Also, the result by Kitaev [77], who has shown that local Hamiltonian problem is QMA-complete, is considered as the quantum analogue of Cook-Levin theorem. Thus, the same as classical case, one would expect that this theorem is the beginning of a rich theory for QMA-complete problems.

After Kitaev's theorem, few problems have been shown to be QMA-complete. Circuit identity testing [68], consistency of local density matrices [85], and N-representability problem [86] are some examples. However, this class is not as rich as the class of NP-complete problems.

In this chapter, after reviewing some basic properties of the complexity class QMA, I define the quantum clique problem and show that it is QMA-complete. Motivated by this result, I prove that the problem of computing Holevo capacity of quantum channels is NP-complete.

2.1 Quantum Merlin-Arthur Games

2.1.1 Definition of QMA

Definition 2.1.1 *A language L is said to be in* QMA = QMA$(2/3, 1/3)$ *if there exists a quantum polynomial time verifier V such that*

- *Completeness: if $x \in L$, there exists $|\xi\rangle$ such that the probability that V accepts under the input $|x\rangle|\xi\rangle$ is at least $2/3$.*
- *Soundness: if $x \notin L$, for any $|\xi\rangle$, the probability that V accepts under the input $|x\rangle|\xi\rangle$ is at most $1/3$.*

In such a protocol, Merlin is the prover and Arthur is the verifier. Merlin sends a quantum witness to Arthur, and Arthur who is a quantum polynomial-time verifier, decides to whether accept or reject.

The complexity class QMA has been defined by Watrous in [119], in which he has shown that the *group non-membership problem* is in QMA. Group non-membership is the following problem: suppose we are given a group G by a *black box* that multiplies and inverts its elements. Also, we are given elements $g, h_1, \ldots h_k$ of G. The question is that whether g is inside the subgroup generated by $\{h_1, \ldots h_k\}$ or not. This problem is in QMA, but it is not known to be in MA.

2.1.2 Gap Amplification

The completeness and soundness bounds, 2/3 and 1/3, respectively, are not crucial in the definition, and can be replaced by any functions $a(n)$ and $b(n)$ provided that, they are far from 0 and 1 by an inverse exponential gap, and also there is an inverse polynomial gap between them. In other words, if $a(n)$ and $b(n)$ are two functions[1] such that $0 < b(n) < a(n) < 1$, and for some constant c,

$$a(n) < 1 - e^{-n^c}, \quad b(n) > e^{-n^c}, \tag{2.1}$$

and

$$a(n) - b(n) > n^{-c}, \tag{2.2}$$

then $\mathsf{QMA}(a(n), b(n)) = \mathsf{QMA}(2/3, 1/3) = \mathsf{QMA}$ [6]. Here is a simple *gap amplification protocol* to prove this equality: Arthur asks Merlin to send polynomially many copies of witness; then, he applies the verification procedure polynomially many times, and takes the majority vote of all verifications as the output. Therefore, by the Chernoff bound the above equality holds.

Note that *parallel repetition* is a common technique for gap amplification in probabilistic complexity classes. For example, we can amplify the gap in MA exactly by the same idea described above. However, in MA Arthur does not need polynomially many copies of the witness; he can simply copy the witness before each round of verification, and use it for the next round. Thus, we can amplify the gap in MA *without increasing* the size of witness, but in the quantum world by the no-cloning theorem (see Section 1.2.3), Arthur cannot copy the witness, and he needs more copies to apply verification procedure many times.

Although in parallel repetition for gap amplification of QMA we should increase the size of witness, Marriott and Watrous in [88] have introduced another gap amplification protocol for QMA without increasing the size of witness. More precisely, they have proved the following theorem.

Theorem 2.1.1 [88] *Let* $\mathsf{QMA}_{s(n)}(a, b)$ *be the same complexity class as* $\mathsf{QMA}(a, b)$ *except that the witness sent by Merlin is an* $s(n)$*-qubit state. Assume that* a *and* b *satisfy Eqs. (2.1) and (2.2). Then*

$$\mathsf{QMA}_{s(n)}(a, b) = \mathsf{QMA}_{s(n)}(2/3, 1/3).$$

[1] n denotes the size of input string x.

2.1.3 Local Hamiltonian Problem

The first known QMA-complete problem is the local Hamiltonian problem introduced by Kitaev [77].

Definition 2.1.2 *k-Local Hamiltonian problem* $(H_1, \ldots H_s, a, b)$

- *Input: An integer n, functions* $a(n), b(n)$ *such that* $b(n)-a(n) > n^{-c}$ *for some constant c, and polynomially many hermitian positive semidefinite matrices* $H_1, \ldots H_s$ *with infinite-norm at most one*[2]*, such that each of them acts only on k of the n qubits.*
- *Promise: The smallest eigenvalue of* $H_1 + \ldots H_s$ *is either less than* $a(n)$ *or greater than* $b(n)$*.*
- *Output: Decide which one is the case.*

Kitaev has proved that local Hamiltonian problem for $k = 5$ is QMA-complete [77]. Latter this result has been improved in [74] and [70].

Theorem 2.1.2 [70] *2-local Hamiltonian problem is* QMA*-complete.*

2.1.4 Other QMA-Complete Problems

In addition to local Hamiltonian problem, there are a few other problems shown to be QMA-complete. The problem of *consistency of local density matrices* [85], *N-representability problem* [86], and *circuit identity testing* [68] are the only such problems based on my knowledge. Among them I will discuss consistency of local density matrices that can be seen as the dual of local Hamiltonian problem.

Suppose we have an n-qubit system, and we are given a collection of local density matrices $\rho_{c_1}, \ldots, \rho_{c_m}$, where each ρ_{c_i} describes a subset c_i of the qubits. We say that the states $\rho_{c_1}, \ldots, \rho_{c_m}$ are "consistent" if there exists some state $\sigma_{1\ldots n}$ (on all qubits) that matches each of the ρ_{c_i} on the subset c_i, i. e. $\sigma_{c_i} = \mathrm{Tr}_{1\ldots n \setminus c_i} \sigma_{1\ldots n} = \rho_{c_i}$. The problem is to decide whether $\rho_{C_1}, \ldots, \rho_{C_m}$ are consistent or not. Liu [85] has shown that this problem is QMA-complete.

2.1.5 QMA $\subseteq$ PP

In Section 1.6.5 I mentioned that QMA is a subset of PP (an unpublished result by Kitaev and Watrous). Here, I can present a proof idea based on [88].

Consider a QMA protocol in which Merlin sends an $s(n)$-qubit state $|\psi\rangle$ to Arthur. By Theorem 2.1.1 we can amplify the gap, and assume that the probability of error is $2^{-2s(n)}$, while the size of witness is still $s(n)$. Now note that Arthur's verification procedure can be formulated as preparing some ancilla qubits, applying some unitary operator which depends on the input x, and then measurement of the first qubit. Thus, the probability of acceptance by Arthur is equal to

$$\mathrm{Tr}[(|1\rangle\langle 1| \otimes I) \cdot (U_x|\psi\rangle\langle\psi| \otimes |0\ldots 0\rangle\langle 0\ldots 0|U_x^\dagger)] = \langle\psi|H_x|\psi\rangle,$$

where $H_x = \langle 0\ldots 0|U_x^\dagger(|1\rangle\langle 1|\otimes I)U_x|0\ldots 0\rangle$.. It means that x is a yes-instant if the maximum eigenvalue of H_x is greater than $1 - 2^{-2s(n)}$, and is a no-instant if all eigenvalues of H_x are

[2]The maximum eigenvalue of H_i is at most 1.

less than $2^{-s(n)}$. Now note that, H_x is a square matrix of size $2^{s(n)}$ and then has $2^{s(n)}$ eigenvalues; therefore, in the no-case $\mathrm{Tr}\, H_x \leq 2^{s(n)}2^{-2s(n)} = 2^{-s(n)}$, while in the yes-case $\mathrm{Tr}\, H_x \geq 1 - 2^{-2s(n)}$. Hence, if we could estimate $\mathrm{Tr}\, H_x$, we would be able to solve the problem. It is well-known in the literature that such a estimation is possible in PP (see [117]). Thus QMA $\subseteq$ PP.

2.1.6 QMA_1

QMA_1 is the same complexity class as QMA except that the completeness bound is 1: L is a language in QMA_1 if for any $x \in L$, Arthur always accepts, and if $x \notin L$ Arthur accept with probability at most 1/3. Of course, the same as QMA, we can amplify the gap, and then the soundness gap is not important (it should be far from 1 by an inverse polynomial).

Here I should point out that in the implementation of a quantum algorithm we usually assume that we are able to apply quantum gates up to an approximation; therefore, any quantum algorithm contains some probability of error anyway, and then the definition of QMA_1 is meaningless. However, this problem can be resolve by emphasizing that the verifier can implement all 3-qubit quantum gates, *exactly*. It means that, by a quantum verifier for QMA_1 protocols we mean the one that has all 3-qubit quantum gates in hand. Bravyi has pointed out this assumption in [35] and has proved the following lemma.

Lemma 2.1.1 [35] *Let U be a unitary operator acting on k qubits. Then U can be exactly represented by a quantum circuit of size poly$(k)2^{2k}$ with three-qubit gates.*

Bravyi also has introduced a complete problem for QMA_1 called *quantum* k-SAT. It is basically the same as k-local Hamiltonian problem, except that all terms in the local Hamiltonian are projections, and the question is that whether there is a vector orthogonal to the support of all projections or not. Bravyi in [35] has shown that quantum 4-SAT is QMA_1-complete[3]. This result has been improved by Nagaj and Mozes [94].

2.1.7 QCMA

QCMA, quantum-classical Merlin-Arthur, is another invariant of QMA in which the message sent by Merlin is a classical message, but Arthur remains a quantum polynomial-time verifier [6].

By definition, QCMA is in QMA, but we do not know whether this inclusion is strict or not; however, there is an oracle separation between them, [3].

Two QCMA-complete problems has been found in [121]. Also, Brandão in [32] has shown that *gapped local Hamiltonian* is QCMA-complete. Gapped local Hamiltonian is the same problem as local Hamiltonian with the extra assumption that there is an inverse polynomial gap between the first and second eigenvalues of the local Hamiltonian.

2.2 Zero-Error Channel Capacity and Clique Problem

2.2.1 Zero-Error Channel Capacity

A classical discrete memoryless channel consists of an input set X, an output set Y, and probability distributions $p(y|x)$, for every $x \in X$ and $y \in Y$, meaning that if we send x

[3]He also, in the same paper, has shown that quantum 2-SAT can be solve efficiently on a classical computer.

through channel, we get y as output with probability $p(y|x)$. Since we want to define the zero-error capacity of this channel, the exact value of $p(y|x)$ is not important for us, but whether it is zero or not. Hence, to get a clearer representation we correspond to the channel a graph G on the vertex set X in which two vertices $x, x' \in X$ are adjacent if there exists $y \in Y$ such that $p(y|x), p(y|x')$ are both non-zero. In other words, x, x' are adjacent in G if they can be confused after passing through channel. Therefore, messages $x_1, \ldots x_k$ can be sent through channel with no error iff there is no edges between them, i. e. $\{x_1, \ldots x_k\}$ is an independent set.

Definition 2.2.1 *In a graph G, a subset of vertices no two of which are adjacent is called an independent set. $\alpha(G)$ denotes the maximum size of an independent set in G.*

By the above discussion, if we want to code our messages in words of length one (one use of channel), the best way is to code them in an independent set of maximum size. In this case, we get to the rate $\alpha(G)$ for transmitting information. However, we may use words of length two to code the messages, so we get to another graph denoted $G \otimes G$.

Definition 2.2.2 *Assume G and H are two graphs on vertex sets V and U, respectively. Their tensor product $G \otimes H$ is a graph on the vertex set $V \times U$ such that (v_1, u_1) and (v_2, u_2) are adjacent if v_1, v_2 are either equal or adjacent in G, and also u_1, u_2 are either equal or adjacent in H.*

It is not hard to see that the graph corresponding to words of length two is $G \otimes G$. Thus, the best way to code the messages in words of length two, is to use an independent set in $G \otimes G$ of size $\alpha(G \otimes G)$, so we get to the rate $\alpha(G \otimes G)^{1/2}$ for transmitting information[4]. Repeating this argument for longer input lengths, we get to the following Theorem due to Shannon [107].

Theorem 2.2.1 [107] *$\Theta(G)$, the zero-error capacity of a channel with the corresponding graph G, is equal to*

$$\Theta(G) = \lim_{n \to \infty} \alpha(G^{\otimes n})^{\frac{1}{n}}.$$

These definitions can all be generalized for quantum channels [89, 90]. The zero-error capacity of a quantum channel is the maximum rate of classical information that can be sent through a quantum channel without using entanglement. To get a closed-form expression for this quantity, suppose Φ is a quantum channel, and we code k messages in quantum states $\rho_1, \ldots \rho_k$. If we want to decode the outputs of channel with no error, we should be able to distinguish states $\Phi(\rho_1), \ldots \Phi(\rho_k)$ without error. It is easy to see that some quantum states can be distinguished by a measurement with no error, iff they have orthogonal supports.

Definition 2.2.3 *For a quantum channel Φ, $\alpha(\Phi)$ is the maximum number of quantum states $\rho_1, \ldots \rho_k$ such that $\Phi(\rho_1), \ldots \Phi(\rho_k)$ have orthogonal supports.*

Also, $\alpha(\Phi^{\otimes n})$ is the maximum number of product states $\rho_{i1} \otimes \cdots \otimes \rho_{in}$, $i = 1, \ldots k$, such that all states $\Phi^{\otimes n}(\rho_{i1} \otimes \cdots \otimes \rho_{in})$, $i = 1, \ldots k$, have orthogonal supports.

Here I should remark on two points. First, if $\Phi(\rho)$ and $\Phi(\rho')$ have orthogonal supports, then ρ and ρ' have orthogonal supports as well. Therefore, $\alpha(\Phi)$ is at most equal to the

[4]Square root is for normalization.

dimension of input states, and then is finite. Second, I emphasize that the input states of $\Phi^{\otimes n}$ should be product states since we do not want to use entanglement. Here, by abuse of notation we denote the maximum number of such product states by $\alpha(\Phi^{\otimes n})$.

By the above definition it is clear what the zero-error capacity of a quantum channel should be.

Proposition 2.2.1 *$\Theta(\Phi)$, the zero-error capacity of the quantum channel Φ, is*

$$\Theta(\Phi) = \lim_{n\to\infty} \alpha(\Phi^{\otimes n})^{\frac{1}{n}}.$$

2.2.2 Computing $\alpha(\Phi)$

Suppose $\alpha(\Phi) = n$, and $\rho_1, \ldots \rho_n$ are n states such that the supports of $\Phi(\rho_1), \ldots \Phi(\rho_n)$ are orthogonal. For $i = 1, \ldots n$, let $|\psi_i\rangle$ be a pure state in the support of ρ_i. Then, since the support of $\Phi(|\psi_i\rangle)$ is a subspace of the support of $\Phi(\rho_i)$, the states $\Phi(|\psi_1\rangle), \ldots \Phi(|\psi_n\rangle)$ have orthogonal supports as well. It means that, to compute $\alpha(\Phi)$ it suffices to restrict ourselves to pure states.

Now assume that the operator sum representation of Φ is

$$\Phi(\rho) = \sum_{k=1}^{r} E_k \rho E_k^\dagger, \tag{2.3}$$

where $\sum_{k=1}^{r} E_k^\dagger E_k = I$. Thus, the support of $\Phi(|\psi_i\rangle)$ is spanned by vectors $E_1|\psi_i\rangle, \ldots E_r|\psi_i\rangle$. It means that, $\Phi(|\psi_1\rangle), \ldots \Phi(|\psi_n\rangle)$ have orthogonal supports, iff these vectors, for different indices i and j, are orthogonal. Summarizing these two statements, we get to the following proposition[5].

Proposition 2.2.2 *For a quantum channel Φ with operator sum representation (2.3), we have $\alpha(\Phi) \geq n$, if and only if there exist pure states $|\psi_1\rangle, \ldots |\psi_n\rangle$ such that $\langle\psi_i|E_k^\dagger E_l|\psi_j\rangle = 0$, for every k, l and i, j, where $i \neq j$.*

2.2.3 Quantum Clique Problem

Deciding whether a given graph has a clique of size k is NP-complete. Considering this problem in the complement graph we find that deciding whether $\alpha(G) \geq k$ is NP-complete. Based on our notation, it means that having a classical channel, deciding whether we can get to the rate k for transmitting information with zero-error by coding messages in words of length one, is NP-complete. Since we have all these notions for the quantum case we can define the *quantum clique problem.*

Basically, the quantum version of clique problem is also to decide whether $\alpha(\Phi) \geq k$, for a given quantum channel Φ, or not. It is equivalent to decide whether there exist quantum states $\rho^1, \ldots \rho^k$ such that $\Phi(\rho^1), \ldots \Phi(\rho^k)$ have orthogonal supports or not.

For any two states σ^1, σ^2, we have $\mathrm{Tr}(\sigma^1\sigma^2) \geq 0$, and equality holds if and only if σ^1, σ^2 have orthogonal supports. Let $\sigma^{12} = \sigma^1 \otimes \sigma^2$. Thus $\mathrm{Tr}(\sigma^1\sigma^2) = \mathrm{Tr}(S\,\sigma^{12})$, where S is the swap gate defined by

$$S|\psi\rangle|\phi\rangle = |\phi\rangle|\psi\rangle.$$

[5]In Appendix A more results can be found regarding the computation of $\alpha(\Phi)$, and its relation to the zero-error capacity of graphs.

Therefore, by applying the swap gate we can estimate $\mathrm{Tr}(\sigma^1\sigma^2)$. However, if σ^{12} is not separable, this equality does not hold and the orthogonality of σ^1 and σ^2 is not implied by $\mathrm{Tr}(S\,\sigma^{12}) = 0$. To resolve this problem we can restrict ourselves to *entanglement breaking channels* to ensure that the output states of the channel are not entangled.

A quantum channel Φ is called entanglement breaking if there is a POVM $\{M_i\}$ and quantum states $\{\sigma_i\}$ such that

$$\Phi(\rho) = \sum_i tr(M_i\rho)\sigma_i.$$

In this case, $\Phi^{\otimes 2}(\rho^{12})$ is always separable. Also, $\mathrm{Tr}(S\,\Phi^{\otimes 2}(\rho^{12})) \geq 0$ and equality implies $\Phi(\rho^1)$ and $\Phi(\rho^2)$ are orthogonal.

Definition 2.2.4 *Quantum clique problem* (Φ, k, a, b)

- *Input: Integer numbers n and k, non-negative real numbers a_n, b_n with an inverse polynomial gap $b_n - a_n > n^{-c}$, and an entanglement breaking channel Φ that acts on n-qubit states.*
- *Promise: Either there exists $\rho^1 \otimes \cdots \otimes \rho^k$ such that $\sum_{i,j} Tr(S\,\Phi(\rho^i) \otimes \Phi(\rho^j)) \leq a_n$ or for any state $\rho^{12\ldots k}$ we have $\sum_{i,j} Tr(S\,\Phi^{\otimes 2}(\rho^{i,j})) \geq b_n$.*
- *Output: Decide which one is the case.*

2.3 Quantum Clique Problem is QMA-Complete

2.3.1 SWAP Test

To prove that quantum clique is in QMA we should use a well-known protocol called SWAP test.

SWAP test is a protocol for deciding whether two given quantum states are the same or not. The protocol is as follows: given two states $|\psi_1\rangle$ and $|\psi_2\rangle$, prepare an ancilla qubit $|0\rangle$. Apply the Hadamard gate on the ancilla, and then the controlled-swap gate on the two registers[6], and again, Hadamard on the ancilla. At the end, measure the ancilla qubit in the computational basis. It is easy to see that this protocol computes the channel

$$\Phi_{swap}(|\psi_1\rangle|\psi_2\rangle) = \frac{1}{2}(1 + |\langle\psi_1|\psi_2\rangle|^2)|0\rangle\langle 0| + \frac{1}{2}(1 - |\langle\psi_1|\psi_2\rangle|^2)|1\rangle\langle 1|. \tag{2.4}$$

In fact, in the measurement we get $|0\rangle$ with probability $\frac{1}{2}(1 + |\langle\psi_1|\psi_2\rangle|^2)$, and $|1\rangle$ with probability $\frac{1}{2}(1 - |\langle\psi_1|\psi_2\rangle|^2)$. Therefore, if we correspond the output $|0\rangle$ to $+1$ and output $|1\rangle$ to -1, the expected value of this number is equal to $|\langle\psi_1|\psi_2\rangle|^2$. In general, when the input state is σ^{12} we can compute $\mathrm{Tr}(S\,\sigma^{12})$, where S is the swap gate.

2.3.2 Main Theorem

Here is the main theorem of this section.

[6]Controlled-swap gate sends $|0\rangle|\psi_1\rangle|\psi_2\rangle$ to itself, and $|1\rangle|\psi_1\rangle|\psi_2\rangle$ to $|1\rangle|\psi_2\rangle|\psi_1\rangle$.

Theorem 2.3.1 *Quantum clique problem (Φ, k, a, b), where Φ is an entanglement breaking channel on n-qubit states and has the operator sum representation*

$$\Phi(\rho) = \sum_{i=1}^{r} E_i \rho E_i^\dagger, \tag{2.5}$$

where $\sum_i E_i^\dagger E_i = I$ and $r = poly(n)$, is QMA*-complete.*

Proof: First, we show that (Φ, k, a, b) is in QMA. Note that, Φ can be written as $\Phi(\rho) = \mathrm{Tr}_2\big(U\rho \otimes |1\rangle\langle 1| U^\dagger\big)$, where U is a unitary operator and

$$U|\psi\rangle|1\rangle = \sum_{i=1}^{r} E_i |\psi\rangle |i\rangle. \tag{2.6}$$

Since $r = poly(n)$, a polynomial time verifier can implement U and then Φ, with arbitrary small error. Therefore, given witness $\rho^{1\ldots k}$, verifier can randomly choose i, j, $1 \leq i, j \leq k$, compute $\Phi^{\otimes 2}(\rho^{i,j})$, and then apply the SWAP test on the outcome. As we mentioned in Section 2.3.1, the expected value of the outcome of SWAP test for fixed i, j, is $\mathrm{Tr}\big(S\,\Phi^{\otimes 2}(\rho^{i,j})\big)$, and for random choices of i, j is equal to

$$\frac{1}{\binom{k}{2}} \sum_{i,j} \mathrm{Tr}\big(S\,\Phi^{\otimes 2}(\rho^{i,j})\big),$$

which is either less than $\frac{2}{k(k-1)}a$ or greater than $\frac{2}{k(k-1)}b$. Hence, there is an inverse polynomial gap between them and the verifier can recognize them in polynomial time. Thus, quantum clique problem is in QMA.

To prove the hardness, we establish a polynomial time reduction from local Hamiltonian problem to quantum clique. Let $(H_1, \ldots H_s, a, b)$ be an instance of the local Hamiltonian problem. Since $H_i \leq I$, we have $\frac{1}{s}H \leq I$, where $H = H_1 + \ldots H_s$, so $M = I - \frac{1}{s}H$ is a positive operator and we can define the following quantum channel

$$\Phi(\rho) = \frac{1}{s}\mathrm{Tr}(H \otimes I\,\rho)|00\rangle\langle 00| + \mathrm{Tr}\big(M \otimes |0\rangle\langle 0|\,\rho\big)|11\rangle\langle 11| + \mathrm{Tr}\big(M \otimes |1\rangle\langle 1|\,\rho\big)|10\rangle\langle 10|.$$

Note that, $s = poly(n)$ and then Φ is of the form of Eq. (2.5). Now consider $(\Phi, k = 2, \frac{1}{s^2}a^2, \frac{1}{s^2}b^2)$ as an instance of quantum clique problem. We prove that $(H_1, \ldots H_s, a, b)$ is a yes-instance of local Hamiltonian if and only if $(\Phi, k = 2, \frac{1}{s^2}a^2, \frac{1}{s^2}b^2)$ is a yes-instance of quantum clique.

Suppose $(H_1, \ldots H_s, a, b)$ is a no-instance. Thus, for any state σ, $\mathrm{Tr}(H\sigma) \geq b$, and then, for any state ρ^{12} we have

$$\mathrm{Tr}\big(S\,\Phi^{\otimes 2}(\rho^{1,2})\big) \geq \frac{1}{s^2}\mathrm{Tr}(H \otimes I\,\rho^1)\mathrm{Tr}(H \otimes I\,\rho^2) \geq \frac{1}{s^2}b^2.$$

It means that $(\Phi, k = 2, \frac{1}{s^2}a^2, \frac{1}{s^2}b^2)$ is also a no-instance. Now, assume that there is $|\psi\rangle$ such that $\langle\psi|H|\psi\rangle \leq a$. Let $\rho^1 = |\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|$, and $\rho^2 = |\psi\rangle\langle\psi| \otimes |1\rangle\langle 1|$. Hence, $\mathrm{Tr}\big(S\,\Phi(\rho^1) \otimes \Phi(\rho^2)\big)$ is equal to

$$\mathrm{Tr}\Big((\frac{1}{s}\langle\psi|H|\psi\rangle|00\rangle\langle 00| + \langle\psi|M|\psi\rangle|11\rangle\langle 11|)(\frac{1}{s}\langle\psi|H|\psi\rangle|00\rangle\langle 00| + \langle\psi|M|\psi\rangle|10\rangle\langle 10|)\Big)$$

$$= \frac{1}{s^2}\langle\psi|H|\psi\rangle^2 \leq \frac{1}{s^2}a^2.$$

Therefore, $(\Phi, k = 2, \frac{1}{s^2}a^2, \frac{1}{s^2}b^2)$ is also a yes-instance. We are done.

□

2.4 Complexity of Computing Holevo Capacity

We proved that computing the zero-error capacity of a quantum channel is a QMA-complete problem. In this section we consider the same problem for the capacity of channels with arbitrary small error.

In the classical case, there is an algorithm called the *Arimoto-Blahut algorithm* that given a classical discrete memoryless channel, computes its capacity [9, 27]. Indeed, computing the capacity of a classical channel involves maximization of some mutual information. In the Arimoto-Blahut algorithm this maximization problem is converted to an alternating maximization one, that tends to the channel capacity and is more tractable. Using the same idea, Nagaoka in [95] proposed the same algorithm for computing the capacity of quantum channels. However, it does not work because in the quantum case there can be a local maximum which is not a global one, so in the quantum Arimoto-Blahut algorithm the alternate maximum value may tend to a local maximum, and not to the channel capacity.

In this section we prove that computing the capacity of a quantum channel, even for entanglement breaking ones, is NP-complete.

2.4.1 Holevo Capacity

The *Holevo capacity of a quantum channel* is the maximum rate of classical information that can be sent through a quantum channel without using entanglement [59]. $\chi(\Phi)$, the Holevo capacity of the quantum channel Φ, is equal to

$$\chi(\Phi) = \max_{\{p_i\},\{\rho_i\}} S(\sum_i p_i\Phi(\rho_i)) - \sum_i p_i S(\Phi(\rho_i)), \tag{2.7}$$

where $S(\rho) = -tr(\rho\log\rho)$ denotes the *von Neumann* entropy, and the maximum is taken over all probability distributions $\{p_i\}$ and quantum states $\{\rho_i\}$. Using the convexity of von Neumann entropy, we can assume that states ρ_i are pure. Also, if Φ acts on an n-dimensional Hilbert space, we may assume that the number of ρ_i's is at most n^2. However, these are not enough information on what the maximum point is, and how we can compute $\chi(\Phi)$.

2.4.2 Minimum Output Entropy

Minimum output entropy of a quantum channel is equal to the minimum entropy of its output states

$$\min_{\rho} S(\Phi(\rho)). \tag{2.8}$$

Again, using the convexity of von Neumann entropy, the minimum is achieved on pure states. The minimum output entropy is an important invariant of quantum channels. Indeed, it has been proved that the famous conjecture of the additivity of Holevo capacity[7] is equivalent to the additivity of minimum output entropy[8], [108]. This result is important for us because it somehow expresses the minimum output entropy in terms of Holevo capacity, and using this idea we can convert the problem of computing the minimum output entropy to the problem of computing Holevo capacity. Indeed, to prove that computing Holevo capacity is NP-complete, we first state the NP-completeness of computing the minimum output entropy.

2.4.3 Main Theorem

Here is the main theorem of this section.

Theorem 2.4.1 *Suppose Φ is a quantum channel that acts on an n-dimensional Hilbert space, and is given by poly(n) number of bits. Also, let c be a real number. Then deciding whether $\chi(\Phi) > c$, is* NP*-complete.*

To prove this theorem we show that this problem is "harder" than the problem of computing the minimum output entropy of quantum channels, and then prove computing the minimum output entropy is NP-complete. In fact, the minimum output entropy of quantum channels is a more tractable quantity than the Holevo capacity, and then proving the NP-completeness of this problem is simpler.

Theorem 2.4.2 *Assume that Φ is a quantum channel acting on an n-dimensional Hilbert space, and is given by polynomially many bits. Also, let c be a real number. Then deciding whether the minimum output entropy of Φ is less than c, is* NP*-complete.*

Using Theorem 2.4.2 we first prove Theorem 2.4.1, and then show the NP-hardness of computing minimum output entropy in the next section.

Proof of Theorem 2.4.1: First of all if Φ is a channel and $\chi(\Phi) \geq c$, then there are probability distribution $\{p_i\}$ and states $\rho_1, \ldots \rho_s$ such that

$$S(\sum_i p_i \Phi(\rho_i)) - \sum_i p_i S(\Phi(\rho_i)) \geq c. \tag{2.9}$$

Note that, we may assume $s \leq n^2$; therefore, given the probability distribution and the quantum states $\rho_1, \ldots \rho_s$, the verifier can check whether (2.9) holds or not, and then this is a problem in NP.

To prove the hardness, since by Theorem 2.4.2 computing the minimum output entropy is NP-complete, if we establish a reduction from the problem of computing the minimum output entropy to computing Holevo capacity, we are done.

Let (Φ, c) be an instance of minimum output entropy problem as in Theorem 2.4.2. Let $|1\rangle, \ldots |n\rangle$ be an orthonormal basis for the Hilbert space, and also let $X_0, \ldots X_{n^2-1}$ be the n-dimensional generalized Pauli matrices

$$X_{mn+d} = T^m R^d,$$

[7] Recently, Hastings has proved that the additivity conjecture is false [58].

[8] Additivity of minimum output entropy has been appeared in [76].

where $T|j\rangle = |j+1 \mod n\rangle$ and $R|j\rangle = e^{2ij\pi/n}|j\rangle$. Define the channel Ψ by

$$\Psi(\rho \otimes |i\rangle\langle i|) = X_i \Phi(\rho) X_i^\dagger.$$

It is clear that

$$\chi(\Psi) = \max_{p_i,\rho_i} S\Big(\sum_i p_i \Psi(\rho_i)\Big) - \sum_i p_i S(\Psi(\rho_i)) \leq \log n - \min_\rho S(\Psi(\rho)). \tag{2.10}$$

Also, it is easy to see that the minimum output entropy of Ψ is equal to the minimum output entropy of Φ. On the other hand, if the minimum output entropy of Φ is taken on $|\phi\rangle$, and we let $\rho_i = |\phi\rangle\langle\phi| \otimes |i\rangle\langle i|$ and $p_i = 1/n$, for $i = 1, \ldots n$, then equality holds in Eq. (2.10). It means that, the minimum output entropy of Φ is less than c if and only if the Holevo capacity of Ψ is greater than $\log n - c$. We are done.

□

2.5 Complexity of Computing Minimum Output Entropy

The only remaining step is the proof of Theorem 2.4.2. To get a clearer proof it would be helpful to first state some lemmas.

2.5.1 Some Lemmas

In this section we study some properties of the points that a channel achieves its minimum output entropy. Before stating the lemmas, remember that the von Neumann entropy is convex, and then the minimum output entropy of a channel is attained on pure states.

Lemma 2.5.1 *Suppose $\Phi_1, \ldots \Phi_k$ are k channels with the same input and output state spaces. Also, assume that output states of every two of them are orthogonal. In other words, for any i, j, $1 \leq i < j \leq k$, and any states ρ, ρ',*

$$Tr(\Phi_i(\rho)\Phi_j(\rho')) = 0.$$

Then

$$\min_\rho S\Big(\sum_{i=1}^k p_i \Phi_i(\rho)\Big) \geq \sum_{i=1}^k p_i \min_\rho S(\Phi_i(\rho)) + H(p_1, \ldots p_k),$$

where, $\{p_1, \ldots p_k\}$ is a probability distribution and $H(p_1, \ldots p_k)$ is its entropy. In particular, the minimum output entropy of $\sum_{i=1}^k p_i \Phi_i$ is at least $H(p_1, \ldots p_k)$, and equality holds iff there is ρ such that all states $\Phi_i(\rho)$ are pure.

Proof: Since $\Phi_i(\rho)$'s have orthogonal supports

$$\begin{aligned}
\min_\rho S\big(\sum_{i=1}^k p_i\Phi_i(\rho)\big) &= \min_\rho \sum_i -\mathrm{Tr}\big(p_i\Phi_i(\rho)\log(p_i\Phi_i(\rho))\big) \\
&= \min_\rho \sum_i -p_i\,\mathrm{Tr}\big(\Phi_i(\rho)\log\Phi_i(\rho) + \log p_i\Phi_i(\rho)\big) \\
&= \min_\rho \sum_i p_i S(\Phi_i(\rho)) - \sum_i p_i\log p_i \\
&\geq \sum_{i=1}^k p_i \min_\rho S(\Phi_i(\rho)) + H(p_1,\ldots p_k).
\end{aligned}$$

□

Lemma 2.5.2 *Let Φ_{trace} be the channel that acts on the Hilbert space $\mathcal{H}\otimes\mathcal{H}$, and traces out the second register:*

$$\Phi_{trace}(\rho^{12}) = Tr_2(\rho^{12}) = \rho^1. \tag{2.11}$$

Then the minimum output entropy of Φ_{trace} is zero, and it is achieved at the product states $|\psi_1\rangle|\psi_2\rangle$.

Proof: Let $|\psi_{12}\rangle$ be a pure state in $\mathcal{H}\otimes\mathcal{H}$. By the *Schmidt decomposition* [96], there are orthonormal bases $\{|i\rangle\}$, $\{|i'\rangle\}$, and real non-negative numbers λ_i, such that

$$|\psi_{12}\rangle = \sum_i \lambda_i|i\rangle|i'\rangle. \tag{2.12}$$

Hence, $\Phi(|\psi_{12}\rangle) = \sum_i \lambda_i^2|i\rangle\langle i|$, and it is a pure state if only if only one of λ_i's is non-zero, or equivalently $|\psi_{12}\rangle$ is a product state.

□

The next lemma is on the minimum output entropy of the SWAP test, described in Section 2.3.1.

Lemma 2.5.3 *Let Φ_{swap} be the channel defined in Eq. (2.4). Then the minimum output entropy of the channel*

$$\Phi(\rho) = \frac{1}{2}\Phi_{trace}(\rho)\otimes|u\rangle\langle u|\otimes|000\rangle\langle 000| + \frac{1}{2}|u'_{12}\rangle\langle u'_{12}|\otimes|10\rangle\langle 10|\otimes\Phi_{swap}(\rho),$$

where $|u\rangle\in\mathcal{H}$ and $|u'_{12}\rangle\in\mathcal{H}\otimes\mathcal{H}$ are arbitrary states, is equal to $H(2)=1$, and is attained at the pure states of the form $|\psi\rangle|\psi\rangle$.

Proof: By Lemma 2.5.1, it is sufficient to show that states of form $|\psi\rangle|\psi\rangle$ are the only states ρ such that $\Phi_{trace}(\rho)$ and $\Phi_{swap}(\rho)$ are simultaneously pure.

Using Lemma 2.5.2, such a state ρ should be a product state $|\psi_1\rangle|\psi_2\rangle$. On the other hand, by Eq. (2.4), it is clear that $\Phi_{swap}(|\psi_1\rangle|\psi_2\rangle)$ is pure iff $|\langle\psi_1|\psi_2\rangle| = 1$, or equivalently $|\psi_1\rangle = |\psi_2\rangle$.

□

For the next lemma, it is helpful to fix some notations. Let $\mathcal{H}$ be an n-dimensional Hilbert space with the orthonormal basis $\{|1\rangle, \ldots |n\rangle\}$. For any $1 \leq i < j \leq n$, let Π_{ij} be the projection over $\frac{1}{\sqrt{2}}(|i\rangle + |j\rangle)$,

$$\Pi_{ij} = \frac{1}{2}(|i\rangle + |j\rangle)(\langle i| + \langle j|).$$

Also, let Π'_{ij} be the projection over $\frac{1}{\sqrt{2}}(|i\rangle - |j\rangle)$,

$$\Pi'_{ij} = \frac{1}{2}(|i\rangle - |j\rangle)(\langle i| - \langle j|).$$

Thus, $\sum_{ij} \Pi_{ij} \otimes \Pi'_{ij}$ is a hermitian matrix and its eigenvalues are at most $\binom{n}{2}$. Therefore,

$$M = I \otimes I - \frac{1}{n(n-1)} \sum_{ij} \Pi_{ij} \otimes \Pi'_{ij},$$

is a positive semidefinite matrix, and does not have zero eigenvalue. It means that, $|v'_{12}\rangle$ is always in the support of the following channel.

$$\Phi_{cube}(\rho) = \frac{1}{n(n-1)} \sum_{ij} \mathrm{Tr}\big(\Pi_{ij} \otimes \Pi'_{ij}\, \rho\big)|v_{12}\rangle\langle v_{12}| + \mathrm{Tr}\big(M\rho\big)|v'_{12}\rangle\langle v'_{12}|. \tag{2.13}$$

Lemma 2.5.4 *Let $|v_{12}\rangle, |v'_{12}\rangle \in \mathcal{H} \otimes \mathcal{H}$ be two orthogonal states, and define Φ_{cube} as in Eq. (2.13). Then the minimum output entropy of channel*

$$\begin{aligned}\Phi(\rho) \;=\; & \frac{1}{3}\Phi_{trace}(\rho) \otimes |u\rangle\langle u| \otimes |000\rangle\langle 000| + \frac{1}{3}|u'_{12}\rangle\langle u'_{12}| \otimes |10\rangle\langle 10| \otimes \Phi_{swap}(\rho) \\ & + \frac{1}{3}\Phi_{cube}(\rho) \otimes |110\rangle\langle 110|\end{aligned}$$

is equal to $H(3) = \log 3$ and is attained at the states $|\psi_{12}\rangle = |\psi\rangle|\psi\rangle$, where

$$|\psi\rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i |i\rangle, \tag{2.14}$$

and $x_i \in \{+1, -1\}$.

Proof: Again, using Lemma 2.5.1, it suffices to show that the only states ρ such that $\Phi_{trace}(\rho)$, $\Phi_{swap}(\rho)$ and $\Phi_{cube}(\rho)$ are pure, are the states $|\psi\rangle|\psi\rangle$ where $|\psi\rangle$ is of the form of Eq. (2.14).

In Lemma 2.5.3 we showed that if $\Phi_{trace}(\rho)$ and $\Phi_{swap}(\rho)$ are pure, then ρ is a pure state of the form $|\psi\rangle|\psi\rangle$, so it remains to show that if $\Phi_{cube}(|\psi\rangle|\psi\rangle)$ is pure, then $|\psi\rangle$ is of the form of Eq. (2.14).

As we mentioned, $|v'_{12}\rangle$ is always in the support of $\Phi_{cube}(\rho)$. Hence, if $\Phi_{cube}(|\psi\rangle|\psi\rangle)$ is pure, then it is equal to $|v'_{12}\rangle$. It means that, $\Phi_{cube}(\rho)$ is pure if and only if

$$\mathrm{Tr}\big(\Pi_{ij} \otimes \Pi'_{ij}\, \rho\big) = 0,$$

for any i, j. Let $|\psi\rangle = \sum_i \lambda_i |i\rangle$, and suppose $\Phi_{cube}(|\psi\rangle|\psi\rangle)$ is pure. We have

$$0 = \langle\psi|\langle\psi|\Pi_{ij} \otimes \Pi'_{ij}|\psi\rangle|\psi\rangle = \langle\psi|\Pi_{ij}|\psi\rangle\langle\psi|\Pi'_{ij}|\psi\rangle = \frac{1}{4}|\lambda_i + \lambda_j|^2|\lambda_i - \lambda_j|^2.$$

In other words, for any i, j, either $\lambda_i = \lambda_j$ or $\lambda_i = -\lambda_j$, so $|\psi\rangle$ should be one of the states in Eq. (2.14).

□

2.5.2 Proof of Theorem 2.4.2

To prove the NP-hardness of the problem of computing the minimum output entropy, we should find a reduction from an NP-complete problem to this one. The most convenient such problem for us is the 2-out-of-4-SAT problem [75]. We can formulate this problem as follows: given $m = poly(n)$ vectors of the form

$$|A_k\rangle = \sum_{i=1}^{n} a_i^k |i\rangle,$$

where for each k, $1 \leq k \leq m$, there are a constant number of non-zero a_i^k, decide whether there exists a vector $|\psi\rangle$ of the form of Eq. (2.14) orthogonal to all $|A_k\rangle$'s, $\langle A_k|\psi\rangle = 0$[9].

Now we are ready to prove the theorem. Given a witness state ρ, we can check whether $S(\Phi(\rho)) < c$, in polynomial time. Therefore, this problem is in NP.

To prove the hardness, let $|A_1\rangle, \ldots |A_m\rangle$ be an instance of 2-out-of-4-SAT. Let

$$H = \frac{1}{m}\sum_{k=1}^{m} |A_k\rangle\langle A_k| \otimes |A_k\rangle\langle A_k|,$$

and define

$$\Phi_H(\rho) = \frac{1}{2}\mathrm{Tr}(H\rho)|w_{12}\rangle\langle w_{12}| + \mathrm{Tr}\big((I \otimes I - \frac{1}{2}H)\,\rho\big)|w'_{12}\rangle\langle w'_{12}|,$$

where $|w_{12}\rangle$ and $|w'_{12}\rangle$ are two orthogonal states in $\mathcal{H} \otimes \mathcal{H}$. Since $H \leq I$, $|w'_{12}\rangle$ is always in the support of Φ_H, so the minimum output entropy of Φ_H is zero, and is achieved at the states $|\psi_{12}\rangle$ that are orthogonal to all $|A_k\rangle|A_k\rangle$, $k = 1, \ldots, m$.

Define the channel

$$\begin{aligned}\Phi(\rho) \;=\; & \frac{1}{4}\Phi_{trace}(\rho) \otimes |u\rangle\langle u| \otimes |000\rangle\langle 000| + \frac{1}{4}|u'_{12}\rangle\langle u'_{12}| \otimes |10\rangle\langle 10| \otimes \Phi_{swap}(\rho) \\ & + \frac{1}{4}\Phi_{cube}(\rho) \otimes |110\rangle\langle 110| + \frac{1}{4}\Phi_H(\rho) \otimes |111\rangle\langle 111|.\end{aligned}$$

By Lemma 2.5.1, the minimum output entropy of Φ is at least $H(4) = 2$, and equality holds if there exists $|\psi_{12}\rangle$ for which all the states $\Phi_{trace}(|\psi_{12}\rangle)$, $\Phi_{swap}(|\psi_{12}\rangle)$, $\Phi_{cube}(|\psi_{12}\rangle)$ and $\Phi_H(|\psi_{12}\rangle)$ are pure. By Lemma 2.5.4 such a state should be of the form $|\psi_{12}\rangle = |\psi\rangle|\psi\rangle$, where $|\psi\rangle$ is of the form of Eq. (2.14). Also, for this state $\Phi_H(|\psi\rangle|\psi\rangle)$ is pure iff $\langle\psi|A_k\rangle = 0$, $k = 1, \ldots, m$.

[9] 2-out-of-4-SAT is the same as 3-SAT except that each clause contains 4 variables, and it is satisfied if exactly 2 of those variables are true. If we represent each variable by a number which is either +1 or −1, then a clause is satisfied if the sum of its variables is equal to 0. This condition is the orthogonality constraint.

Therefore, the minimum output entropy of Φ is $H(4)$, if and only if $(|A_1\rangle, \ldots |A_m\rangle)$ is a yes-instance of 2-out-of-4-SAT problem. Notice that, using the integrality of the problem, there exists $\epsilon > 1/poly(n)$, such that $(|A_1\rangle, \ldots |A_m\rangle)$ is a yes-instance, if and only if the minimum output entropy of Φ is less than $2 + \epsilon$. We are done.

2.5.3 Restriction to Entanglement Breaking Channels

We proved that computing the minimum output entropy, and then, Holevo capacity are NP-complete. In these two theorems, we considered general quantum channels, but one may expect that if we restrict ourselves to a special class of quantum channels, then we get to simpler problems.

For example, let Φ be a *classical-quantum channel* (c-q channel) of the form

$$\Phi(\rho) = \sum_{i=1}^{n} \langle i|\rho|i\rangle \sigma_i, \tag{2.15}$$

where $|1\rangle, \ldots |n\rangle$ is an orthonormal basis and $\sigma_1, \ldots \sigma_n$ are arbitrary states. Then, obviously, the minimum output entropy of Φ is equal to

$$\min_i S(\sigma_i),$$

and can be computed in polynomial time. Also, it is easy to see that computing the Holevo capacity of Φ is a convex optimization problem and can be solved efficiently.

Therefore, to get a non-trivial problem we should consider a more general class of quantum channels. Indeed, c-q channels that we considered in Eq. (2.15) are special cases of Entanglement breaking channels. An entanglement breaking channel is a channel Φ of the form

$$\Phi(\rho) = \sum_{i=1}^{r} \mathrm{Tr}(M_i \rho)\sigma_i, \tag{2.16}$$

where $\{M_i\}$ is a POVM and $\sigma_1, \ldots \sigma_r$ are arbitrary states. Although, it seems that the problem of computing the minimum output entropy and Holevo capacity of entanglement breaking channels, is simpler than the general case, we prove that these are also NP-complete.

Theorem 2.5.1 *Assume Φ is an entanglement breaking channel of the form (2.16) acting on an n-dimensional Hilbert space, and is given by polynomially many bits. Also let c be a real number. Then the problems of bounding the Holevo capacity and the minimum output entropy of Φ,*

$$\chi(\Phi) > c,$$

and

$$\min_\rho S(\Phi(\rho)) < c,$$

are NP-complete.

If we show that computing the minimum output entropy for entanglement breaking channels is NP-hard, then by the same argument as in the proof of Theorem 2.4.1, we can prove the hardness of computing the Holevo capacity as well. Also, recall that, in the proof of Theorem 2.4.2 all the channels that we used, are entanglement breaking except Φ_{trace}.

Therefore, if we replace Φ_{trace} with an entanglement breaking channel that captures the same properties, we are done.

The key idea is the following observation first appeared in [34]. Suppose ρ is the density matrix of a two-qubit state. Let $\sigma_0 = I, \sigma_1, \sigma_2, \sigma_4$ be the Pauli matrices (see Section 1.6.1). Also, for $i = 1, 2, 3$ let $P_i^{\pm} = \frac{1}{2}(I \pm \sigma_i)$ be density matrices of the $+1$ and -1 eigenstates of σ_i. For $0 \leq i, j \leq 3$ define $c_{ij} = \mathrm{Tr}(\sigma_i \otimes \sigma_j \rho)$. Then, we have

$$\rho = \frac{1}{4} \sum_{i,j=0}^{3} c_{ij}\, \sigma_i \otimes \sigma_j.$$

If we rewrite this equation in terms of $P_i^{\pm}$, we get

$$\begin{aligned} \rho = \frac{1}{4} \sum_{i,j=1}^{3} \quad & (\tfrac{1}{9} + \tfrac{1}{3}c_{i0} + \tfrac{1}{3}c_{0j} + c_{ij}) P_i^+ \otimes P_j^+ \\ & + (\tfrac{1}{9} - \tfrac{1}{3}c_{i0} + \tfrac{1}{3}c_{0j} - c_{ij}) P_i^- \otimes P_j^+ \\ & + (\tfrac{1}{9} + \tfrac{1}{3}c_{i0} - \tfrac{1}{3}c_{0j} - c_{ij}) P_i^+ \otimes P_j^- \\ & + (\tfrac{1}{9} - \tfrac{1}{3}c_{i0} - \tfrac{1}{3}c_{0j} + c_{ij}) P_i^- \otimes P_j^-. \end{aligned}$$

Suppose all the coefficients in this expression are non-negative. Then

$$\begin{aligned} \mathrm{Tr}_2(\rho) = \frac{1}{4} \sum_{i,j=1}^{3} \quad & (\tfrac{1}{9} + \tfrac{1}{3}c_{i0} + \tfrac{1}{3}c_{0j} + c_{ij}) P_i^+ \\ & + (\tfrac{1}{9} - \tfrac{1}{3}c_{i0} + \tfrac{1}{3}c_{0j} - c_{ij}) P_i^- \\ & + (\tfrac{1}{9} + \tfrac{1}{3}c_{i0} - \tfrac{1}{3}c_{0j} - c_{ij}) P_i^+ \\ & + (\tfrac{1}{9} - \tfrac{1}{3}c_{i0} - \tfrac{1}{3}c_{0j} + c_{ij}) P_i^-. \end{aligned}$$

In other words, $\mathrm{Tr}_2(\rho)$ can be written as a linear combination of states $P_i^{\pm}$ with coefficients of the form $\mathrm{Tr}(M\rho)$, where $\{M\}$ is some POVM.

It means that, if the coefficients were always non-negative, then $\rho \mapsto \mathrm{Tr}_2(\rho)$ was an entanglement breaking channel. To satisfy this extra assumption we can replace ρ with $\rho_\epsilon = 1/4(1-\epsilon)I + \epsilon\rho$, where $0 < \epsilon < 1/16$, and observe that the coefficients for ρ_ϵ are all non-negative. In general, we have the following lemma, proved in [34].

Lemma 2.5.5 [34] *Let ρ be a state in $\mathcal{H} \otimes \mathcal{H}$, where $\mathcal{H}$ is an n-dimensional Hilbert space. Also, let $1/n^2 I$ be the maximally mixed state in $\mathcal{H} \otimes \mathcal{H}$ and $0 < \epsilon < 1/n^2$. Then, $1/n^2(1-\epsilon)I + \epsilon\rho$ is a separable state. As a consequence,*

$$\Phi'_{trace}(\rho) = Tr_2(1/n^2(1-\epsilon)I + \epsilon\rho) \tag{2.17}$$

is an entanglement breaking channel.

Using this lemma, the proof of Theorem 2.5.1 follows immediately.

Proof of Theorem 2.5.1: All steps of the proof are the same as in Theorem 2.4.2, except that we replace the channel Φ_{trace} with Φ'_{trace}, which is an entanglement breaking one. The only property that we should check is that the minimum output entropy of Φ'_{trace} is achieved at product states. It holds because $\mathrm{Tr}_2(1/n^2(1-\epsilon)I + \epsilon\rho) = 1/n(1-\epsilon)I + \epsilon\mathrm{Tr}_2(\rho)$, and $S(\mathrm{Tr}_2(1/n^2(1-\epsilon)I + \epsilon\rho))$ achieve their minimum for a given ρ, if and only if $S(\mathrm{Tr}_2(\rho))$ takes its minimum at ρ.

□

2.6 Summary

In this chapter I have proved that the quantum clique problem is QMA-complete. This is obtained by considering an NP-complete problem, and somehow translating it into the language of quantum information theory. The key point is that clique problem in graphs can be stated in terms of zero-error capacity, so this translation is straightforward. It is interesting to consider other NP-hard problems and try to define the corresponding QMA-hard problem. Notice that, this idea is first captured in the QMA-completeness of local Hamiltonian problem.

I have also considered the problem of computing the Holevo capacity, and then the minimum output entropy of a quantum channel, and have proved that they are NP-complete. Since, there are a few results on the computational complexity of invariants of quantum channels, it would be a natural question to consider the complexity of other such quantities for channels as well as quantum states.

Chapter 3

Multiple-Merlin-Arthur Games

In a Merlin-Arthur game suppose there are more than one Merlin, so that each Merlin sends his witness to Arthur, and Arthur in polynomial time decides to whether accept or reject. In the classical case, there is no advantage in having multiple Merlins since only one Merlin can send all the messages. However, In the quantum case, messages sent by different Merlins are separable, and if we ask one Merlin to send all of them, he may cheat by sending an entangled state. Therefore, the complexity class multiple-Merlin-Arthur may be different from QMA. In this section I define the complexity class QMA(k), describe a gap amplification protocol for that, and give some results regarding its relation to QMA.

3.1 QMA(k) and its Basic Properties

3.1.1 Definition

The complexity class QMA(k) is first defined by Kobayashi, Matsumoto, and Yamakami [81].

Definition 3.1.1 *Let k be an integer and $0 < b < a < 1$. Then the complexity class* QMA(k, a, b) *consists of languages L for which there exists a polynomial-time quantum verifier V such that for all inputs $x \in \{0,1\}^n$:*

(i) If $x \in L$, there exist witnesses $|\psi_1\rangle, \ldots, |\psi_k\rangle$ on polynomially many qubits, such that V accepts with probability at least a given $|x\rangle \otimes |\psi_1\rangle \otimes \cdots \otimes |\psi_k\rangle$.

(ii) If $x \notin L$, V accepts with probability at most b given $|x\rangle \otimes |\psi_1\rangle \otimes \cdots \otimes |\psi_k\rangle$, for all $|\psi_1\rangle, \ldots, |\psi_k\rangle$.

As a convention, we denote QMA = QMA(1).

Notice that the difference between QMA and QMA(k), for $k \geq 2$, is that in QMA(k) there are k Merlins and the state sent by them is *separable*.

3.1.2 Three Basic Questions

In Chapter 2 we saw that we can amplify the gap in QMA, there are complete problems for this complexity class, and also there is a non-trivial classical upper bound for that. To

understand QMA(k) we should check the same properties for $k \geq 2$. Indeed, there are three main questions regarding QMA(k).

First, is it possible to amplify the gap in QMA(k) protocols? Recall that in QMA = QMA(1) we can amplify the gap by both parallel repetition and also strong amplification (Theorem 2.1.1).

Second, is more than one Merlin ever helpful? In other words, maybe, the same as in the classical case, all QMA(k) complexity classes be the same. Note that, by definition we have QMA(1) $\subseteq$ QMA(2) $\subseteq \cdots$. Now the question is that, does there exist k for which QMA($k+1$) = QMA(k)?

Third, NEXP is an upper bound for QMA(k) because the prover can send the classical description of witnesses and the verifier can simulate a quantum polynomial time algorithm in exponential-time. Now the question is that, can we improve this upper bound?

As we will see in this chapter, unlike for the case $k = 1$, all of these questions are non-trivial for $k \geq 2$.

3.1.3 Pure State N-Representability Problem

A partial answer to the question of whether QMA is the same as QMA(2), has been given by Liu, Christandl, and Verstraete in [86]. They have shown a problem called *pure state N-representability* is in QMA(2), while it is not known to be in QMA. We can state this problem in terms of the problem of consistency of local density matrices mentioned in Section 2.1.4.

Suppose we have local density matrices $\rho_{c_1}, \ldots, \rho_{c_m}$, and want to find a state σ consistent with all of these local matrices, such σ is *pure*. We know that if we remove the purity assumption, we can solve this problem in QMA. However, purity of a given state cannot be checked in QMA. Instead, if we have two copies of σ by applying the SWAP test (see Section 2.3.1) we can check whether σ is pure or not.

This observation is an evidence on QMA(2) being different from QMA.

3.1.4 Quantum Clique Problem for General Channels

In the previous chapter we saw that computing $\alpha(\Phi)$, the maximum number of states that can be distinguished with no error after passing through channel, for an entanglement breaking channel is in QMA. Remember, we should restrict the set of quantum channels to entanglement breaking ones because we need the output states of channel, acting on entangled states, to be unentangled. However, if we assume that the given states by Merlin are not entangled, then we do not need to restrict the set of channels. In other words, the problem of whether $\alpha(\Phi) \geq k$, for an arbitrary quantum channel Φ, is in QMA(k).

This observation is another evidence on the fact that QMA(k), $k \geq 2$, and QMA are different.

3.2 Gap Amplification Implies QMA(2) = QMA(k), for $k \geq 3$

Kobayashi, Matsumoto and Yamakami in [82] have shown that the question of whether QMA($k+1$) = QMA(k), and gap amplification are related. Indeed, they proved that if we could amplify the gap in QMA(k), then QMA(k)-hierarchy collapses to QMA(2). In this section I present a stronger result. I show that if we could amplify the gap only in QMA(2), then QMA(k) = QMA(2), for any $k \geq 2$.

3.2.1 Some Lemmas

Before getting to the main results we need to review some lemmas that will be needed.

Recall that for two mixed states ρ and σ, their trace distance is denoted by $\|\rho - \sigma\|_{\mathrm{Tr}}$ (see Section 1.3.1). We say σ is *ε-close* to ρ if $\|\rho - \sigma\|_{\mathrm{Tr}} \leq \varepsilon$, and *$\varepsilon$-far* otherwise. The following lemma is a direct consequence of Eq. 1.9.

Lemma 3.2.1 *Suppose σ is ε-close to ρ. Then any measurement that accepts ρ with probability p, accepts σ with probability at most $p + \varepsilon$.*

□

Lemma 3.2.2 *Given a k-partite state $\rho^{A_1 A_2 \cdots A_k}$, suppose there are k states $|\psi_1\rangle, \ldots, |\psi_k\rangle$ such that $\langle \psi_i | \rho^{A_i} | \psi_i \rangle \geq 1 - \varepsilon_i$, for all i. Let $|\Psi\rangle := |\psi_1\rangle \otimes \cdots \otimes |\psi_k\rangle$ and $\varepsilon := \varepsilon_1 + \cdots + \varepsilon_k$. Then $\langle \Psi | \rho^{A_1 A_2 \cdots A_k} | \Psi \rangle \geq 1 - \varepsilon$.*

Proof: We can assume without loss of generality that $|\psi_i\rangle = |0\rangle$ for all i. Then each ρ^{A_i}, when measured in the standard basis, yields the outcome $|0\rangle$ with probability at least $1 - \varepsilon_i$. By the union bound, it follows that $\rho^{A_1 A_2 \cdots A_k}$, when measured in the standard basis, yields the outcome $|\Psi\rangle = |0\rangle^{\otimes k}$ with probability at least $1 - \varepsilon$; hence, $\langle \Psi | \rho^{A_1 A_2 \cdots A_k} | \Psi \rangle \geq 1 - \varepsilon$.

□

The following lemma is a direct consequence of the definition of SWAP test in Section 2.3.1.

Lemma 3.2.3 *Suppose $\langle \psi | \rho | \psi \rangle < 1 - \varepsilon$ for all pure states $|\psi\rangle$. Then, the SWAP test between ρ and any other state rejects with probability at least $\varepsilon/2$.*

Proof: Choose a basis that diagonalizes ρ, so that $\rho = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_N)$ where $\lambda_1, \ldots, \lambda_N$ are ρ's eigenvalues. By assumption, $\lambda_i < 1 - \varepsilon$, for every i, so given any mixed state σ, a SWAP test between ρ and σ accepts with probability

$$\frac{1 + \mathrm{Tr}\,(\rho\sigma)}{2} = \frac{1}{2} + \frac{1}{2} \sum_{i=1}^{N} \lambda_i \sigma_{ii} < \frac{1}{2} + \frac{1 - \varepsilon}{2} \sum_{i=1}^{N} \sigma_{ii} = 1 - \frac{\varepsilon}{2}.$$

□

3.2.2 QMA(2) Contains All QMA(k)-Hierarchy

In this section I prove that for any k and a, b, there are a', b' such that $\mathsf{QMA}(k, a, b) \subseteq \mathsf{QMA}(2, a', b')$. Before proving this result it is helpful to state a lemma.

Lemma 3.2.4 $\mathsf{QMA}\,(k, a, b) \subseteq \mathsf{QMA}\left(k, 1 - 2^{-p(n)}, 1 - (a - b)\right)$ *for all k, all $b < a < 1$, and all polynomials $p(n)$.*

Proof: We use the following protocol. Each Merlin provides $m = C \cdot \frac{p(n)}{(a-b)^2}$ registers for some constant C. Then Arthur runs his verification procedure m times in parallel, once with each k-tuple of registers, and accepts if and only if at least a d fraction of invocations accept, for some d slightly less than a.

To show completeness, we use a Chernoff bound. Assuming the Merlins are honest, each one simply provides m copies of his witness. Then on each invocation, Arthur accepts with

independent probability at least a. Thus, assuming we chose a sufficiently large constant C, the probability that Arthur accepts less than dm times is at most $2^{-p(n)}$.

To show soundness, we use Markov's inequality. The expected number of accepting invocations is at most bm (by linearity of expectation, this is true even if the registers are entangled). Hence, the probability that this number exceeds dm is at most b/d, which we can ensure is less than $1-(a-b)$ by choosing $d \in \left(\frac{b}{1-(a-b)}, a\right)$ (note that such a d must exist by the assumption $b < a < 1$).

□

Theorem 3.2.1 $\mathsf{QMA}(k,a,b) \subseteq \mathsf{QMA}\left(2, 1-2^{-n}, 1-\frac{(a-b)^2}{8k}\right)$.

Proof: We will show that for all k and all $\delta = \Omega(1/\operatorname{poly}(n))$,

$$\mathsf{QMA}\left(k, 1-2^{-n}, 1-\delta\right) \subseteq \mathsf{QMA}\left(2, 1-2^{-n}, 1-\frac{\delta^2}{8k}\right).$$

This will suffice to prove the theorem since Lemma 3.2.4 implies that for all k and all a, b, we have $\mathsf{QMA}(k,a,b) \subseteq \mathsf{QMA}(k, 1-2^{-n}, 1-(a-b))$.

Our protocol is as follows. Merlin_A and Merlin_B send k-partite states $\rho^{A_1 A_2 \cdots A_k}$ and $\rho^{B_1 B_2 \cdots B_k}$ respectively. Given these states, Arthur performs one of the following two tests, each with probability 1/2:

(1) Choose $1 \leq i \leq k$ uniformly at random, perform a SWAP test between ρ^{A_i} and ρ^{B_i}, and accept if and only if the SWAP test accepts.

(2) Simulate the $\mathsf{QMA}(k, 1-2^{-n}, 1-\delta)$ protocol, using $\rho^{A_1 A_2 \cdots A_k}$ in place of the k witness registers.

We first show completeness of the above protocol. If the Merlins are honest, they can both simply send k unentangled accepting witnesses for the $\mathsf{QMA}(k)$ protocol being simulated. In that case, step (1) accepts with probability 1, while step (2) accepts with probability at least $1-2^{-n}$.

We now show soundness. Suppose any set of unentangled witnesses causes the $\mathsf{QMA}(k)$ protocol to reject with probability at least δ. Then we need to show that any pair of witnesses $\rho^{A_1 A_2 \cdots A_k}$ and $\rho^{B_1 B_2 \cdots B_k}$ causes the $\mathsf{QMA}(2)$ protocol to reject with probability at least $\frac{\delta^2}{8k}$. We consider two cases.

First suppose $\rho^{A_1 A_2 \cdots A_k}$ is ε-close in trace distance to some separable pure state $|\Psi\rangle$. Then by Lemma 3.2.1, step (2) rejects with probability at least $\delta - \varepsilon$.

Next suppose $\rho^{A_1 A_2 \cdots A_k}$ is ε-far in trace distance from any separable pure state. Then by Eq. 1.12, we have $\langle \Psi | \rho^{A_1 A_2 \cdots A_k} | \Psi \rangle < 1 - \varepsilon^2$ for all separable pure states $|\Psi\rangle$. Thus, taking the contrapositive of Lemma 3.2.2, for all pure states $|\psi_1\rangle, \ldots, |\psi_k\rangle$ we have

$$\sum_{i=1}^{k} \left(1 - \langle \psi_i | \rho^{A_i} | \psi_i \rangle\right) > \varepsilon^2.$$

Hence, step (1) rejects with probability greater than $\frac{\varepsilon^2}{2k}$ by Lemma 3.2.3. Setting $\varepsilon = 3\delta/4$, we thus find that the protocol rejects with probability at least $\frac{1}{2}\max\left\{\frac{\delta}{4}, \frac{(3\delta/4)^2}{2k}\right\} \geq \frac{\delta^2}{8k}$.

□

This theorem shows that any QMA(k) protocol can be simulated in QMA(2), with a smaller gap. Therefore, if we could amplify the gap in QMA(2), we basically have two complexity classes QMA(1) and QMA(2).

3.3 Gap Amplification in QMA(2)

3.3.1 Parallel Amplification

Let us try to apply the usual parallel amplification idea on QMA(2).

Arthur asks Merlin_A and Merlin_B to send polynomially many copies of witnesses. Denote these two states $\rho^{A_1 \dots A_{q(n)}}$ and $\rho^{B_1 \dots B_{q(n)}}$. In each round, Arthur chooses random $1 \leq j, k \leq q(n)$, that are not chosen yet, and applies the verification procedure on $\rho^{A_j B_k}$. Of course, there is no problem for the first round; however, after that, the two registers $A_1 \dots A_{q(n)}$ and $B_1 \dots B_{q(n)}$ become entangled, and cannot be used for the second round. On the other hand, intuitively, the amount of induced entanglement is comparable to the number of qubits in A_j and B_k. Therefore, if $q(n)$ is large compared to this number, the amount of entanglement between randomly chosen registers $A_{j'}, B_{k'}$ for the second round, is small. More precisely, if the produced entanglement is distributed between all pairs $A_{j'}, B_{k'}$, then a randomly chosen $\rho^{A_{j'} B_{k'}}$ is almost separable. Thus, in the second round the probability of acceptance is almost the same as in the first round. By the same argument we can repeat verification procedure, say, $\sqrt{q(n)}$ times and take the majority vote, so by Chernoff bound the probability of error decreases exponentially.

The analysis of our amplification protocol involves showing that Arthur cannot create *too much* entanglement during his verification procedure. To make this precise, we need some way to measure the entanglement of mixed states. Based on the argument, such a measure of entanglement (denoted $E(\cdot)$) should satisfy some properties.

(1) If $E(\rho^{AB}) \leq \varepsilon$, then ρ^{AB} is δ-close to a separable state, for some δ that tends to 0 if ε tends to 0.

(2) If Φ is a quantum operation acting on n qubits, then $E(\Phi(\rho^A \otimes \rho^B)) \leq cn$, for some constant c.

(3) $E\left(\rho^{A_1 A_2 \cdots A_k, B_1 B_2 \cdots B_k}\right) \geq \frac{c}{k} \sum_{i,j=1}^{k} E(\rho^{A_i B_j})$, for some constant c.

Our choice for such an entanglement measure is entanglement of formation[1] defined in Section 1.4.2.

3.3.2 Properties of Entanglement of Formation

Recall that for a bipartite state ρ^{AB}, the entanglement of formation $E_F(\rho^{AB})$ is the minimum of $\sum_i p_i S(\text{Tr}_A |\psi_i\rangle\langle\psi_i|)$ over all decompositions $\rho^{AB} = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. We are going to check the three properties mentioned in the previous section for entanglement of formation.

[1] One my consider other entanglement measures; for example, the *squashed entanglement* E_{sq} defined by Christandl and Winter [39] might be a choice since it is known to be superadditive, and then satisfies the third property. However, the trouble with E_{sq} is that it badly violates the second property: there exist $n \times n$-dimensional bipartite states ρ^{AB} such that $E_{sq}(\rho^{AB}) = O\left(\frac{\log n}{n}\right)$, yet ρ^{AB} has trace distance $\Omega(1)$ to any separable state. This is why we cannot use squashed entanglement here, and must instead use entanglement of formation.

By definition, E_F is convex: for all ρ^{AB} and σ^{AB},

$$E_F\left(p\rho^{AB}+(1-p)\,\sigma^{AB}\right)\leq pE_F\left(\rho^{AB}\right)+(1-p)\,E_F\left(\sigma^{AB}\right).$$

It is also easy to see that $E_F\left(\rho^{AB}\right)=0$ if and only if ρ^{AB} is separable. To check the first property in Section 3.3.1 we need a stronger version of this statement.

Lemma 3.3.1 *Suppose $E_F(\rho^{AB})\leq\varepsilon$. Then there exists a separable state that is $\sqrt{2\varepsilon}$-close to ρ^{AB} in trace distance.*

Proof: Let $S\left(\rho||\sigma\right)$ be the quantum relative entropy between mixed states ρ and σ (see Section 1.4.5). Vedral and Plenio in [114] showed that

$$E_F(\rho^{AB})\geq\min S\left(\rho^{AB}||\sigma^{AB}\right),$$

where the minimum is taken over all separable states σ^{AB}. Also, it is known (see Klauck et al. [79] and Ohya and Petz [98]) that

$$S(\rho^{AB}||\sigma^{AB})\geq\frac{1}{2}\left\|\rho^{AB}-\sigma^{AB}\right\|_{\mathrm{Tr}}^{2}.$$

Putting these results together, if $E_F(\rho^{AB})\leq\varepsilon$, there exists a separable state σ^{AB} such that $S\left(\rho^{AB}||\sigma^{AB}\right)\leq\varepsilon$, and hence $\left\|\rho^{AB}-\sigma^{AB}\right\|_{\mathrm{Tr}}\leq\sqrt{2\varepsilon}$.

□

The next lemma proves the second property.

Lemma 3.3.2 *Let ρ^{AB} be a separable state, and suppose σ^{AB} is obtained from ρ^{AB} by applying an arbitrary entangled measurement on at most n qubits from each register, and then possibly conditioning on the outcome. Then $E_F\left(\sigma^{AB}\right)\leq 2n$.*

Proof: By convexity, we can assume without loss of generality that ρ^{AB} is a pure state, $|\psi_A\rangle\otimes|\psi_B\rangle$. Thus, we can write σ^A as $\Phi\left(|\psi_A\rangle\langle\psi_A|\right)/\left\|\Phi\left(|\psi_A\rangle\langle\psi_A|\right)\right\|$, where Φ is some non-trace-increasing operator acting on at most n qubits. In the operator-sum representation we have,

$$\sigma^A=\frac{\sum_{i=1}^{M}E_i\,|\psi_A\rangle\langle\psi_A|\,E_i^\dagger}{\mathrm{Tr}\sum_{i=1}^{M}E_i\,|\psi_A\rangle\langle\psi_A|\,E_i^\dagger}$$

where $\sum_{i=1}^{2^{2n}}E_i^\dagger E_i\leq I$ and $M\leq 2^{2n}$. We then have

$$\begin{aligned}E_F\left(\sigma^{AB}\right)&\leq S\left(\sigma^A\right)\\&=S\left(\frac{\sum_{i=1}^{M}E_i\,|\psi_A\rangle\langle\psi_A|\,E_i^\dagger}{\mathrm{Tr}\sum_{i=1}^{M}E_i\,|\psi_A\rangle\langle\psi_A|\,E_i^\dagger}\right)\\&\leq\log_2 M\\&\leq 2n\end{aligned}$$

where the first line follows from the concavity of the von Neumann entropy.

□

Given an entanglement measure E, we call E *superadditive* if for any state $\rho^{AA',BB'}$ on four registers,

$$E(\rho^{AA',BB'})\geq E\left(\rho^{AB}\right)+E(\rho^{A'B'}).$$

It is clear that if E is superadditive, then it satisfies the third property of Section 3.3.1. It had been a conjecture for a while that entanglement of formation is superadditive, but in a spectacular recent development, Hastings [58] has shown that this conjecture is false. (More precisely, Hastings shows a failure of additivity for the minimum output entropy of a quantum channel. By a result of Shor [108], this is equivalent to the superadditivity of entanglement of formation.) However, the violation of additivity found by Hastings is extremely small, and is perfectly consistent with additivity being true in a weaker or asymptotic sense. Indeed, the third property may still be true.

We call an entanglement measure E *weakly superadditive* if it satisfies the relation

$$E\left(\rho^{A_1A_2\cdots A_k,B_1B_2\cdots B_k}\right) \geq \frac{c}{k}\sum_{i,j=1}^{k} E(\rho^{A_iB_j}),$$

for some constant c independent of k. Weak superadditivity is, in particular, implied by the following inequality:

$$E(\rho^{AA',BB'}) \geq \frac{1}{2}\left[E\left(\rho^{AB}\right) + E(\rho^{AB'}) + E(\rho^{A'B}) + E(\rho^{A'B'})\right]$$

which in turn is implied by ordinary superadditivity. Then we conjecture the following:

Conjecture 3.3.1 (Weak Additivity Conjecture) *E_F is weakly superadditive.*

3.3.3 Weak Additivity Conjecture Implies Gap Amplification

Theorem 3.3.1 *Assume the Weak Additivity Conjecture holds. Then*

$$\mathsf{QMA}\left(2,a,b\right) = \mathsf{QMA}\left(2,1-2^{-p(n)},2^{-p(n)}\right)$$

for all $a-b=\Omega\left(1/\operatorname{poly}\left(n\right)\right)$ and all polynomials $p(n)$.

Proof: Let L be a language in $\mathsf{QMA}\left(2,a,b\right)$. Let Q be Arthur's verification algorithm in the original $\mathsf{QMA}\left(2,a,b\right)$ protocol, and let the original Merlins' messages have $r\left(n\right)$ qubits each for some polynomial r. Also, let $T\left(n\right)$ be a number of repetitions of Q that suffices to amplify it to error probability $2^{-p(n)}$, assuming no entanglement among Merlin_A's or Merlin_B's registers. By Chernoff bound, we can take $T\left(n\right) := C\cdot p\left(n\right)/\left(a-b\right)^2$ for some constant C.

Our amplified protocol is the following.

(1) Arthur asks Merlin_A and Merlin_B to supply $q\left(n\right)$ copies each of their respective witnesses, where $q\left(n\right) := C'\cdot T\left(n\right)r\left(n\right)/\left(a-b\right)^2$ for some constant C'. Denote by $\rho^{A_1A_2\cdots A_{q(n)}}$ and $\rho^{B_1B_2\cdots B_{q(n)}}$ the states on $q\left(n\right)r\left(n\right)$ qubits that Arthur actually receives.

(2) For all $t := 1$ to $T\left(n\right)$, Arthur chooses registers A_j and B_k uniformly and independently from among those not already chosen, and runs Q on the state $\rho^{A_jB_k}$.

(3) Arthur accepts if at least $\frac{a+b}{2}T\left(n\right)$ of the $T\left(n\right)$ invocations of Q accepted, and rejects otherwise.

Completeness: If the Merlins are honest, they can simply send $|\psi_A\rangle^{\otimes q(n)}$ and $|\psi_B\rangle^{\otimes q(n)}$ respectively, where $|\psi_A\rangle \otimes |\psi_B\rangle$ is a witness that Q accepts with probability at least a. Then by assumption, Arthur will accept with probability at least $1 - 2^{-p(n)}$.

Soundness: Our central claim is the following: *At every one of the $T(n)$ iterations, Arthur can be considered to be running Q on a bipartite state ρ^{AB} that is ε-close to a separable state, where $\varepsilon = O\left(\sqrt{T(n) r(n) / q(n)}\right)$.*

Let us first see why soundness follows from the above claim. Suppose $x \notin L$. Then Q accepts every separable state with probability at most b. By Lemma 3.2.1, then, Q also accepts every state that is ε-close to separable with probability at most $a + \varepsilon$, but

$$\varepsilon = O\left(\sqrt{\frac{T(n) r(n)}{q(n)}}\right) \leq \frac{a-b}{4},$$

provided we chose a sufficiently large constant C' when defining $q(n)$. Thus, every invocation of Q accepts with probability at most $b + \frac{a-b}{4}$. Therefore, provided we choose a sufficiently large constant C when defining $T(n)$, Arthur will accept with probability at most $2^{-p(n)}$ by a Chernoff bound.

We now prove the claim. By Lemma 3.3.2, the entanglement of formation between Merlin_A's registers and Merlin_B's registers is at most $2vr(n)$ after the v-th iteration. Hence

$$E_F\left(\rho^{A_1 A_2 \cdots A_{q(n)}, B_1 B_2 \cdots B_{q(n)}}\right) \leq 2T(n) r(n)$$

throughout the protocol. Also, let S_A and S_B be the sets of A-registers and B-registers, respectively, that Arthur has not yet chosen. Then, $|S_A| = |S_B| = q(n) - T(n)$. Assuming the Weak Additivity Conjecture, we therefore have

$$\begin{aligned}\sum_{A_j \in S_A, B_k \in S_B} E_F\left(\rho^{A_j B_k}\right) &= O\left((q(n) - T(n)) E_F\left(\rho^{A_1 A_2 \cdots A_{q(n)}, B_1 B_2 \cdots B_{q(n)}}\right)\right) \\ &= O\left(T(n) r(n) (q(n) - T(n))\right),\end{aligned}$$

so if we define

$$\sigma := \frac{1}{|S_A| |S_B|} \sum_{A_j \in S_A, B_k \in S_B} \rho^{A_j B_k},$$

then the convexity of E_F implies that

$$\begin{aligned}E_F(\sigma) &\leq \frac{1}{|S_A| |S_B|} \sum_{A_j \in S_A, B_k \in S_B} E_F\left(\rho^{A_j B_k}\right) \\ &= O\left(\frac{T(n) r(n)}{q(n) - T(n)}\right) \\ &= O\left(\frac{T(n) r(n)}{q(n)}\right),\end{aligned}$$

using the fact that $T(n) \leq q(n)/2$. Therefore, by Lemma 3.3.1, σ is ε-close to a separable state, where $\varepsilon = O\left(\sqrt{T(n) r(n) / q(n)}\right)$. We are done.

□

3.4 Nonexistence of Perfect Disentanglers

The problem of pure state N-representability problem, and also deciding whether $\alpha(\Phi) \geq 2$ for a general channel, are in QMA(2) and are not known to be in QMA. Thus, we expect that QMA(2) $\neq$ QMA; however, let us consider the other direction and try to prove the converse.

Suppose there exists a quantum operation Φ such that $\Phi(\rho)$ is always separable, and also for any separable state σ there is a ρ such that $\Phi(\rho) = \sigma$. In that case, we could simulate QMA(2) inside QMA efficently[2]: Arthur asks Merlin to send ρ, then he applies Φ on ρ, and then simulates the QMA(2) verification algorithm on $\Phi(\rho)$. In this section we show that such a Φ does not exist.

Definition 3.4.1 *Let $\mathcal{H}$ and $\mathcal{K}$ be two finite-dimensional Hilbert spaces. Then given a quantum operation that maps density matrices over $\mathcal{H}$ to density matrices over $\mathcal{K} \otimes \mathcal{K}$, we say Φ is an (ε, δ)-disentangler if*

(i) $\Phi(\rho)$ is ε-close to a separable state for every ρ, and

(ii) for every separable state σ, there exists a ρ such that $\Phi(\rho)$ is δ-close to σ.

Watrous (personal communication) has proposed the following fundamental conjecture.

Conjecture 3.4.1 (Watrous) *For all constants $\varepsilon, \delta < 1$, any (ε, δ)-disentangler requires $dim\mathcal{H} = 2^{\Omega(dim\mathcal{K})}$.*

A proof of Conjecture 3.4.1 would be an important piece of formal evidence that QMA(2) $\neq$ QMA, and might even lead to a "quantum oracle separation" (as defined by Aaronson and Kuperberg [3]) between the two classes.

Here, we show that at least in the case $\varepsilon = \delta = 0$, no disentangler exists in *any* finite dimension.

Theorem 3.4.1 *Let Φ be a perfect disentangler sending density matrices over $\mathcal{H}$ to density matrices over $\mathcal{K} \otimes \mathcal{K}$. Then $dim\mathcal{K} \geq 2$ implies $dim\mathcal{H} = \infty$.*

Proof: For any pure state $|\alpha\rangle \in \mathcal{K}$, by assumption there exists a state ρ_α such that $\Phi(\rho_\alpha) = |\alpha\rangle\langle\alpha| \otimes |\alpha\rangle\langle\alpha|$. By linearity, we can assume $\rho_\alpha = |\phi_\alpha\rangle\langle\phi_\alpha|$ is pure. Also, suppose $\dim\mathcal{H}$ is finite. Then Φ admits an operator-sum representation $\Phi(\rho) = \sum_{i=1}^{k} E_i \rho E_i^\dagger$ where $\sum_{i=1}^{k} E_i^\dagger E_i = I$. We then have

$$\Phi(|\phi_\alpha\rangle\langle\phi_\alpha|) = \sum_{i=1}^{k} E_i |\phi_\alpha\rangle\langle\phi_\alpha| E_i^\dagger = |\alpha\rangle\langle\alpha| \otimes |\alpha\rangle\langle\alpha| .$$

Thus, we find that $E_i |\phi_\alpha\rangle$ must be a multiple of $|\alpha\rangle|\alpha\rangle$ for all i and α; that is, there exist constants $c_{\alpha,i}$ such that $E_i |\phi_\alpha\rangle = c_{\alpha,i} |\alpha\rangle|\alpha\rangle$.

Now let $|\alpha\rangle, |\beta\rangle$ be any two pure states in $\mathcal{K}$ with $|\alpha\rangle \neq |\beta\rangle$. Also let $|\psi\rangle = a|\phi_\alpha\rangle + b|\phi_\beta\rangle$ for some nonzero real numbers a, b. Then

$$\begin{aligned}
\Phi(|\psi\rangle\langle\psi|) &= a^2\Phi(|\phi_\alpha\rangle\langle\phi_\alpha|) + b^2\Phi(|\phi_\beta\rangle\langle\phi_\beta|) + ab\Phi(|\phi_\alpha\rangle\langle\phi_\beta|) + ab\Phi(|\phi_\beta\rangle\langle\phi_\alpha|) \\
&= a^2 |\alpha\rangle\langle\alpha| \otimes |\alpha\rangle\langle\alpha| + b^2 |\beta\rangle\langle\beta| \otimes |\beta\rangle\langle\beta| + abc |\alpha\rangle\langle\beta| \otimes |\alpha\rangle\langle\beta| \\
&\quad + abc^* |\beta\rangle\langle\alpha| \otimes |\beta\rangle\langle\alpha| ,
\end{aligned}$$

[2]We should also assume that Φ can be applied in polynomial-time.

where

$$c = \sum_{i=1}^{k} c_{\alpha,i} c^*_{\beta,i}.$$

We claim that $c = 0$. To see this, recall that $\Phi(|\psi\rangle\langle\psi|)$ is a separable mixed state, and consider any decomposition of $\Phi(|\psi\rangle\langle\psi|)$ into separable pure states. Since $\Phi(|\psi\rangle\langle\psi|)$ is a mixed state in the subspace spanned by $|\alpha\rangle|\alpha\rangle$ and $|\beta\rangle|\beta\rangle$, every pure state in the support of $\Phi(|\psi\rangle\langle\psi|)$ must have the form $x|\alpha\rangle|\alpha\rangle + y|\beta\rangle|\beta\rangle$, but by the assumption $|\alpha\rangle \neq |\beta\rangle$, such a state cannot be separable unless $x = 0$ or $y = 0$. Hence, the only separable pure states in the support of $\Phi(|\psi\rangle\langle\psi|)$ are $|\alpha\rangle|\alpha\rangle$ and $|\beta\rangle|\beta\rangle$. Therefore $abc = 0$, but a and b were nonzero, so $c = 0$ as claimed.

This means in particular that $\Phi(|\phi_\alpha\rangle\langle\phi_\beta|) = 0$ for all $|\alpha\rangle \neq |\beta\rangle$. Hence

$$\begin{aligned}
\langle\phi_\beta|\phi_\alpha\rangle &= \sum_{i=1}^{k} \langle\phi_\beta| E_i^\dagger E_i |\phi_\alpha\rangle \\
&= \mathrm{Tr}\left(\sum_{i=1}^{k} E_i |\phi_\alpha\rangle\langle\phi_\beta| E_i^\dagger\right) \\
&= \mathrm{Tr}\left(\Phi(|\phi_\alpha\rangle\langle\phi_\beta|)\right) \\
&= 0.
\end{aligned}$$

Thus for different $|\alpha\rangle$'s, the states $|\phi_\alpha\rangle$ are all orthogonal, and since the number of $|\alpha\rangle$'s is infinite, $\dim\mathcal{H}$ must be infinite as well.

□

3.5 Summary

In this Chapter I have shown that any QMA(k) protocol, $k > 2$, can be simulated inside QMA(2) with a smaller gap. I have also shown that if the Weak Additivity Conjecture holds, then we can amplify the gap in QMA(2), and then the whole QMA(k)-hierarchy collapses to QMA$(2, 2/3, 1, 3)$, and we basically have two quantum Merlin-Arthur complexity classes: QMA and QMA(2). I have also, by proving the non-existence of a perfect disentangler, given an evidence that these two classes are not the same.

There are two important open problems regarding QMA(2). First, in there any QMA(2)-complete problem? Second, can we find a non-trivial classical upper bound for QMA(2)? Of course, NEXP contains QMA(2), and QMA $\subseteq$ QMA(2), but these two upper and lower bounds are far from each other.

Chapter 4

Gap in QMA(2) Protocols

In Chapter 3, we showed that if we ignore that gap in QMA(2), then it contains the whole QMA(k)-hierarchy. The focus of this chapter is to study the relation between gap and size of witnesses in QMA(2) protocols.

4.1 $\mathsf{QMA}_{\log}(2)$

In Merlin-Arthur games, other than the number of Merlins, we can consider the case where the size of the witnesses is less than $poly(n)$. For example, in the classical case $\log(n)$-size witnesses never help the verifier to solve a problem beyond P because he can check all such witnesses in polynomial time, but this argument fails in the quantum case and we can define the complexity classes $\mathsf{QMA}_{\log}(k)$.

Definition 4.1.1 $\mathsf{QMA}_{\log}(k,a,b)$ *is the same as* $\mathsf{QMA}(k,a,b)$ *except that the size of witnesses sent by Merlins is* $O(\log n)$.

Based on the strong gap amplification protocol for $\mathsf{QMA} = \mathsf{QMA}(1)$, for $k = 1$, we have $\mathsf{QMA}_{\log} = \mathsf{BQP}$.

Theorem 4.1.1 [88] $\mathsf{QMA}_{\log} = \mathsf{BQP}$[1].

This theorem shows that in the case of $k = 1$, we have the same situation as in the classical case; however, we do not know any non-trivial upper bound for $\mathsf{QMA}_{\log}(2)$.

Recently, Blier and Tapp [28] have shown that $\mathsf{QMA}_{\log}(2)$ with perfect completeness and soundness $1 - \frac{1}{24n^6}$ contains NP.

Theorem 4.1.2 [28] $\mathsf{NP} \subseteq \mathsf{QMA}_{\log}(2, 1, 1 - \frac{1}{24n^6})$.

This result shows that $\mathsf{QMA}_{\log}(2)$ is a non-trivial complexity class because $\mathsf{QMA}_{\log}$ is the same as BQP, and it is widely believed that BQP does not contain NP. This observation is an strong evidence on the fact that $\mathsf{QMA}(2) \neq \mathsf{QMA}$, and it turns the complexity class $\mathsf{QMA}_{\log}(2)$ to an interesting one which contains both BQP and NP. However, the gap in this theorem is very small, and it would be a stronger result if we could prove the same containment with constant gap. Here, I should mention that it has been proved that $\mathsf{QMA}_{\log}(\sqrt{n})$ contains NP with constant gap [2]. In this section I show that $\mathsf{QMA}_{\log}(2)$ with the gap $n^{3+\epsilon}$, for any $\epsilon > 0$, contains NP.

[1] This theorem can be proved by the same idea as in the proof of Theorem 4.2.2.

4.1.1 $\mathsf{QMA}_{\log}(2)$ as a Maximization Problem

Consider a $\mathsf{QMA}_{\log}(2)$ protocol. Arthur after receiving a separable state ρ^{AB}, prepares some ancilla qubits, applies a unitary operation U, and then measures the first qubit; he accepts if the outcome of measurement is $|0\rangle$. Therefore, the probability of acceptance can be written in the form

$$\mathrm{Tr}[(|1\rangle\langle 1| \otimes I) \cdot (U\rho^{AB} \otimes |0\ldots 0\rangle\langle 0\ldots 0|U^\dagger)] = \mathrm{Tr}\left(H\rho^{AB}\right),$$

where

$$H = \langle 0\ldots 0|U^\dagger(|1\rangle\langle 1| \otimes I)U|0\ldots 0\rangle.$$

Thus, the problem of $\mathsf{QMA}_{\log}(2)$ is to compute

$$\max \mathrm{Tr}\left(H\rho^{AB}\right),$$

where the maximum is taken over all separable states ρ^{AB}.

This formulation is the same as what we used in Section 2.1.5. However, it is more useful here because if the size of witnesses is logarithmic, then H is a polynomial size matrix and is more tractable.

4.1.2 Complexity of Recognizing Entanglement

Let H be a hermitian matrix of polynomial size. Then, the problem of maximizing $\langle\phi|H|\phi\rangle$ over all states $|\phi\rangle$ is an eigenvalue problem and can be solved in polynomial time. Now assume that we are restricting $|\phi\rangle$ to be a separable state. Then this problem is NP-hard due to the following observation by Gurvits [57].

Let H be of the form

$$H = \begin{pmatrix} 0 & B_1 & \cdots & B_s \\ B_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ B_s & 0 & \ldots & 0 \end{pmatrix}, \tag{4.1}$$

where B_i, $1 \leq i \leq s$, is a hermitian matrix. Then

$$\langle\psi|\langle\phi|H|\phi\rangle|\psi\rangle = \langle\phi|H(|\psi\rangle)|\phi\rangle,$$

where

$$H(|\psi\rangle) = \begin{pmatrix} 0 & \langle\psi|B_1|\psi\rangle & \cdots & \langle\psi|B_s|\psi\rangle \\ \langle\psi|B_1|\psi\rangle & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \langle\psi|B_s|\psi\rangle & 0 & \ldots & 0 \end{pmatrix}. \tag{4.2}$$

It means that the maximum of $\langle\psi|\langle\phi|H|\phi\rangle|\psi\rangle$, for a fixed $|\psi\rangle$ is equal to the maximum eigenvalue of $H(|\psi\rangle)$. $H(|\psi\rangle)$ is a rank 2 matrix with eigenvalues 0 and $\pm\big(\langle\psi|B_1|\psi\rangle^2 + \cdots + \langle\psi|B_s|\psi\rangle^2\big)^{1/2}$. Therefore,

$$\max_{|\phi\rangle|\psi\rangle} \langle\psi|\langle\phi|H|\phi\rangle|\psi\rangle = \max_{|\psi\rangle} \left[\, \langle\psi|B_1|\psi\rangle^2 + \cdots + \langle\psi|B_s|\psi\rangle^2 \,\right]^{1/2}. \tag{4.3}$$

Now note that, by the calculation of previous section, any problem in $\mathsf{QMA}_{\log}(2)$ can be

formulated as the maximization problem

$$\max_{\rho} \mathrm{Tr}(H\rho),$$

where H is a polynomial size positive semi-definite matrix, and the maximum is taken over the set of bipartite separable states. Therefore, given the maximum point $|\phi\rangle|\psi\rangle$, a quantum polynomial time verifier can compute $\langle\psi|\langle\phi|H|\phi\rangle|\psi\rangle$, or equivalently the left hand side of (4.3). Hence, proving that computing the left hand side is NP-hard is equivalent to $\mathsf{NP} \subseteq \mathsf{QMA}_{\log}(2)$. Although this is known by Theorem 4.1.2, using this idea we can get the same result with a larger gap.

4.1.3 2-out-of-4-SAT

To prove the containment $\mathsf{NP} \subseteq \mathsf{QMA}_{\log}(2)$ we should find a protocol to solve some NP-complete problem in $\mathsf{QMA}_{\log}(2)$. Although the most well-known such problem is 3-SAT, it is convenient for us to use another version of this problem called 2-out-of-4-SAT[2]. We can formulate this problem as follows. Let $|a_1\rangle, |a_2\rangle, \ldots, |a_m\rangle$ be vectors of the form

$$|a_k\rangle = \sum_{i=1}^{n} c_{ki} |i\rangle, \tag{4.4}$$

where $c_{ki} = 0$ or $\pm\frac{1}{2}$, and for each k there are exactly 4 non-zero c_{ik}, $1 \leq i \leq n$. Now the problem is to decide whether there exists a vector $|\psi\rangle$ orthogonal to all $|a_k\rangle$'s which is of the form

$$|\psi\rangle = \sum_{i=1}^{n} \pm\frac{1}{\sqrt{n}} |i\rangle. \tag{4.5}$$

To get a larger gap we will need our 2-out-of-4-SAT instance to be a PCP one, and to be bounded-literal (every variable should appear in a constant number of clauses). The following lemma shows how to get everything we want with only a poly-logarithmic blowup in the number of variables and clauses.

Lemma 4.1.1 *There exists a polynomial time Karp reduction that maps a* 3-SAT *instance* α *to a* 2-out-of-4-SAT *instance* β *such that*

- *If* α *has* n *variables and* $m \geq n$ *clauses, then* β *has* $O(mpoly\log(m))$ *variables and* $O(mpoly\log(m))$ *clauses.*
- *Every variable of* β *occurs in at most* c *clauses, for some constant* c*.*
- *The reduction is a PCP, meaning that satisfiable instances map to satisfiable instances, while unsatisfiable instances map to instances in which at most a constant fraction of the clauses can be satisfied at the same time.*

Proof: Given a 3-SAT instance φ, we first amplify its soundness gap to a constant using the celebrated method of Dinur [44]. Next we use a reduction due to Papadimitriou and Yannakakis [101], which makes every variable occur in exactly 29 clauses, without destroying

[2]See Section 2.4.2

the soundness gap. Finally we use a gadget due to Khanna et al. [75], which converts from 3-SAT to 2-out-of-4-SAT, without destroying either the soundness gap or the bounded literal property. Note that the reduction of Dinur [44] incurs only a poly-logarithmic blowup in the total size of the instance, while the other two reductions incur a constant blowup.

□

4.1.4 The Proper State Case

Suppose Arthur has applied Lemma 4.1.1, to obtain a bounded-literal 2-out-of-4-SAT instance ϕ with $N = O(m \operatorname{polylog} m)$ variables, $M = O(m \operatorname{polylog} m)$ clauses, and a constant soundness gap $\varepsilon > 0$. Now suppose Merlin sends Arthur a $\log N$-qubit state of the form

$$|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (-1)^{x_i} |i\rangle,$$

where $x_1, \ldots, x_N \in \{0,1\}^N$ is a claimed satisfying assignment for ϕ. Call a state having the above form (for some Boolean x_i's) a *proper* state. Then we claim the following:

Lemma 4.1.2 *Assuming $|\psi\rangle$ is proper, Arthur can check whether ϕ is satisfiable with perfect completeness and constant soundness.*

Proof: To perform the check, Arthur uses the following *Satisfiability Test.* First he partitions the clauses of ϕ into a constant number of blocks $B_1, \ldots, B_s$, such that within each block, no two clauses share a variable. Such a partition clearly exists by the assumption that ϕ is bounded-literal, and furthermore can be found efficiently (e.g., using a greedy algorithm). Next he chooses one of the blocks B_r uniformly at random, and measures $|\psi\rangle$ in an orthonormal basis with one projector for each clause in B_r. Because a single block in the partition of clauses does not necessarily cover all the variables, it is possible that the measurement result will not correspond to any clause in B_r, in which case Arthur accepts. However, suppose that the measurement yields the following reduced state, for some random clause $C_{ijk\ell} := (i,j,k,\ell)$ in B_r:

$$|\psi_{ijkl}\rangle := \frac{1}{2}\left[(-1)^{x_i}|i\rangle + (-1)^{x_j}|j\rangle + (-1)^{x_k}|k\rangle + (-1)^{x_\ell}|\ell\rangle\right].$$

Notice that, of the 16 possible assignments to the variables (x_i, x_j, x_k, x_ℓ), six of them satisfy $C_{ijk\ell}$, and those six lead to three states $|\psi_{ijk\ell}\rangle$ that are orthogonal to one another (as well as the negations of those states, which are essentially the same). It follows that Arthur can perform a projective measurement on $|\psi_{ijk\ell}\rangle$, which accepts with probability 1 if $C_{ijk\ell}$ is satisfied, and rejects with constant probability if $C_{ijk\ell}$ is unsatisfied. Furthermore, because the number of blocks B_r is a constant, each of the M clauses of ϕ is checked in this test with probability $\Omega(1/M)$. And we know that, if $x_1, \ldots, x_N$ is *not* a satisfying assignment for ϕ, then a constant fraction of the clauses will be unsatisfied.

Putting everything together, we find that if ϕ is satisfiable, then the Satisfiability Test accepts $|\psi\rangle$ with probability 1; while if ϕ is unsatisfiable, then it rejects with constant probability.

□

4.1.5 $\mathsf{NP} \subseteq \mathsf{QMA}_{\log}(2)$

In this section we prove our main result.

Theorem 4.1.3 *For every constant $\epsilon > 0$, $\mathsf{NP} \subseteq \mathsf{QMA}_{\log}(2, a, a - \frac{1}{n^{3+\epsilon}})$, for some a independent of ϵ.*

To prove this theorem we give a Merlin-Arthur protocol for the 2-out-of-4-SAT problem. This protocol consists of two parts. First, given a satisfying assignment we should check whether this state is a proper state, that is a state the form (4.5). Second, we should check whether it is a satisfying assignment of 2-out-of-4-SAT instance. The following lemma is direct consequence of Lemma 4.1.2 and also Lemma 3.2.1.

Corollary 4.1.1 *Let us assume that Merlin is restricted to send a state that is δ-close, in trace distance, to a proper state, for a constant $\delta > 0$. Then Arthur can solve* 3-SAT *with perfect completeness and constant soundness.*

□

Lemma 4.1.3 *Let $\epsilon > 0$ be a constant. Then there exists a Merlin-Arthur protocol in which Merlins send state $|\phi\rangle|\psi\rangle$ and Arthur can check whether $|\psi\rangle$ is $(5n^{-\epsilon/4})$-close, in trace distance, to a proper state. More precisely, if $|\psi\rangle$ is proper then Arthur accepts with probability*

$$\frac{1}{2} + \frac{1}{3n}(2 - \frac{2}{n})^{1/2},$$

and if it is not $(5n^{-\epsilon/4})$-close to a proper state then he rejects with probability

$$\frac{1}{2} + \frac{1}{3n}(2 - \frac{2}{n})^{1/2} - \frac{1}{20n^{3+\epsilon}}.$$

Proof of Theorem 4.1.3: Given a 3-SAT instance α Arthur reduces it to a 2-out-of-4-SAT instance β over m variables using Lemma 4.1.1, and asks Merlins to send him $|\phi\rangle|\psi\rangle$ where $|\psi\rangle$ is a satisfying assignment for β. Then he applies one of the tests in Lemmas 4.1.3 or 4.1.2, each with probability 1/2.

If $|\psi\rangle$ is not $(5m^{-\epsilon'/4})$-close to a proper state, for some $\epsilon > \epsilon' > 0$, Arthur rejects the test in Lemma 4.1.3 with probability $\frac{1}{2} + \frac{1}{3m}(2 - \frac{2}{m})^{1/2} - \frac{1}{20m^{3+\epsilon'}}$. On the other hand, if it is $(5m^{-\epsilon'/4})$-close to a proper state while not a satisfying assignment then Arthur rejects the test of Lemma 4.1.2 and Corollary 4.1.1 with constant probability. Therefore, $\mathsf{NP} \subseteq \mathsf{QMA}_{\log}(2, a, b)$, where

$$a = \frac{1}{2}\big[1 + \frac{1}{2} + \frac{1}{3m}(2 - \frac{2}{m})^{1/2}\big],$$

and

$$b = \frac{1}{2}\big[1 + \frac{1}{2} + \frac{1}{3m}(2 - \frac{2}{m})^{1/2}\big] - \frac{1}{n^{3+\epsilon}}.$$

Here we replace ϵ' with ϵ, to consider the poly-logarithmic blowup in the size of problem by reducing it from a 3-SAT instance to a 2-out-of-4-SAT instance, and also to get ride of the constants in Lemma 4.1.3.

□

The only remaining part is the proof of Lemma 4.1.3.

4.1.6 Proof of Lemma 4.1.3

For any $1 \leq i < j \leq n$, define the hermitian matrices

$$B_{ij} = |i\rangle\langle j| + |j\rangle\langle i|.$$

Let

$$H = \begin{pmatrix} 0 & B_{1,2} & \cdots & B_{(n-1)n} \\ B_{1,2} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ B_{(n-1)n} & 0 & \cdots & 0 \end{pmatrix}, \tag{4.6}$$

be a hermitian matrix. A simple calculation shows that λ is an eigenvalue of H iff λ^2 is an eigenvalue of $\sum_{i,j} B_{ij}^2$. Then, $\|H\|_\infty$, the infinity-norm of matrix H, is

$$\|H\|_\infty^2 = \|\sum_{i,j} B_{ij}^2\|_\infty \leq \sum_{i,j} \|B_{ij}\|_\infty^2 = \binom{n}{2} \leq n^2.$$

Therefore, $\frac{1}{2}I + \frac{1}{3n}H$ is a hermitian positive semi-definite matrix, and in fact a $O(\log(n))$-local Hamiltonian, with norm $\|\frac{1}{2}I + \frac{1}{3n}H\|_\infty < 1$. Then by the idea of [77, 6] having the state $|\phi\rangle|\psi\rangle$ Arthur can throw a coin with probability of head being $\langle\psi|\langle\phi|\frac{1}{2}I + \frac{1}{3n}H|\phi\rangle|\psi\rangle$ and accept if it was head. Then by (4.3) the probability of accepting is at most

$$\frac{1}{2} + \frac{1}{3n} \max_{|\psi\rangle} \left[\sum_{i,j} \langle\psi|B_{ij}|\psi\rangle^2 \right]^{1/2}. \tag{4.7}$$

Now we need the following lemma.

Lemma 4.1.4 *$\sum_{i,j}\langle\psi|B_{ij}|\psi\rangle^2 \leq 2 - \frac{2}{n}$, and equality holds iff $|\psi\rangle$ is a proper state. Also if*

$$\sum_{i,j}\langle\psi|B_{ij}|\psi\rangle^2 \geq 2 - \frac{2}{n} - \frac{1}{n^{2+\epsilon}}, \tag{4.8}$$

then $|\psi\rangle$ is $(5n^{-\epsilon/4})$-close to a proper state in trace distance.

Using this lemma the probability of accepting is at most

$$\frac{1}{2} + \frac{1}{3n}(2 - \frac{2}{n})^{1/2},$$

and if it is greater than

$$\frac{1}{2} + \frac{1}{3n}(2 - \frac{2}{n})^{1/2} - \frac{1}{20n^{3+\epsilon}},$$

then by Lemma 4.1.4, $|\psi\rangle$ is $(5n^{-\epsilon/4})$-close to a proper state.

It remains to prove Lemma 4.1.4.

Proof of Lemma 4.1.4: Let

$$|\psi\rangle = \sum_{i=1}^{n} x_i |i\rangle.$$

Then

$$\begin{aligned}\sum_{i,j}\langle\psi|B_{ij}|\psi\rangle^2 &= \sum_{i<j}(\overline{x}_i x_j + x_i\overline{x}_j)^2 \\ &= \sum_{i<j}\overline{x}_i^2 x_j^2 + x_i^2\overline{x}_j^2 + 2|x_i|^2|x_j|^2 \\ &= (\sum_i \overline{x}_i^2)(\sum_i x_i^2) - \sum_i |x_i|^4 + 2\sum_{i<j}|x_i|^2|x_j|^2 \\ &= |\sum_i x_i^2|^2 + (\sum_i |x_i|^2)^2 - 2\sum_i |x_i|^4.\end{aligned}$$

Now using $\sum_{i=1}^n |x_i|^2 = 1$,

$$\sum_i |x_i|^4 \geq \frac{1}{n}, \tag{4.9}$$

and

$$|\sum_i x_i^2|^2 \leq 1, \tag{4.10}$$

we find that $\sum_{i,j}\langle\psi|B_{ij}|\psi\rangle^2 \leq 2 - \frac{2}{n}$, and equality holds iff $x_j^2 = e^{i\theta}\frac{1}{n}$, $1 \leq j \leq n$, for a constant θ, or equivalently iff $|\psi\rangle$ is a proper state.

Now assume that (4.8) holds. Then by (4.9) and (4.10) we have

$$\sum_i |x_i|^4 \leq \frac{1}{n} + \frac{1}{n^{2+\epsilon}}. \tag{4.11}$$

and

$$|\sum_i x_i^2|^2 \geq 1 - \frac{1}{n^{2+\epsilon}}. \tag{4.12}$$

Note that

$$\sum_i (|x_i|^2 - \frac{1}{n})^2 = \sum_i \left(|x_i|^4 + \frac{1}{n^2} - \frac{2}{n}|x_i|^2\right) = \sum_i |x_i|^4 - \frac{1}{n}.$$

Therefore by (4.11), for every i

$$||x_i|^2 - \frac{1}{n}| \leq \frac{1}{n^{1+\delta}}, \tag{4.13}$$

where $\delta = \epsilon/2$. Then for sufficiently large n

$$||x_i| - \frac{1}{\sqrt{n}}| \leq \frac{\sqrt{n}}{n^{1+\delta}}. \tag{4.14}$$

Also using (4.12) we have

$$|\sum_i |x_i|^2|^2 - |\sum_i x_i^2|^2 \leq \frac{1}{n^{2+\epsilon}},$$

and then

$$|x_i x_j|^2 - Re\, x_i^2 \bar{x}_j^2 \leq \frac{1}{n^{2+\epsilon}}, \tag{4.15}$$

for any i and j.

Now let $x_j = s_j r_j e^{i\theta_j}$, where $s_j \in \{+1, -1\}$, r_j is a non-negative real number, and $-\frac{\pi}{2} < \theta_j \leq \frac{\pi}{2}$. Then by (4.14)

$$|r_i - \frac{1}{\sqrt{n}}| \leq \frac{\sqrt{n}}{n^{1+\delta}}. \tag{4.16}$$

Also by (4.13) and (4.15)

$$1 - Re\, e^{2i(\theta_k - \theta_j)} \leq \frac{1}{n^{2+\epsilon}} \left(\frac{1}{n} - \frac{1}{n^{1+\delta}}\right)^{-2} = (n^\delta - 1)^{-2} \leq \frac{2}{n^\epsilon}, \tag{4.17}$$

for sufficiently large n. Without loss of generality, we may assume that $\theta_1 = 0$ and then for any j we have

$$1 - Re\, e^{2i\theta_j} \leq \frac{2}{n^\epsilon}, \tag{4.18}$$

and since $-\frac{\pi}{2} < \theta_j \leq \frac{\pi}{2}$,

$$1 - Re\, e^{i\theta_j} \leq \frac{2}{n^\epsilon}. \tag{4.19}$$

Now using $(Re\, e^{i\theta_j})^2 + (Im\, e^{i\theta_j})^2 = 1$, it is easy to see that

$$|1 - e^{i\theta_j}| \leq \frac{4}{n^\delta}. \tag{4.20}$$

Therefore using (4.16) and (4.20)

$$\begin{aligned} |\, r_j e^{i\theta_j} - \frac{1}{\sqrt{n}}| &\leq |r_j - \frac{1}{\sqrt{n}}| + |r_i(1 - e^{i\theta_j})| && (4.21) \\ &\leq \frac{\sqrt{n}}{n^{1+\delta}} + (\frac{1}{\sqrt{n}} + \frac{\sqrt{n}}{n^{1+\delta}})\frac{4}{n^\delta}, && (4.22) \end{aligned}$$

or

$$|\, r_j e^{i\theta_j} - \frac{1}{\sqrt{n}}| \leq \frac{10\sqrt{n}}{n^{1+\delta}}. \tag{4.23}$$

Now define the proper state

$$|\phi\rangle = \sum_j \frac{s_j}{\sqrt{n}} |j\rangle.$$

We have

$$|\langle\phi|\psi\rangle| = |\sum_j \frac{1}{\sqrt{n}} s_j^2 r_j e^{i\theta_j}| \tag{4.24}$$

$$= \frac{1}{\sqrt{n}} |\sum_j r_j e^{i\theta_j}| \tag{4.25}$$

$$\geq \frac{1}{\sqrt{n}} (\sqrt{n} - |\sum_j r_j e^{i\theta_j} - \frac{1}{\sqrt{n}}|). \tag{4.26}$$

Using (4.23) we get

$$|\langle\phi|\psi\rangle| \geq 1 - \sqrt{n}\frac{10\sqrt{n}}{n^{1+\delta}} = 1 - \frac{10}{n^\delta}. \tag{4.27}$$

Therefore,

$$\||\psi\rangle - |\phi\rangle\|_{\mathrm{Tr}} = \left(1 - |\langle\phi|\psi\rangle|^2\right)^{1/2} \leq \left(\frac{20}{n^\delta}\right)^{1/2} < 5n^{-\delta/2}, \tag{4.28}$$

and we are done.

□

4.2 Gap vs Size of Witnesses

By Theorem 2.1.1 we can amplify the gap in QMA without increasing the size of witness . However, it seems that if number of Merlins is more than one, by changing the gap versus size of witnesses we get to different complexity classes. In this section I bring two evidences to support this claim.

4.2.1 QMA(k) with Exponentially Small Gap

According to the computation in Section 2.1.5 any QMA problem can be formulated as an exponential-size eigenvalue problem and then can be solved in EXP. This containment is independent of the gap in QMA. Indeed, QMA even with exponentially small gap in inside EXP. Equivalently, we have PostQMA $\subseteq$ EXP, where PostQMA is the same as QMA except that Arthur has the ability of *post-selection*[3]. On the other hand, if we scale the containment in either Theorem 4.1.2 or Theorem 4.1.3 to exponential-size problems, we find that QMA(2) with exponentially small gap contains NEXP.

Theorem 4.2.1 PostQMA(2) = NEXP.

Proof: By Theorems 4.1.2 and 4.1.3, NEXP is in QMA(2) with exponentially small gap. Also QMA(2), independent of gap, is in NEXP because the prover can send the classical description of quantum state, and the verifier can simulate a quantum verifier in exponential-time.

□

By this theorem there is a huge gap between PostQMA and PostQMA(2).

[3]See Appendix B for the exact definition, and also for a proof

4.2.2 $\mathsf{QMA}_{\log}(2)$ with Exponentially Small Error

In this section, based on an observation by Brandão (see [2]), it is proved that $\mathsf{QMA}_{\log}(2)$ with exponentially small error is equal to BQP.

Theorem 4.2.2 *For any function $s(n)$ that is bounded by a polyanomial, $\mathsf{QMA}_{s(n)}(2, 1 - 2^{-2s(n)}, 2^{-2s(n)}) = \mathsf{QMA}_{s(n)}$. and therefore, $\mathsf{QMA}_{\log}(2, 1 - 2^{-n}, 2^{-n}) = \mathsf{BQP}$.*

Proof: Since we can amplify the gap in $\mathsf{QMA}_{s(n)}$ without increasing the size of witness (Theorem 2.1.1), we have $\mathsf{QMA}_{s(n)} \subseteq \mathsf{QMA}_{s(n)}(2, 1 - 2^{-2s(n)}, 2^{-2s(n)})$.

For the other direction, if Merlin is honest, he sends a separable state to Arthur, and there is no need for two Merlins. For the soundness, assume the probability of acceptance of Arthur's verification on any separable state is at most $2^{-2s(n)}$. Now assume Merlin sends the bipartite state $|\psi_{AB}\rangle$ to Arthur. Consider the Schmidt decomposition of $|\psi_{AB}\rangle$.

$$|\psi_{AB}\rangle = \sum_{i=1}^{2^{s(n)}} \lambda_i |\phi_i\rangle_A |\varphi_i\rangle_B .$$

Note that Arthur's acceptance probability can be written of the form $\langle\psi_{AB}|H|\psi_{AB}\rangle$, for some positive semi-definite matrix H. Therefore, Arthur's acceptance probability on $|\psi_{AB}\rangle$ is at most

$$\sum_{i,j=1}^{2^{s(n)}} \lambda_i^* \lambda_j \langle\phi_i|_A \langle\varphi_i|_B H |\phi_j\rangle_A |\varphi_j\rangle_B \leq 2^{-2s(n)} \sum_{i,j=1}^{2^{s(n)}} \lambda_i^* \lambda_j \leq 2^{-2s(n)} \left(\sum_{i=1}^{2^{s(n)}} |\lambda_i| \right)^2 \leq 2^{-s(n)},$$

and then we are done.

By letting $s(n) = O(\log n)$, and using Theorem 4.1.1, we get $\mathsf{QMA}_{\log}(2, 1 - 2^{-n}, 2^{-n}) = \mathsf{BQP}$.

□

This theorem shows that if $\mathsf{QMA}_{\log}(2)$ with inverse polynomial gap is the same as $\mathsf{QMA}_{\log}(2)$ with exponentially small error, then by Theorems 4.1.2 and 4.1.3, BQP contains NP, which seems unlikely.

4.3 Summary

In this section we have seen that $\mathsf{QMA}_{\log}(2)$ with the gap $n^{-(3+\epsilon)}$, for any $\epsilon > 0$, contains NP. This result turns $\mathsf{QMA}_{\log}(2)$ to an interesting complexity class which contains both BQP and NP. Based on our results in this chapter, it is unlikely that NP is in $\mathsf{QMA}_{\log}(2)$ with constant gap; however it is interesting to see whether $\mathsf{QMA}_{\log}(2)$ with the gap n^{-2} or even n^{-1} contains NP or not. In the next chapter I will discuss the complexity of separability problem with constant gap to get more intuition on $\mathsf{QMA}_{\log}(2)$ with constant gap.

Chapter 5

Separability Problem

5.1 Introduction

The problem of detecting entanglement has been focused in quantum information theory for many years. The problem is: given a bipartite mixed state ρ_{AB}, decide whether this state is entangled or separable. The first attack toward solving this problem is the following observation due to Peres and the Horodeckis, [102, 62]. If $\rho_{AB} = \sum_i p_i \rho_{A_i} \otimes \rho_{B_i}$ is separable, then $[\rho_{AB}]^{T_B} = \sum_i p_i \rho_{A_i} \otimes [\rho_{B_i}]^T$, where M^T denotes the transpose of matrix M, is also a quantum state, and thus is a positive semi-definite matrix. Therefore, if ρ_{AB} is separable, its partial transpose, $[\rho_{AB}]^{T_B}$, should be positive semi-definite. The Horodeckis have proved that this criterion characterizes all separable states in dimensions 2×2 and 2×3, [62]. However, there are entangled states in dimension 3×3 with a positive partial transpose, [23].

Although the set of positive partial transpose states (PPT states) does not coincide with the set of separable states, it is usually considered as an approximation of this set. For example in [43] the distance of an arbitrary state from PPT states has been computed to estimate the distance from separable states. Also in [112] the geometry of the set of PPT states has been studied to understand the properties of the set of separable states. However, we do not know how efficient these approximations are. For instance, given an upper bound on the distance of a state from PPT states, does it give an upper bound on the distance of the state from separable states?

We can think of this problem in the point of view of complexity theory. Gurvits [57] has proved that given a bipartite density matrix ρ_{AB}, it is NP-hard to decide whether this state is separable or entangled. An approximate formulation of this problem is the following. Given a bipartite density matrix ρ_{AB} and $\epsilon > 0$, decide whether there exists a separable state in the ϵ-neighborhood (in trace distance) of ρ_{AB}. Gurvits has established a reduction from Knapsack to this problem, and has proved the NP-hardness of the separability problem only for exponentially small ϵ. However, as mentioned in Chapter 4, by replacing Knapsack with 2-out-of-4-SAT, we can get to the NP-hardness for an inverse polynomial ϵ. Also, Gharibian [54] has shown the same result using a reduction from the Clique problem. Now, the question is that how large ϵ can be while getting to the same result. For example, is there an efficient algorithm to decide whether the distance of a given state from separable states is less than $1/3$, or it is an NP-hard problem? Equivalently, is there a separability test such that if a state passes the test then it is $1/3$-close to the set of separable states?

The converse of this question is what we are looking for in this chapter: given a sepa-

rability criterion, if a state passes this test can we claim a non-trivial upper bound on the distance of this state from separable states? I show that the answer to this question is no for PPT test, and also for some other separability criteria.

Before getting to the details of the results, I like to mention that the problems of the geometry and the relation between the sets of whole bipartite states, PPT states, and separable states have been studied in [12, 13, 113, 124]. It has been shown that if we consider these sets in the Euclidean space, then the ratio of the volume of the set of separable states and the set of PPT states tends to 0 as the dimension of the quantum states tends to infinity. Although this statement and the results of this chapter are based on different metrics (Hilbert-Schmidt norm versus trace norm) and also different parameters (volume versus distance), they give the same intuition on the relation between separable states and PPT states.

5.1.1 Separable States

Recall that a pure state $|\psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ is called *separable* if it can be written of the form $|\psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$, where $|\psi_A\rangle \in \mathcal{H}^A$ and $|\psi_B\rangle \in \mathcal{H}^B$. A density matrix acting on $\mathcal{H}^A \otimes \mathcal{H}^B$ is called separable if it can be written as a convex combination of separable pure states $|\psi\rangle\langle\psi|$. We denote the set of separable states by SEP.

5.1.2 Positive Partial Transpose Test

Assume that $\dim\mathcal{H}_A = \dim\mathcal{H}_B = d$, and fix an orthonormal basis $|1\rangle, \ldots, |d\rangle$ for both of Hilbert spaces. Then the partial transpose of matrices acting on $\mathcal{H}^A \otimes \mathcal{H}^B$ is a linear map defined by $(M_A \otimes N_B)^{T_B} = M_A \otimes N_B^T$. It is clear that if ρ_{AB} is a separable state then $\rho_{AB}^{T_B}$ is also a density matrix and then positive semi-definite. However, it does not hold for an arbitrary state. For example, the partial transpose of the maximally entangled state is not positive semi-definite. To see that, let $\Phi(d)$ to be the maximally entangled state on $\mathcal{H}$

$$\Phi(d) = \frac{1}{d}\sum_{i,j=1}^{d} |i,i\rangle\langle j,j|. \tag{5.1}$$

We have

$$\begin{aligned}
\Phi(d)^{T_B} &= \frac{1}{d}\sum_{i,j} |i\rangle\langle j| \otimes |j\rangle\langle i| \\
&= \frac{1}{d}I - \frac{1}{d}\sum_{i\neq j} |i\rangle\langle i| \otimes |j\rangle\langle j| + \frac{1}{d}\sum_{i\neq j} |i\rangle\langle j| \otimes |j\rangle\langle i| \\
&= \frac{1}{d}I - \frac{2}{d}\sum_{i<j} |\phi_{ij}\rangle\langle\phi_{ij}|,
\end{aligned}$$

where

$$|\phi_{ij}\rangle = \frac{1}{\sqrt{2}}(|i\rangle|j\rangle - |j\rangle|i\rangle). \tag{5.2}$$

Therefore, positive partial transpose is a test to detect entanglement [102, 62]. More formally, if we denote the set of density matrices with a positive semi-definite partial transpose

by PPT, then SEP $\subseteq$ PPT.

5.1.3 Some Other Separability Criteria

Here is a list of some other separability criteria, see [65].

- Reduction criterion, [60]: $I \otimes \rho_B \geq \rho_{AB}$, where $\rho_B = \mathrm{Tr}_A(\rho_{AB})$. Here, by $M \geq N$ we mean $M - N$ is a positive semi-definite matrix.

- Entropic criterion, [61]: $S_\alpha(\rho_{AB}) \geq S_\alpha(\rho_A)$ for $\alpha = 2$ and in the limit $\alpha \to 1$, where $S_\alpha(\rho) = \frac{1}{1-\alpha} \log \mathrm{Tr}(\rho^\alpha)$.

- Majorization criterion, [97]: $\lambda^\downarrow_{\rho_A} \succ \lambda^\downarrow_{\rho_{AB}}$, where $\lambda^\downarrow_\rho$ is the list of eigenvalues of ρ in non-increasing order, and $y \succ x$ means that, for any k, the sum of the first k entries of list x is less than or equal to that of list y.

- Cross norm criterion, [105, 37]: $\mathrm{Tr}|\mathcal{U}(\rho_{AB})| \leq 1$, where $\mathcal{U}$ is a linear map defined by $\mathcal{U}(M \otimes N) = v(M)v(N)^T$, relative to a fixed basis, and

$$v(X) = (col_1(X)^T, \ldots, col_d(X)^T)^T,$$

  where $col_i(X)$ is the i-th column of X.

All of these criteria for separability are necessary conditions but not sufficient. Doherty et al. [46, 47] have introduced a hierarchy of separability criteria which are both necessary and sufficient. Let $\rho_{AB} = \sum_i p_i \sigma_i \otimes \tau_i$ be a separable state. Then

$$\rho_{AB_1B_2\cdots B_k} = \sum_i p_i \sigma_i \otimes \tau_i^{\otimes k}$$

is an extension of ρ^{AB}, meaning that $\rho_{AB} = \mathrm{Tr}_{B_2\cdots B_k}(\rho_{AB_1\cdots B_k})$. Also it is symmetric, meaning that it is unchanged under any permutation of subsystems B_i. More precisely, for any permutation π of k objects, if we define the linear map P_π by $P_\pi|\psi_1\rangle \otimes \cdots \otimes |\psi_k\rangle = |\psi_{\pi(1)}\rangle \otimes \cdots \otimes |\psi_{\pi(k)}\rangle$, we have

$$P_\pi^{B_1\ldots B_k} \rho_{AB_1B_2\cdots B_k} P_\pi^{B_1\ldots B_k} = \rho_{AB_1B_2\cdots B_k}. \tag{5.3}$$

If such an extension exists, we say that ρ_{AB} has a symmetric extension to k copies. Doherty el al. have proved that a quantum state is separable iff it has a symmetric extension to k copies for any number k, [46, 47]. Also, they have shown that the problem of checking whether a given state has a symmetric extension to k copies, for a fixed k, can be expressed as a semi-definite programming, and can be solved efficiently[1]. So we get to another separability test.

- Symmetric extension criterion: If ρ_{AB} is separable, then it has a symmetric extension to k copies.

[1]Notice that knowing that a state has a symmetric extension to k copies, for a fixed k, gives us no upper bound on the distance of the state from separable states. Indeed, to get a non-trivial upper bound k has to be of the order of the dimension of the state. It is because the upper bound on the distance from separable states comes from the finite quantum de Finetti theorem, and this theorem gives a trivial bound for a constant k. See [38] and [83] for finite de Finetti theorem.

5.1.4 Quantum State Tomography

An informationally complete POVM on $\mathcal{H}$ is a set of positive semi-definite operators $\{M_n\}$ forming a basis for the space of hermitian matrices on $\mathcal{H}$, and such that $\sum_n M_n = I$. In [83] there is an explicit construction of an informationally complete POVM in any dimension. Such a POVM is useful for quantum state tomography.

Suppose $\{M_n^*\}$ is the dual of basis $\{M_n\}$. That is $\mathrm{Tr}(M_n M_m^*) = \delta_{mn}$, where δ_{mn} is the Kronecker delta function. For any hermitian operator X we have

$$X = \sum_n \mathrm{Tr}(X M_n)\, M_n^*.$$

Therefore, having some copies of the state ρ, by measuring ρ using the POVM $\{M_n\}$, we can approximate $\mathrm{Tr}(\rho M_n)$ and then find the matrix representation of ρ.

Assume that $\mathcal{H} = \mathcal{H}^A \otimes \mathcal{H}^B$ is a bipartite Hilbert space. If $\{P_n\}$ and $\{Q_m\}$ are informationally complete POVM's on $\mathcal{H}^A$ and $\mathcal{H}^B$, respectively, then it is easy to see that $\{P_n \otimes Q_m\}$ is an informationally complete POVM on $\mathcal{H}$. Therefore, if the state ρ_{AB} is shared between two far apart parties A and B, they still can perform quantum state tomography. Also, if the state ρ_{AB} is separable, then all the states during the process are separable as well.

5.1.5 Quantum de Finetti Theorem

As in Eq. (5.3), a quantum state $\rho^{(n)}$ acting on $\mathcal{H}^{\otimes n}$ is called symmetric if $P_\pi \rho^{(n)} P_\pi = \rho^{(n)}$ for any permutation π of n objects. A symmetric state is called *k-exchangeable* if it has a symmetric extension to $n+k$ registers. That is a symmetric state $\rho^{(n+k)}$ such that $\mathrm{Tr}_{1,\ldots,k}\rho^{(n+k)} = \rho^{(n)}$. Clearly, any state of the form $\rho^{\otimes n}$ is k-exchangeable, for any k. Also any convex combination of these states is k-exchangeable. *Quantum de Finetti theorem* says that the converse of this observation holds. That is, if a state is k-exchangeable, for any k, it is in the convex hall of symmetric product states.

Quantum de Finetti theorem gives a characterization of infinitely-exchangeable states. The following theorem, known as the finite quantum de Finetti theorem, says that if a state is k-exchangeable (but not necessarily $(k+1)$-exchangeable), then an approximation of the above result holds.

Theorem 5.1.1 [38] *Assume that $\rho^{(n+k)}$ is a symmetric state acting on $\mathcal{H}^{\otimes n+k}$. Let $\rho^{(n)} = Tr_{1\ldots k}\,\rho^{(n+k)}$ be the state obtained by tracing out the first k registers. Then there exists a probability measure μ on the set of density matrices on $\mathcal{H}$ such that*

$$\|\rho^{(n)} - \int \mu(d\sigma)\sigma^{\otimes n}\|_{Tr} \leq 2 dim\mathcal{H}\, \frac{n}{n+k}.$$

5.2 Main Result

The main result of this chapter in the following theorem.

Theorem 5.2.1 *Let $\mathcal{H}$ be a bipartite Hilbert space. If the dimension of each subsystem of $\mathcal{H}$ is large enough, there exists a PPT state acting on $\mathcal{H}$ whose trace distance from separable states is greater than $1-\epsilon$, for an arbitrary $\epsilon > 0$.*

5.2.1 Main Ideas

Let $\mathcal{H} = \mathcal{H}^A \otimes \mathcal{H}^B$ be a bipartite Hilbert space. We want to find PPT states $\rho^{(n)} \in \mathcal{H}^{\otimes n}$ such that the trace distance of $\rho^{(n)}$ from separable states is close to 1, for enough large numbers n. Suppose ρ is an entangled PPT state. Then $\rho^{\otimes n}$ is entangled and also PPT. We claim that the sequence of states $\rho^{(n)} = \rho^{\otimes n}$ works for us. The intuition is that for two different quantum states ρ and σ, the trace distance of $\rho^{\otimes n}$ and $\sigma^{\otimes n}$ tends to 1 as n tends to infinity. However, in this problem σ is not a fixed state and ranges over all separable states. Also, it is not obvious (and may not hold)[2] that the closest separable state to $\rho^{\otimes n}$ is of the form $\sigma^{\otimes n}$.

Another idea is to use entanglement distillation. Suppose the state ρ is distillable. It means that, having arbitrary many copies of ρ, using an LOCC map, we can obtain arbitrary many EPR pairs. Notice that LOCC maps send separable states to separable states, and the trace distance decreases under trace preserving quantum operations. Therefore, the distance of $\rho^{\otimes n}$ from separable states is bounded from below by the distance of $\text{EPR}^{\otimes m}$ from separable states, which we know is close to 1 for large numbers m. Therefore, if ρ is distillable then the trace distance of $\rho^{\otimes n}$ from separable states tends to 1.

It is well-known that PPT states are not distillable under LOCC maps. So we cannot use this idea directly. On the other hand, in this argument, the only property of LOCC maps that we use, is that they send separable states to separable states. So, we may replace LOCC maps with *non-entangling maps*, the maps that send every separable state to a separable state. Due to the seminal work of Brandao and Plenio [32, 33] every entangled state is distillable under *asymptotically* non-entangling maps[3]. Hence, by replacing LOCC maps with asymptotically non-entangling maps and repeating the previous argument, we conclude that the trace distance of $\rho^{\otimes n}$ from separable states tends to 1.

Although this idea gives a full proof of Theorem 5.2.1, we do not present it here. Instead, we use more fundamental techniques, namely, *quantum state tomography* and *quantum de Finetti theorem* [38, 83]. In fact, these two techniques are the basic ideas of the results of [32, 33] that we mentioned above. Since $\rho^{\otimes(n+k)}$ is a symmetric state, we may assume that the closest separable state to $\rho^{\otimes(n+k)}$ is also symmetric. Then by tracing out k registers[4] and using the finite quantum de Finetti theorem we conclude that the trace distance of $\rho^{\otimes(n+k)}$ from separable states is lower bounded by the trace distance of $\rho^{\otimes n}$ from separable states of the form

$$\sum_i p_i \, \sigma_i^{\otimes n}. \tag{5.4}$$

Since such a state is separable and ρ is not separable, the sum of p_i's for which σ_i is close to ρ cannot be large. On the other, if n is large, using quantum state tomography one can distinguish $\rho^{\otimes n}$ from $\sigma_i^{\otimes n}$, where σ_i is far from ρ. Therefore, the trace distance of $\rho^{\otimes n}$ and a separable state of the form of Eq. (5.4) is close to 1 for enough large n.

Notice that, in both of these arguments the only property of PPT states that we use, is that if ρ and σ are PPT, then $\rho \otimes \sigma$ is also PPT. So, we can conclude the same result for

[2] If we replace the trace distance with $E_R(\rho)$, the relative entropy of entanglement, this property does not hold [116].

[3] This is because the entanglement of distillation under asymptotically non-entangling maps is equal to the regularized relative entropy of entanglement, and this measure of entanglement is faithful, meaning that it is non-zero for every entangled state.

[4] Notice that the trace distance is monotonic decreasing under CPTP maps 1.3.1.

any separability test the satisfies this property.

5.2.2 Proof of Theorem 5.2.1

Let $\mathcal{H} = \mathcal{H}^A \otimes \mathcal{H}^B$ and assume that $d = \dim\mathcal{H} > 6$. Then there exists a PPT state $\rho_{AB} = \rho$ acting on $\mathcal{H}$ which is not separable (For example see [23]). Let

$$\epsilon = \min_{\sigma \in \mathrm{SEP}} \|\rho - \sigma\|_{\mathrm{Tr}}. \tag{5.5}$$

Since ρ is not separable, $\epsilon > 0$.

For any number n, $\rho^{\otimes n}$ can be considered as a bipartite state acting on $(\mathcal{H}^A)^{\otimes n} \otimes (\mathcal{H}^B)^{\otimes n}$, and it is a PPT state. Therefore, if we prove that the trace distance of $\rho^{\otimes n}$ from separable states tends to 1, as n goes to infinity, we are done.

Let $\sigma^{(n)}$ be the closest separable state to $\rho^{\otimes n}$. Since $\rho^{\otimes n}$ is a symmetric state, for any permutation π we have

$$\|\rho^{\otimes n} - P_\pi \sigma^{(n)} P_\pi\|_{\mathrm{Tr}} = \|\rho^{\otimes n} - \sigma^{(n)}\|_{\mathrm{Tr}}.$$

Hence, by triangle inequality

$$\|\rho^{\otimes n} - \frac{1}{n!}\sum_\pi P_\pi \sigma^{(n)} P_\pi\|_{\mathrm{Tr}} \leq \frac{1}{n!}\sum_\pi \|\rho^{\otimes n} - P_\pi \sigma^{(n)} P_\pi\|_{\mathrm{Tr}} = \|\rho^{\otimes n} - \sigma^{(n)}\|_{\mathrm{Tr}},$$

and then $\|\rho^{\otimes n} - \frac{1}{n!}\sum_\pi P_\pi \sigma^{(n)} P_\pi\|_{\mathrm{Tr}} = \|\rho^{\otimes n} - \sigma^{(n)}\|_{\mathrm{Tr}}$. This means that, we may assume that the closest separable state to $\rho^{\otimes n}$ is symmetric.

Let $\sigma^{(n+n^2)}$ be the closest symmetric separable state to $\rho^{\otimes(n+n^2)}$, and let $\mathrm{Tr}_{1\ldots n^2}\,\sigma^{(n+n^2)}$ be the state obtained by tracing out n^2 registers. We have

$$\|\rho^{\otimes(n+n^2)} - \sigma^{(n+n^2)}\|_{\mathrm{Tr}} \geq \|\rho^{\otimes n} - \mathrm{Tr}_{1\ldots n^2}\,\sigma^{(n+n^2)}\|_{\mathrm{Tr}}. \tag{5.6}$$

Using the finite quantum de Finetti theorem (Theorem 5.1.1), there exists a measure μ such that

$$\mathrm{Tr}_{1\ldots n^2}\,\sigma^{(n+n^2)} = \int \mu(d\tau)\tau^{\otimes n} + X_n, \tag{5.7}$$

where $\|X_n\|_{\mathrm{Tr}} \leq 2d\frac{n}{n+n^2}$. Hence, using Eq. (5.6), if we prove that $\|\rho^{\otimes n} - \big(\int \mu(d\tau)\tau^{\otimes n} + X_n\big)\|_{\mathrm{Tr}}$ tends to 1, as n goes to infinity, we are done.

Consider an informationally complete POVM on $\mathcal{H}^A$ and $\mathcal{H}^B$, and by taking their pairwise tensor product extend them to an informationally complete POVM on $\mathcal{H}$. Now apply quantum state tomography on $(n-1)$ copies of ρ. The outcomes of the measurements give an approximation of ρ. To be more precise, let $\{M_i\}$ be the informationally complete POVM on $\mathcal{H}$. For any sequence of outcomes $(M_{l_1}, \ldots, M_{l_{(n-1)}})$ we get to the approximation

$$\sum_i \frac{r_i}{n-1} M_i^*, \tag{5.8}$$

where r_i is the number of repetition of M_i in $(M_{l_1}, \ldots, M_{l_{(n-1)}})$. Let A_n be the sum of $(n-1)$-tuple tensor products $M_{l_1} \otimes \cdots M_{l_{(n-1)}}$ for sequences $(M_{l_1}, \ldots, M_{l_{(n-1)}})$ whose approximations, according to Eq. (5.8), are in $B_{\epsilon/2}(\rho)$, the ball of radios $\epsilon/2$ in trace distance

around ρ. Therefore, by the law of large numbers [51], $\mathrm{Tr}(A_n\rho^{\otimes(n-1)}) \to 1$ as n goes to infinity. Also for any τ far from ρ, $\mathrm{Tr}(A_n\tau^{\otimes(n-1)})$ tends to zero.

Notice that $A_n \leq I$. Hence,

$$\|\rho^{\otimes n} - \big(\int \mu(d\tau)\tau^{\otimes n} + X_n\big)\|_{\mathrm{Tr}} \geq \mathrm{Tr}\,(I \otimes A_n \cdot \rho^{\otimes n}) - \mathrm{Tr}\big[(I \otimes A_n) \cdot \big(\int \mu(d\tau)\tau^{\otimes n} + X_n\big)\big],$$

and since $\mathrm{Tr}\,(I \otimes A_n \cdot \rho^{\otimes n}) \to 1$, if we prove that

$$\mathrm{Tr}\big[(I \otimes A_n) \cdot \big(\int \mu(d\tau)\tau^{\otimes n} + X_n\big)\big] \to 0,$$

as n goes to infinity, we are done.

By Eq. (5.7), $\int \mu(d\tau)\tau^{\otimes n} + X_n$ is a separable state. Also, since we can apply quantum state tomography locally (see Section 5.1.4), at the end the outcome is a separable state. We can write the outcome, before normalization, in the form

$$\int \mu(d\tau)\mathrm{Tr}[A_n\tau^{\otimes(n-1)}]\,\tau + \widetilde{X}_n,$$

where $\|\widetilde{X}_n\|_{\mathrm{Tr}} \leq 2d\,\frac{n}{n+n^2}$. Let

$$Y_n = \int_{\tau \notin B_{\epsilon/2}(\rho)} \mu(d\tau)\mathrm{Tr}[A_n\tau^{\otimes(n-1)}]\,\tau + \widetilde{X}_n,$$

and

$$c_n = \int_{\tau \in B_{\epsilon/2}(\rho)} \mu(d\tau)\mathrm{Tr}[A_n\tau^{\otimes(n-1)}].$$

By the law of large numbers [51] there exists δ_n such that for any $\tau \notin B_{\epsilon/2}(\rho)$ we have

$$\mathrm{Tr}[A_n \cdot \tau^{\otimes(n-1)}] \leq \delta_n,$$

and $\delta_n \to 0$ as n goes to infinity. Then $\|Y_n\|_{\mathrm{Tr}} \leq \delta_n + 2d\frac{n}{n+n^2}$.

Now, the state

$$\widetilde{\tau} = \frac{1}{c_n + \mathrm{Tr}(Y_n)}\big[\int_{\tau \in B_{\epsilon/2}(\rho)} \mu(d\tau)\mathrm{Tr}[A_n\tau^{\otimes(n-1)}]\,\tau + Y_n\big]$$

is separable. On the other hand, by definition

$$\widetilde{\rho} = \frac{1}{c_n}\int_{\tau \in B_{\epsilon/2}(\rho)} \mu(d\tau)\mathrm{Tr}[A_n\tau^{\otimes(n-1)}]\,\tau$$

is in the $\epsilon/2$-neighborhood of ρ. Using Eq. (5.5) we have

$$\begin{aligned}
\epsilon &\leq \|\rho - \widetilde{\tau}\|_{\mathrm{Tr}} \\
&\leq \frac{c_n}{c_n + \mathrm{Tr}(Y_n)} \cdot \|\rho - \widetilde{\rho}\|_{\mathrm{Tr}} + \frac{|\mathrm{Tr}(Y_n)|}{c_n + \mathrm{Tr}(Y_n)} \cdot \|\rho\|_{\mathrm{Tr}} + \frac{1}{c_n + \mathrm{Tr}(Y_n)} \cdot \|Y_n\|_{\mathrm{Tr}} \\
&\leq \frac{c_n}{c_n + \mathrm{Tr}(Y_n)} \cdot \frac{\epsilon}{2} + \frac{2}{c_n + \mathrm{Tr}(Y_n)} \cdot \|Y_n\|_{\mathrm{Tr}}.
\end{aligned}$$

Hence,

$$\epsilon c_n + \epsilon \mathrm{Tr}(Y_n) \leq \frac{\epsilon}{2} c_n + 2\|Y_n\|_{\mathrm{Tr}},$$

and then

$$c_n \leq \frac{2(2+\epsilon)\|Y_n\|_{\mathrm{Tr}}}{\epsilon} \leq 6\epsilon^{-1}[\delta_n + 2d\frac{n}{n+n^2}].$$

Putting everything together we find that

$$\begin{aligned} \mathrm{Tr}\big[(I \otimes A_n) \cdot (\int \mu(d\tau)\tau^{\otimes n} + X_n)\big] &= \mathrm{Tr}\big[\int_{\tau \in B_{\epsilon/2}(\rho)} \mu(d\tau)\mathrm{Tr}[A_n \tau^{\otimes(n-1)}]\,\tau + Y_n\big] \\ &\leq c_n + \|Y_n\|_{\mathrm{Tr}} \\ &\leq (6\epsilon^{-1}+1)\cdot(\delta_n + 2d\frac{n}{n+n^2}). \end{aligned}$$

Therefore

$$\mathrm{Tr}\big[(I \otimes A_n) \cdot (\int \mu(d\tau)\tau^{\otimes n} + X_n)\big] \to 0,$$

as n goes to infinity. We are done.

5.3 Geometry of the Set of Separable States

Theorem 5.2.1 tells us that estimating the distance of a bipartite state from separable state by the distance from PPT states is not a good approximation. However, one may say the set of PPT states may be a reasonable approximation for the set of separable states in a geometrical point of view. For instance, two spheres centered at origin with radiuses 1 and 2 are far from each other, while they have the same geometric properties up to a scaler factor. In the following theorem we show that the set of separable states relative to the set of PPT states is not of this form.

By Theorem 5.2.1 the maximum distance of a PPT state from the boundary of the set of separable states is close to 1. We can think of this problem in another direction. What is the maximum distance of a state on the boundary of separable states from the boundary of PPT states? To get an intuition on this problem, we can think of the unit sphere centered at origin in $\mathbb{R}^n$, and the cube with vertices $(\pm 1, \ldots, \pm 1)$. It is easy to see that the distance of any point on the sphere from points of the cube is less than 2. However, the distance of $(1, \ldots, 1)$ from sphere is $\sqrt{n}-1$. It is because sphere and cube have totally different shapes.

Theorem 5.3.1 *Assume that $\mathcal{H} = \mathcal{H}^A \otimes \mathcal{H}^B$, and $dim\mathcal{H}^A = dim\mathcal{H}^B = d$. Then for any separable state ρ acting on $\mathcal{H}$ there exists a state σ on the boundary of the set of PPT states such that $\|\rho - \sigma\|_{Tr} \leq \frac{1}{\sqrt{d}}$.*

Proof: Let σ be an arbitrary PPT state, and $\Phi(d)$ be the maximally entangled state defined in Eq. (5.1). Then the fidelity of σ and $\Phi(d)$ is

$$F(\sigma, \Phi(d)) = [\mathrm{Tr}\,\sigma\,\Phi(d)]^{1/2} = [\mathrm{Tr}\,\sigma^{T_B}\,\Phi(d)^{T_B}]^{1/2} = [\mathrm{Tr}\,\sigma^{T_B}\,(\frac{1}{d}I - \frac{2}{d}\sum_{i<j}|\phi_{ij}\rangle\langle\phi_{ij}|)]^{1/2},$$

where $|\phi_{ij}\rangle$ is defined in Eq. (5.2). Now, using the fact that ρ^{T_B} is positive semi-definite we have

$$F(\sigma, \Phi(d)) \leq \frac{1}{\sqrt{d}}.$$

Therefore, by the well-known inequality between fidelity and trace distance, see Section 1.3.2, we have

$$\|\sigma - \Phi(d)\|_{\mathrm{Tr}} \geq 1 - F(\sigma, \Phi(d)) \geq 1 - \frac{1}{\sqrt{d}}. \tag{5.9}$$

Let ρ be an arbitrary separable state. Define $\rho_t = (1-t)\rho + t\Phi(d)$. Then $\rho_0 = \rho$ is separable and then PPT, and $\rho_1 = \Phi(d)$. Hence, there exists $0 \leq c \leq 1$ such that ρ_c is on the boundary of PPT states. Then we have

$$\|\rho - \rho_c\|_{\mathrm{Tr}} = \|\rho - \Phi(d)\|_{\mathrm{Tr}} - \|\rho_c - \Phi(d)\|_{\mathrm{Tr}} \leq 1 - (1 - \frac{1}{\sqrt{d}}) = \frac{1}{\sqrt{d}},$$

where in the last inequality we use Eq. (5.9).

□

5.4 Generalization to Other Separability Criteria

By the result of Section 5.2.2, if the dimension of the space is enough large, there exists a PPT state arbitrary far from separable states. In the proof, our candidate for such a state is $\rho^{\otimes n}$, where ρ is an entangled PPT state. Indeed, the only property of the set of PPT states that we use, is that this set is closed under tensor product. Therefore, the same argument as in the proof of Theorem 5.2.1, gives us the following general theorem.

Theorem 5.4.1 *Assume that C is a necessary but not sufficient separability criterion such that if ρ and σ satisfy C, then $\rho \otimes \sigma$ satisfies C as well. Then for any $\epsilon > 0$ there exists a state ρ that satisfies C, and whose trace distance from separable states is at least $1 - \epsilon$.*

Proof: Let ρ be an entangled state that satisfies C. Then $\rho^{\otimes n}$ satisfies C, and by the proof of Theorem 5.2.1, the trace distance of $\rho^{\otimes n}$ from separable states, tends to 1 as n goes to infinity. □

In the following theorem we prove that all separability criteria mentioned in Section 5.1.3 satisfy the assumption of Theorem 5.4.1.

Theorem 5.4.2 *For any of the separability criteria mentioned in Section 5.1.3 there exists an entangled state that passes that test while it is arbitrary far, in trace distance, from separable states.*

Proof: By Theorem 5.4.1 it is sufficent to prove that those separability criteria are closed under tensor product.

- Reduction criterion: Let X, Y, Z and W be positive semi-definite matrices such that $X \geq Y$ and $Z \geq W$. Then $(X-Y) \otimes (Z+W)$ and $(X+Y) \otimes (Z-W)$ are positive semi-definite. Therefore $X \otimes Z - Y \otimes W = \frac{1}{2}[(X-Y) \otimes (Z+W) + (X+Y) \otimes (Z-W)]$

is positive semi-definite. It means that if $X \geq Y$ and $Z \geq W$, then $X \otimes Z \geq Y \otimes U$. Now assume that ρ_{AB} and σ_{AB} pass reduction criterion. Therefore $\rho_A \otimes I \geq \rho_{AB}$ and $\sigma_A \otimes I \geq \sigma_{AB}$, and then $\rho_A \otimes \sigma_A \otimes I \geq \rho_{AB} \otimes \sigma_{AB}$. Hence, $\rho_{AB} \otimes \sigma_{AB}$ passes reduction criterion.

- Entropic criterion: It follows easily from $S_\alpha(\rho \otimes \sigma) = S_\alpha(\rho) + S_\alpha(\sigma)$.

- Majorization criterion: $x \prec y$ if and only if there exists a doubly-stochastic matrix[5] D such that $x = Dy$, see [96] page 575. Therefore, if $x \prec y$ and $x' \prec y'$, there exist D and D' such that $x = Dy$ and $x' = D'y'$. Hence $x \otimes x' = (D \otimes D')(y \otimes y')$ and then $x \otimes x' \prec y \otimes y'$. The proof follows easily using this property.

- Cross norm criterion: Using $v(X \otimes X') = v(X) \otimes v(X')$ we have $\mathcal{U}((X \otimes X') \otimes (Y \otimes Y')) = \mathcal{U}(X \otimes Y) \otimes \mathcal{U}(X' \otimes Y')$. The proof follows from this equation.

- Symmetric extension criterion: If $\rho^{(k)}$ and $\sigma^{(k)}$ are symmetric extensions of ρ and σ to k copies, respectively, then $\rho^{(k)} \otimes \sigma^{(k)}$ is a symmetric extension of $\rho \otimes \sigma$ to k copies.

□

5.5 Summary

In this Chapter we have proved that for any separability criterion that is closed under tensor product, meaning that $\rho \otimes \sigma$ passes the test if ρ and σ pass the test, the set of states that pass the test is not a good approximation of the set of separable states. In other words, all well-known algorithms for detecting entanglement, give no bound on the distance of a state from separable states. For the special case of positive partial transpose test, using Theorem 5.3.1, we have shown that the set of PPT states and separable states have totally different shapes. An interesting question to answer is to find a separability criterion that is stronger than the known ones, and also is not closed under tensor product. This problem may clarify the complexity of separability problem: is it NP-hard to decide whether there exists a separable state whose trace distance from a given state is less than a constant c or not?

[5] A matrix is called doubly-stochastic if all of whose entries are positive, and the sum of entries on any row and column is equal to 1.

Chapter 6

Conclusion

Properties of the complexity classes QMA(k) are related, on the one hand, to the theory of NP-completeness, and on the other hand, to entanglement theory. These two points of view turn these complexity classes to interesting ones. In this thesis I tried to understand these two views in order to use techniques of one of them into another.

In Chapter 2 I introduced a new QMA-complete problem, called quantum clique, which is the problem of computing the maximum number of states that are distinguishable after passing through an entanglement breaking channel. This problem is defined by a translation of clique problem in terms of zero-error information theory. It is interesting to translate other NP-complete problems in the language of quantum physics in order to find more QMA-complete problems.

In Chapter 2 I also showed that computing the Holevo capacity and minimum output entropy of quantum channels are NP-complete even for entanglement breaking channels. Indeed, computing an inverse polynomial approximation of these quantities is NP-hard. It is important that whether these two problems are hard in the case of a constant approximation or not because this question is related to the separability problem with constant gap.

In Chapter 3 I considered the multiple prover version of quantum Merlin-Arthur games, denoted QMA(k). I presented an important relation between the weak additivity of entanglement of formation and this complexity class. I showed that we can amplify the gap in QMA(2) if the weak additivity conjecture holds, and concluded that assuming the conjecture, all the QMA(k)-hierarchy collapses to QMA(2).

In Chapter 4 I improved the result of [28] by showing that $\mathsf{QMA}_{\log}(2)$ with the gap $n^{-(3+\epsilon)}$, contains NP. I proved this result based on Gurvits idea [57], who showed that separability problem is NP-complete. Indeed, $\mathsf{NP} \subseteq \mathsf{QMA}_{\log}(2)$ can be shown even with the ideas in Chapter 2 for proving the NP-hardness of computing the minimum output entropy; however, Gurvits's construction together with the PCP theorem gives a larger gap. Although it seems unlikely, it is interesting to see whether $\mathsf{QMA}_{\log}(2)$ with constant gap contains NP or not.

In Chapter 5, motivated by results of the previous chapter, I considered the problem of detecting entanglement. I proved that non of the well-known separability criteria give any bound on the distance of a bipartite state from separable states. More precisely, I showed that for any $\epsilon > 0$, there exists a PPT state whose trace distance from the set of separable states is at lease $1 - \epsilon$. This result is an evidence that separability problem even with constant gap is not easy; however, we still do not know whether it is NP-hard or not.

Appendix A

Zero-Error Capacity

In Section 2.2.1 the zero-error capacity of graphs (classical channels) as well as quantum channels have been defined. Here, I present some properties of the function $\alpha(\Phi)$, and also give a new bound on the capacity of graphs.

A.1 Computing $\alpha(\Phi)$

To compute $\Theta(\Phi)$ first we should be able to compute $\alpha(\Phi)$. Fix a quantum channel Φ with the operator sum representation

$$\Phi(\rho) = \sum_k E_k \rho E_k^\dagger,$$

where $\sum_k E_k^\dagger E_k = I$. Suppose $\alpha(\Phi) = n$. It means that, there are states $\rho_1, \dots \rho_n$ such that we can distinguish $\Phi(\rho_1), \dots \Phi(\rho_n)$ without error. Equivalently, any pair of $\Phi(\rho_i)$, $i = 1 \dots, n$ have orthogonal supports. Therefore, if we choose an arbitrary pure state $|\psi_i\rangle$ in the support of ρ_i, $i = 1 \dots, n$, all the states $\Phi(|\psi_i\rangle\langle\psi_i|)$ have orthogonal supports as well, and then they are distinguishable. Thus, instead of sending ρ_i's we can send pure states $|\psi_1\rangle, \dots |\psi_n\rangle$ without error, and we get to the following proposition appeared in [90].

Proposition A.1.1 [90] *For a quantum channel Φ, $\alpha(\Phi) \geq n$ if and only if there exist pure states $|\psi_i\rangle$, $i = 1 \dots, n$, such that they can be transmitted through Φ without error. In fact, to find $\alpha(\Phi)$ we can restrict ourselves to pure states.*

By Proposition A.1.1, the problem of finding $\alpha(\Phi)$ is equivalent to finding maximum number of pure states $|\psi_i\rangle$ such that all states $\Phi(|\psi_i\rangle\langle\psi_i|)$ have orthogonal supports. Assume that $\Phi(|\psi_i\rangle\langle\psi_i|)$, $i = 1 \dots, n$, are such states and have orthogonal supports. We know that

$$\Phi(|\psi_i\rangle\langle\psi_i|) = \sum_k E_k |\psi_i\rangle\langle\psi_i| E_k^\dagger,$$

and the support of $\Phi(|\psi_i\rangle\langle\psi_i|)$ is spanned by vectors $E_k|\psi_i\rangle$, so $\Phi(|\psi_i\rangle\langle\psi_i|)$, $i = 1, \dots n$ have orthogonal supports if and only if

$$\langle\psi_i|E_k^\dagger E_l|\psi_j\rangle = 0, \qquad \forall k, l \;\; \forall i, j \; i \neq j \tag{A.1}$$

Note that, by Eq. (A.1) for any $i \neq j$ we have

$$0 = \sum_k \langle\psi_i|E_k^\dagger E_k|\psi_j\rangle = \langle\psi_i|\sum_k E_k^\dagger E_k|\psi_j\rangle = \langle\psi_i|\psi_j\rangle,$$

and vectors $|\psi_1\rangle, \ldots |\psi_n\rangle$ should be orthonormal. Therefore, $P = \sum_i |\psi_i\rangle\langle\psi_i|$ is a projection and by Eq. (A.1), for any k, l, we have

$$PE_k^\dagger E_l P = \sum_{i,j} |\psi_i\rangle\langle\psi_i|E_k^\dagger E_l|\psi_j\rangle\langle\psi_j| = \sum_i \langle\psi_i|E_k^\dagger E_l|\psi_i\rangle|\psi_i\rangle\langle\psi_i|. \tag{A.2}$$

Theorem A.1.1 *$\alpha(\Phi) \geq n$ if and only if there is an orthogonal projection P such that $rank(P) = n$ and all operators $PE_k^\dagger E_l P$ commute.*

Before proving this theorem, note that the condition in this theorem is a weakened version of the condition in the *quantum error correction code* theorem (see [96]). In that theorem, we wanted to find a code subspace and the projection P on that subspace such that we would be able to recover *every state* in the code space after passing through channel. In this case, the condition for P is that $PE_k^\dagger E_l P = c_{kl}P$, for some constants c_{kl}. It means that, the restriction of operators $E_k^\dagger E_l$ on the subspace P should be a multiple of identity operator on P, but in Theorem A.1.1 the condition on $E_k^\dagger E_l$'s is that their restrictions on P should commute, which of course is a weakened version of previous condition. Indeed, in the quantum error correction code theorem we want to recover every state, but here we want to recover only a finite number of states. Thus, we get to a simpler condition.

Proof: First suppose $\alpha(\Phi) \geq n$. Then there are states $|\psi_i\rangle$, $i = 1 \ldots n$, that satisfy the condition of Proposition A.1.1. Hence, as we showed, $|\psi_i\rangle$'s are orthonormal and we can define the projection $P = \sum_i |\psi_i\rangle\langle\psi_i|$. rank$P = n$, and also by Eq. (A.2), all $PE_k^\dagger E_l P$'s are diagonal in the orthonormal basis $|\psi_i\rangle$, $i = 1 \ldots n$, and then they commute.

Now, suppose there is an orthogonal projection P of rank n such that $PE_k^\dagger E_l P$'s commute. It means that there is a basis $|\psi_i\rangle$, $i = 1 \ldots, n$, of unit vectors for the subspace P such that all $PE_k^\dagger E_l P$'s are diagonal in that basis, see [84]. In other words, there are constants λ_i^{kl} such that

$$PE_k^\dagger E_l P|\psi_i\rangle = \lambda_i^{kl}|\psi_i\rangle, \tag{A.3}$$

for very i and k, l. We show that we can assume $|\psi_i\rangle$'s are orthogonal. Suppose $\langle\psi_i|\psi_j\rangle \neq 0$, for some i, j, where $i \neq j$. We have

$$\langle\psi_j|PE_k^\dagger E_l P|\psi_i\rangle = \lambda_i^{kl}\langle\psi_j|\psi_i\rangle$$

and also

$$\langle\psi_i|PE_l^\dagger E_k P|\psi_j\rangle = \lambda_j^{lk}\langle\psi_i|\psi_j\rangle.$$

Therefore, $\lambda_i^{kl}\langle\psi_j|\psi_i\rangle = \overline{\lambda_j^{lk}}\langle\psi_j|\psi_i\rangle$, and then $\lambda_i^{kl} = \overline{\lambda_j^{lk}}$. Also, since each vector is not orthogonal to itself, we have $\lambda_j^{kl} = \overline{\lambda_j^{lk}}$, and then $\lambda_i^{kl} = \lambda_j^{kl}$. It means that, if $|\psi_i\rangle$ and $|\psi_j\rangle$ are not orthogonal, they belong to the same eigenspace of $PE_k^\dagger E_l P$, for every k, l. So, by replacing $|\psi_i\rangle$ and $|\psi_j\rangle$ with a linear combination of them we can assume that they are orthogonal.

Hence, without loss of generality, we may assume that all $PE_k^\dagger E_l P$'s are diagonal in the orthonormal basis $|\psi_i\rangle$, $i = 1 \ldots, n$, and they satisfy Eq. (A.3). Then

$$\langle\psi_j|E_k^\dagger E_l|\psi_i\rangle = \langle\psi_j|PE_k^\dagger E_l P|\psi_i\rangle = \lambda_i^{kl}\langle\psi_j|\psi_i\rangle = 0,$$

for all i, j, $i \neq j$. Therefore, states $\Phi(|\psi_i\rangle\langle\psi_i|)$ have orthogonal supports, and $|\psi_1\rangle, \ldots |\psi_n\rangle$ are distinguishable after passing through channel. Hence $\alpha(\Phi) \geq n$.

□

Corollary A.1.1 *Suppose Ψ is a channel with operator sum representation*

$$\Psi(\rho) = \sum_l F_l \rho F_l^\dagger,$$

and each F_l is a linear combination of E_k's, where E_k's are operators in the operator sum representation of another channel Φ. Then $\alpha(\Psi) \geq \alpha(\Phi)$.

Proof: Suppose $\alpha(\Phi) = n$, then by Theorem A.1.1 there is a projection P of rank n such that $PE_k^\dagger E_l P$'s commute. F_l's are linear combination of E_k's. Therefore $PF_k^\dagger F_l P$'s are linear combination of $PE_k^\dagger E_l P$'s, and then they all commute. So by Theorem A.1.1, $\alpha(\Psi) \geq n$.

□

Using this corollary, in order to estimate $\alpha(\Phi)$ we can restrict ourselves to the channels in which their operators in the operator sum representation (errors) are in *Pauli group*. It is because Pauli operators consist a basis for the space of all operators. This is exactly the same idea as we have in the theory of quantum error correcting codes.

Here, by the same idea, we want to restrict everything to Pauli group and try to estimate $\alpha(\Phi)$. Namely, we assume that E_k's, errors in the quantum channels Φ, are Pauli operators, and also we assume that the projection P in Theorem A.1.1 is a projection on a stabilizer code subspace. We know that such a projection P can be written as

$$P = \prod_i \frac{I + g_i}{2},$$

where g_i's are Pauli operators, $g_i^2 = I$, for every i, and also they all commute ($g_i g_j = g_j g_i$).

Suppose for some k and l, $E_k^\dagger E_l$ does not commute with all g_i's. So, there is j such that $E_k^\dagger E_l$ and g_j do not commute, and because they are in Pauli group they anti-commute, i.e. $E_k^\dagger E_l g_j = -g_j E_k^\dagger E_l$. Then

$$\begin{aligned} PE_k^\dagger E_l P &= \prod_{i,i\neq j} \frac{I+g_i}{2} \left(\frac{I+g_j}{2} E_k^\dagger E_l \frac{I+g_j}{2}\right) \prod_{i,i\neq j} \frac{I+g_i}{2} \\ &= \prod_{i,i\neq j} \frac{I+g_i}{2} \left(\frac{I+g_j}{2}\frac{I-g_j}{2}\right) E_k^\dagger E_l \prod_{i,i\neq j} \frac{I+g_i}{2} = 0,, \end{aligned} \tag{A.4}$$

because $(I+g_j)(I-g_j) = I - g_j^2 = 0$. So $PE_k^\dagger E_l P = 0$ and it commutes with any operator. In order to satisfy the condition of Theorem A.1.1, we should focus on $E_k^\dagger E_l$'s that commute with all g_i's.

Suppose for some k, l and k', l', $E_k^\dagger E_l$ and $E_{k'}^\dagger E_{l'}$ anti-commute, but they both commute with g_i's. We have

$$(PE_k^\dagger E_l P)(PE_{k'}^\dagger E_{l'} P) = P(E_k^\dagger E_l)(E_{k'}^\dagger E_{l'})$$
$$= -P(E_{k'}^\dagger E_{l'})(E_k^\dagger E_l) = -(PE_{k'}^\dagger E_{l'} P)(PE_k^\dagger E_l P). \quad (A.5)$$

Then $PE_k^\dagger E_l P$ and $PE_{k'}^\dagger E_{l'} P$ anti-commute.

Theorem A.1.2 *Suppose in a channel Φ all the errors E_k are in Pauli group. Also assume $g_1 \ldots, g_n$ are Pauli operators, where $g_i^2 = I$, for every i, and they all commute. Let $S = \langle g_1 \ldots, g_n \rangle$ be the subgroup generated by g_i's, and $N(S)$ be its normalizer group, i.e. set of Pauli operators that commute with all g_i's*

$$N(S) = \{h \in \textit{Pauli group} : hg_i = g_i h,\ i = 1 \ldots n\}.$$

Then $P = \prod_i \frac{I+g_i}{2}$, the projection over the common eigenspace one of all g_i's, satisfies Theorem A.1.1 if and only if all operators in $\{E_k^\dagger E_l : k,\ l\} \cap N(S)$ commute.

Again, note that the condition in this theorem is a weakened version of the condition in the theory of stabilizer codes. In that case, the subspace P is a protected subspace if

$$\{E_k^\dagger E_l : k,\ l\} \cap N(S) \subseteq S,$$

and because S is abelian, operators in $\{E_k^\dagger E_l : k,\ l\} \cap N(S)$ commute automatically.
Proof: For some k, l if $E_k^\dagger E_l \notin N(S)$, there is at least one j such that $g_j E_k^\dagger E_l = -E_k^\dagger E_l g_j$. Then by Eq. (A.4), $PE_k^\dagger E_l P = 0$, and it commutes with every operator. On the other hand, for any two $E_k^\dagger E_l, E_{k'}^\dagger E_{l'} \in N(S)$ using Eq. (A.5), $PE_k^\dagger E_l P$ and $PE_{k'}^\dagger E_{l'} P$ commute if and only $E_k^\dagger E_l$ and $E_{k'}^\dagger E_{l'}$ commute. So, $\{PE_k^\dagger E_l P : k,\ l\}$ is a commutative set if and only if $\{E_k^\dagger E_l : k,\ l\} \cap N(S)$ is a commutative set.

□

A.2 Zero-Error Capacity of Graphs and C-Q Channels

Computing the zero-error capacity is a hard problem, both in the classical case and the quantum case. In this section, we restrict ourselves to a special case of channels, called *classically-quantum channels* (c-q channels), and find a relation between the zero-error capacity of these channels and capacity of graphs.

In a c-q channel, first we measure the input state in an orthonormal basis, and then send a state as the output based on the outcome of the measurement. If the basis of the measurement is $|k\rangle$, $k = 1, \ldots m$, and output states are σ_k, the channel is

$$\Phi(\rho) = \sum_k \langle k|\rho|k\rangle \sigma_k.$$

Again, for simplicity, assume that states σ_k are pure, and let $\sigma_k = |\phi_k\rangle\langle\phi_k|$, for $1 \leq k \leq m$. Thus Φ can be written as

$$\Phi(\rho) = \sum_k E_k \rho E_k^\dagger,$$

where $E_k = |\phi_k\rangle \otimes \langle k|$. Then $E_k^\dagger E_l = \langle \phi_k|\phi_l\rangle |k\rangle\langle l|$, and depending on whether $|\phi_k\rangle$ and $|\phi_l\rangle$ are orthogonal or not, $E_k^\dagger E_l$ is either zero or a non-zero multiple of $|k\rangle\langle l|$. To keep

these numbers, let us define a graph G on the vertex set $\{v_1, \ldots v_m\}$ and say v_k and v_l are adjacent if $\langle\phi_k|\phi_l\rangle \neq 0$.

Assume $\alpha(\Phi) = n$. Then, there are pure states $|\psi_i\rangle$, $i = 1, \ldots n$, such that they can be transmitted through channel with no error. This is equivalent to $\langle\psi_i|E_k^\dagger E_l|\psi_j\rangle = 0$, for every k, l and every i, j, $i \neq j$. It means that, for every two adjacent vertices v_k and v_l we have $\langle\psi_i|k\rangle\langle l|\psi_j\rangle = 0$, for every i, j, $i \neq j$.

If $k = l$, we find that for each k there is at most one i where $|\psi_i\rangle$ is not orthogonal to $|k\rangle$. So if we define $C_i = \{v_k : \langle\psi_i|k\rangle \neq 0\}$, then $C_i \cap C_j = \emptyset$, for $i \neq j$. Note that all C_i's are non-empty, because $\{|1\rangle, \ldots |m\rangle\}$ is a basis for the vector space of input states.

Now consider two adjacent vertices v_k and v_l. For every i, j, $i \neq j$, we have $\langle\psi_i|k\rangle\langle l|\psi_j\rangle = 0$. Hence, either $\langle\psi_i|k\rangle = 0$ or $\langle\psi_j|l\rangle = 0$. Equivalently there is no edges between vertices in C_i and C_j.

By the above discussion, we conclude that there are disjoint subsets $C_1, \ldots C_n$ of vertices of G such that there is no edge between them. If there are such subsets, we can pick one vertex from each C_i, and get to an independent subset of vertices of size n in G, and then $\alpha(G) \geq \alpha(\Phi)$. On the other hand, if $W \subseteq \{v_1, \ldots v_m\}$ is an independent set of size $\alpha(G)$, define $|\psi_i\rangle = |i\rangle$ for any i where $v_i \in W$. Then for any two different $|\psi_i\rangle, |\psi_j\rangle$, and any edge $v_k v_l$, since v_i and v_j are not adjacent, either $i \neq k$ or $j \neq l$, and therefore $\langle\psi_i|k\rangle\langle l|\psi_j\rangle = 0$. So, we can transmit $|\psi_i\rangle$'s without error, and $\alpha(\Phi) \geq \alpha(G)$.

Theorem A.2.1 *Consider a c-q channel Φ, where*

$$\Phi(\rho) = \sum_{k=1}^{m} \langle k|\rho|k\rangle |\phi_k\rangle\langle\phi_k|.$$

Define a graph on the vertex set $\{v_1, \ldots v_m\}$, in which two vertices v_k and v_l are adjacent if $\langle\phi_k|\phi_l\rangle \neq 0$. Then $\alpha(\Phi) = \alpha(G)$ and $\Theta(\Phi) = \Theta(G)$.

Proof: By the above discussion $\alpha(\Phi) = \alpha(G)$. To show $\Theta(\Phi) = \Theta(G)$, notice that $\Phi^{\otimes r}$ is also a c-q channel with pure outcome states and the associated graph of $\Phi^{\otimes r}$, as defined in the theorem, is $G^{\otimes r}$. Also, by the construction discussed above, the $\alpha(G^{\otimes r})$ input states that can be sent through $\Phi^{\otimes r}$ with no error, are all product states. In fact, they are states of the form $|i_1\rangle|i_2\rangle \cdots |i_n\rangle$, where $|i_j\rangle$, $1 \leq j \leq n$, are vectors in the basis. So $\alpha(\Phi^{\otimes r}) = \alpha(G^{\otimes r})$, for every r, and then $\Theta(\Phi) = \Theta(G)$.

□

This theorem reduces the problem of computing the zero-error capacity of c-q channels to the problem of graph capacity.

A.3 A New Bound on the Capacity of Graphs

It it clear that the zero-error capacity of a quantum channel is not greater than its Holevo capacity. Holevo capacity is the maximum rate of transmitting classical information through a quantum channel with arbitrary small error, using product states. Thus, it contains transmitting information with zero-error, and then zero-error capacity of a quantum channel is not greater than its Holevo capacity

$$\Theta(\Phi) \leq \chi(\Phi) = \max_{p_i, |\psi_i\rangle} S\big(\Phi(\sum_i p_i |\psi_i\rangle\langle\psi_i|)\big) - \sum_i p_i S\big(\Phi(|\psi_i\rangle\langle\psi_i|)\big), \tag{A.6}$$

where the maximum is taken over all ensemble of pure states $|\psi_i\rangle$ and positive numbers p_i such that $\sum_i p_i = 1$. Let us apply this inequality on the special case of c-q channels.

Consider the c-q channel Φ,

$$\Phi(\rho) = \sum_{k=1}^{m} \langle k|\rho|k\rangle |\phi_k\rangle\langle\phi_k|, \tag{A.7}$$

We have

$$\chi(\Phi) = \max_{p_i,|\psi_i\rangle} S\big(\sum_k (\sum_i p_i |\langle\psi_i|k\rangle|^2)|\phi_k\rangle\langle\phi_k|\big) - \sum_i p_i S\big(\sum_k |\langle\psi_i|k\rangle|^2 |\phi_k\rangle\langle\phi_k|\big). \tag{A.8}$$

In order to get to the maximum point, let $\lambda_1, \ldots \lambda_m$ be non-negative real numbers such that $\lambda_k^2 = \sum_i p_i |\langle\psi_i|k\rangle|^2$. Then $\sum_k \lambda_k^2 = 1$ and for the ensemble $\{\lambda_k^2, |k\rangle\}$ we have

$$S\big(\Phi(\sum_k \lambda_k^2 |k\rangle\langle k|)\big) - \sum_k \lambda_k^2 S(\Phi(|k\rangle\langle k|)) = S\big(\sum_k \lambda_k^2 |\phi_k\rangle\langle\phi_k|\big) - \sum_k \lambda_k^2 S(|\phi_k\rangle\langle\phi_k|)$$

$$= S\big(\sum_k \lambda_k^2 |\phi_k\rangle\langle\phi_k|\big) = S\big(\sum_k (\sum_i p_i |\langle\psi_i|k\rangle|^2)|\phi_k\rangle\langle\phi_k|\big).$$

Comparing to Eq. (A.8), we conclude that

$$\chi(\Phi) = \max_{\lambda_k} S(\sum_k \lambda_k^2 |\phi_k\rangle\langle\phi_k|),$$

where the maximum is taken over all non-negative real numbers λ_k such that $\sum_k \lambda_k^2 = 1$.

Now assume that $\rho^{12} = |v\rangle\langle v|$ is a pure state in a composite system, where $|v\rangle = \sum_k \lambda_k |k\rangle|\phi_k\rangle$. Since ρ^{12} is pure $S(\rho^1) = S(\rho^2)$ (see Section 1.4.1), and we have

$$\rho^2 = \sum_k \lambda_k^2 |\phi_k\rangle\langle\phi_k|,$$

and

$$\rho^1 = \sum_{k,l} \lambda_k \lambda_l \langle\phi_l|\phi_k\rangle |k\rangle\langle l|.$$

Hence,

$$\chi(\Phi) = \max_{\lambda_k} S(\rho^2) = \max_{\lambda_k} S(\rho^1) = \max_{\lambda_k} S\Big(\sum_{k,l} \lambda_k \lambda_l \langle\phi_l|\phi_k\rangle |k\rangle\langle l|\Big).$$

But $\rho^1 = \sum_{k,l} \lambda_k \lambda_l \langle\phi_l|\phi_k\rangle |k\rangle\langle l|$ as a matrix in the basis $|1\rangle, \ldots |m\rangle$ is equal to $\rho^1 = \Lambda B \Lambda$, where B is an $m \times m$ matrix such that $B_{kl} = \langle\phi_l|\phi_k\rangle$, and Λ is a diagonal matrix with λ_k's on its diagonal. Thus we have

$$\chi(\Phi) = \max_{\Lambda} S(\Lambda B \Lambda), \tag{A.9}$$

and then

$$\Theta(\Phi) \leq \max_{\Lambda} S(\Lambda B \Lambda).$$

Using the above inequality, we want to state an upper bound on the capacity of a fixed

graph G. Before that we need a definition.

Definition A.3.1 *For a graph G on the vertex set $\{v_1, \ldots v_m\}$, define $\mathcal{M}(G)$ to be set of all positive semidefinite $m \times m$ matrices B, where all of whose diagonal entries are one, i.e. $B_{kk} = 1$, for every k, $1 \leq k \leq m$, and $B_{kl} = 0$ if v_k and v_l are not adjacent.*

Theorem A.3.1 *For a graph G on the vertex set $\{v_1, \ldots v_m\}$ we have*

$$\Theta(G) \leq \vartheta(G) = \min_B \max_\Lambda S(\Lambda B \Lambda), \tag{A.10}$$

where minimum is taken over all $B \in \mathcal{M}(G)$, and Λ ranges over all $m \times m$ diagonal matrices, with λ_k on the k-th entry of the diagonal such that $\lambda_k \geq 0$, for every k, $1 \leq k \leq m$, and $Tr\Lambda^2 = \sum_k \lambda_k^2 = 1$.

Proof: First, suppose we have proved that

$$\Theta(G) \leq \max_\Lambda S(\Lambda B \Lambda),$$

for every B and Λ having the conditions in the theorem together with the extra condition that $B_{kl} \neq 0$ if v_k and v_l are adjacent, ($B_{kl} \neq 0$ if and only if v_k and v_l are adjacent). Then we have

$$\Theta(G) \leq \inf_B \max_\Lambda S(\Lambda B \Lambda),$$

where infimum is taken over all such matrices B. But note that the set of these matrices is not closed, and its closure is $\mathcal{M}(G)$. Therefore we get to the statement in the theorem.

Now, suppose B is a positive semidefinite matrix such that $B_{kl} \neq 0$ if and only if v_k and v_l are adjacent, and $B_{kk} = 1$. It is well-known that for any positive semidefinite matrix B there is a matrix A such that $B = A^\dagger A$. In fact, if we let $|\phi_k\rangle$ to be the conjugate of k-th column of A then $B_{kl} = \overline{\langle \phi_k | \phi_l \rangle} = \langle \phi_l | \phi_k \rangle$. Also, since the diagonal of B is one, we get $\langle \phi_k | \phi_k \rangle = 1$ and $|\phi_k\rangle$'s are unit vectors. Hence, if we define the channel Φ as in Eq. (A.7), then by Theorem A.2.1 we have $\Theta(G) = \Theta(\Phi)$. On the other hand, by Eq. (A.6), $\Theta(\Phi) \leq \chi(\Phi)$, and using Eq. (A.9) we get to

$$\Theta(G) = \Theta(\Phi) \leq \chi(\Phi) = \max_\Lambda S(\Lambda B \Lambda).$$

Now taking infimum over all such matrices B, and equivalently taking minimum over all $B \in \mathcal{M}(G)$ we get to

$$\Theta(G) \leq \vartheta(G).$$

□

This theorem gives us an upper bound on the capacity of a graph. It is important because computing $\Theta(G)$ is a hard problem even for simple graphs. For instance, computing $\Theta(C_5)$, cycle of length five, was open for many years until Lovász found an upper bound for the capacity of graphs, [87]. This upper bound in terms of our notation is

$$\Theta(G) \leq \max_{B \in \mathcal{M}(G^c)} \log(\mathrm{Tr}(BJ)). \tag{A.11}$$

Here, G^c is the complement of graph G and J is a matrix all of whose entries are equal to one. But we do not know any relation between Lovász's bound and our bound in Theorem A.3.1.

A.4 Computing $\vartheta(G)$

One the most well-known methods for computing the capacity of a graph is Lovász's bound, Eq. (A.11), but using this bound we can not even find $\Theta(C_7)$, the capacity of the cycle of length 7. Thus, in this theory other bounds for $\Theta(G)$ would be helpful. Here, we introduced the bound $\vartheta(G)$. Thus computing it efficiently would be a great advantage in this theory.

It seems that finding $\vartheta(G)$ is an easier problem than finding $\Theta(G)$. Because it has an algebraic expression, but $\Theta(G)$ is in term of $\alpha(G^{\otimes n})$ that may have unusual behaviors, see [8]. We do not know how to find $\vartheta(G)$ in general. But we show that at least for a subclass of graphs we can compute it, and show that it is equal to $\Theta(G)$. Meaning that, at least for this subclass the bound (A.10) is tight. First the definition of this subclass.

Definition A.4.1 *Let G be a graph with vertex set V. A subset $U \subseteq V$ is called a clique if the subgraph induced by U is a complete graph.*

A clique cover of size k is a partition of V into $V_1 \cup V_2 \cup \ldots V_k$ such that each V_i, $1 \leq i \leq k$, is a clique. Also, the clique cover number of G is the number of cliques in a smallest clique cover of G.

It is obvious that in a clique cover of G two non-adjacent vertices can not be in a same clique. Therefore, all $\alpha(G)$ vertices in the maximal independent set are in different cliques of a clique cover. It means that clique cover number of G is at least $\alpha(G)$.

Theorem A.4.1 *Assume the clique cover number of G is equal to $\alpha(G)$. Then $\log \alpha(G) = \Theta(G) = \vartheta(G)$, and for this subclass of graphs the bound of (A.10) is tight.*

Before getting to the proof note that, this subclass of graphs are exactly those graphs that Shannon in [107] could compute their capacity. In fact, all graphs with at most four vertices have the same clique cover number and independence number, and Shannon using this fact could find the capacity of all these graphs. But C_5, cycle of length five does not have this property, and its capacity was open until Lovász found his famous bound.

Proof: First of all note that, if G is equal to the disjoint union of k cliques then $\Theta(G) = \log k$. It is because $G^{\otimes n}$ is also equal to the disjoint union of k^n cliques, and then $\alpha(G^{\otimes n}) = k^n$. Hence $\Theta(G) = \log \alpha(G) = \log k$.

Now suppose the clique cover number of G is equal to $\alpha(G)$, and $V = V_1 \cup V_2 \cup \ldots V_{\alpha(G)}$ is a partition of V into cliques. In this case, by deleting edges between V_i and V_j where $i \neq j$, we get to an other graph G' which is the union of $\alpha(G)$ cliques, and obviously $\Theta(G') \geq \Theta(G)$. Therefore

$$\log \alpha(G) = \log \alpha(G') = \Theta(G') \geq \Theta(G) \geq \log \alpha(G),$$

and then $\Theta(G) = \log \alpha(G)$.

So it remains to show that for such a graph $\vartheta(G) = \log \alpha(G)$. Let B_0 be a block diagonal matrix where the i-th block of B_0 is a $|V_i| \times |V_i|$ matrix all of whose entries are one. Then B_0 is positive semidefinite and $B_0 \in \mathcal{M}(G)$. Let Λ be a diagonal matrix with non-negative entries such that $\mathrm{Tr}(\Lambda^2) = 1$. It is easy to see that $\Lambda B_0 \Lambda$ is again a block diagonal matrix and all of whose blocks are rank-one. Indeed, if we let Λ_i to be the restriction of Λ on the entries in V_i then all eigenvalues of the i-th block of $\Lambda B_0 \Lambda$ are zero except one, which is $\mathrm{Tr}(\Lambda_i^2)$. Therefore

$$S(\Lambda B_0 \Lambda) = \sum_{i=1}^{\alpha(G)} -\mathrm{Tr}(\Lambda_i^2) \log(\mathrm{Tr}(\Lambda_i^2)),$$

and since $\sum_i \mathrm{Tr}(\Lambda_i^2) = \mathrm{Tr}(\Lambda_2) = 1$ by letting $p_i = \mathrm{Tr}(\Lambda_i^2)$ we have

$$\max_{\Lambda} S(\Lambda B_0 \Lambda) = \max_{\{p_1,\dots p_{\alpha(G)}\}} \sum_{i=1}^{k} -p_i \log(p_i),$$

where the maximum is taken over all non-negative $p_1, \dots p_{\alpha(G)}$ such that $\sum_i p_i = 1$. It is a well-known result in information theory that the maximum is on the point $p_1 = \dots p_{\alpha(G)} = 1/\alpha(G)$ and

$$\max_{\Lambda} S(\Lambda B_0 \Lambda) = \max_{\{p_1,\dots p_{\alpha(G)}\}} \sum_{i=1}^{\alpha(G)} -p_i \log(p_i) = \log \alpha(G).$$

Hence

$$\log \alpha(G) \leq \Theta(G) \leq \vartheta(G) = \min_{B \in \mathcal{M}(G)} \max_{\Lambda} S(\Lambda B \Lambda) \leq \max_{\Lambda} S(\Lambda B_0 \Lambda) = \log \alpha(G),$$

and then $\Theta(G) = \vartheta(G) = \log \alpha(G)$.

□

This theorem is an evidence that $\vartheta(G)$ can be helpful for computing the capacity of graphs. But we should be able to compute $\vartheta(G)$ itself. Indeed, we should find the points $B \in \mathcal{M}(G)$ and Λ that $\min_B \max_\Lambda S(\Lambda B \Lambda)$ takes its optimum value.

Proposition A.4.1 *Suppose G is a graph with at least one edge. Then in*

$$\vartheta(G) = \min_{B} \max_{\Lambda} S(\Lambda B \Lambda)$$

the optimum point is taken at a singular (non-invertible) matrix B.

Proof: Let B', B'' be two positive semidefinite matrices. As we saw in Section A.3, there are c-q channels Φ' and Φ'' such that $\chi(\Phi') = \max_\Lambda S(\Lambda B' \Lambda)$, and $\chi(\Phi') = \max_\Lambda S(\Lambda B' \Lambda)$. In fact, there are states $|\phi'_1\rangle, \dots |\phi'_m\rangle$ and $|\phi''_1\rangle, \dots |\phi''_m\rangle$ such that $B'_{kl} = \langle \phi'_l | \phi'_k \rangle$, $B''_{kl} = \langle \phi''_l | \phi''_k \rangle$, and

$$\Phi'(\rho) = \sum_{k=1}^{m} \langle k|\rho|k\rangle |\phi'_k\rangle\langle \phi'_k|,$$

$$\Phi''(\rho) = \sum_{k=1}^{m} \langle k|\rho|k\rangle |\phi''_k\rangle\langle \phi''_k|.$$

Let $|\phi_k\rangle = |\phi'_k\rangle \otimes |\phi''_k\rangle$, $k = 1, \dots m$, and define the channel

$$\Phi(\rho) = \sum_{k=1}^{m} \langle k|\rho|k\rangle |\phi_k\rangle\langle \phi_k|,$$

with the corresponded matrix B where

$$B_{kl} = \langle \phi_l | \phi_k \rangle = \langle \phi'_l | \phi'_k \rangle \langle \phi''_l | \phi''_k \rangle = B'_{kl} B''_{kl}. \tag{A.12}$$

It is not hard to see that $\chi(\Phi) \geq \chi(\Phi')$. Indeed, for any state ρ, $\Phi(\rho)$ describes the state of two particles together, and $\Phi'(\rho)$ is the state of the first one, i.e. $\Phi'(\rho) = \mathrm{Tr}_2(\Phi(\rho))$. Then any protocol for Φ' also works for Φ just by discarding the second particle. Therefore $\chi(\Phi) \geq \chi(\Phi')$, and by Eq. (A.9)

$$\max_{\Lambda} S(\Lambda B \Lambda) \geq \max_{\Lambda} S(\Lambda B' \Lambda). \tag{A.13}$$

On the other hand, by Eq. (A.12) if $B' \in \mathcal{M}(G)$, $B \in \mathcal{M}(G)$. So that, if B is the the optimum point in $\vartheta(G)$, then B' is also optimal.

Now assume $B \in \mathcal{M}(G)$ is non-singular and is the optimum point in $\vartheta(G)$. Since G is not the empty graph, B contains at least one non-zero off diagonal entry, say B_{kl}. Also, since B is non-singular if we replace B_{kl} and B_{lk} by other numbers enough closed to them, then B is still non-negative, and then is in $\mathcal{M}(G)$. Let B' be a matrix which is equal to B except on entries kl, lk. Also let B'' be a matrix all of whose entries are one except kl, lk. It is not hard to see that B'_{kl} and B''_{kl} can be chosen in such a way that $B_{kl} = B'_{kl} B''_{kl}$, and both B' and B'' be non-negative, and B' be singular. Then by the definition of B' it is in $\mathcal{M}(G)$, and Eq. (A.13) holds. So B' is singular and takes the optimum.

□

Appendix B

Post Selection

PostBQP is the usual complexity class BQP together with the extra ability of post-selection, meaning that in PostBQP if we have a state of the form $\alpha|0\rangle|\psi_0\rangle + \beta|1\rangle|\psi_1\rangle$, where β is non-zero, then regardless of how large is $|\beta|$, we can post-select the state to be $|1\rangle|\psi_1\rangle$. A more precise definition follows.

Definition B.0.2 PostBQP *is the class of languages L that have a quantum polynomial time circuit such that for every $x \in \{0,1\}^n$*

- *The first qubit of the final state has a non-zero probability to be measured $|1\rangle$.*
- *If $x \in L$, then conditioned on the first qubit being $|1\rangle$, the second qubit is $|1\rangle$ with probability at least $2/3$.*
- *If $x \notin L$, then conditioned on the first qubit being $|1\rangle$, the second qubit is $|1\rangle$ with probability at most $1/3$.*

Since, post-selection has a natural physical meaning, we may define an oracle Post that given a quantum state $\alpha|0\rangle|\psi_0\rangle + \beta|1\rangle|\psi_1\rangle$, where β is non-zero, it returns the state $|1\rangle|\psi_1\rangle$. By the definition of PostBQP we have $\mathsf{PostBQP} = \mathsf{BQP}^{\mathsf{Post}}$.

PostBQP has been first defined by Aaronson [1], who has shown the following.

Theorem B.0.2 [1] $\mathsf{PostBQP} = \mathsf{PP}$.

The idea of post-selection can be naturally extended to any quantum complexity class such as QMA. Thus, we define $\mathsf{PostQMA}(k)$ to be the class of languages L that have a QMA protocol with k Merlins, together with the ability of post-selection for Arthur.

Post-selection is a physical tool for implementing quantum algorithms that have exponentially small gaps. This point of view is highlighted in the proof of Theorem B.0.2, so in order to handle post-selection in quantum complexity theory, it would be helpful to identify it with the ability to recognize exponentially small gaps.

Theorem B.0.3 $\mathsf{PostQMA}(k, 2/3, 1/3) = \mathsf{QMA}(k, 1/2 + 2^{-n}, 1/2 - 2^{-n})$.

Proof: Consider a language L in $\mathsf{QMA}(k, 1/2 + 2^{-n}, 1/2 + 2^{-n})$. In the $\mathsf{QMA}(k)$ protocol we may assume that the outcome of algorithm is determined by the measurement of the first qubit. Let $x \in \{0,1\}^n$, and suppose that the final state of the algorithm (before measurement), applied to x, is of the form

$$\alpha|0\rangle|\psi_0\rangle + \beta|1\rangle|\psi_1\rangle.$$

We have $|\beta|^2 \geq 1/2 + 2^{-n}$ if $x \in L$, and $|\beta|^2 < 1/2 - 2^{-n}$ if $x \notin L$. To amplify the gap, add m qubits to the system, all prepared in $|0\rangle$, and conditioned on the first qubit apply Hadamard gate to all of them. We get to the state

$$\frac{1}{2^{m/2}}\alpha|0\rangle|\psi_0\rangle \sum_{y\in\{0,1\}^m} |y\rangle + \beta|1\rangle|\psi_1\rangle|00\ldots0\rangle.$$

Now, post-select the last m qubits to be $|00\ldots0\rangle$, and prepare the following state

$$\frac{1}{\sqrt{\frac{|\alpha|^2}{2^m} + |\beta|^2}}\left(\frac{1}{2^{m/2}}\alpha|0\rangle|\psi_0\rangle + \beta|1\rangle|\psi_1\rangle\right).$$

A careful choice of m and a straightforward calculation show that in the above state the gap has been amplified to a constant. Hence, $\supseteq$ holds.

By the same argument we can amplify the gap in any PostQMA protocol, and then $\subseteq$ follows by the definition.

□

According to the proof, in fact, this theorem says that PostQMA(k, a, b) does not change for different values of a, b, provided that they have at least an inverse exponential gap. Hence, PostQMA(k) is an enough robust complexity class, and some questions arise naturally about that. First of all, it is clear that PostQMA$(k) \subseteq$ PostQMA$(k+1)$, but are these inclusions strict or some of them are equality? Second, what is the relation between these complexity classes and others? Here, we state two results to answer these questions.

Theorem B.0.4 PostQMA $\subseteq$ EXP.

Proof: By Theorem B.0.3, PostQMA is QMA with exponentially small gap. On the other hand, it is well known that QMA is a semidefinite programing of exponential size, [6, 88]. We are done.

□

Theorem B.0.5 PostQMA(k) = NEXP, *for any* $k \geq 2$.

Proof: By definition PostQMA(2) $\subseteq$ PostQMA$(k) \subseteq$ NEXP, for any $k \geq 2$. Now using Theorem 4.2.1 (in fact Theorem 4.1.3), we have NEXP $\subseteq$ PostQMA(2). We are done.

□

Bibliography

[1] Scott Aaronson, *Quantum Computing, Postselection, and Probabilistic Polynomial-Time*, quant-ph/0412187

[2] Scott Aaronson, Salman Beigi, Andrew Drucker, Bill Fefferman and Peter W. Shor, *The Power of Unentanglement*, Proceedings of IEEE Complexity 2008, pp. 223-236, arXiv:0804.0802

[3] S. Aaronson and G. Kuperberg. Quantum versus classical proofs and advice. *Theory of Computing*, 3(7):129–157, 2007. Previous version in Proceedings of CCC 2007. quant-ph/0604056.

[4] L. Adleman, J. DeMarrais, and M. Huang, *Quantum computability*, SIAM Journal on Computing 26:1524-1540, 1997.

[5] M. Agrawal, N. Kayal, N. Saxena, *PRIMES is in P*, Annals of Mathematics 160 (2004), no. 2, pp. 781793.

[6] Dorit Aharonov and Tomer Naveh, *Quantum NP - A Survey*, quant-ph/0210077

[7] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, O. Regev, *Adiabatic Quantum Computation is Equivalent to Standard Quantum Computation*, SIAM Journal of Computing, Vol. 37, Issue 1, p. 166-194 (2007)

[8] N. Alon and E. Lubetzky, The Shannon capacity of a graph and the independence numbers of its powers, *IEEE Transactions on Information Theory*, 52 (2006), 2172-2176.

[9] Suguru Arimoto, *An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels*, IEEE Tkans. on Inform. Theory, IT-18, pp. 14-20, January 1972.

[10] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy, *Proof verification and hardness of approximation problems*, Journal of the ACM 45(3):501-555, 1998

[11] S. Arora and S. Safra, Probabilistic checking of proofs: A new characterization of NP, Journal of the ACM, 45(1):70-122, 1998

[12] Stanislaw Szarek, *The volume of separable states is super-doubly-exponentially small*, Phys. Rev. A 72, 032304 (2005)

[13] Guillaume Aubrun and Stanislaw J. Szarek, *Tensor products of convex sets and the volume of separable states on N qudits*, Phys. Rev. A. 73, 022109 (2006)

[14] Laszlo Babai L., *Trading Group Theory for Randomness*, Proc. 17th STOC, 1985, pp. 421429.

[15] L. Babai, L. Fortnow, and C. Lund, *Nondeterministic exponential time has two-prover interactive protocols*, Computational Complexity 1:3-40, 1991

[16] Salman Beigi, NP *vs* $QMA_{\log}(2)$, arXiv:0810.5109

[17] Salman Beigi and Peter W. Shor, *On the Complexity of Computing Zero-Error and Holevo Capacity of Quantum Channels*, arXiv:0709.2090

[18] Salman Beigi and Peter W. Shor, *Approximating the Set of Separable States Using the Positive Partial Transpose Test*, arXiv:0902.1806

[19] C.H. Bennett, H.J. Bernstein, S. Popescu, and B. Schumacher, *Concentrating Partial Entanglement by Local Operations*, Phys. Rev. A 53, 2046 (1996).

[20] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, W. K. Wootters, *Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels*, Phys. Rev. Lett. 70, 1895-1899 (1993)

[21] C.H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. Smolin, and W.K. Wootters. *Purifcation of Noisy Entanglement and Faithful Teleportation via Noisy Channels*, Phys. Rev. Lett. 78, 2031 (1996).

[22] C.H. Bennett, D.P. DiVincenzo, J. Smolin, and W.K. Wootters, *Mixed state entanglement and quantum error correction*, Phys. Rev. A 54, 3824 (1996).

[23] C. H. Bennett, D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin and B. M. Terhal, *Unextendible Product Bases and Bound Entanglement*, Phys. Rev. Lett. 82 (1999) 5385

[24] M. Ben-Or, S. Goldwasser, J. Kilian, and A. Wigderson, *Multi-prover interactive proofs: how to remove intractability*, Proceedings of ACM STOC'88, pp. 113-131, 1988.

[25] A. Ben-Tal and A. Nemirovski, *Robust convex optimization*, Mathematics of Operational Research, Vol. 23, 4 (1998), 769-805.

[26] E. Bernstein and U. Vazirani, *Quantum complexity theory*, SIAM Journal on Computing, 26(5):1411-1473, 1997.

[27] Richard E. Blahut, *Computation of Channel Capacity and Rate-Distortion Functions*, IEEE Thins. on Inform. Theory, IT-18, pp. 460-473, July 1972.

[28] Hugue Blier and Alain Tapp, *All languages in NP have very short quantum proofs*, quant-ph/0709.0738

[29] A. Borodin, *On relating time and space to size and depth*, SIAM J. Comput., 6(4):733744, 1977.

[30] A. Borodin, S. Cook, and N. Pippenger, *Parallel computation for well-endowed rings and space-bounded probabilistic machines*, Information and Control, 58(1-3):113136, 1983.

[31] P. O. Boykin, T. Mor, M. Pulver and V. Roychowdhury, and F. Vatan, *On universal and fault-tolerant quantum computing*, In Proc. 40th FOCS, 1999, 486-494,

[32] Fernando G.S.L. Brandao, *Entanglement Theory and the Quantum Simulation of Many-Body Physics*, PhD thesis, arXiv:0810.0026

[33] Fernando G.S.L. Brandao, Martin B. Plenio, *Entanglement Theory and the Second Law of Thermodynamics*, Nature Physics 4, 873 (2008)

[34] S.L. Braunstein, C.M. Caves, R. Jozsa, N. Linden, S. Popescu, R. Schack, *Separability of very noisy mixed states and implications for NMR quantum computing*, Phys. Rev. Lett., 83 (1999) 1054-1057

[35] Sergey Bravyi, *Efficient algorithm for a quantum analogue of 2-SAT*, quant-ph/0602108

[36] H. Buhrman, R. Cleve, and W. van Dam, *Quantum Entanglement and Communication Complexity*, SIAM J.Comput. 30 (2001) 1829-1841

[37] K. Chen and L.-A. Wu, *A matrix realignment method for recognizing entanglement* Quant. Inf. Comp., 3:193, 2003.

[38] M. Christandl,. R. König, G. Mitchison and R. Renner, *One-and-a-Half Quantum de Finetti Theorems*, Communications in Mathematical Physics, Volume 273, Issue 2, pp.473-498

[39] M. Christandl and A. Winter. "Squashed entanglement" - an additive entanglement measure. *J. Math. Phys.*, 45(3):829–840, 2004. quant-ph/0308088.

[40] Clay Mathematics Institute, Millennium Problems, `http://www.claymath.org/millennium/P_vs_NP/`

[41] Richard Cleve, Wim van Dam, Michael Nielsen, Alain Tapp, *Quantum Entanglement and the Communication Complexity of the Inner Product Function*, Lect.Notes Comput.Sci. 1509 (1998) 61-74

[42] S. A. Cook, The complexity of theorem-proving procedures, Proceedings of ACM STOC'71, pp. 151-158, 1971.

[43] J. Dehaene, B. De Moor and F. Verstraete, *On the geometry of entangled states*, Journal of Modern Optics, Volume 49, Number 8, July 10, 2002 , pp. 1277-1287(11)

[44] I. Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12, 2007.

[45] David P. Divincenzo, *Title: Two-Bit Gates are Universal for Quantum Computation*, Phys. Rev. A 51, 1015 (1995)

[46] A. C. Doherty, P. A. Parrilo and F.M. Spedalieri, *Distinguishing separable and entangled states*, Phys. Rev. Lett., 88:187904, 2002.

[47] A. C. Doherty, P. A. Parrilo and F. M. Spedalieri, *Complete family of separability criteria*, Phys. Rev. A, 69:022308, 2004.

[48] D. Deutsch, A. Barenco and Artur Ekert, *Universality in quantum computation*, Computer Bulletin, 1995 v. 449, 669-677.

[49] David Deutsch, *Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer*, Proc. R. Soc. London A, Vol. 400, No. 1818. (1985), pp. 97-117.

[50] David Deutsch, *Quantum Computational Networks*, Proc. R. Soc. London A, Volume 425, Issue 1868, pp. 73-90

[51] R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press (2002).

[52] A. Einstein, B. Podolsky, and N. Rosen, *Can quantum-mechanical description of physical reality be considered complete?*, Phys. Rev. 47 777 (1935)

[53] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, 1979.

[54] Sevag Gharibian, *Strong NP-Hardness of the Quantum Separability Problem*, arXiv:0810.4507

[55] J. Gill, *Computational complexity of probabilistic Turing machines*, SIAM Journal on Computing 6(4):675-695, 1977.

[56] J. P. Gordon, *Noise at optical frequencies: Information theory*, in Proceedings of the International School of Physics Enrico Fermi. Course XXXI: Quantum Electronics and Coherent Light, (P. A. Mills, ed., Academic Press, New York, 1964) pp. 156181.

[57] Leonid Gurvits, *Classical deterministic complexity of Edmonds' problem and Quantum Entanglement*, quant-ph/0303055

[58] M. B. Hastings. A counterexample to additivity of minimum output entropy. arXiv:0809.3972, 2008.

[59] A. S. Holevo, *Bounds for the quantity of information transmitted by a quantum communication channel*, Probl. Peredachi Inf. 9(3), 311 (1973) [in Russian; English translation in Probl. Inf. Transm. (USSR) 9, 177183 (1973)].

[60] M. Horodecki and P. Horodecki, *Reduction criterion of separability and limits for a class of distillation protocols*, Phys. Rev. A, 59:4206, 1999.

[61] R. Horodecki, P. Horodecki and M. Horodecki, *Quantum α-entropy inequalities: independent condition for local realism?* Phys. Lett. A, 210:377381, 1996.

[62] M. Horodecki, P. Horodecki and R. Horodecki, *Separability of mixed states: necessary and sufficient conditions*, Physics Letters A, v. 223, p. 1-8, 1996.

[63] M. Horodecki, P. Horodecki, and R. Horodecki, *Mixed-State Entanglement and Distillation: Is there a Bound Entanglement in Nature?*, Phys. Rev. Lett. 80, 5239 (1998).

[64] P.M. Hayden, M. Horodecki, and B.M. Terhal, *The asymptotic entanglement cost of preparing a quantum state*, J. Phys. A: Math. Gen. 34, 6891 (2001).

[65] Lawrence M. Ioannou, *Computational complexity of the quantum separability problem*, Quantum Information and Computation, Vol. 7, No. 4 (2007) 335-370

[66] Peter Lax, *Functional Analysis*, Wiley-Interscience, 2002

[67] L. B. Levitin, *On the quantum measure of the amount of information*, in Proceedings of the Fourth All-Union Conference on Information Theory, Tashkent (1969), pp. 111 115, in Russian.

[68] Dominik Janzing, Pawel Wocjan, Thomas Beth, *Identity check is QMA-complete*, quant-ph/0305050

[69] R. Jozsa, *Fidelity for Mixed Quantum States*, IN J. Mod. Opt. 41.23152324 (1994).

[70] J. Kempe, A. Kitaev, and O. Regev, *The Complexity of the Local Hamiltonian Problem*, SIAM Journal of Computing, Vol. 35(5), p. 1070-1097 (2006)

[71] J. Kempe, H. Kobayashi, K. Matsumoto, B. Toner and T. Vidick, *On the Power of Entangled Provers: Immunizing games against entanglement*, quant-ph/0704.2903

[72] J. Kempe, H. Kobayashi, K. Matsumoto, T. Vidick, *Using Entanglement in Quantum Multi-Prover Interactive Proofs*, arXiv:0711.3715

[73] J. Kempe, H. Kobayashi, K. Matsumoto, B. Toner, T. Vidick, *Entangled games are hard to approximate*, arXiv:0704.2903

[74] J. Kempe and O. Regev, *3-Local Hamiltonian is QMA-complete*, Quantum Inf. Comput., 3(3):258-264, 2003

[75] S. Khanna, M. Sudan, L. Trevisan and D. P. Williamson, *The approximability of constraint satisfaction problems*, SIAM J. Comput., 30(6):1863-1920, 2000

[76] C. King and M. B. Ruskai, *Minimal entropy of states emerging from noisy quantum channels*, IEEE Trans. Info. Theory, 47, pp. 192-209 (2001)

[77] A. Kitaev, A. Shen, and M. N. Vyalyi, *Classical and Quantum Computation*, American Mathematical Society, 2002

[78] A. Kitaev and J. Watrous, *Parallelization, amplification, and exponential time simulation of quantum interactive proof systems*, Proceedings of ACM STOC'2000, pp. 608-617, 2000

[79] H. Klauck, A. Nayak, A. Ta-Shma, and D. Zuckerman. Interaction in quantum communication. *IEEE Trans. Information Theory*, 53(6):1970–1982, 2007. Earlier version in STOC'2001. quant-ph/0603135.

[80] H. Kobayashi and K. Matsumoto, *Quantum multi-prover interactive proof systems with limited prior entanglement*, Proceedings of ISAAC'2002, pp. 115-127, 2002.

[81] Hirotada Kobayashi, Keiji Matsumoto, Tomoyuki Yamakami, *Quantum Certificate Verification: Single versus Multiple Quantum Certificates*, quant-ph/0110006

[82] Hirotada Kobayashi, Keiji Matsumoto, Tomoyuki Yamakami, *Quantum Merlin-Arthur Proof Systems: Are Multiple Merlins More Helpful to Arthur?*, quant-ph/0306051

[83] R. König and R. Renner, *A de Finetti representation for finite symmetric quantum states* J. Math. Phys. 46, 122108 (2005).

[84] Serge Lang, Introduction to Linear Algebra, Springer-Verlag, New York, 1986.

[85] Yi-Kai Liu, *Consistency of Local Density Matrices is QMA-complete*, quant-ph/0604166

[86] Y.-K. Liu, M. Christandl, F. Verstraete, *N-representability is QMA-complete*, Phys. Rev. Lett. 98, 110503 (2007)

[87] L. Lovász, On the Shannon capacity of a graph, *IEEE Transactions on Information Theory*, 25(1) (1979), 1-7.

[88] Chris Marriott, John Watrous, *Quantum Arthur-Merlin Games*, Computational Complexity, 14(2): 122 - 152, 2005

[89] R. A. C. Medeiros and F. M. de Assis, *Quantum zero-error capacity*, Int. J. Quant. Inf., vol. 3, no. 1, pp. 135139, 2005

[90] R. A. C. Medeiros, R. Alleaume, G. Cohen, F. M. de Assis, *Quantum states characterization for the zero-error capacity*, quant-ph/0611042

[91] R. A. C. Medeiros and F. M. de Assis, *Quantum Zero-Error Capacity and HSW Capacity*, AIP Conference Proceedings, volume 734 (2004), 52-54

[92] R. A. C. Medeiros, R. Alleaume, G. Cohen and F. M. de Assis, *Zero-error capacity of quantum channels and noiseless subsystems*, (ITS2006), SEPTEMBER 3-6

[93] Ashley Montanaro, Andreas Winter, *A lower bound on entanglement-assisted quantum communication complexity*, In Proc. ICALP'07, 2007

[94] Daniel Nagaj, Shay Mozes, *A new construction for a QMA complete 3-local Hamiltonian*, J. Math. Phys. 48, 072104 (2007)

[95] Hiroshi Nagaoka, *Algorithms of Arimoto-Blahut type for computing quantum channel-capacity*, Proc. of 1998 IEEE International Symposium on Information Theory, p.354, 1998.

[96] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000

[97] M. Nielsen and J. Kempe, *Separable states are more disordered globally than locally*, Phys. Rev. Lett., 86:51847, 2001.

[98] M. Ohya and D. Petz. *Quantum Entropy and its Use.* Springer, 1993.

[99] Susumu Osawa, Hiroshi Nagaoka, *Numerical Experiments on The Capacity of Quantum Channel with Entangled Input States*, quant-ph/0007115

[100] C. H, Papadimitriou, *Computational Complexity*, Addison-Wesley Publishing Company, Inc., 1994.

[101] C. H. Papadimitriou and M. H. Yannakakis. Optimization, approximation, and complexity classes. *J. Comput. Sys. Sci.*, 43(3):425–440, 1991.

[102] A. Peres, *Separability criterion for density matrices*, Phys. Rev. Lett., 77:14131415, 1996.

[103] John Preskill, *Lecture notes on Quantum Computation*, `http://www.theory.caltech.edu/people/preskill/ph229/`

[104] R. Rivest, A. Shamir and L. Adleman, *A Method for Obtaining Digital Signatures and Public-Key Cryptosystems*, Communications of the ACM 21 (2): pp.120126, (1978).

[105] O. Rudolph, *Further results on the cross norm criterion for separability*, quant-ph/0202121.

[106] A. Shamir, *IP=PSPACE*, Proceedings of IEEE FOCS'90, pp. 11-15, 1990

[107] C. E. Shannon, The zero-error capacity of a noisy channel, *IRE Transactions on Information Theory*, 2(3) (1956), 8-19.

[108] P. W. Shor. Equivalence of additivity questions in quantum information theory. *Communications in Mathematical Physics*, 246(3):453–472, 2004. quant-ph/0305035.

[109] P. W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM J. Computing 26, pp. 1484-1509 (1997).

[110] P. W. Shor, *Algorithms for quantum computation: Discrete logarithms and factoring*, Proc. 35nd Annual Symposium on Foundations of Computer Science, IEEE Computer Society Press (1994), 124-134.

[111] Michael Sipser, *Introduction to the Theory of Computation*, PWS Publishing Company, 2005

[112] S. Szarek, I. Bengtsson and K. Zyczkowski, *On the structure of the body of states with positive partial transpose*, J. Phys. A 39 L119-L126 (2006)

[113] S. J. Szarek, E. Werner, and Karol Zyczkowski, *Geometry of sets of quantum maps: A generic positive map acting on a high-dimensional system is not completely positive*, J. of Mathematical Physics 49, 032113 (2008).

[114] V. Vedral and M. B. Plenio. Entanglement measures and purification procedures. *Phys. Rev. A*, 57:1619–1633, 1998. quant-ph/9707035.

[115] V. Vedral, M.B. Plenio, M.A. Rippin, P.L. Knight, *Quantifying Entanglement*, Phys. Rev. Lett. 78, 2275 (1997).

[116] K.G.H. Vollbrecht and R.F. Werner, *Entanglement Measures under Symmetry*, Phys. Rev. A 64, 062307 (2001).

[117] M. Vyalyi, *QMA=PP implies that PP contains PH*, ECCC TR03-021, 2003.

[118] J. Watrous, *Quantum Computational Complexity*, arXiv:0804.3401

[119] J. Watrous, *Succinct quantum proofs for properties of finite groups*, Proceedings of IEEE FOCS'2000, pp. 537-546, 2000

[120] J. Watrous, *PSPACE has constant-round quantum interactive proof systems*, Proceedings of IEEE FOCS'99, pp. 112-119, 1999.

[121] Pawel Wocjan, Dominik Janzing and Thomas Beth, *Two QCMA-complete problems*, quant-ph/0305090

[122] W.K. Wootters and W.H. Zurek, A Single Quantum Cannot be Cloned, Nature 299 (1982), pp. 802-803.

[123] D. Yang, M. Horodecki, R. Horodecki, and B. Synak-Radtke, *Irreversibility for all bound entangled states*, Phys. Rev. Lett. 95, 190501 (2005).

[124] Deping Ye, *On the Bures Volume of Separable Quantum States*, arXiv:0902.1505