Department of Physics

National Taiwan Normal University

PhD Thesis

Information causality and its tests for quantum communications

I-Ching Yu

Advisor: Feng-Li Lin, Ph.D.

February. 25. 2013

to

my family and teachers

Acknowledgements

In the past six years, there were many times that I wanted to give up. Because of the help and the encouragement from teachers, my family and friends, I could hold on straight to the last second and finish this thesis.

First, I would like to thank Prof. Feng-Li Lin, my supervisor. Prof. Lin not only taught me how to do the research but also reminded me the correct attitude for doing researches. It is practical and down-to-earth. When I made mistakes in writings or I was not on the right track, Prof. Lin used his patience to modify them again and again. Without his insistence, this thesis could not have the present form.

I also want to thank Prof. Hong-Yi Chen and Prof. Li-Yi Hsu. Prof. Chen gave me a lot of help for writing the computer program and building the machine for Message Passing Interface (MPI). Due to Prof. Hsu's guidance, I have deeper and broader understanding of quantum information. Furthermore, I want to appreciate the committee members of my oral defence, Prof. Chi-Yee Cheung, Prof. Zheng-Yao Su, Prof. Hsi-Sheng Goan and Prof. Min-Hsiu Hsieh. They took time to read my thesis and provided many valuable comments. It benefited me a lot from their comments.

My gratitude would go to my friend Ching-Yu. We were classmates over ten years. I can discuss all the worry and problems with her and always get solutions. I will never forget the moment which we went through. I also want to appreciate the assistant, Chun-Ping and the members in room A201, Chao-Chun, Hsuan-Hao and Pei-Hua. With their suggestions and help, I can have the different point of view to solve my problems and my oral defence can be held without a hitch.

In addition, I would like to express my gratitude to my husband, Hsin-Ping for all that he has given to our family. With his support, I did not have the financial stress such that I can concentrate on my study. Finally, I want to thank my parents, Mr. Kuo-Wen Yu and Mrs. Chiu-Feng Huang. Thank them for giving my life and all the supports without conditions. My gratitude can not be written in words. I hope that I can pay them back and let them be proud of me on some day.

Abstract

Information causality has been proposed to constrain the maximal mutual information shared between sender and receiver in a communication protocol based on physical theories such as quantum mechanics.

We reformulate the information causality in a more general framework by adopting the results of signal propagation and computation in a noisy circuit. In our framework, the information causality leads to a broad class of Tsirelson inequalities for the two-level quantum systems. This fact allows us to subject the information causality to the experimental scrutiny. A no-go theorem for reliable nonlocal computation is also derived. Information causality prevents any physical circuit from performing reliable computations.

Moreover, we test the information causality for the more general quantum communication protocols with multi-level and (non-)symmetric channels by directly evaluating the mutual information. Our results support the information causality which is never violated for the more general settings discussed in this work. For the two-inputs/two-outputs cases, we also find that the information causality is saturated not for the channels with the maximal quantum non-locality associated with the Tsirelson inequality but for the marginal cases saturating the Bell's inequality. This indicates that the more quantum non-locality may not always yield the more mutual information.

Keywords: Information causality, quantum communication, quantum computation, quantum non-locality

中文摘要

在基於物理理論的通訊協定中，例如:量子力學，訊息因果論限制傳送者與接受者之間的最大共有訊息量.

我們通過更廣泛的框架，即使用討論訊號傳送及錯誤計算的結果，來重新了解訊息因果論與量子力學的關係. 在我們的框架中，訊息因果論將導致一組在二態量子系統下的 Tsirelson 不等式（量子系統的極限值). 基於這樣的結果，訊息因果論對使用物理系統的實驗產生限制. 此外，在我們的框架中，可信賴的非定域性的計算是不可行的. 訊息因果論的限制將使得物理系統的計算線路無法進行可信賴的計算.

另外，我們直接計算共有訊息量，藉以測試訊息因果論在更普遍的通訊協定中的正確性，這些普遍的通訊協定包含多態的系統及非對稱的通訊管道. 我們的結果支持訊息因果論，意思是在這些普遍的通訊協定中，共有訊息量不會超過訊息因果論給的限制. 此外，如果通訊管道包含兩個輸出及兩個輸入，我們發現共有的量子系統擁有最大非定域性時（滿足 Tsirelson 不等式的限制)，共有訊息量的值不是最大的. 最大的共有訊息量出現在共有的系統恰好滿足定域性理論給出的極大值時 (Bell 不等式給的限制)，且此時共有的訊息量和訊息因果論給的限制相同. 這個結果指出共享一個量子非定域系統，並不一定產生較多的共有訊息量.

關鍵字： 訊息因果論、量子通訊、量子計算、量子非定域性

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 No-signaling theory and quantum non-locality

Quantum information science has been developed for nearly a half century. The tasks in the quantum information science such as commutation, computation and cryptography are operated with quantum systems. Since the famous Shor's factoring algorithm [1] was proposed, people were aware that the quantum computation may be efficient. Therefore, quantum information science became a very important topic in the 1980s. Of course, we cannot ignore the beautiful idea of the BB84 key distribution protocol [2]. Using the uncertainty principle of quantum mechanics, the BB84 protocol keeps the security of the communication.

From the above examples, we could know the importance of quantum mechanics to the information science. Nowadays, we know that the quantum mechanics is inherently non-local and it cannot violate the no-signaling theory which states that superluminal signaling is impossible. Due to these properties, here comes a problem: could the quantum correlations be singled out by considering both the non-locality and the no-signaling theory? In this section, we will review the theory of quantum non-locality and the no-signaling theory, and then try to answer this question.

1.1.1 Quantum non-locality

In 1935, Albert Einstein and his collaborators Boris Podolsky and Nathan Rosen (known collectively as EPR) proposed a thought experiment to show the conflict between locality and physical reality in quantum mechanics. Their arguments were latter called the EPR paradox. The paradox implies that quantum mechanics is an incomplete physicals theory [3]. They also gave the definition for a complete physical theory. That is, if every element of the physical reality have a counterpart in the physical theory, we then could judge this physical theory is complete. Furthermore, what is the element of physical reality? The authors said if one could predict the quantity of a measurement with certainly for an

undisturbed system, then there must exist an element of physical reality determining this quantity. However, there are two ways to solve the EPR paradox, the first one is the local hidden variable theory and the other one is the quantum non-locality.

Almost thirty years after the proposition of EPR paradox, in 1964, J. S. Bell proposed the famous Bell inequality [5]. The construction of the Bell inequality is based on the local hidden variable theory. Bell then asked, no matter how Alice and Bob measure their particles, could the probability distribution of quantum measurement's outcomes always be reproduced by the local hidden variable theory? In the other words, could the Bell inequality always be satisfied? The answer is no. Bell found an example to show this. Due to the negative answer, one may think that the local hidden variable theory may be wrong. Therefore, the failure of the local hidden variable theory will imply the validity of the other solution for EPR paradox, quantum non-locality.

In this subsection, we will explain what is the EPR paradox and therefore how to use the local hidden variable theory and quantum non-locality to solve the paradox. Moreover, we will introduce the more general Bell inequalities and show the quantum violation, so that we have more confidence to judge that quantum mechanics is a non-local theory.

1.1.1.1 The EPR paradox and the local hidden variable theory

Now, we will study the EPR paradox. We know quantum mechanics can describe physical phenomena at microscopic scales very well, but it still has some limitations. According to the uncertainty principle in quantum mechanics, we cannot precisely obtain two quantities simultaneously if they correspond to non-commute operators such as spin angular momentum operators for x- and z-axis. Therefore, by the definition of the element of physical reality, the elements of physical reality related to the spin angular momentum for x- and z-axis could not exist simultaneously.

Bohm's thought experiment [4] may challenge the uncertainty principle. This experiment is as follows. Suppose Alice and the distant party Bob hold an electron from the emitted pair of electrons separately. The quantum state of the electrons is the singlet state, i.e.,

$$|\Psi\rangle_{EPR} = \frac{1}{\sqrt{2}}(|+z,-z\rangle_{AB} - |-z,+z\rangle_{AB}), \tag{1.1}$$

where $|+z\rangle$ and $|-z\rangle$ are used to denote the eigenstates of spin operator along z-axis

(up or down). Note that, $|+z\rangle$ corresponds to the state with the z-component of spin angular momentum being $+\frac{\hbar}{2}$, and similarly $|-z\rangle$ to $-\frac{\hbar}{2}$. If Alice measures the spin angular momentum along z-axis, once she obtains $|+z\rangle$, the quantum state then will collapse to $|+z,-z\rangle_{AB}$. Alice can predict that Bob will get $|-z\rangle$ with certainty if Bob then makes a measurement along the z-axis. Similarly, if Alice obtains $|-z\rangle$, Bob then gets $|+z\rangle$. Of course, they can choose x-axis to make the measurement. Since the total spin angular momentum for the singlet state is conserved to be zero, Bob will get the opposite result to Alice's outcome. These cases show us that Alice can predict Bob's measurement results along z- and x-axis with certainty before Bob measure his electron. Furthermore, the corresponding elements of physical reality to these quantities must exist simultaneously. Recall the constraint of the uncertainty principle, we cannot predict the spin angular momentum along x- and z-axis with certainty simultaneously. Obviously, this implies that quantum mechanics is not a self-consistent theory. This dilemma is the so-called the EPR paradox. Due to the EPR paradox, the authors of [3] gave a conclusion that quantum mechanics is an incomplete physical theory while the wave function does not provide the complete description to reveal the elements of physical reality.

There are two ways to resolve the EPR paradox. The first one is the local hidden variable theory. The other one is the quantum non-locality. The first one was proposed by EPR [3]. The local hidden variable theory satisfies the following conditions.

- In a complete physical theory, the measurement results are determined before measurements.
- Bob's measurement will not be disturbed by Alice's measurement if Alice is far from Bob, i.e., the action in a distance is impossible. Therefore, the non-locality between distant partite is impossible.

In the local hidden variable theory, the hidden variable is like a secret code, it can determine the result of measurement. Therefore, if one can know the hidden variable and the wave function of a system, one then can predict the measurement result.

Now, we can discuss how to use the local hidden variable to resolve the EPR paradox. Since the hidden variable theory is a local theory, the hidden variable for Bob's particle will not be modified while Alice measures her own particle. Thus, Alice can obtain the information about the value of the hidden variable when she measures her own particle.

She can then predict the outcome of Bob's measurement. Let's take Bohm's thought experiment for example. Before measuring the spin angular momentum along x- and z-axis, the outcomes of Alice's and Bob's measurements are already determined by the hidden variables. Due to the hidden variables, Bob always gets the opposite to Alice's outcome while Alice and Bob choose the same axes. Thus, before the measurements, the two emitted electrons may be one of the following four types:

Alice's particle	Bob's particle
$(\|+z\rangle, \|+x\rangle, \lambda_1)$	$(\|-z\rangle, \|-x\rangle, \lambda_1)$
$(\|+z\rangle, \|-x\rangle, \lambda_2)$	$(\|-z\rangle, \|+x\rangle, \lambda_2)$
$(\|-z\rangle, \|+x\rangle, \lambda_3)$	$(\|+z\rangle, \|-x\rangle, \lambda_3)$
$(\|-z\rangle, \|-x\rangle, \lambda_4)$	$(\|+z\rangle, \|+x\rangle, \lambda_4)$

Here λ_i are used to denote the hidden variable with different values. If Alice and Bob measure their own electron along the different axes, Bob will obtain $+\frac{\hbar}{2}$ or $-\frac{\hbar}{2}$ with equal probability no matter what outcome Alice obtains. However, one can find Bob's measurement result is determined no matter what axis Alice chooses to measure. It is consistent with the property of the locality, i.e., the impossibility of the action at a distance. Thus, the local hidden variable theory not only satisfies the locality but also obtains the same prediction of measurement outcomes as predicted by quantum mechanics. It seems that we have already obtained a complete physical theory.

1.1.1.2 The Bell's inequality and the CHSH inequality

Bell's experiment is similar as Bohm's thought experiment, but Alice and Bob have three choices (axes a, b and c) to measure the spin angular momentum. According to the local hidden variable theory, the determined measurement results for a singlet state (1.1) can be divided into eight types as Tab.1.1. Alice's and Bob's measurement results are opposite to each other for each types in order to preserve the angular momentum. We use N_i to denote the number of times for type i while we repeat the same experiment for $N = \sum_{i=1}^{8} N_i$ times. Suppose Alice obtains $+\frac{\hbar}{2}$ for the measurement along a-axis and Bob also obtains $+\frac{\hbar}{2}$ when measuring the spin along b-axis. Let $\Pr(+,+|a,b)$ be the joint probability of this situation. Obviously, the determined result of particles could be type 3 or 4. Therefore,

$$\Pr(+,+|a,b) = \frac{N_3 + N_4}{N}. \tag{1.2}$$

Table 1.1: The determined measurement results in the local hidden variable theory

Number	Alice's particle	Bob's particle
N_1	$(\|+a\rangle, \|+b\rangle, \|+c\rangle)$	$(\|-a\rangle, \|-b\rangle, \|-c\rangle)$
N_2	$(\|+a\rangle, \|+b\rangle, \|-c\rangle)$	$(\|-a\rangle, \|-b\rangle, \|+c\rangle)$
N_3	$(\|+a\rangle, \|-b\rangle, \|+c\rangle)$	$(\|-a\rangle, \|+b\rangle, \|-c\rangle)$
N_4	$(\|+a\rangle, \|-b\rangle, \|-c\rangle)$	$(\|-a\rangle, \|+b\rangle, \|+c\rangle)$
N_5	$(\|-a\rangle, \|+b\rangle, \|+c\rangle)$	$(\|+a\rangle, \|-b\rangle, \|-c\rangle)$
N_6	$(\|-a\rangle, \|+b\rangle, \|-c\rangle)$	$(\|+a\rangle, \|-b\rangle, \|+c\rangle)$
N_7	$(\|-a\rangle, \|-b\rangle, \|+c\rangle)$	$(\|+a\rangle, \|+b\rangle, \|-c\rangle)$
N_8	$(\|-a\rangle, \|-b\rangle, \|-c\rangle)$	$(\|+a\rangle, \|+b\rangle, \|+c\rangle)$

Similarly,

$$\Pr(+,+|a,c) = \frac{N_2+N_4}{N}$$
$$\Pr(+,+|c,b) = \frac{N_3+N_7}{N}. \tag{1.3}$$

Since we know each number N_i should be non-negative, the following inequality will hold, i.e.,

$$N_3 + N_4 \leq N_2 + N_4 + N_3 + N_7. \tag{1.4}$$

Therefore, when both sides of the above inequality are divided by N, we can obtain

$$\Pr(+,+|a,b) \leq \Pr(+,+|a,c) + \Pr(+,+|c,b). \tag{1.5}$$

This is the so-called Bell inequality.

Now, by checking the Bell inequality (1.5), we can check if the prediction of quantum mechanics is always consistent with the local hidden variable theory. Suppose Alice and Bob share a singlet state (1.1) and Alice measures the spin angular momentum along a-axis. In quantum mechanics, Alice has half a chance to obtain $+\frac{\hbar}{2}$ along a-axis and therefore the singlet state will then collapse to $|+a,-a\rangle$. Suppose Bob then measures his particle along b-axis, Bob will obtain $+\frac{\hbar}{2}$ with probability $|\langle +b|-a\rangle|^2 = \sin^2(\frac{\theta_{ab}}{2})$. Here θ_{ab} is the angle between a- and b-axis. Put this value and the other similar twos into the Bell inequality (1.5), one can rewrite the (1.5) as

$$\sin^2(\frac{\theta_{ab}}{2}) \leq \sin^2(\frac{\theta_{ac}}{2}) + \sin^2(\frac{\theta_{bc}}{2}). \tag{1.6}$$

Assume these three axes (a, b and c) are in the x-z plane, i.e., $\theta_{ab} = \frac{\pi}{2}$ and $\theta_{ac} = \theta_{bc} = \frac{\pi}{4}$. One can find the left hand side of (1.6) is 0.5 and the right hand side is 0.2929. Obviously,

the Bell inequality is violated. Thus, the local hidden variable theory and quantum mechanics are not always consistent with each other. Moreover, the non-locality property is revealed by quantum mechanics.

In 1969, another form of the Bell inequality named the CHSH inequality was proposed [6]. The CHSH inequality is more convenient for experiments to test the non-locality. As the Bell inequality, the construction of the CHSH inequality is based on the local hidden variable theory, and therefore one could find the quantum violation.

Suppose both Alice and Bob has two observables and each observable has two outcomes $+1$ or -1. We denote Alice's outcomes for her two observables as A_0 and A_1, respectively. Similarly, B_0 and B_1 are Bob's outcomes for his observables, respectively. These outcomes would satisfy one of the following conditions.

1. If $A_0 + A_1 = 0$, then $A_0 - A_1 = \pm 2$.
2. If $A_0 + A_1 = \pm 2$, then $A_0 - A_1 = 0$.

One could then find the equality

$$C = (A_0 + A_1)B_0 + (A_0 - A_1)B_1 = \pm 2. \tag{1.7}$$

Since we know

$$|\langle C \rangle| \leq \langle |C| \rangle = 2, \tag{1.8}$$

where $\langle . \rangle$ is used to denote the expectation value, so that

$$-2 \leq \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle = C_{0,0} + C_{0,1} + C_{1,0} - C_{1,1} \leq 2. \tag{1.9}$$

Note that $C_{i,j}$ is the expectation value for outcomes $A_i B_j$. This inequality (1.9) is the well-known CHSH inequality.

Now, we want to translate the above into the language to quantum mechanics. In quantum mechanics, the observable is a measurement operator with ± 1 eigenvalues. Therefore, the expected value of the outcomes is relied on the type of the measurement operators and the quantum state. Suppose Alice and Bob share a singlet state and Alice's observables are x and x' with the outcomes A_0 and A_1 respectively. Similarly, Bob's observables are y and y' with the outcomes B_0 and B_1 respectively. Here, the observables r could be expressed as $\vec{r} \cdot \vec{\sigma}$, where $\vec{r}$ is a real three-dimensional unit vector and each element of

vector $\vec{\sigma}$ corresponds to different Pauli matrices, i.e., $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$. Therefore, the expectation value for outcomes A_0 and B_0 is

$$\langle A_0 B_0 \rangle =_{EPR} \langle \Psi | (\vec{x} \cdot \vec{\sigma})(\vec{y} \cdot \vec{\sigma}) | \Psi \rangle_{EPR} = -\vec{x} \cdot \vec{y} = -\cos\theta, \tag{1.10}$$

where θ is the angle between $\vec{x}$ and $\vec{y}$. We know that the expectation value is related to the direction of the measurements. Consider the case where Alice and Bob perform measurements with the following obserables:

$$\begin{aligned} x = \sigma_z, x' = \sigma_x; \\ y = -\frac{\sigma_z + \sigma_x}{\sqrt{2}}, y' = \frac{\sigma_z - \sigma_x}{\sqrt{2}}. \end{aligned} \tag{1.11}$$

Thus,

$$|C_{0,0} + C_{0,1} + C_{1,0} - C_{1,1}| = 2\sqrt{2}. \tag{1.12}$$

Obviously, the prediction of quantum mechanics exceeds 2, i.e., the bound from the local hidden variable theory. Therefore, the CHSH inequality is violated by quantum mechanics.

Although we have known that the quantum mechanics violates the CHSH inequality, what is the maximal violation? B. S. Cirel'son used the quantum properties of the observables to obtain the answer, $2\sqrt{2}$ [7]. Thus, the maximally quantum violation of the CHSH inequality is called the Tsirelson bound.

One can obtain the Tsirelson bound with the following properties [59]:

- Since these observables for quantum measurements should be the hermitian operators with eigenvalue ± 1, thus the square of each observable is equal to the identity operator, i.e.,

$$x^2 = x'^2 = y^2 = y'^2 = I. \tag{1.13}$$

- Alice's observables commute with Bob's since they only measure their own particles. Thus,

$$[x, y] = [x, y'] = [x', y] = [x', y'] = 0. \tag{1.14}$$

Using the above conditions (1.13) and (1.14), one can obtain

$$(xy + x'y + xy' - x'y')^2 = 4I + [x, x'][y, y']. \tag{1.15}$$

Using the property of *sup norm* of a matrix M, i.e., $sup_{|\psi\rangle}(\frac{\|M|\psi\rangle\|}{\||\psi\rangle\|})$, we can obtain

$$\| [x, x'] \| \leq 2 \| x \| \| x' \| = 2. \tag{1.16}$$

Therefore,

$$\| xy + x'y + xy' - x'y' \| \leq 2\sqrt{2}. \tag{1.17}$$

This means that the maximal expectation value of the operator $xy + x'y + xy' - x'y'$ cannot exceed $2\sqrt{2}$.

1.1.1.3 More general Bell-type inequalities

After the publication of the CHSH inequality, more general Bell-type inequalities was proposed such as the Bell-type inequality for higher dimensional system [8] and the case of multi-setting per site [9]. We can have more understanding about quantum non-locality through these more general Bell-type inequalities. As the construction of the Bell inequality, the constructions of these Bell-type inequalities are based on the local hidden variable theory. Therefore, one may find the quantum violation of these Bell-type inequalities since the quantum correlation is non-local and cannot be described by the local hidden variable theory.

Here, we would like to review the multi-level Bell-type inequality which is called the CGLMP inequality [8]. In this case, both Alice and Bob have two observables. Alice's outcomes are A_x, and similarly Bob's outcomes are B_y $(x, y \in \{0, 1\})$. Therefore, in the multi-level system A_x and $B_y \in \{0, 1, ..., d-1\}$.

Recall the local hidden variable theory, assuming Alice and Bob's measurement outcomes are determined by the local hidden variables j, k, l and m, so that the value of observable A_0 is j and the one of A_1 is k. Similarly, B_0 gives l and B_1 gives m. The probability for the local hidden variables j, k, l, m is denoted by c_{jklm}, and therefore the summation of the probabilities should be one, i.e., $\sum_{jklm} c_{jklm} = 1$. Consider the special case

$$\begin{aligned} r' &= B_0 - A_0 = l - j \\ s' &= A_1 - B_0 = k - l \\ t' &= B_1 - A_1 = m - k \\ u' &= A_0 - B_1 = j - m, \end{aligned} \tag{1.18}$$

which automatically yield

$$r' + s' + t' + u' = 0. \tag{1.19}$$

This means the relations among the local hidden variables cannot be arbitrary. Although we could determine the relations r', s' and t' by choosing variables j, k, l and m, but the last term u' should be constrained by (1.19). This plays the central role in deriving the CGLMP inequality.

Consider a function for a $d = 3$ system as follows.

$$\begin{aligned} I_3 = \quad & [\Pr(A_0 = B_0) + \Pr(B_0 = A_1 + 1) + \Pr(A_1 = B_1) + \Pr(B_1 = A_0)] \\ & -[\Pr(A_0 = B_0 - 1) + \Pr(B_0 = A_1) + \Pr(A_1 = B_1 - 1) + \Pr(B_1 = A_0 - 1)]. \end{aligned} \tag{1.20}$$

According to the local hidden variable theory, the maximal value of (1.20) is 2, i.e., $I_3(local) \leq 2$. This is because when the three probabilities with the " $+$ "-sign are satisfied, the term with the " $-$ "-sign will also be satisfied. On the other hand, once all the four terms with " $+$ "-sign could be satisfied by non-local correlations, the maximal value of the function (1.20) would be 4.

One can then generalize the function from $d = 3$ to arbitrary d. The generalized function for the multi-level Bell-type inequality is

$$\begin{aligned} I_d = & \sum_{k=0}^{[\frac{d}{2}]-1} (1-\tfrac{2k}{d-1})[\Pr(A_0{=}B_0{+}k){+}\Pr(B_0{=}A_1{+}k{+}1){+}\Pr(A_1{=}B_1{+}k){+}\Pr(B_1{=}A_0{+}k)] \\ & -[\Pr(A_0{=}B_0{-}k{-}1){+}\Pr(B_0{=}A_1{-}k){+}\Pr(A_1{=}B_1{-}k{-}1){+}\Pr(B_1{=}A_0{-}k{-}1)]. \end{aligned} \tag{1.21}$$

As for the $d = 3$ case, the maximal value of I_d from the local hidden variable theory is 2, i.e., $I_d(local) \leq 2$. Therefore, the maximal value achieved by the non-local correlations is still 4. For more detailed proof, please see [8].

Now, we can study the quantum violations respected to the maximum of $I_d(local)$. Assuming Alice and Bob share a maximally entangled state

$$|\psi\rangle = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} |j\rangle_A \otimes |j\rangle_B, \tag{1.22}$$

and Alice's observables are x and y', similarly y and y' for Bob. One can obtain optimal

value of (1.21) when these observables correspond to the following eigenvectors [8]:

$$|k\rangle_{A,x} = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} \exp(i\frac{2\pi}{d} j(k+\alpha_x))|j\rangle_A$$
$$|l\rangle_{B,y} = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} \exp(i\frac{2\pi}{d} j(l-\beta_y))|j\rangle_B, \tag{1.23}$$

where $\alpha_0 = 0$, $\alpha_1 = 1/2$, $\beta_0 = 1/4$ and $\beta_0 = -1/4$. Thus, one can obtain the joint probability for the outcomes

$$\begin{aligned} \Pr(A_x = k, B_y = l) = & \quad |\langle\psi|(|k\rangle_{A,x A,x}\langle k| \otimes |l\rangle_{B,y B,y}\langle l|)|\psi\rangle| \\ = & \quad \frac{1}{2d^3 \sin^2[\pi(k-l+\alpha_x+\beta_y)/d]}. \end{aligned} \tag{1.24}$$

With this joint probability, one can obtain the bound achieved by the maximally entangled state

$$I_d(QM) = 4d \sum_{k=0}^{[\frac{d}{2}]-1} (1 - \frac{2k}{d-1})(\Pr(A_1 = B_1 + k) - \Pr(A_1 = B_1 - k - 1)) > 2 \geq I_d(local). \tag{1.25}$$

According to this quantum violation, we have another example such that the local hidden variable theory is not consistent with quantum mechanics. Thus, quantum mechanics is intrinsically non-local.

1.1.2 No-signaling theory

The quantum mechanics is constrained by some natural principle such as the relativistic causality. The relativistic causality gives many important results, one of them is the speed of propagation cannot be faster than the light speed in any reference frame. Due to the fact, here comes the no-signaling theory which states that the speed of the propagating information cannot be faster than the light speed, i.e., the superluminal signaling is impossible.

Consider the constraint of no-signaing theory, many related works have been studied. It is interesting that the constraints from the no-signaling theory and quantum mechanics lead to the same behaviors of the correlations such as no-cloning [11, 12], no-broadcasting [13] and monogamy of entanglement [11]. As shown [14, 15], both the no-signaling theory and quantum mechanics preserve the security of the key distribution.

We will study the specific constraint of the no-signaling theory in this subsection. First, we define the measurement scenario for producing the bi-partite correlations. Second, we introduce the correlations satisfying the no-signaling theory. Furthermore, these no-signaling correlation could be divided into two subsets: the local correlations and the non-local ones. We will also discuss the properties of these correlations.

1.1.2.1 The measurement scenario and the box version

First, let us define the measurement scenario for the bi-partite correlations. Both Alice and Bob hold a physical system. They can choose to measure arbitrary observables to probe their systems. We denote the observable for Alice as x and denote the outcome of the observable x as A_x. Similarly, Bob's observable is denoted as y and the corresponding outcome is B_y. Therefore, the bi-partie correlation is defined by the conditional joint probability, i.e.,

$$\Pr(A_x, B_y|x, y). \tag{1.26}$$

Note that, the conditional joint probability should be nonnegative and satisfy the normalization conditions, thus the summation of the conditional joint probabilities with fixed observables should be one, i.e.,

$$\sum_{A_x, B_y} \Pr(A_x, B_y|x, y) = 1. \tag{1.27}$$

One can reformulate the scenario in a more abstract way. Instead of giving a specific physical system, one assumes a bi-partite box shared by Alice and Bob. Both of them put the inputs into the box, he (she) then will obtain the corresponding outputs. The shared box is characterized by the joint probability (1.26). Note that, the number of Alice and Bob's inputs may not be the same, similarly for the number of their outputs. According to the property of the bi-partite box, one can divide them into many types such as the no-signaling box and the signaling one. We will discuss the property of the bi-partite box later. Hereafter, we will use the terms "box" and "correlation" interchangeably when we discuss the property of the joint probability (1.26).

For the necessary, one can generalize the above bi-partite box to the multi-partite case.

1.1.2.2 No-signaling correlations

Now, we can discuss the no-signaling condition for the bi-partie correlations. Note that, If the shared box can simulate no-signaling correlations, we call the box as a no-signaling box [20]. Otherwise, the shared box is a signaling one.

Despite of any non-local correlations previously shared between them, Alice cannot signal to the distant Bob by her choice of inputs due to the no-signaling theory. This indicates that Bob's marginal probability $\Pr(B_y|y)$ is independent with Alice's input x. Similarly, Alice's marginal probability is independent with Bob's input y. Therefore, the condition for the correlations (box) to be no-signaling is

$$\begin{aligned}\Pr(B_y|y) &= \sum_{A_x} \Pr(A_x, B_y|x, y) = \sum_{A_{x'}} \Pr(A_{x'}, B_y|x', y), \quad \forall B_y, y, x, x', \\ \Pr(A_x|x) &= \sum_{B_y} \Pr(A_x, B_y|x, y) = \sum_{B_{y'}} \Pr(A_x, B_{y'}|x, y'), \quad \forall A_x, x, y, y'. \end{aligned} \tag{1.28}$$

The no-signaling box can be divided into two types, the local and non-local ones. The local box can simulate the local correlations which can be described by the local hidden variable theory. From the information theoretic point of view, the local correlations can be simulated by the non-communicating observers with the pre-shared classical random data. Therefore, the local correlations should satisfy

$$\Pr(A_x, B_y|x, y) = \sum_{\lambda} \Pr(\lambda) \Pr(A_x|x, \lambda) \Pr(B_y|y, \lambda), \tag{1.29}$$

where λ is the pre-shared classical random data and the probability for its occurrence is $\Pr(\lambda)$. The probability of occurrence for Alice's output A_x with the given random data λ and the input x is denoted as $\Pr(A_x|x, \lambda)$. The marginal probability $\Pr(B_y|y, \lambda)$ has the similar definition. If the box cannot simulate the local correlation, we will call it the non-local box.

Besides using the the local and the non-local as the categories for the correlations, we want to discuss a another way of categorizing, the quantum correlartions. Quantum correlations can be obtained by sharing the quantum resource such as the quantum state. These correlations can be written as

$$\Pr(A_x, B_y|x, y) = \mathrm{Tr}(E_{A_x} \otimes E_{B_y} \rho), \tag{1.30}$$

where ρ is the density matrix of the shared quantum state, and E_{A_x} and E_{B_y} are the projection operators corresponding to Alice and Bob's measurements, respectively. Note

that, these operators are the elements of a positive-operator valued measure (POVM) [60] and therefore the operators should satisfy

$$\Sigma_{A_x} E_{A_x} = \Sigma_{B_y} E_{B_y} = I, \quad \forall x, y. \tag{1.31}$$

1.1.3 Could the no-signaling theory single out quantum correlations?

Let us recall how to obtain the bound of quantum non-locality. It comes from the mathematic constraint. Therefore, we are curious if there is a natural principle which could imply quantum mechanics. After reviewing the constraint of the no-signaling theory, we may ask a question: could the no-signaling theory be the physical principle which we are looking forward to? More precisely, does the non-locality and the no-signaling theory together imply quantum theory? This question was firstly asked and answered by S. Popescu and D. Rohrlich [21]. Their answer is negative, since quantum mechanics is not the only theory which satisfies the no-signaling constraints and also violates the CHSH inequality. Furthermore, under the constraints of the no-signaling theory, the maximal violation of the CHSH inequality is bigger than the maximally quantum violation. These extremal non-local correlations are called *super-quantum* correlations in [21].

We will discuss the properties of the *super-quantum* correlations. We then review the geometric picture of all the no-signaling correlations in [20]. It is interesting that the Bell-type inequality and the facet of the local polytope are related. Finally, under the constraints of the no-signaling theory, one can demonstrate that the extremal non-local correlations lead to the trivial communication complexity from information theoretic point of view.

1.1.3.1 Beyond the quantum correlations

Does the non-locality and the no-signaling theory together imply quantum mechanics? In [21], the authors used the following process to show the negative answer. First, one could take the non-locality and the no-signaling theory as two axioms. Second, one could try to find the maximal violation of the CHSH inequality and notice that the maximal violation cannot be achieved by quantum mechanics. Therefore, the axiom of non-locality implies the quantum correlation is not the most non-local one. One may then guess that the other axiom, the no-signaling theory, might give the constraint of quantum non-locality. However, there is a set of joint probabilities achieves the maximal violation of the CHSH

inequality and satisfies the no-signaling theory (1.28). This means, the non-signaling constraints cannot single out quantum correlations.

To be specific, we want to find the maximal non-locality. In [21], they used the CHSH function to characterize the non-locality, where the CHSH function is

$$|C_{0,0} + C_{0,1} + C_{1,0} - C_{1,1}|. \tag{1.32}$$

Note that, each term $C_{x,y}$ in (1.32) is the correlation function for the specific inputs x and y. The definition of $C_{x,y}$ is

$$C_{x,y} = \sum_{A_x,B_y} (-1)^{A_x+B_y} \Pr(A_x, B_y|x, y). \tag{1.33}$$

Here x, y, A_x and $B_y \in \{0, 1\}$. With the definition of $C_{x,y}$, we know that each $C_{x,y}$ lies in $[-1, 1]$. Therefore, the maximal value of the CHSH function (1.32) should be 4. Since we have known that the local bound and the quantum bound (Tsirelson bound) of CHSH function are 2 and $2\sqrt{2}$, respectively. Obviously, the maximal value of CHSH function is bigger than the Tsirelson bound and cannot be achieved by the quantum correlations.

A question then arises: why quantum mechanics could not be more non-local? Is this because of the constraint from the no-signaling theory? Therefore, one could consider the no-signaling correlations (1.28) and then calculate the corresponding value of CHSH function (1.32). With the no-signaling theory, Alice cannot signal to Bob her choice of inputs, and vise versa. In [21], they found the following joint probabilities such that the no-signaling theory is satisfied and the maximum of the CHSH function (4) is achieved:

$$\Pr(A_x, B_y|x, y) = \begin{cases} \frac{1}{2} & :A_x + B_y \pmod 2 = xy; \\ 0 & : \text{otherwise.} \end{cases} \tag{1.34}$$

The no-signaling theory is satisfied, this is because $\Pr(A_x|x)$ and $\Pr(B_y|y)$ are equal to $\frac{1}{2}$. Obviously, this set of joint probabilities is not consistent with quantum correlations. Therefore, the no-signaling theory and the non-locality together cannot imply quantum mechanics.

Besides finding the above fact, one could also find the extremal non-local correlations over the non-signaling correlations (1.34). It is called *super-quantum* correlations in [21]. Once the shared box can simulate the *super-quantum* correlations, we would say Alice and Bob share a PR-box.

1.1.3.2 The no-signaling polytope

In [20], the authors gave a geometric picture of the bi-partie no-signaling correlations. The dimensions of the no-signaling polytope could be obtained by the following calculation.

Suppose Alice and Bob have k_x and k_y kinds of inputs, respectively. Each input could have d kinds of outputs. Therefore, we could have $k_x k_y d^2$ kinds of joint probabilities $\Pr(A_x, B_y|x, y)$'s. Since we consider the no-signaling box, the constraints of the no-signaling theory (1.28) and the normalization conditions should be imposed on these joint probabilities. This leads to

1. The normalization conditions give $k_x k_y$ equalities.
2. The no-signaling constraints give $\sum_x d + \sum_y d = (k_x + k_y)d$ equalities.

However, these two types of conditions are not independent. For a fixed x, consider the normalization condition, one of the marginal probabilities $\Pr(A_x|x)$'s could be expressed by the others. Therefore, for a fixed x, one no-signaling condition can be deduced by the normalization condition and the other $(d-1)$ no-signaling conditions. So that, there are $k_x k_y + \sum_x (d-1) + \sum_y (d-1)$ linear independent equalities. Thus, the number of joint probabilities will be reduced to

$$dim = k_x k_y d^2 - (k_x k_y + \sum_x (d-1) + \sum_y (d-1)). \tag{1.35}$$

This means that the no-signaling boxes form a dim-dimensional polytope [20].

For the two-inputs/two-outputs case, both Alice and Bob have two kinds of inputs, and each input could have two outputs. Therefore, the bi-partite no-signaling boxes yield a 8-dimensional polytope.

The no-signaling polytope has 24 vertices, 16 of them correspond to local deterministic boxes and other 8 vertices correspond to the extremal non-local boxes.

The local deterministic box implies that the value of the marginal probabilities $\Pr(A_x|x)$ and $\Pr(B_y|y)$ is either 0 or 1. The joint probabilities of the local deterministic box should satisfy

$$\Pr(A_x, B_y|x, y) = \begin{cases} 1 & : A_x = \alpha x + \beta \pmod 2 \text{ and } B_y = \gamma y + \delta \pmod 2; \\ 0 & : \text{otherwise.} \end{cases}, \tag{1.36}$$

where α, β, γ and $\delta \in \{0, 1\}$. Actually, these local deterministic boxes can yield another convex polytope of locality by themselves. The facets of this polytope can be divided

into two types. The first type of facets restricting the joint probabilities in the polytope should be non-negative. The second type of facets correspond to the Bell-type inequality. Therefore, whenever the Bell-type inequality is violated by a set of joint probabilities, these joint probabilities will lie outside this polytope of locality. In the two-inputs/two-outputs case, the Bell-type inequalities correspond to the CHSH inequalities [22], i.e., (1.9) and its three symmetric partners by shifting the minus sign. There are 8 kinds of the CHSH inequalities.

The remainder vertices of the no-signaling polytope correspond to the extremal non-local boxes. The joint probabilities should satisfy

$$\Pr(A_x, B_y|x, y) = \begin{cases} \frac{1}{2} & :A_x + B_y = xy + \alpha x + \beta y + \gamma \pmod 2; \\ 0 & : \text{otherwise.} \end{cases}, \tag{1.37}$$

where α, β and $\gamma \in \{0, 1\}$. When $\alpha = \beta = \gamma = 0$, the extremal non-local box is the PR-box (1.34). Note that, each extremal non-local box will violate one of the CHSH inequalities. Furthermore, they can achieve the maximal violation, i.e., 4 or -4. It is interesting that one non-local vertex of the no-signaling polytope corresponds to one facet of polytope of locality.

There is another example for this connection. Suppose Alice and Bob have two inputs and each input has d outputs, one of the non-local vertex of the no-signaling polytope can achieve the maximal violation of the CGLMP inequality [8], i.e., the value of (1.21) is 4. The corresponding joint probabilities are as follows.

$$\Pr(A_x, B_y|x, y) = \begin{cases} \frac{1}{d} & :B_y - A_x(\text{mod } 2)=xy \ ; \\ 0 & : \text{otherwise.} \end{cases}, \tag{1.38}$$

Moreover, one could use this relations to find the complete set of the Bell-type inequalities by other extremal non-local boxes.

1.1.3.3 The communication complexity and the extremal non-local correlations

The communication complexity [23] is used to discuss how much of the communication is needed to solve a distributed decision problem. More specifically, if Alice has a n-bit string $\vec{x} = (x_1, x_2, ..., x_n)$ and Bob also has a n-bit string $\vec{y} = (y_1, y_2, ..., y_n)$. Their goal is to obtain the value of the function $f(\vec{x}, \vec{y})$. Note that the function f is a Boolean function: $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$. The definition of the communication complexity is the

minimum number of bits exchanged between Alice and Bob in the worst case in order to achieve such a task. If the value of the function $f(\vec{x}, \vec{y})$ can be obtained with only one bit of communication, one could say that the distributed decision problem has the trivial communication complexity. In practical, Alice and Bob could share some non-local resources to reduce the communication complexity such as an entangle state [26].

One show the power of the extremal non-local correlations in terms of communication complexity. When Alice and Bob share PR-boxes, for any distributed decision problem, the communication complexity is trivial [24]. Let us take an example to show the fact. If Alice and Bob want to determine the inner-product function ($IP_n : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}$), where

$$IP_n(\vec{x}, \vec{y}) = \sum_{i=1}^{n} x_i y_i. \tag{1.39}$$

Recall the PR-box (1.34), it keeps the inputs and the outputs perfectly correlated, i.e., $\Pr(A_x + B_y = xy|x, y) = 1$, $\forall x, y$. If Alice and Bob share n PR-boxes, they could take the i-th bit of the string x_i and y_i as the input of the i-th PR-box, and obtain the corresponding outputs A_{x_i} and B_{y_i}, respectively. Thus,

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} A_{x_i} + B_{y_i} = \sum_{i=1}^{n} A_{x_i} + \sum_{i=1}^{n} B_{y_i}. \tag{1.40}$$

This means that, Bob only needs to send the bit $b = \sum_{i=1}^{n} B_{y_i}$ to Alice, Alice can then determine the value of the Inner Product function $IP_n(\vec{x}, \vec{y})$ by her outputs and the bit b. Thus, the communication complexity is trivial in determining the inner-product function.

In [24], van Dam showed if one can determine the inner-product function $f(\vec{x}, \vec{y}) = \vec{x} \cdot \vec{y}$ of 2^n inputs by one bit communication, then one can also determine any Boolean function $f(\vec{x}, \vec{y})$ of n inputs with trivial communication complexity. According to the previous example, the inner-product function could have the trivial communication complexity for any number of inputs, we then know that for any distributed decision problem the communication complexity is trivial by sharing the PR-boxes. Moreover, some works [27, 28] show that the trivial communication complexity can indeed be achieved by sharing other non-local boxes. These non-local boxes are not the extremal non-local ones. However, the correlations of these boxes are still more no-local than the quantum correlations, i.e., $|C_{0,0} + C_{0,1} + C_{1,0} - C_{1,1}| \geq 2\sqrt{2}$. Therefore, we could not use the trivial communication complexity to single out the quantum correlations.

1.2 Information Causality

1.2.1 Information Causality single out quantum correlation?

In the previous section, we have shown that the no-signaling criterion can not be used to single out the quantum correlations. What principle could be our next candidate? The answer is Information Causality (IC) [29]. IC is presented through the following task which is equivalent to the random access code (RAC) [30]. Alice has a database with k elements, denoted by the vector $\vec{a} = (a_0, a_1, ..., k-1)$. Each element a_i is a binary random variable and is only known to Alice. A second distant party, Bob is given a random variable $b \in \{0, 1, 2, ..., k-1\}$. The task of Bob is to optimally guess Alice's database a_b after receiving m-bit string $\vec{\alpha}$ sent from Alice via the pre-shared correlation between Alice and Bob.

Before Alice sends the $\vec{\alpha}$, Bob cannot obtain the information about Alice's database, this is the no-signaling constraint. However, we are concerned about the information gain after Bob receives Alice's m-bit string $\vec{\alpha}$. IC states that the maximal mutual information shared between Alice and Bob cannot exceed the amount of classical communication, even they have pre-shared physical correlations, this condition can be expressed as follows:

$$I = \sum_{i=0}^{k-1} I(a_i; \beta | b = i) \leq m \ , \tag{1.41}$$

where $I(a_i; \beta | b = i)$ is Shannon's mutual information between a_i and Bob's guess d-bit β under the condition $b = i$.

The condition of IC (1.41) is not theory-independent [29]. It holds for quantum information theory. Let us define the pre-shared quantum state for Bob's part as ρ_B. In the task, after Alice sends the m-bit string $\vec{\alpha}$ to Bob, the information possessed by Bob about Alice's database includes $\vec{\alpha}$ and ρ_B. Note that, the mutual information between Alice's database and Bob's own information can not be more than m bit, i.e.,

$$I(\vec{a}; \vec{\alpha}, \rho_B) \leq m \ . \tag{1.42}$$

(1.42) can be proved by some elementary properties and the no-signaling condition, i.e., $I(\vec{a}; \rho_B) = 0$. On the other hand, one can prove $I \leq I(\vec{a}; \vec{\alpha}, \rho_B)$ by the elementary properties. Therefore, one can obtain that $I \leq I(\vec{a}; \vec{\alpha}, \rho_B) \leq m$. Thus, (1.41) holds for the pre-shared quantum correlations.

IC is not only satisfied by the pre-shared quantum correlations, but also violated by the pre-shared PR-box which can simulate the extremal non-local correlations under the no-signaling constraint [29]. This shows IC is more suitable to single out quantum correlations. Moreover, the Tsirelson bound could be derived from IC. For the multi-setting case, we show that more general Tsirelson-type inequalities can be derived by the signal decay theory [50, 51] and IC. One can see Chapter 2 for more detailed proof. According to these results, IC may be the natural principle which we are looking for.

Actually, one should prove that IC can exclude all non-quantum correlations so that IC could be used to single out quantum correlations. Therefore, we still need more checks. Since there are some non-quantum correlations below the Tsirelson bound, and therefore we cannot express the entire boundary of quantum correlations by the no-signaling boxes which achieving the Tsirelson bound. Thus, we have to check if the entire quantum boundary can be recovered by IC. In [35], the relation between IC and the quantum boundary for two-inputs/two-outputs case has been studied. In this case, IC is not violated by the quantum correlations, therefore IC is supported. Besides, the relation between bi-partite IC and the more general no-signaling boxes was studied. In the multi-level no-signaling box [36], the relation between bi-partite IC and quantum correlations is unclear. In the multi-partite no-signaling box [37, 38], bi-partite IC is violated by the most non-quantum correlations, but not all the non-quantum correlations violate bi-partite IC. The reason may be, the set of quantum correlations for an arbitrary number of parties can not be singled out by using the principle with bi-partite information concepts [39]. Therefore, although bi-partite IC is satisfied by some non-quantum correlations, it does not mean IC cannot single out the quantum correlations. It just means that we need to generalize the bi-partite IC to the multi-partite one.

However, even in this simplest case (the two-inputs/two-outputs no-signaling box) [35], it is still unclear that if IC could exclude all the non-quantum correlations or not. Why it is so difficult to prove the criterion? The reason is that there are too many strategies for using the no-signaling boxs. Even sharing the same no-signaling box, different strategies for using the box will yield different probabilities to win the task. Therefor, the mutual information between Alice's database a_b and Bob's guessing bit β will be different. So, it is hard to ascertain if the non-quantum correlations violate IC or not.

However, we can have more understanding about the relation between IC and quantum

correlations by testing IC for different quantum communication protocols or more general settings. We will study these more general cases in chapter 3.

In the following subsections, we will review some important relations between IC and quantum correlations which we mentioned in this subsection. They include how to derive the Tsirelson bound from IC and how to compare quantum boundary and IC for the two-inputs/two-outputs case.

1.2.2 Information Causality and the Tsirelson bound

1.2.2.1 The extremal non-locality violates Information Causality

As shown in [29], IC will be violated by the PR-box which is the extremal non-local box under the constraint of no-signaling theory. One can use the specific RAC protocol to show the fact.

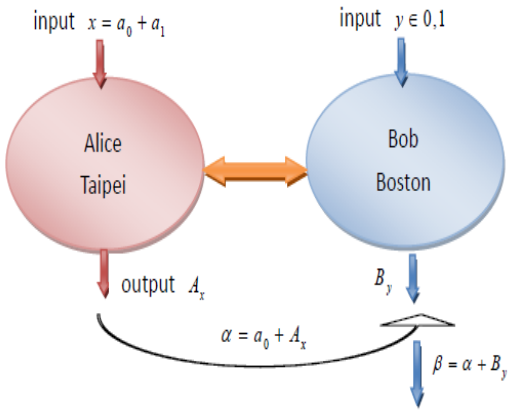

Figure 1.1: The communication protocol for $k = 2$ task

Let us consider the simplest case, it is described in Fig1.1. The protocol is used to solve $k = 2$ task. To be specific, first, Alice encodes her database as the input x of the no-signaling box, where $x = a_0 + a_1$. On the other hand, the distant Bob takes his given bit b as the input y of the no-signaling box. They then obtain the outputs A_x and B_y, respectively. Note that, in this case x, y, A_x and $B_y \in \{0, 1\}$. Second, after obtaining their outputs, Alice then sends a bit $\alpha = A_x + a_0$ to Bob. Finally, Bob can decode Alice's database a_b by calculating the bit $\beta = \alpha + B_y$. Since $\beta = \alpha + B_y = A_x + B_y + a_0$, if

Alice's and Bob's inputs and outputs are perfectly correlated, i.e., $A_x + B_y = xy$, $\forall x, y$, Bob can then obtain $\beta = xy + a_0 = (a_0 + a_1)y + a_0 \ (mod \quad 2)$. Obviously, either if $y = 0$, $\beta = a_0$ or if $y = 1$, $\beta = a_1$. Thus, Bob can guess Alice's database perfectly. If not the PR-box, Bob's successful probability depends on the value of the joint probability $\Pr(A_x + B_y = xy|x, y)$.

In this case, the successful probability for the task is related to the CHSH function (1.32). The successful probability to guess a_0 right under the condition $b = 0$ is denoted as P_0, similarly P_1 for a_1. If Alice's input is unbiased, i.e., the marginal probabilities $\Pr(x = i) = \frac{1}{2}$ $i \in \{0, 1\}$, the successful probabilities can be rewritten as

$$\begin{aligned} P_0 &= \frac{1}{2}(\Pr(A_x + B_y = 0|x = 0, y = 0) + \Pr(A_x + B_y = 0|x = 1, y = 0)) \\ P_1 &= \frac{1}{2}(\Pr(A_x + B_y = 0|x = 0, y = 1) + \Pr(A_x + B_y = 1|x = 1, y = 1)). \end{aligned} \tag{1.43}$$

Note that, one can obtain $\Pr(A_x + B_y = 0|x = 0, y = 0) = \frac{1+C_{0,0}}{2}$ by the definition (1.33) of the correlation function $C_{0,0}$ and the normalization condition of the joint probability. Similarly, $\Pr(A_x + B_y = 0|x = 1, y = 0) = \frac{1+C_{1,0}}{2}$, $\Pr(A_x + B_y = 0|x = 0, y = 1) = \frac{1+C_{0,1}}{2}$ and $\Pr(A_x + B_y = 1|x = 1, y = 1) = \frac{1-C_{1,1}}{2}$. Consider these relations, if the marginal probability $\Pr(b)$ is uniform, one can then find that the successful probability is equivalent to the CHSH function, i.e.,

$$P = \frac{1}{2}(P_0 + P_1) = \frac{1}{8}(4 + C_{0,0} + C_{0,1} + C_{1,0} - C_{1,1}). \tag{1.44}$$

Since we have already know the locality bound and the Tsirelson bound for the CHSH function, we can estimate the local bound and the quantum bound for the function P. The local bound for P is $\frac{3}{4}$ and the quantum one is $\frac{2+\sqrt{2}}{4}$. Under the no-signaling condition, we can also know that the bound for the PR-box is 1. In this case, Bob can always guess Alice's database with certainty. Therefore, the mutual information between a_i and β under the condition $b = i$ is 1, i.e., $I(a_0; \beta|b = 0) = I(a_1; \beta|b = 1) = 1$. That is, the mutual information I is 2. Thus, IC is violated by sharing a PR-box.

1.2.2.2 Information Causality derives the Tsirelson bound

The Tsirelson bound emerges from IC, it is one of the important results in [29]. In order to show the fact, one has to connect the mutual information I and the CHSH function. Recall the CHSH function is equivalent to the successful probability of the task. Therefore,

one should rewrite the mutual information I in terms of the successful probability of the task. First, one can rewrite the mutual information I as $\sum_i H(a_i|b=i) - H(a_i|\beta, b=i)$, where H is the binary entropy function. Assuming that the marginal probability $\Pr(a_i)$ is uniform, thus the entropy function $H(a_i|b=i)=1$ for each a_i. On the other hand, one can use the chain rule to obtain $H(a_i|\beta, b=i) \leq H(a_i+\beta(mod \quad 2)|b=i) = H(P_i)$, where P_i is the successful probability for Bob to guess a_i. Therefore, the mutual information becomes

$$I \geq \sum_{i=0}^{k-1}[1 - H(P_i)]. \tag{1.45}$$

Here, in order to reveal the Tsirelson bound from IC, one needs a specific strategy to use the no-signaling box. One way is to nest many two-inputs/two-outputs no-signaling boxes. In this way, Alice's database $\vec{a} = (a_0, a_1, ..., a_{k-1})$ has $k = 2^n$ elements with n an integer. Guessing a_b perfectly is Bob's task. Thus, Bob is given n-bit string $\vec{y} = (y_0, y_1, ..., y_{n-1})$ used to encode $b = \Sigma_{i=0}^{n-1} y_i 2^i$. Alice's output of the i-th box is denoted as $A^{(i)}$, similarly $B^{(i)}$ for Bob. One may need to nest $k-1$ two-inputs/two-outputs no-signaling boxes to solve the task. Let us take $k=4$ and $m=1$ task for example. The RAC protocol is described in Fig. 1.2. Since there are 4 elements in Alice's database, she can divide them into two subset and therefore each subset has two elements. As for the $k=2$ simplest RAC protocol, she can then encode each subset as the input of the no-signaling box. In this case, Alice can input $a_0 + a_1$ in the first box and then send $\alpha^{(1)} = A^{(1)} + a_0$ to make Bob able to guess a_0 and a_1. Similarly, she inputs $a_2 + a_3$ in the second box and sends $\alpha^{(2)} = A^{(2)} + a_2$ for the same purpose. Note that, Alice only allows to send 1 bit, thus she needs the third box to make Bob able to guess $\alpha^{(1)}$ and $\alpha^{(2)}$. Similarly, Alice inputs $\alpha^{(1)} + \alpha^{(2)}$ and sends the bit $\alpha = A^{(3)} + \alpha^{(1)}$ to Bob. On the other hand, suppose Bob's task is to guess a_0, the bit $b=0$ is encoded as $\vec{y} = (y_0 = 0, y_1 = 0)$. Therefore, Bob inputs y_0 and y_1 in the first and the third box, respectively. Bob can get $\alpha^{(1)}$ by calculating $\alpha + B^{(3)}$. Similarly, he can then use the output $B^{(1)}$ of the first box to get a_0 by calculating $\alpha^{(1)} + B^{(1)}$. Therefore, Bob's optimal guessing bit $\beta = \alpha + B^{(3)} + B^{(1)}$. This communication protocol can be generalized to any $k = 2^n$ case.

In this RAC protocol, the successful probability for Bob to guess a_b right is given by

$$P_b^n = \frac{1}{2}(1 + E_1^{n-s} E_2^s), \tag{1.46}$$

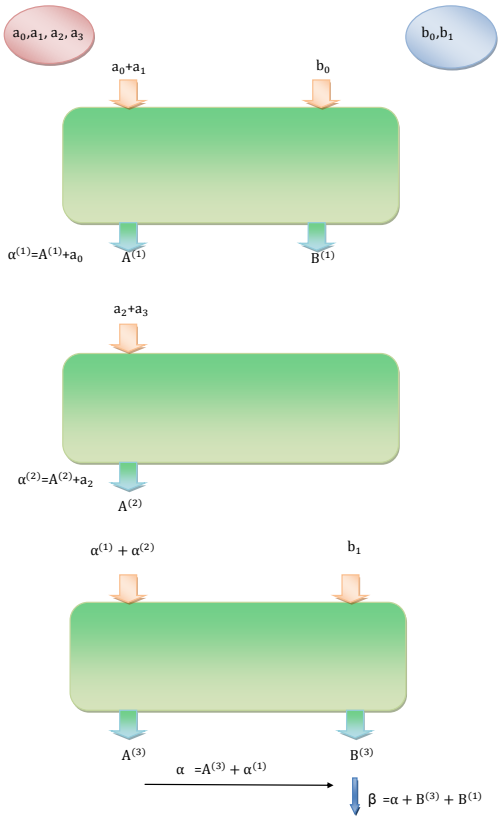

Figure 1.2: The RAC protocol for $k = 4$ task

where E_i is related to the successful probability P_i (1.43) in the $k = 2$ RAC protocol, the relation is $E_i = 2P_i - 1$. The value of s depends on the encoded vector $\vec{y}$, the definition of s is $s = \sum_{i=0}^{n-1} y_i$.

Now, we are ready to calculate the bound from IC. Put (1.46) into I (1.45) and impose

the condition $1 - H(\frac{1+Z}{2}) \geq \frac{Z^2}{2ln2}$, we can then obtain

$$\begin{aligned} I &\geq \sum_{b=0}^{k-1}[1 - H(P_b^n)], \\ &= \sum_{s=0}^{n} \begin{pmatrix} n \\ s \end{pmatrix} [1 - H(\frac{1}{2}(1 + E_1^{n-s}E_2^s))], \\ &\geq \frac{1}{2\ln 2} \sum_{s=0}^{n} \begin{pmatrix} n \\ s \end{pmatrix} (E_1^2)^{n-s}(E_2^2)^s, \\ &= \frac{1}{2\ln 2}(E_1^2 + E_2^2)^n. \end{aligned} \tag{1.47}$$

One can find that if

$$E_1^2 + E_2^2 > 1, \tag{1.48}$$

then the mutual information I will be bigger than 1 for some n, i.e., IC is violated.

Recall the definition (1.43) of the successful probability P_i , one can rewrite the constraint as

$$(C_{0,0} + C_{1,0})^2 + (C_{0,1} - C_{1,1})^2 > 4. \tag{1.49}$$

In fact, this is nothing but the violation of Uffink's quadratic inequality [41]. Consider the special case, let $P_1 = P_2 = P$ and therefore $E_1 = E_2 = E$. In this case, IC will be violated while $2E^2 > 1$. Recall the relation between the successful probability and the CHSH function (1.44), one can find that the Tsirelson bound is violated while $E > \frac{1}{\sqrt{2}}$. This case shows that the Tsirelson bound emerges from IC.

1.2.3 Information Causality and the boundary of quantum correlations

The authors in [35] tried to answer that if IC can recover the entire quantum boundary or not. In their proof, they considered the simplest case, $n = 1$, $k = 2$ RAC protocol and the quantum correlations for two-inputs/two-outputs no-signaling box. They studied the no-signaling correlations by several two dimensional slices of the non-signalling polytope, that is equivalent to considering the following noisy PR-box.

$$PR_{\lambda,\mu} = \lambda PR + \mu B + (1 - \lambda - \mu)I, \tag{1.50}$$

where PR is the PR-box and B could be another extremal non-local box (1.37) or local deterministic box (1.36). Thus, the noisy PR-boxes can be grouped into two families. In

[35], the Uffink's inequality (1.49) is taken as the condition for the violation of IC. On the other hand, they imposed two quantum constrains to obtain the quantum boundary. These constraints correspond to different families of noisy PR boxes. Finally, one may compare the condition for the violation of IC and the quantum boundary. Two results were arrived in [35]. First, the bi-partite quantum correlations satisfy IC. Second, only parts of the quantum boundary emerges from IC.

For the first family, B is one of the extremal non-local box (1.37) except the PR one. In this family, the marginal probabilities $\Pr(A_x|x)$ and $\Pr(B_y|y)$ are uniform. The correlation functions $C_{x,y}$ can be rewritten as

$$C_{0,0} = \lambda + (-1)^{\gamma}\mu,\ C_{0,1} = \lambda + (-1)^{\beta+\gamma}\mu,\ C_{1,0} = \lambda + (-1)^{\alpha+\gamma}\mu,\ C_{1,1} = \lambda + (-1)^{\alpha+\beta+\gamma}\mu, \tag{1.51}$$

where α, β and γ are used to label the types of extremal non-local boxes (1.37). For fixed α, β and γ (or equivalently, a chosen extremal non-local box B), one could plug (1.51) into the condition (1.49) for the violation of IC so that IC and the noisy PR-box are related. For example, if $(\alpha, \beta, \gamma) = (0, 1, 0)$, one may find that IC is violated when

$$\lambda^2 + \mu^2 > \frac{1}{2}. \tag{1.52}$$

On the other hand, if a set of correlation functions $C_{x,y}$ can be reproduced by the quantum system (quantum operators and quantum state), this set of correlation functions need to satisfy the following quantum condition,

$$|C_{0,0}C_{1,0} - C_{0,1}C_{1,1}| \leq \sum_{j=0,1} \sqrt{(1 - C_{0,j}^2)(1 - C_{1,j}^2)}. \tag{1.53}$$

The above condition was proposed by Landau [42], and an equivalent set of conditions were proposed by Masanes [43] expressed in different form. Note that, if (1.53) is satisfied by a set of correlation functions $C_{x,y}$, the corresponding marginal probabilities $\Pr(A_x|x)$ and $\Pr(B_y|y)$ must be uniform. As for the condition for the violation of IC, for fixed α, β and γ, one can plug (1.51) into the quantum condition (1.53) and find the associated quantum boundary. For example, when $(\alpha, \beta, \gamma) = (0, 1, 0)$, the boundary of quantum correlations and the condition for the violation of IC are the same. In this case, the quantum boundary emerges from IC. However, this is not always true, for example for $(\alpha, \beta, \gamma) = (1, 1, 1)$. There are some non-quantum correlations satisfying IC. This does

not mean IC cannot exclude these non-quantum correlations, because if one use another strategy for the protocol such as nesting some non-quantum boxes mentioned in the previous subsection, these non-quantum boxes may violate IC.

In the second family of the noisy PR-box, B is the local deterministic box considered in (1.36). Among these local boxes, one could consider the special ones which lie on the CHSH facets of the polytope of locality discussed in section 1.3.2. Therefore, the parameters of these special boxes given by (1.36) should satisfy $\alpha\gamma + \beta + \delta = 0 \mod 2$. As for the first family, one can use the parameters of the noisy PR-box given in (1.50) to rewrite the correlation function $C_{x,y}$. For example, if $(\alpha, \beta, \gamma, \delta) = (0, 0, 0, 0)$, one then obtains

$$C_{0,0} = C_{0,1} = C_{1,0} = \lambda + \mu, C_{1,1} = \mu - \lambda. \tag{1.54}$$

Note that, the marginal probabilities $\Pr(A_x|x)$ and $\Pr(B_y|y)$ are no longer uniform. Therefore, the marginal correlations C_x and C_y are not zero. The marginal correlations are the linear combinations of the marginal probabilities. The definition for marginal correlations is $C_x = \Pr(A_x = 0|x) - \Pr(A_x = 1|x)$ and $C_y = \Pr(B_y = 0|y) - \Pr(B_y = 1|y)$. Here, the marginal correlations $C_{x=0} = C_{x=1} = C_{y=0} = C_{y=1} = \mu$. In this case, the quantum constraint (1.53) is not suitable for picking up the quantum correlations. Instead, one may use the following condition and its three symmetric partners by shifting the minus sign. These conditions include the marginal correlations, i.e.,

$$|\arcsin(D_{0,0}) + \arcsin(D_{0,1}) + \arcsin(D_{1,0}) - \arcsin(D_{1,1})| \leq \pi, \tag{1.55}$$

where $D_{x,y} = \frac{C_{x,y} - C_x C_y}{\sqrt{(1-C_x^2)(1-C_y^2)}}$. This condition was proposed in [44, 45]. Note that, (1.55) is the quantum constraint in the first step of the hierarchal semidefinite programming [45]. This means, if a set of correlation functions $C_{x,y}$, C_x and C_y satisfies (1.55), we could not make sure if it could be reproduced by a quantum system, unless it can satisfy all the quantum constraint in each step of the hierarchical semidefinite programming. Plug all the correlation functions (1.54) into the quantum constraint (1.55), one can obtain the associated quantum boundary. Similarly, plug (1.54) into the condition (1.49) for the violation of IC, one may find that IC is violated when $(\lambda + \mu)^2 + \lambda^2 > \frac{1}{2}$. In this case, the quantum boundary from the quantum constraint (1.55) and the condition of IC are not the same. There are some non-quantum correlations satisfying IC. As for the case in

the first family, one may use another strategy such that these non-quantum correlations will still violate IC.

However, with the above examples, one may only know that quantum correlations satisfy IC and could not make sure if IC can exclude all the non-quantum correlations. Moreover, one may find there are some assumptions and approximations in arriving the condition (1.48) for the violation of IC. Therefore, instead of using this condition, we directly calculate the mutual information I of the quantum correlations and then compare with IC. See Chapter 3 for more details.

1.3 Signal propagating and noisy computation

Signal propagating suffers signal decay. It is a very important issue when considering communication systems. On the other hand, one could image the noisy computation as a sequence of steps [49]. Each step has an input and an output. The output has some information about the associated input, and the output will be the input of the next step. Once some step produces a random noise, one can interleave some parallel step to control the error rate of the entire computation. In this way, each step of the noisy computation can be seen as a noisy communication channel. Thus, one may expect that the issue of signal decay will also appear for the noisy computation. von Neumann was the first to be aware of this fact [47]. He suggested that the error of the computation should be treated by thermodynamical method as the treatment for the communication in Shannon's work [46]. This means that, for the noisy computation, one should use some information-theoretic methods related to the noisy communication. After thirty years from the publication of von Neumann's work, Pippenger took some information theoretic arguments to study the reliable noisy computation [48]. After that, Evans and Schulman proposed a new result for the efficient signal propagating and then use it to study the threshold of noisy computation [50, 51]. In this section, we want to review Evans and Schulman's work and also compare it with the previous results obtained by von Neumann and Pippenger.

1.3.1 The efficient propagation through a noisy channel

In the process of signal propagation, we will face the problem of signal decay. To be specific, it is manifested through the data processing inequality. Consider two communication channels and three random variables. Let variable X be the input of the first channel, the variable Y is the output of the first channel and also is the input of the second one, and the variable Z is the output of the second channel, i.e., the cascade of two communication channels: $X \to Y \to Z$. No matter what are the properties of the communication channels, the data processing inequality states that

$$I(X;Z) \leq I(X;Y). \tag{1.56}$$

The mutual information $I(X;Y) = H(Y) - \sum_i \Pr(X = i) H(Y|X = i)$, where $H(Y)$ and $H(Y|X)$ are the Shannon entropies for the probability $\Pr(Y)$ and the conditional

probability $\Pr(Y|X)$, respectively.

Here are some properties of the communication channel which we will use later. If a communication channel has m inputs and n outputs, this communication channel could be expressed by a $m \times n$ matrix. The (i, j)-th element of this matrix is the conditional probability of obtaining the output j under the condition of the input i. Let us take the binary channel for example. A general binary channel is specified by

$$A = \begin{pmatrix} \sin^2 \alpha & \cos^2 \alpha \\ \cos^2 \beta & \sin^2 \beta \end{pmatrix}. \tag{1.57}$$

Let the random variables G and F be the input and output of the binary channel A, respectively. Since the output probability can be expressed as $\Pr(F = j) = \sum_{i=0}^{1} \Pr(F = j|G = i) \Pr(G = i)$, we can plug in the elements of the channel to obtain $\Pr(F = 0) = \sin^2 \alpha \Pr(G = 0) + \cos^2 \beta \Pr(G = 1)$ and $\Pr(F = 1) = \cos^2 \alpha \Pr(G = 0) + \sin^2 \beta \Pr(G = 1)$. Note that, there is a special channel called the symmetric channel. The rows of the symmetric channel are cyclic under permutations, so are the columns. The binary symmetric channel has the following form:

$$A = \begin{pmatrix} \frac{1+\xi}{2} & \frac{1-\xi}{2} \\ \frac{1-\xi}{2} & \frac{1+\xi}{2} \end{pmatrix}, \tag{1.58}$$

where ξ is the noisy parameter and $0 \leq \xi \leq 1$.

Now, we can study the new result proposed by Evans and Schulman. Naively, by the data processing inequality one may expect the upper bound of the ratio $\frac{I(X;Z)}{I(X;Y)}$ to be 1. In [48], Pippenger proposed that the upper bound of the ratio $\frac{I(X;Z)}{I(X;Y)}$ for the binary symmetric channel should be tighter than 1. In [50, 51], Evans and Schulman further gave the upper bound of the ratio $\frac{I(X;Z)}{I(X;Y)}$ for the general binary channels. It is amazing that this bound is a function of the second communication channel in the the process of signal propagation. If the second channel is specified by (1.57), one can obtain the upper bound of the ratio $\frac{I(X;Z)}{I(X;Y)}$ to be $\sin^2(\alpha - \beta)$. For the binary symmetric channel (1.58), the maxima of the ratio $\frac{I(X;Z)}{I(X;Y)}$ is ξ^2. Note that, this bound is tighter than ξ proposed by Pippenger [48].

Evans and Schulman used the following fact to reduce the complexity of the proof. They found that the maxima of the ratio $\frac{I(X;Z)}{I(X;Y)}$ is achieved when the conditional probabilities $\Pr(Y|X = 0)$ and $\Pr(Y|X = 1)$ are almost indistinguishable. This condition implies that

the mutual information between the variables X and Y are very small. Therefore, we can know that propagating weak signal is the most efficient way to avoid signal decay in a noisy channel. After reducing to the weak signal case, the problem of maximizing the ratio $\frac{I(X;Z)}{I(X;Y)}$ becomes much simpler because the ratio $\frac{I(X;Z)}{I(X;Y)}$ can be expressed as the ratio of the relative entropy $D(p\|q) := \sum_x \Pr(p=x) \log \frac{\Pr(p=x)}{\Pr(q=x)}$. Thus, the ratio $\frac{I(X;Z)}{I(X;Y)}$ does not depend on $\Pr(X)$. We then have

$$\frac{I(X;Z)}{I(X;Y)} = \frac{D((\vec{p}+\vec{\epsilon})\cdot A \parallel \vec{p}\cdot A)}{D(\vec{p}+\vec{\epsilon}\parallel \vec{p})} + O(\epsilon), \tag{1.59}$$

where $\vec{p}$ is the probability $\Pr(Y)$ and $\vec{p}+\vec{\epsilon}$ is the conditional probability $\Pr(Y|X=0)$. Consider the normalization condition, $\vec{\epsilon}$ could have the form of $(\epsilon, 1-\epsilon)$. Note that, for the weak signal case, ϵ should be sufficiently small. One can then maximize the leading term of the expansion of (1.59) over the probabilities $\vec{p}$ (or equivalently $\Pr(Y=0)$). Finally, we can obtain the maxima of the ratio $\frac{I(X;Z)}{I(X;Y)}$ to be $\sin^2(\alpha-\beta)$ as proposed by Evans and Schulman.

In Chapter 2 and 3, we will combine the bounds on the ratio $\frac{I(X;Z)}{I(X;Y)}$ and IC to obtain the Tsirelson-type inequalities for the two-level systems and also the form of the Bell-type inequality for the multi-level systems.

1.3.2 The noisy computation

1.3.2.1 The model of noisy computation

Before studying how to implement the new result for signal propagation in noisy computation, we need to discuss the model of noisy computation in a more precise way. Modeling the computation as a sequence of steps is more convenient for us to connect the computation and the communication but is not precise enough. The more precise model was proposed by von Neumann in 1952 [47]. It is described in Fig.1.3.

The model of noisy computation is a noisy circuit. The circuit has n Boolean inputs and one Boolean output. Note that, the entire circuit is used to compute the Boolean function $f : \{0,1\}^n \to \{0,1\}$ and is constructed by many gates which correspond to the vertices in Fig.1.3. Each gate has a fixed number of Boolean inputs and one Boolean output. In Fig.1.3, the edge from vertex a to b corresponds to the variable as the output of gate a and the input of gate b. Assuming these connections are directional and do not allow any feedback. Therefore, the entire circuit is a directional and acyclic graph. Note that, the

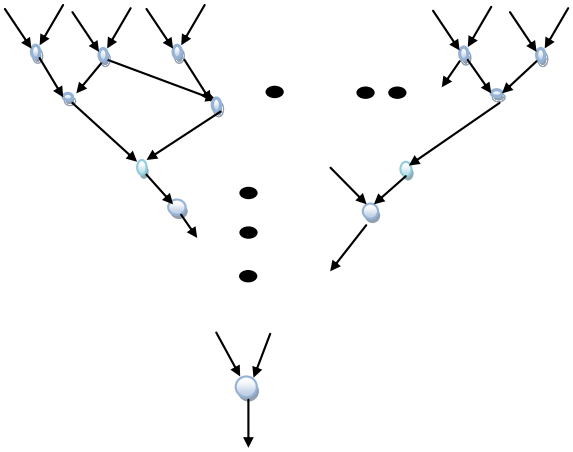

Figure 1.3: The model of noisy computation

number of the outgoing edges for each gate does not have to be one. If each gate only has only one outgoing edge, the circuit then has the tree structure. One may call the circuit a formula. The following jargons are used to characterize the model of computation.

- Depth: The depth is the number of gates in the longest path of the circuit.
- Size: The size is the number of the gates in the entire circuit.

In this model, for each gate, the unsuccessful probability for obtaining the Boolean function is ϵ. This means that, we cannot always have the perfect computation for the entire circuit. However, we can study the reliable computation. The circuit is the reliable computation, if we can obtain the Boolean function of the n inputs $f : \{0,1\}^n \to \{0,1\}$ with the $(1-\delta) \geq \frac{1}{2}$ successful probability.

1.3.2.2 The tolerable error rate and the depth for a reliable computation.

For the reliably computational circuit, the error rate for each gate and the depth of the circuit are the essential values because both of them will affect the successful probability for the entire computation.

In [47], von Neumann calculated the tolerable error rate for the reliable computation. In von Neumann's computation circuit, each gate has three inputs. The reliable computation can be achieved by the noisy gates when the error rate ϵ of each gate is independent and less than $\frac{1}{6}$. In this case, the error rate could not be arbitrary for reliable computation. On the other hand, one may increases the depth of the circuit to control the successful probability for the entire computation. Since the computational time is increased with bigger depth of the circuit, the price to pay for the reliable noisy computation is the computational time.

Moreover, in [48] Pippenger used the information-theoretic method to prove that the limit of ϵ and the longer time are required for the reliably noisy formulas (tree structure of the circuits). Let $x_1, x_2, ..., x_n$ be the Boolean inputs of the circuit. In Pippenger's proof, for any input x_m, one can obtain that the output of the noiseless circuit is equal to x_m or the complement of x_m while the other $n-1$ inputs correspond to a specific setting so that the circuit is noiseless. Therefore, with this specific setting, the output O of the reliable circuit should be strongly correlated with the input x_m. In this information theoretical setup, it implies that the mutual information between x_m and the output O of the reliable circuit should be high enough. Since we know that the structure of the noisy gates affects the mutual information between the input x_m and the output O, the total information flow from the input x_m to the output O is bounded by the summation of information flow for each path p from x_m to the output O. According to this condition, Pippenger obtained the upper bound of the mutual information

$$I(x_m; O) \leq \Sigma_p \xi^{|p|}, \tag{1.60}$$

where $|p|$ is the number of the gates in the path p from x_m to O and the error rate for each gate is $\epsilon = \frac{1-\xi}{2}$.

With this inequality, Pippenger then obtained the lower bound of the depth for the reliable circuit. Assuming n inputs of the circuit and that each gate has at most k inputs. Note that, the error rate for each gate $\epsilon = \frac{1-\xi}{2}$. Pippenger then obtained the constraint for the depth c of the $1-\delta \geq \frac{1}{2}$-reliable circuit as follows:

- (i) if $\xi > \frac{1}{k}$ then $c \geq \log(n\Delta)/\log(k\xi)$,
- (ii) if $\xi \leq \frac{1}{k}$ then $n \leq 1/\Delta$,

where $\Delta := 1 + \delta \log \delta + (1-\delta)\log(1-\delta)$. For $k = 3$ case, one may find the tolerable error rate $\frac{1}{3}$ is higher than the one in the von Neumann's argument. On the other hand, the lower bound of the depth must increase. Here, the lower bound of the noiseless circuit is $\log_k n$ and one may multiply it by at least $\frac{1}{(1+\log_k \xi)}$ to obtain the lower bound of the depth for the noisy circuit. Moreover, when the error rate is higher than the tolerable one, the number of inputs cannot be arbitrary and it will be limited by Δ. However, since Pippenger's argument only works for the formulas, we are not sure if the above bounds on the error rate and the depth also hold for the other circuits.

Pippenger used three inequalities to obtain the upper bound (1.60) on the mutual information. The first one is the data processing inequality. Assuming the gate G produces the output Y with inputs Y_i for $i \in \{1, ..., k\}$. The mutual information between the essential input x_m and Y is smaller than the one between x_m and $Y_1, Y_2, ..., Y_k$, i.e., $I(x_m; Y) \leq I(x_m; Y_1, Y_2, ..., Y_k)$. The second one is $I(x_m; Y_1, Y_2, ..., Y_k) \leq \sum_i I(x_m; Y_i)$. This inequality holds when we can obtain the information about x_m from Y_i independently. Therefore, the second inequality holds for the formulas but may not be satisfied by other circuits. In [51], one example (in Fig.1.4) can show that the inequality does not hold any more. In this example, Y_2 is the input for both of the gates which are noiseless NOR gates [1]. Therefore, one can know x_m by Y_1 and Y_2, i.e., $I(x_m; Y_1, Y_2) = 1$. Note that, x_m could not be obtained when one only knows Y_1 or Y_2, i.e., $I(x_m; Y_1) = I(x_m; Y_2) = 0$. Obviously, in this case, $I(x_m; Y_1, Y_2) > \sum_{i=1}^{2} I(x_m; Y_i)$.

The third ingredient is $I(x_m, Z) \leq \xi I(x_m; Y)$, where Z is the output for the other gate with the input Y. One can image $Z = Y + V$, where V is a Boolean variable and its value is 1 with probability $\frac{1-\xi}{2}$ and 0 with probability $\frac{1+\xi}{2}$ [48]. With this assumption, one may obtain the third inequality. Note that, this inequality also implies $\frac{I(x_m,Z)}{I(x_m;Y)} \leq \xi$.

In order to obtain the lower bound of the depth for more general circuits, Evans and Schulman overcame the difficulty for the second inequality and modified the third inequality [50, 51]. Instead of summing over the individual mutual information between x_m and Y_i, Evans and Schulman directly calculated the bound of the mutual information between x_m and a set of random variables. On the other hand, for the third inequality, they had an impressive modification. They found that the noisy gate could correspond a binary symmetric channel given by (1.58). According to the proof in the previous

[1] When both the inputs of the NOR gate are 0, the output will be 1, otherwise the output is 0

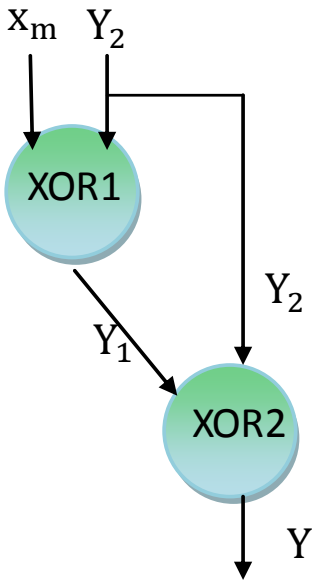

Figure 1.4: The example for the failure of inequality $I(x_m; Y_1, Y_2, ..., Y_k) \leq \sum_i I(x_m; Y_i)$

subsection, instead of ξ implied by the third inequality, one could obtain the upper bound of the ratio $\frac{I(x_m;Z)}{I(x_m;Y)}$ to be ξ^2. Therefore, one can modify the third inequality to be $I(x_m, Z) \leq \xi^2 I(x_m; Y)$.

Using these modifications, one can then obtain the upper bound of the mutual information, i.e.,

$$I(x_m; W) \leq \Sigma_p \xi^{2|p|}, \tag{1.61}$$

where W is used to denote the set of variables and $|p|$ is the number of the gates on the path p from x_m to W. Note that, if W is the output O, one may find this upper bound is tighter than (1.60). Similar to Pippenger's argument, one may use this upper bound to obtain the lower bound of the depth c for the $1 - \delta \geq \frac{1}{2}$-reliable circuit using the gates with at most k inputs. The results are as follows:

- (i) if $\xi^2 > \frac{1}{k}$ then $c \geq \log(n\Delta)/\log(k\xi^2)$,
- (ii) if $\xi^2 \leq \frac{1}{k}$ then $n \leq 1/\Delta$,

where $\Delta := 1 + \delta \log \delta + (1-\delta)\log(1-\delta)$, and the error rate for each gate is $\epsilon = \frac{1-\xi}{2}$.

Compare this result with the ones from von Neumann's and Pippenger's arguments, one can find two things. The first one is that, the threshold of the error rate for each gate becomes looser. For the reliable computation with $k = 3$ gates, von Neumann showed it can be achieved when $\epsilon < \frac{1}{6}$, and the error rate requires $\epsilon < \frac{1}{3}$ for Pippenger's argument. For the same case, Evans and Schulman showed that the threshold error rate is $\frac{1}{2}(1 - \frac{1}{\sqrt{3}})$ which is bigger than the ones from von Neumann and Pippenger's arguments. The second one is that, the lower bound of the depth is bigger than the one from Pippenger's arguments. As known that in order to achieve the reliable computation using the noisy gates, one needs to increase the depth of the noisy circuit. Here, the lower bound of the depth for the noiseless circuit is $\log_k n$. From Pippenger's arguments, one needs to multiply it by $\frac{1}{(1+\log_k \xi)}$ to obtain the lower bound of the depth for the noisy circuit. On the other hand, Evans and Schulman obtained that one needs to multiply it by $\frac{1}{(1+\log_k \xi^2)}$. Thus, one may need more time to realize the reliably noisy computation.

In Chapter 2, we will use the bound on the error rate to check if IC allows the $1-\delta \geq \frac{1}{2}$-reliable computation.

1.4 Semidefinite programming and the quantum correlations for the bi-partite systems

Since the joint probability (1.26) can describe the correlation shared by two distant partites, therefore if one want to ensure the shared correlation is quantum, one need to check if the corresponding joint probabilities can be reproduced from quantum mechanics or not. This is a very important issue because one can know the quantum limitation for some information-theoretic tasks and the non-local nature of the quantum correlations. However, the quantum correlations cannot form a polytope. Therefore, unlike the no-signaling and the local correlations, the quantum correlations cannot be described by a finite set of joint probabilities which correspond to the vertices of the polytope.

On the other hand, in the previous sections, we have seen that quantum correlations may violate the Bell-type inequalities. However, not all the maximally quantum violation of these inequalities could be found analytically. Therefore, one may need some numerical methods to find the maximally quantum violation of these Bell-type inequality.

Semidefinite programming (SDP) is the procedure of optimizing a linear function subjected to certain conditions associated with a positive semidefinite matrix X. A positive semidefinite matrix X should satisfy the condition $v^\dagger X v \geq 0$ for $v \in \mathbb{C}^n$, and is denoted by $X \succeq 0$. The standard primal problem of SDP is as follows. Given the $n \times n$ symmetric matrices C and D_q's with $q = 1, \cdots, m$, we like to optimize the $n \times n$ positive semidefinite matrix $X \succeq 0$ such that we can achieve the following:

$$\text{minimize} \qquad \text{Tr}(C^T X) \tag{1.62a}$$

$$\text{subject to} \qquad \text{Tr}(D_q^T X) = b_q, \quad q = 1, \cdots, m\ . \tag{1.62b}$$

Each primal problem of SDP has a corresponding dual problem. One can solve the maximization problem whenever both the optimal solutions for the primal and the dual problems are the same. One can see the Appendix C for more details about the corresponding dual problem.

In this subsection, we want to review how to find the maximally quantum violation of the Bell-type inequalities and how to bound the quantum correlations by this numerical tool, SDP.

1.4.1 The quantum correlations for two-level quantum systems

In [56], Wehner used SDP to find the maximally quantum violation of more general CHSH-type inequalities. These inequalities are based on the measurements for the bipartite quantum systems. Each partite could have many measurement settings and each measurement setting produces two outcomes. The essential point of Wehner's work is to use the theorem proposed by Tsirelson [53, 55].

1.4.1.1 Characterizing quantum correlations by Tsirelson's theorem

In order to formulate the problem of finding the Tsirelson bound as a SDP, one needs to use Tsirelson's theorem [53, 55] to characterize the quantum correlations. Assuming there is a quantum state $|\Psi\rangle \in \mathbb{A} \bigotimes \mathbb{B}$ shared by two distant observers Alice and Bob. Alice's observable X_x ($x \in \{0, 1, 2, ..., t-1\}$) could produce the outcome $A_x \in \{-1, 1\}$. Similarly, Bob's observable Y_y ($y \in \{0, 1, 2, ..., v-1\}$) has the outcome $B_y \in \{-1, 1\}$. The correlation function $C_{x,y}$ is the expectation value for the product of their outcomes, i.e., $A_x B_y$, i.e., $C_{x,y} = \langle\Psi|X_x Y_y|\Psi\rangle$. Tsirelson's theorem states that the correlation function $C_{x,y}$ can be expressed by the inner product of two real vectors of unit norm α_x, $\beta_y \in \mathbb{R}^{t+v}$, i.e.,

$$C_{x,y} = \alpha_x \cdot \beta_y. \tag{1.63}$$

Note that, t and v are the numbers of Alice's and Bob's measurement settings, respectively. With this statement, we could rewrite CHSH-type inequalities in terms of vectors.

On the contrary, let α_x ($x \in \{0, 1, 2, ..., t-1\}$) and β_y ($y \in \{0, 1, 2, ..., v-1\}$) be two sets of operators of unit norm in $\mathbb{R}^n$. Let $d = 2^{[\frac{n}{2}]}$, and $|\Psi_{max}\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i, i\rangle$ be the maximally entangled state. There then exists the quantum observables such that

$$\alpha_x \cdot \beta_y = \langle\Psi_{max}|X_x Y_y|\Psi_{max}\rangle = C_{x,y}. \tag{1.64}$$

Note that, making measurement on the maximally entangled state with observables X_x and Y_y implies the zero of the marginal expectation value [57], i.e., $\langle\Psi_{max}|X_x \otimes I|\Psi_{max}\rangle = \langle\Psi_{max}|I \otimes Y_y|\Psi_{max}\rangle = 0$. However, it means that these sets of unit vectors allow quantum representation and could be reproduced by the quantum systems.

1.4.1.2 The maximally quantum violation for the CHSH-type inequalities

With the Tsirelson's theorem, we can cast the problem of finding the Tsirelson bound for multi-setting CHSH-type inequalities as a problem of SDP. To be specific, let us take the CHSH inequality as an example. According to the Tsirelson's theorem, one may rewrite the CHSH function $C_{0,0} + C_{0,1} + C_{1,0} - C_{1,1}$ in terms of vectors of unit norm α_x and β_y. After rewriting, the optimal problem is as follows:

$$\begin{aligned} \text{maximize} \quad & \alpha_0\beta_0 + \alpha_1\beta_0 + \alpha_1\beta_0 - \alpha_1\beta_1, && \text{(1.65a)} \\ \text{s.t.} \quad & \|\alpha_0\| = \|\alpha_1\| = \|\beta_0\| = \|\beta_1\| = 1. && \text{(1.65b)} \end{aligned}$$

One should rewrite this problem into the primal problem of SDP (1.62) by constructing the matrices X, C and A_i's from the vectors of unit norm α_x and β_y.

As shown in [56], the vectors α_x and β_y can be used to construct a positive semidefinite matrix G. Note that, this matrix G is the Gram matrix of the vectors α_0, α_1, β_0 and β_1, i.e.,

$$G = \begin{pmatrix} \alpha_0 \cdot \alpha_0 & \alpha_0 \cdot \alpha_1 & \alpha_0 \cdot \beta_0 & \alpha_0 \cdot \beta_1 \\ \alpha_1 \cdot \alpha_0 & \alpha_1 \cdot \alpha_1 & \alpha_1 \cdot \beta_0 & \alpha_1 \cdot \beta_1 \\ \beta_0 \cdot \alpha_0 & \beta_0 \cdot \alpha_1 & \beta_0 \cdot \beta_0 & \beta_0 \cdot \beta_1 \\ \beta_1 \cdot \alpha_0 & \beta_1 \cdot \alpha_1 & \beta_1 \cdot \beta_0 & \beta_1 \cdot \beta_1 \end{pmatrix}.$$

Let the matrix $B = (\alpha_0\alpha_1\beta_0\beta_1)$, where the rows of B correspond the vectors α_0, α_1, β_0 and β_1, respectively. Therefore, one can rewrite the matrix $G = B^T B$. For any $v \in \mathbb{C}^n$, $v^\dagger G v = v^\dagger B^T B v = w^\dagger w \geq 0$, where we denote the vector Bv as w. Obviously, G is indeed a positive semidefinite matrix.

If one rewrite the matrix G in terms of correlation function $C_{x,y}$, one can obtain the necessary and sufficient condition for the quantum correlation as follows:

$$G = \begin{pmatrix} 1 & \theta_1 & C_{0,0} & C_{0,1} \\ \theta_1 & 1 & C_{1,0} & C_{1,1} \\ C_{0,0} & C_{1,0} & 1 & \theta_2 \\ C_{0,1} & C_{1,1} & \theta_2 & 1 \end{pmatrix} \succeq 0, \tag{1.66}$$

where θ_1 and θ_2 correspond to the measurements performed on the same partite and therefore they are not commutative. If a set of correlations $C_{x,y}$ allows quantum representation, one can then find the valid θ_1 and θ_2 to make condition (1.66) satisfied. Note

that, this condition and the conditions proposed by Masanes [43] and Landau [42] are equivalent.

Compare our problem (1.65) to the standard form of SDP (1.62), one may define $X = G$, and then one can define a matrix C by

$$C = \frac{1}{2}\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix}.$$

Now, one may consider the constraint on these quantum correlations (or equivalently the vectors of unit norm). Since the vectors α_1, α_2, β_1 and β_2 are unit normed, the diagonal element of matrix G, called g_{ii}, must be one. The primal problem becomes

$$\begin{aligned} \text{maximize} \quad & \mathrm{Tr}(CG), && \text{(1.67a)} \\ \text{s.t.} \quad & g_{ii} = 1\ , \quad \forall i && \text{(1.67b)} \\ & G \succeq 0. && \text{(1.67c)} \end{aligned}$$

With the above procedure, one could formulate the problem of finding the Tsirelson bound as a problem of SDP. Note that, this procedure is only suitable for the two-level quantum systems. After defining the associated matrixes and vectors in the standard SDP, one could use the numerical recipes to solve the SDP such as SeDuMi [62] and CVXOPT [63]. For the well known CHSH inequality, the Tsirelson bound from the numerical evaluations is equal to the analytical solution, $2\sqrt{2}$.

1.4.2 The quantum correlations for more general quantum systems

Navascués, Pironio, and Acín showed how to check if a given set of joint probabilities could be reproduced by quantum mechanics or not [44, 45]. To do this, they found a hierarchy of conditions to test it. Each condition in the hierarchy could be formulated as a SDP. Therefor, once the given set of joint probabilities fails the test at some step of the hierarchical SDP, one may judge that these joint probabilities are non-quantum.

There are two applications for the hierarchy of conditions. The first one is to find the more general condition for entire quantum correlations. The second one is find the maximally quantum violation of the more general Bell-type inequalities. Since we have

already discussed the first application in (1.55), we will just focus the second in the following.

1.4.2.1 The constraints for bi-partite quantum probabilities

Two distant observers Alice and Bob share a bi-partite system. Alice and Bob possess local observables X and Y on their own systems, respectively. They then obtain the corresponding outcomes $a \in A$ and $b \in B$. Here A and B are used to denote the set of all possible Alice's and Bob's measurement outcomes, respectively. We use $X(a)$ and $Y(b)$ to denote the corresponding observables. The correlation between Alice and Bob is characterized by the joint probabilities $\Pr(a,b)$'s. As described in (1.30), once the joint probabilities $\Pr(a,b)$'s allow a quantum representation, they can be determined by the shared quantum state ρ and the projection operators E_a and E_b as following:

$$\Pr(a,b) = \mathrm{Tr}(E_a E_b \rho). \tag{1.68}$$

Note that (1.68) is the abbreviation of (1.30), i.e., $\Pr(A_x, B_y|x,y) = \mathrm{Tr}(E_{A_x} E_{B_y} \rho)$. Therefore, one may ask the question, can we find the corresponding quantum state and projection operators for the given joint probabilities?

Assuming that the given joint probabilities $\Pr(a,b)$'s admit a quantum representation, so that they have the corresponding quantum state and projection operators. Since the corresponding projection operators should satisfy some conditions, the given joint probabilities should also satisfy the associated conditions. Thus, one could study the quantum constraints for the joint probabilities $\Pr(a,b)$ by studying the conditions for the projection operators. The genuine quantum operators E_a and E_b shall satisfy (i) hermiticity: $E_a^\dagger = E_b$ and $E_b^\dagger = E_b$; (ii) orthogonality: $E_a E_{a'} = \delta_{aa'}$ if $X(a) = X(a')$ and $E_b E_{b'} = \delta_{b,b'}$ if $Y(b) = Y(b')$; (iii) completeness: $\Sigma_{a \in X} E_a = \mathbb{I}$ and $\Sigma_{b \in Y} E_b = \mathbb{I}$; and (iv) commutativity: $[E_a, E_b] = 0$.

Using these projection operators $\{E_a : a \in A\}$ and $\{E_b : b \in B\}$, we can construct a set of operators $O = \{O_1, O_2, ..., O_i, ...\}$. Here O_i is some linear function of products of operators in $\{E_a : a \in B\} \cup \{E_b : b \in B\}$. The set O has an associated matrix Γ given by

$$\Gamma_{ij} = \mathrm{Tr}(O_i^\dagger O_j \rho). \tag{1.69}$$

By construction, Γ is non-negative definite, i.e.,

$$\Gamma \succeq 0. \tag{1.70}$$

This can be easily proved as follows. For any vector $v \in \mathbb{C}^n$ (assuming Γ is a n by n matrix), one can have

$$v^\dagger \Gamma v = \Sigma_{s,t} v_s^* \mathrm{Tr}(O_s^\dagger O_t \rho) v_t = \mathrm{Tr}(V^\dagger V \rho) \geq 0. \tag{1.71}$$

Due to the definition (1.69) for the matrix Γ and the relation (1.68) between the joint probabilities $\Pr(a,b)$'s and the projection operators, one may find the joint probabilities $\Pr(a,b)$'s are then encoded in the matrix Γ. Note that, the elements of the matrix Γ should satisfy the following conditions:

$$\Sigma_{i,j} c_{i,j} \Gamma_{i,j} = 0 \quad \text{if} \quad \Sigma_{i,j} c_{i,j} O_i^\dagger O_j = 0. \tag{1.72}$$

$$\Sigma_{i,j} c_{i,j} \Gamma_{i,j} = \Sigma_{a,b} d_{a,b} \Pr(a,b) \quad \text{if} \quad \Sigma_{i,j} c_{i,j} O_i^\dagger O_j = \Sigma_{a,b} d_{a,b} E_a E_b. \tag{1.73}$$

These constraints reflect the conditions on the projection operators. Let us consider the simplest case. Both two distant partites Alice and Bob have two observables and each observables have two kinds of outcomes. We denote the projection operators as E_i ($i \in \{1,2,3,...,8\}$). E_1 and E_2 correspond to Alice's first observable; E_3 and E_4 correspond to Alice's second observable. Similarly, E_5, E_6, E_7 and E_8 correspond to Bob's first and second observables, respectively. Let the operators $O_i = E_i$ ($i \in \{1,2,3,...,8\}$). Therefore, the associated matrix Γ is a 8×8 matrix. When consider the properties of projection operators, the element of Γ should be constrained. These constraints are as follows:

- Consider the aforementioned property (ii) of the operators E_i's, the product of two projection operators must be zero when these two projection operators belong to same observable, i.e., $E_1 E_2 = E_3 E_4 = E_5 E_6 = E_7 E_8 = 0$, therefore $\Gamma_{1,2} = \Gamma_{3,4} = \Gamma_{5,6} = \Gamma_{7,8} = 0$.

- Because $O_i O_j = E_i E_j$, when $i \in \{1,2,3,4\}$ and $j \in \{5,6,7,8\}$, $\Gamma_{i,j}$ should be the joint probability $\Pr(a = i, b = j-4)$.

- Consider the property (ii) $E_i E_i = E_i$ ($i \in \{1,2,...,8\}$), the diagonal elements $\Gamma_{i,i} = \Pr(a = i)$ for $i \in \{1,2,3,4\}$ and $\Gamma_{i,i} = \Pr(b = i-4)$ for $i \in \{5,6,7,8\}$.

- Consider the property (iii), $\sum_{k=1}^{2} E_{k+2g} E_j = E_j$ ($g \in \{0,1,2,3\}$). Therefore, $\sum_{k=1}^{2} O_{k+2g} O_j = O_j$ and thus $\sum_{k=1}^{2} \Gamma_{k+2g,j} = \Pr(a = j)$ for $j \in \{1,2,3,4\}$. Otherwise, $\sum_{k=1}^{2} \Gamma_{k+2g,j} = \Pr(b = j-4)$.

Note that these equations also hold when permuting the operators as $\mathrm{Tr}(E_a E_b \rho) = \mathrm{Tr}(E_b E_a \rho)$. Moreover, we can make the matrix Γ to be real and symmetric by redefining it as $\Gamma = (\Gamma^* + \Gamma)/2$.

With these conditions, one can determinate the corresponding parameter $c_{i,j}$ and $d_{a,b}$ in (1.72) and (1.73) to list all the conditions. Therefore, once the given set of the quantum probabilities $\Pr(a, b)$'s allows the quantum representation (1.68), for any set of operators O the associated Γ matrix should satisfy the quantum constraints (1.70), (1.72) and (1.73).

1.4.2.2 Bounding the quantum correlations with the hierarchal semidefinite programming

We have known that for any set of operator O, the associated matrix Γ should satisfy the quantum constraints (1.70), (1.72) and (1.73). Thus, the existence of such a matrix is the necessary condition for the quantum probability $\Pr(a, b)$. Since the conditions for the matrix Γ is related to the constraints on some positive semidefinite matrix, thus one may use SDP to test the existence of the matrix Γ satisfying all the quantum constraints (1.70), (1.72) and (1.73). Note that, if the joint probability $\Pr(a, b)$ does not allow a quantum representation, the associated SDP problem for some operators O does not have the solution.

Since different operator sets O's yield different quantum constrains (1.70), (1.72) and (1.73), there are too many quantum constraints needed for thhe test. In [44], the authors found that these constraints have a hierarchical structure such that one may use a systematic way to test them. In the first step of this hierarchical SDP, one may take single projection operators $\{E_\mu\}$ to form the operator set O, where $\{E_\mu\} = \{E_a\} \cup \{E_b\}$. In the second step, one can take the product of projection operators $\{E_\mu E_\nu\}$ to form O besides the ones in the first step, where $\{E_\mu\} = \{E_\nu\} = \{E_a\} \cup \{E_b\}$. One can then take $\{E_\mu E_\nu E_\kappa\}$ to form O in the third step besides the ones in the previous steps, and so on. Therefore, the infinite sequence of the operator sets leads to a hierarchical SDP.

When joint probabilities $\Pr(a, b)$ satisfy the n-th step of the hierarchical SDP, we denote the collection of these joint probabilities $\Pr(a, b)$ as Q_n. Since we know that the associated constraints are stronger than ones in the previous steps of the hierarchical sequence, the collection Q_n will become smaller for the higher n. That is, the non-quantum correlations will definitely fail the test at some step in the hierarchical SDP.

1.4.2.3 The maximally quantum violation for the general Bell-type inequality

Any Bell-type inequality can be written as the linear combination of the joint probabilities, thus the hierarchical SDP can be used to approach the maximally quantum violation of the Bell-type inequality. Since the quantum constraint is stronger in the hierarchical SDP and the collection of Q_n will become smaller while n is increasing. We then know that the bound of Bell-type inequality becomes tighter for larger n and it may converge to the quantum bound for large enough n.

One can try to find the quantum violation of the d-level Bell-type inequality (the CGLMP inequality [8]). In [58], the best-known lower bound of the maximally quantum violation for the CGLMP inequality ($d \leq 8$) has been obtained. Using the hierarchal SDP to approach the bound of quantum violation, one may find that the best-known bound of quantum violation in [58] is equal to the bound given by the second step of the hierarchal SDP. Therefore, one may obtain the maximally quantum bound in the finite step of SDP.

In Chapter 3, we will write down the associated SDP for finding the maximally quantum violation for the general Bell-type inequality more precisely.

Chapter 2

Information Causality and Noisy Computations

2.1 Introduction

As a physical theory, quantum mechanics has been extremely successful in describing the microscopic physics. Nevertheless, its current framework is incapable of explaining the nature of quantum entanglement. Attempts to remedy this situation have been made by reconstructing quantum mechanics in terms of physical principles. These physical principles should be able to yield or constrain the non-local correlation implied by the quantum entanglement. One such candidate is the principle of space-time causality. This principle will constrain the possible non-local correlation such that any physical theory must be no-signaling [21], i.e., signal cannot be send in the way of violating causality. However, a broad class of no-signaling theories other than quantum mechanics exist. Certain features, usually thought of as specifically quantum, are common for many of these theories [11, 12]. Clearly, no-signaling is insufficient as a principle to single out quantum mechanics.

Some of these theories are allowed to have more non-local correlation than quantum mechanics [11, 12, 15, 17, 18]. Specifically, the non-local correlation in these theories can violate Bell-type inequalities by more than Tsirelson's bound [21, 53]. From this perspective, we should search for a physical principle as follows. The principle can single out Tsirelson's bound as a limitation on the extent of the allowed correlation for a physical theory. With the advent of quantum information science, some principles of information theoretic flavor have been proposed. These proposed candidates set the constraints on the physically realizable correlations. In this Letter, we focus on a promising candidate — the information causality. Information causality states that, in a bipartite code protocol prepared with any physically local or non-local resources, the accessible information gain cannot exceed the amount of classical communication. In [29, 31] information causality is demonstrated by a generic task similar to random access codes (RAC) and oblivious

transfer. In this task, a database of k bits is prepared: $\vec{a} := (a_0, a_1, \cdots, a_{k-1})$, where each a_i is a random variable, which is only known by the first party, Alice. A second, distant party, Bob, is given a random variable $b \in (0, \cdots, k-1)$ along with a bit α send by Alice. With the bit α and the pre-shared correlation with Alice, Bob's task is to optimally guess the bit a_b. Then, according to information causality the quantity I has an upper bound

$$I = \sum_{i=0}^{k-1} I(a_i; \beta | b = i) \leq 1 \ . \tag{2.1}$$

Here $I(a_i; \beta | b = i)$ is the Shannon mutual information between a_i and Bob's guessing bit β under the condition $b = i$. Classically, I can reach 1 once $\alpha = a_i$ and $I(a_i; a_j) = \delta_{ij}$ (i.e., the Kronecker delta).

To perform the RAC task, Alice and Bob can use (earlier prepared and distributed) correlations among either classical or quantum systems. These no-signaling correlation resources can be simulated by the no-signaling box (NS-box). The NS-box correlates the inputs and outputs of Alice and Bob in an imperfect way subjected to the probabilistic noise. The noise of NS-box is intrinsically inherited from the underlying physical theory such as quantum mechanics. The quantity I in (2.1) is unavoidably affected by the intrinsic noise of NS-box. In this framework, the signal decay theorem in [50, 51] for a noisy circuit is exploited to yield a tight bound for $I(a_b; \beta | b)$ in terms of noise of NS-box. According to information causality, the tight bound should also obey the upper bound in (2.1). By expressing the tight bound in terms of correlation functions between Alice's and Bob's measurement outcomes, this then yields our main result — a broad class of multi-setting Tsirelson-type inequalities. As a result, we can then subject the physical principle of information causality to scrutiny by experimentally verifying or falsifying the generalized Tsirelson's bounds.

Without classical communication, the RAC can be regarded as nonlocal computation. Therein, distant Alice and Bob compute a general Boolean function without knowing the other's input. Here, NS-box can be regarded as a noisy gate for non-local computation [51]. Noise of the gate is closely related to the reliability of non-local computation. The computational noise of the gate is related to the intrinsic reliability of the physically realized NS-box. In this aspect, we can tackle a fundamental question on noisy computation with its nonlocal version. As raised by von Neumann [47], this question is originally stated as follows. Could physical circuits of finite size perform the reliable noisy non-local

computation of any Boolean function? Based on constraint by the information causality for any physical circuit, we will see that the answer is negative in non-local computation.

2.2 Tsirelson-type inequalities from the information causality

We start by reformulating the NS-box as a noisy distributed gate for nonlocal computation. The NS-box is initially distributed between two distant parties, Alice and Bob. Locally, Alice and Bob input bit strings $\vec{x}$ and $\vec{y}$, respectively, into half of the box, which then outputs bits $A_{\vec{x}}$ and $B_{\vec{y}}$, respectively. The lengths of the bit strings can be chosen by design. Our NS-box is further characterized by the conditional joint probabilities $\Pr\left[A_{\vec{x}} + B_{\vec{y}} = f(\vec{x},\vec{y}) \,|\vec{x},\vec{y}\right]$. Therein, $f(\vec{x},\vec{y})$ is the task function. Notably, if the NS-box is physically realizable, these joint probabilities must fulfill the no-signaling conditions.

In the RAC protocol, the chosen task function $f(\vec{x},\vec{y})$ depends on how we encode Alice's database $\vec{a}$ and Bob's given random variable b into $\vec{x}$ and $\vec{y}$, respectively. From now on, we will implicitly use the following protocol. Firstly, Alice encodes her database $\vec{a}$ into the $(k-1)$-bit string $\vec{x} := (x_1, \cdots, x_{k-1})$ by $x_i = a_0 + a_i$. Alice's half of NS-box then produces an outcome $A_{\vec{x}}$. At the same time, Bob encodes the given b in to $(k-1)$-bit string $\vec{y} := (y_1, \cdots, y_{k-1})$ by $y_i = \delta_{b,i}$ for $b \neq 0$, and $\vec{y} = \vec{0}$ for $b = 0$. Bob's half of NS-box then produces an outcome $B_{\vec{y}}$. Secondly, Alice sends Bob a bit $\alpha = a_0 + A_{\vec{x}}$. The optimal strategy for Bob's task is to output a guess bit $\beta = \alpha + B_{\vec{y}}$. As a result, Bob can decode Alice's bit a_b successfully whenever $A_{\vec{x}} + B_{\vec{y}} = \vec{x} \cdot \vec{y}$ (modulo 2) is true. Most of the calculations in this Letter are modulo-2 defined.

In quantum mechanics, Alice's and Bob's outcomes can be produced by performing the corresponding measurement of 2^{k-1} and k settings, respectively. For the above protocol, the success probability of Bob's task in guessing Alice's bit a_b is related to the one for noisy computation as follows

$$\Pr[\beta = a_b|b\,] = \frac{1}{N_{\vec{x}}} \sum_{\{\vec{x}\}} \Pr\left[A_{\vec{x}} + B_{\vec{y}} = f(\vec{x},\vec{y})|\vec{x},\vec{y}\,\right], \tag{2.2}$$

where $N_{\vec{x}}$ is the cardinality of the input space spanned by the encoding $\{\vec{x}\}$. By defining the correlation functions between Alice's and Bob's measurement outcomes as $C_{\vec{x},\vec{y}} := \sum_{A_{\vec{x}}=0,1} \sum_{B_{\vec{y}}=0,1} (-1)^{A_{\vec{x}}+B_{\vec{y}}} \Pr\left[A_{\vec{x}}, B_{\vec{y}} \,|\vec{x},\vec{y}\,\right]$, we find

$$\xi_{\vec{y}} = \frac{1}{N_{\vec{x}}} \sum_{\{\vec{x}\}} (-1)^{f(\vec{x},\vec{y})} C_{\vec{x},\vec{y}}. \tag{2.3}$$

where the coding noise parameter is defined as $\xi_{\vec{y}} := 2\Pr[\beta = a_b \,|b\,] - 1$. The sub-index $\vec{y}$ of $\xi_{\vec{y}}$ is understood to be equivalent to Bob's given parameter b via encoding.

One of the main results of this paper is a broad class of Tsirelson's bound implied by information causality, i.e.,

$$|\sum_{\{\vec{y}\}} \xi_{\vec{y}}\,| = \frac{1}{N_{\vec{x}}}|\sum_{\{\vec{x}\},\{\vec{y}\}} (-1)^{f(\vec{x},\vec{y})} C_{\vec{x},\vec{y}}\,| \leq \sqrt{k}\,. \tag{2.4}$$

For $k = 2$, it is easy to check that (2.4) is the Tsirelson's bound $|C_{0,0}+C_{0,1}+C_{1,0}-C_{1,1}| \leq 2\sqrt{2}$ [53]. For the case of $k > 2$ with $f(\vec{x},\vec{y}) = \vec{x}\cdot\vec{y}$, we have verified (2.4) to be the Tsirelson's bound in quantum mechanics by using the semidefinite programing [61]. Please see appendix D for more detailed discussions.

Indeed, later we will see that information causality will render (2.4). This implies that information causality can be tested by experimental verification or refutation via the measurement of the correlation functions of a quantum system.

In order to arrive the Tsirelson's bound (2.4) from the information causality constraint (2.1), we need to relate $I(a_b, \beta|b)$ to $\xi_{\vec{y}}$. It turns out that this can be done by using the following signal decay theorem on the signal propagation [50, 51].

Theorem 1: Let X, Y and Z be Boolean random variables. Consider a cascade of two communication channels: $X \hookrightarrow Y \hookrightarrow Z$. X and Y are the input and the output of the first channel, respectively. Let Y in turn be the input of a cascading binary symmetric channel C_ϵ with a noise parameter ϵ, i.e.,

$$C_\epsilon = \begin{pmatrix} \frac{1}{2}(1+\epsilon) & \frac{1}{2}(1-\epsilon) \\ \frac{1}{2}(1-\epsilon) & \frac{1}{2}(1+\epsilon) \end{pmatrix}.$$

Let Z be the output of C_ϵ, (i.e., $Z = \overline{Y}$ with the bit-flipping probability $\frac{1}{2}(1-\epsilon)$)

$$\frac{I(X;Z)}{I(X;Y)} \leq \epsilon^2. \tag{2.5}$$

A special case arises if the first channel is noiseless or trivial, i.e., $I(X;Y=X) = 1$ such that $I(X;Z) \leq \epsilon^2$. Note also that regardless of the properties of the second channel, there is a data processing inequality $I(X;Z) \leq I(X;Y)$.

We apply this theorem to our RAC protocol as follows. Because Alice's database $a_0, a_1, \cdots, a_{k-1}$ are random variables and independent of each other, so that all the a_j's with $j \neq i$ can be fixed without disturbing $I(a_i\,;\beta|b)$. Let $X = a_i$, $Y = a_0 + f(\vec{x},\vec{y})$, and

$Z = \beta$. Here Y is Bob's ideal answer and hence $I(X;Y) = 1$. The coding noise ϵ for our protocol is $\xi_{\vec{y}}$, then according to the Theorem 1, we have

$$I(a_i\,;\beta|b=i) \leq \xi_{\vec{y}}^2. \tag{2.6}$$

Therefore, the information causality in Eq. (2.1) yields

$$I \leq \sum_{\{\vec{y}\}} \xi_{\vec{y}}^2 \leq 1. \tag{2.7}$$

In [29, 31], similar inequalities are derived to avoid the divergence of I, which justifies the information causality. However, such trouble does not exist in our reformulation because of the tight bound of Theorem 1. With the help of (2.3) the second inequality in (2.7) becomes a quadratic Tsirelson-type inequality for the correlation function $C_{\vec{x},\vec{y}}$. Moreover, using the Cauchy-Schwarz inequality, we can obtain $|\sum_{\{\vec{y}\}} \xi_{\vec{y}}| \leq \sqrt{k}$, which results in the linear Tsirelson inequality of Eq. (2.4).

2.3 Noisy nonlocal computation

In the previous discussion we have considered the information causality using a single nonlocal NS-box. Instead, we can treat the NS-box as a non-local gate for performing the nonlocal computation, i.e., computing the function $f(\vec{x},\vec{y})$ [51]. Unlike using the same gate for the RAC, no classical communication between Alice and Bob is required to perform the nonlocal computation. In details, Alice's and Bob's local outputs are $A_{\vec{x}}$ and $B_{\vec{y}}$, respectively. The computation is successful if $A_{\vec{x}} + B_{\vec{y}} = f(\vec{x},\vec{y})$. The computational noise parameter is defined as

$$\epsilon_{\vec{x},\vec{y}} := 2\Pr[A_{\vec{x}} + B_{\vec{y}} = f(\vec{x},\vec{y})|\vec{x},\vec{y}] - 1. \tag{2.8}$$

From (2.8) and (2.3) the computational noise of the gate is related to its coding noise by

$$\xi_{\vec{y}} = \frac{1}{N_{\vec{x}}} \sum_{\{\vec{x}\}} \epsilon_{\vec{x},\vec{y}}\,. \tag{2.9}$$

Basically, computational errors inherently come from the gate noise. Information causality constraints the noisy extent of the NS-box as a gate. From this perspective, information causality is deeply connected with nonlocal computation.

Furthermore, we can combine the NS-box gates to form a more complicated circuit without worrying about the coding protocol. Then the total task function for the whole

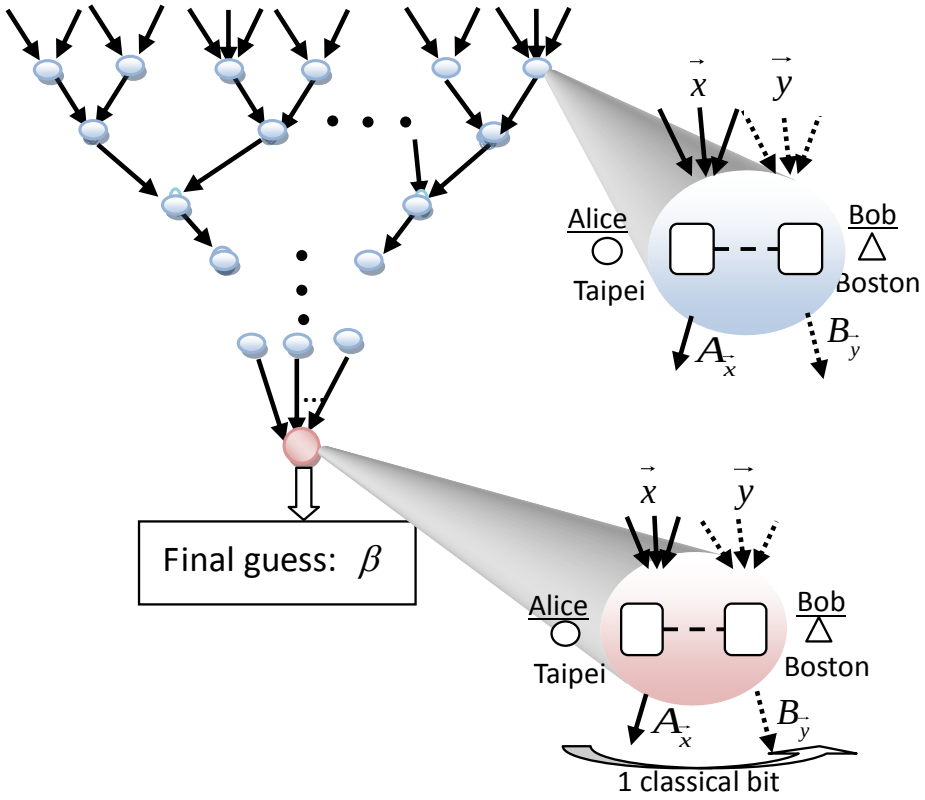

Figure 2.1: RAC protocol for a (n, k, l)-circuit. Each vertex of the circuit corresponds to a NS-box, with its details shown in the big ellipses.

circuit will be a complicated function, i.e., a composite of task functions of all NS-boxes. We can then try to answer the following fundamental question: *could a noiseless (nonlocal) computation be simulated using a noisy nonlocal physical resource?*

Specifically, we consider the so-called (n, k, l)-circuit, G, formed by cascading layers of noisy gates into a circuit in the form of a directed, acyclic tree (see Fig 1). On the top of G, there are n inputs to the NS-boxes — the leaves; at the bottom there is only one NS-box — the root. The longest path from the leaves to the root is called the depth of the circuit, denoted by l. The maximum input number of a gate in G is k. Note that, in [29] G comprises $k = 2$ gates and is exploited to compress n bits of $\vec{x}$ into one bit $A_{\vec{x}}$. However, there is no restriction on the task function for each NS-box, as long as the final circuit is a consistent acyclic tree diagram.

We then use the circuit G to perform the following nonlocal computation. Alice's n-

bit database $\vec{a} := (a_0, a_1, \cdots, a_{n-1})$ is given to the leaves of G, and a conditional input $b \in \{0, 1, \cdots, n-1\}$ is given to the distant Bob. The previous encoding $\vec{a} \to \overrightarrow{x}$ and $b \to \vec{y}$ for the RAC protocol, is also exploited here. Alice's output is properly encoded and then fed into the NS-box at the next layer, again with Bob's conditional input. The same procedure is performed recursively until reaching the root, with its output as the answer to the total task function at the root.

Alternatively, Bob's decoding gates can be thought to be noise free, and the computational noise is only due to Alice's encoding gates, and vice versa. This makes it easier to understand the above procedure of noisy computation. Now we can consider the information flow of G.

Theorem 2: For a noisy, local circuit G with an arbitrary depth, the root outputs at most one-bit information.

Note that the circuit G can perform the RAC if the appropriate protocol is given at each layer and 1-bit communication is allowed for the whole process. Then, the above theorem implies that information causality holds true for the circuit G.

To prove the theorem, we will show that the mutual information between the leaves and the root of G is bounded by one. This can be done by mathematical induction as follows. We begin with a circuit of depth one, which is nothing but a single NS-box; information causality ensures the bound. We then assume that the bound holds true for a circuit of depth ℓ. According to information causality and sub-additivity, the mutual information $I_\ell^{(m)}$ between the leaves and the root obeys $I_\ell^{(m)} \leq \sum_{i_m} I(X_{i_m}; R_m) = \sum_{i_m} I(X_{i_m}; R_m|\text{Bob's knowledge}) \leq 1$, where the index m labels a collection of circuits of depth ℓ with root R_m, and the index i_m labels the inputs of the m-th circuit. Now, we construct a circuit of depth $\ell + 1$ by connecting all roots R_m's to a single NS-box whose output is R. Then, the mutual information $I_{\ell+1}$ between leaves and root R of the final circuit should obey the subadditivity, i.e., $I_{\ell+1} \leq \sum_m \sum_{i_m} I(X_{i_m}; R)$. From Theorem 1, we have $I(X_{i_m}; R) \leq \xi_m^2 I(X_{i_m}; R_m)$ because we have a cascade of two channels: $X_{i_m} \hookrightarrow R_m \hookrightarrow R$ where the second channel is a binary symmetric one with the noise ξ_m. Using this result, we have $I_{\ell+1} \leq \sum_m \xi_m^2 \sum_{i_m} I(X_{i_m}; R_m) \leq \sum_m \xi_m^2 \leq 1$. Q.E.D.

Here, we have only considered the case in which the computational noise is isotropic to $\vec{x}$, denoted by $\epsilon_{\vec{y}}$. From (2.9) we have $\epsilon_{\vec{y}} = \xi_{\vec{y}}$ and the information causality requires $\sum_{\{\vec{y}\}} \epsilon_{\vec{y}}^2 \leq 1$. We would like to know whether the reliable computation is also constrained

by the information causality or not. To check this, we invoke the main Evans-Schulman theorem on the conditions for reliable noisy computation as follows [50, 51].

Evans-Schulman Theorem: A circuit of complete k-ary tree with depth l (i.e., $n = k^l$) can perform δ-reliable noisy computation only

- (i) if $\sum_{\{\vec{y}\}} \epsilon_{\vec{y}}^2 > 1$ then $\ell \geq \log(n\Delta)/\log(\sum_{\{\vec{y}\}} \epsilon_{\vec{y}}^2)$,
- (ii) if $\sum_{\{\vec{y}\}} \epsilon_{\vec{y}}^2 \leq 1$ then $n \leq 1/\Delta$,

where $\Delta := 1 + \delta \log \delta + (1-\delta)\log(1-\delta)$. The computation is called δ-reliable if the root outputs correctly with a probability $1-\delta$ (with $\delta < 1/2$). This theorem provides stricter conditions than the original proposal by Von Neumann [47, 48].

By definition, smaller $\epsilon_{\vec{y}}$ means larger noise, and the condition (ii) is for the cases with larger noise such that only functions with a smaller number of inputs can be reliably computed. Immediately, we see that information causality implies a large computational noise for the RAC circuit such that only condition (ii) for reliable noisy computation can possibly be fulfilled. As a result, Alice's output asymptotically becomes random because $\Delta \to 0$ and hence $\delta \to \frac{1}{2}$ as $n \to \infty$. In summary, this implies that *information causality prevents any physically realizable (n, k, l)-circuit from achieving reliable computations of excessively complicated functions, i.e., with either too many inputs or lengthy steps needed.*

The above result applies only when classical communication between Alice and Bob is disallowed. Under such circumstances, the noise of the gate is intrinsically constrained by the underlying physical theory. Otherwise, the classical communication can be exploited to improve the reliability of the gates so that the no-go result could be lifted.

2.4 Summary

In this chapter, we show how information causality leads to Tsirelson bounds in a much easier way. A series of new Tsirelson bounds are then derived. Deep ramifications concerning non-local quantum computation are also found and discussed.

Chapter 3

Testing Information Causality for General Quantum Communication Protocols

3.1 Introduction

The advantage of quantum information over the classical one in the computing and communication has been well exploited in the past decades. The abstract form of this advantage in the communication process is termed as communication complexity [25]. Despite of the seemingly non-local feature of quantum entanglement employed in the quantum communication process, its communication complexity is bounded and not maximal [1]. Recently, the bound is formulated for general physical theories and is termed as information causality [29], which states that the maximal mutual information shared between the sender and receiver in a communication protocol with the resources based on physical theories cannot exceed the amount of classical communication. The criterion of information causality selects a subset of non-signaling theories, including quantum mechanics.

If the information causality holds for all the realistic communication processes, it can be erected as a new physical principle to formulate the fundamental theories from the information theoretical point of view. Moreover, one may wonder if the quantum mechanics is equivalent to the theories saturating the non-local bound given by information causality. Or, the quantum mechanics cannot saturate the information causality. However, most of the tests on the above questions are performed only for two-level quantum communication protocols. Especially, in [32] it was shown that the non-local bounds from information causality exactly coincide with the generalized Tsirelson inequalities for a particular set of two-level quantum protocols (of multi-settings). Related tests based on information causality and macroscopic locality[2] [34] was done in [36].

[1] In [24], by sharing a PR-box [21], any distributed decision problem can be solved with perfect success with only one bit communication. This means the communication complexity is trivial. Therefore, the communication complexity is related to the non-locality. Thus, in this paper we use the terms "the communication complexity" and "the non-locality" interchangeably when we discuss the non-local correlations.

[2] Macroscopic locality states that any physical theory should recover classical physics in the continuum limit, i.e., the

It is then interesting to test the information causality over quantum mechanics for more general quantum communication protocols, e.g. multi-level ones [3]. In this work, we will make efforts along this direction, namely, we will try to find the maximal bound of the mutual information shared between sender and receiver for multi-level quantum communication protocols, and compare with the bound from information causality.

In our communication protocol, Alice has a database of k elements, denoted by the vector $\vec{a} = (a_0, a_1, , , a_{k-1})$. Each element a_i is a d-level random variable and is only known by Alice. A second distant party, Bob is given a random variable $b \in 0, 1, 2, , , k-1$. The value of b is used to instruct Bob to optimally guess the d-level bit (d-bit) a_b after receiving a d-bit α sent from Alice via the pre-shared correlation between Alice and Bob. In this context, the information causality can be formulated as follows:

$$I = \sum_{i=0}^{k-1} I(a_i; \beta | b = i) \leq \log_2 d \, . \tag{3.1}$$

where $I(a_i; \beta | b = i)$ is Shannon's mutual information between a_i and Bob's guess d-bit β under the condition $b = i$. The bound is the amount of the classical communication encoded in α.

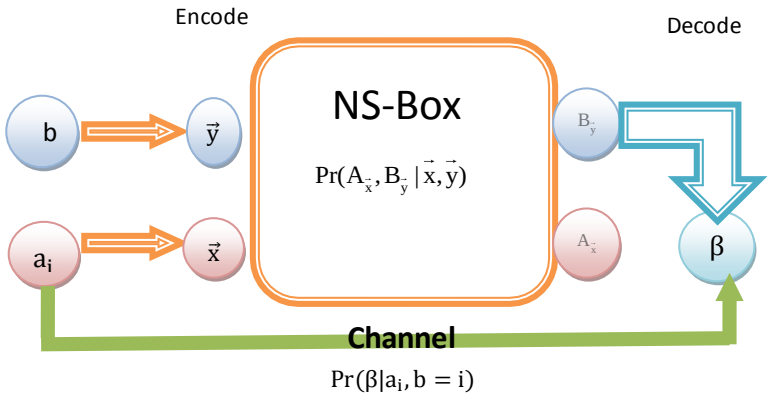

Figure 3.1: The NS-box and the channel

To be specific, we use the random access code (RAC) to encode Alice's data base $\vec{a}$

probability distributions of large number particles should satisfy the Bell's inequalities.

[3] In this paper, using the non-uniform $\Pr(a_i)$ or the more general communication channels such as the asymmetric and anisotropic channels is belonged to the more general communication protocols.

into $\vec{x} := (x_1, \cdots, x_{k-1})$ with $x_i = a_i - a_0$, and Bob's input b into $\vec{y} := (y_1, \cdots, y_{k-1})$ with $y_i = \delta_{b,i}$ for $b \neq 0$ and $\vec{y} = 0$ for $b = 0$. Besides, the communication protocol is also supplemented by the pre-shared correlation, called the non-signaling box (NS-box). The above $\vec{x}$ and $\vec{y}$ are the inputs of the NS-box, and their corresponding outcomes are denoted by $A_{\vec{x}}$ and $B_{\vec{y}}$, respectively. Therefore, the NS-box and thus the communication protocol is characterized by the conditional joint probabilities $\Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x}, \vec{y})$ satisfying the following non-signaling conditions:

$$\sum_{B_{\vec{y}}} \Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x}, \vec{y}) = \Pr(A_{\vec{x}}|\vec{x}) \quad \text{and} \quad \sum_{A_{\vec{x}}} \Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x}, \vec{y}) = \Pr(B_{\vec{y}}|\vec{y}), \qquad \forall \vec{x}, \vec{y}. \tag{3.2}$$

This implies that superluminal signaling is impossible.

Here, we shall mention that the mutual information in (3.1) is referred to the channel characterized by the conditional probability $\Pr(\beta|a_i, b = i)$ with RAC decoding $\beta = \alpha + B_{\vec{y}}$, which relate the outcome β of the channel to its input a_i and b. Note that the channel probability $\Pr(\beta|a_i, b = i)$ cannot be completely determined only by the conditional joint probability $\Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x}, \vec{y})$ but also by the input marginal probability $\Pr(a_i)$. In a sense, the flow of the NS-box is perpendicular to the flow of the channel, this is schematically shown in Fig 3.1. We will see that the difference of these two flows will be relevant in how to maximize the mutual information I in (3.1).

Naively, one will formulate the whole problem as maximizing the mutual information I of the protocol characterized by Alice's input marginal probabilities $\Pr(a_i)$ and the channel derived from the quantum correlations. To proceed, we have to make sure the whole problem can be formulated as a convex optimization problem [61, 64] so that some numerical recipes such as [63] can be utilized for maximizing I. However, we will show that this is not a convex problem if we would like to maximize I by varying over the input marginal probability $\Pr(a_i)$ and the joint probability $\Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x}, \vec{y})$ of NS-box [4]. Therefore, in this way the numerical recipe [63] cannot be applied to finding the mutual information bound for information causality. Observe that our problem here is different from the usual way of determining the channel capacity. The usual way of finding the channel capacity for a given channel, i.e., fixing $\Pr(\beta|a_i, b = i)$, is to maximize the mutual information I over the input marginal probability $\Pr(a_i)$. This is the convex optimization

[4] In fact, the mutual information I defined in (3.1) depends on $\Pr(a_i)$ and only $\Pr(B_{\vec{y}} - A_{\vec{x}}|\vec{x}, \vec{y})$ not the full $\Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x}, \vec{y})$. Thus, later when referring to the joint probability of NS-box, it will in fact mean $\Pr(B_{\vec{y}} - A_{\vec{x}}|\vec{x}, \vec{y})$ though we may use the expression $\Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x}, \vec{y})$.

problem and can be solved numerically by the recipe such as [63].

To by-pass this no-go situation, a naive way is to maximize I over $\Pr(a_i)$ and $\Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x},\vec{y})$ by brutal force numerically without relying on convex optimization. The viability is, however, restricted by the computational power and cannot work for complicated protocols, like the ones with multi-level and multi-setting. Despite that, we will try some simpler cases and verify that the bound by information causality may not be the same as the generalized Tsireslon bound associated with the quantum non-locality.

The other way of by-passing the no-go is to consider some special cases such that the generalized Tsireslon bound agrees with the bound for information causality. With the suitable object for quantum communication complexity, e.g. the CHSH function in Bell inequality, maximizing the object in determining the optimal quantum channel is a convex optimization problem. Thus, we need to find the special conditions such that the mutual information I is monotonically increasing with the object for quantum communication complexity. Therefore, we can then by-pass the no-go and find the bound on information causality for quantum channels. For such cases, maximizing mutual information over quantum channel is the same as finding the generalized Tsireslon bound. This is also the way adopted in this work.

Before the proof, one important issue is to find the appropriate object of communication complexity which is called Bell-type function in [56, 44, 45]. For the two-level protocols, the CHSH correlation function gives the natural object for the communication complexity. Moreover the Tsirelson theorem helps to yield the Tsirelson inequalities to constraint $\Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x},\vec{y})$ of the NS-box. However, for the multi-level protocols, there is neither analogy CHSH correlation as a natural object for the communication complexity, nor the Tsirelson-type inequalities to yield suitable Bell-type inequalities for quantum constraint. Despite that, one found an alternative way to derive the Tsirelson inequalities for two-level protocols in [32], and it will be generalized for the multi-level protocols in this paper. This is based on the signal decay theorem [50, 51]. Consider a cascade of two communication channels: $X \hookrightarrow Y \hookrightarrow Z$ with the second channel being binary and symmetric with a noise parameter ξ. Then,

$$\frac{I(X;Z)}{I(X;Y)} \leq \xi^2. \tag{3.3}$$

Applying this to our protocol of RAC and NS-box, one can arrive

$$I(a_i\,;\beta|b=i)\leq\xi_i^2 \tag{3.4}$$

where ξ_i is the coding noise parameter and can be expressed in terms of the joint probability $\Pr(A_{\vec{x}},B_{\vec{y}}|\vec{x},\vec{y})$ and the input marginal probability $\Pr(a_i)$. Assuming independent and identically distributed(i.i.d.) for a_i's, we can then sum over i to obtain bound on I. Using the Cauchy inequality, one can linearize the above quadratic inequality to arrive

$$\sum_i \xi_i \tag{3.5}$$

as the object for communication complexity, Furthermore, assuming uniform probabilities for all Alice's database a_i's, we showed in [32] that the information causality can be formulated as $|\sum_i \xi_i|\leq\sqrt{k}$, and is exactly equivalent to the Tsirelson-type inequalities for two-level protocols. Recently, similar considerations can also be found in [40, 52].

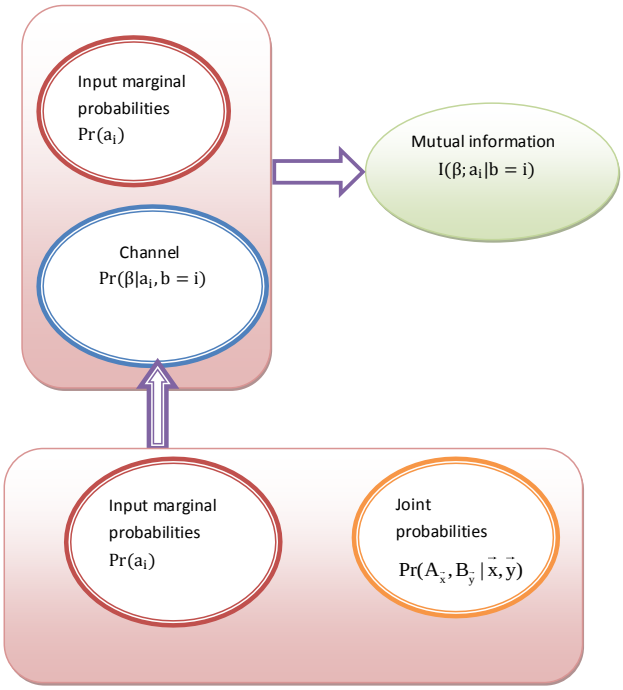

Figure 3.2: Scheme for maximizing the mutual information I over quantum channels.

Instead of deriving as a mathematical theorem, the above derivation of (3.5) for Tsirelson-type inequalities is based on physical content of the communication protocols,

and can be generalized to the multi-level cases. This is one of the main tasks of this paper. To by-pass the no-go as described, we only consider the i.i.d. inputs $\{a_i\}$ with uniform $\Pr(a_i)$, in the symmetric and isotropic channels. The symmetric channel is defined as usual, and the isotropic channel means that the noise parameter for NS-box is uniform, i.e., ξ_i is independent of i. Based on these assumptions, we can show that I and (3.5) are monotonically related so that maximizing I is equivalent to maximizing (3.5). One can then use the standard semidefinite programming (SDP) [61, 64] algorithm which also is a linear convex optimization programming to build up the quantum channel. Based on these quantum channels with uniform $\Pr(a_i)$ one can further calculate the mutual information I. This is schematically shown in Fig 3.2. In this way, we can test the information causality over the the more generic multi-level quantum communication protocols.

The above discussions are for multi-level symmetric channels. Although we can transfer the maximizing mutual information I to the convex optimization problem, it needs many assumptions. If we would like to consider the asymmetric and anisotropic channels with possibly the non-uniform $\Pr(a_i)$, how can we do? Instead of using convex optimization, we are forced to use brutal force method which we mentioned before. That is, we have to pick up all the sets of quantum correlation $\Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x}, \vec{y})$ and input marginal probability $\Pr(a_i)$ and then evaluate the corresponding mutual information I. We compare all of them to find the maximal one. In this way, we find that the boundaries for the information causality and the quantum non-locality may not agree. That means, for general communication channels, maximizing mutual information I over quantum channels is not equivalent to finding the generalized Tsireslon bound. However, due to the demanding computational resources in numerical optimization, we will only consider the two-level and two-setting case.

The paper is organized as follows. In the next section we will define our communication protocols based on RAC and NS-box and then derive the objects for the communication complexity of symmetric quantum channels with i.i.d. and uniform input marginal probabilities. In section 3.3, we will show that maximizing the mutual information I over joint probabilities $\Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x}, \vec{y})$ and input marginal probabilities $\Pr(a_i)$ is not a convex optimization problem and also prove that (3.5) and the mutual information I are monotonically related. In section 3.4, we will briefly review the SDP algorithm proposed in [44, 45] and then apply it to maximize (3.5) numerically for the multi-level symmetric

channels with i.i.d. and uniform input marginal probabilities. Our numerical results for the maximal quantum non locality will be used to evaluate the corresponding mutual information I and compare to the information causality. In 3.5, we will use brutal force to maximize the mutual information I for more general communication channels. Finally, in section 3.6 we will summary our paper with some discussions.

In appendix A, we briefly review the signal decay theorem for binary channels mentioned in [50, 51] and then generate it to multi-nary channels. In appendix B, we give a proof that mutual information I is not a concave function to joint probabilities and input marginal probabilities. In appendix C, the standard primal and dual problem for SDP are defined. We can rewrite our problem as the standard form and use numerical recipes to solve it. In appendix E, we write down the quantum constraints for the first and second step of the convex optimization programming. We also estimate the number of quantum constraints and explain how to write down these constraints efficiently.

3.2 Multi-level Bell-type inequality from signal decay theorem

In the Introduction, we have briefly describe our communication protocol for two distant partite Alice and Bob: Given one encoded d-bit α by Alice and one random number b, Bob needs to optimally guess a_b in Alice's database $\vec{a} := (a_0, ..., a_{k-1})$. In this task, Bob can use the pre-shared correlations simulated by NS-box, whose inputs are the Alice's encoded d-bit-string $\vec{x}$ and Bob's $\vec{y}$ as mentioned previously. The corresponding outputs of the NS-box are $A_{\vec{x}}$ and $B_{\vec{y}}$, respectively. More specifically, the d-bit sent by Alice is $\alpha = A_{\vec{x}} - a_0$, and the pre-shared correlation is defined by the conditional probability $\Pr(B_{\vec{y}} - A_{\vec{x}} = \vec{x}\cdot\vec{y}|\vec{x},\vec{y})$ between the inputs and outputs of the NS-box. Accordingly, Bob's optimal guessing d-bit β can be chosen as $B_{\vec{y}} - \alpha$. This is because $\beta = B_{\vec{y}} - A_{\vec{x}} + a_0 = \vec{x}\cdot\vec{y} + a_0$ as long as $B_{\vec{y}} - A_{\vec{x}} = \vec{x}\cdot\vec{y}$ holds. In this case, Bob guesses a_b perfectly. Take $d = 3$ and $k = 3$ as an example for illustration: Bob's optimal guess bit is

$$\beta = \vec{x}\cdot\vec{y} + a_0 = (a_1 - a_0, a_2 - a_0)\cdot(y_0, y_1) + a_0. \tag{3.6}$$

If Bob's input $\vec{y} = (y_0, y_1) = (0, 0)$, $\beta = a_0$; if $\vec{y} = (y_0, y_1) = (1, 0)$, $\beta = a_1$; and if $\vec{y} = (y_0, y_1) = (0, 1)$, $\beta = a_2$. Bob can guess a_b perfectly.

Using the above communication protocol, Alice and Bob have d^{k-1} and k measurement settings respectively, and each of the measurement settings will give d kinds of

outcomes. However, the noise of the NS-box affects the communication complexity so that Bob can not always guess the d-bit a_b correctly with the pre-shared correlation. If the NS-box is a quantum mechanical one, then the conditional probability $\Pr(B_{\vec{y}} - A_{\vec{x}} = \vec{x} \cdot \vec{y} | \vec{x}, \vec{y}])$ should be constrained by the quantum non-locality, so is the joint probability $\Pr(A_{\vec{x}}, B_{\vec{y}} | \vec{x}, \vec{y})$. Then the question is how? For $d = 2$ and $k = 2$, the quantum constraint comes from the well-known Tsirelson inequality. That is, the CHSH correlation function $C_{\vec{x},\vec{y}}$, which can be expressed in terms of joint probability as $\Pr(00|\vec{x}, \vec{y}) - \Pr(01|\vec{x}, \vec{y}) - \Pr(10|\vec{x}, \vec{y}) + \Pr(11|\vec{x}, \vec{y})$ for uniform output marginal probabilities, is the object for quantum communication complexity bounded by $2\sqrt{2}$. This is the constraint for $\Pr(A_{\vec{x}}, B_{\vec{y}} | \vec{x}, \vec{y})$ to be consistent with quantum mechanics.

However, there is no known Tsirelson inequality for the cases with $d > 2$. Despite that, in [32], we find a systematic way to construct $d = 2$ multi-setting Tsirelson inequalities by the signal decay theorem [50, 51]. We will generalize this method to $d > 3$ case to yield suitable objects for quantum communication complexity. To proceed, we first recapitulate the derivation for $d = 2$ cases.

Signal decay theory tells the loss of mutual information when processing the data through a noisy channel. Consider a cascade of two communication channels: $X \hookrightarrow Y \hookrightarrow Z$, then intuitively we have

$$I(X; Z) \leq I(X; Y). \tag{3.7}$$

Moreover, if the second channel is a binary symmetric one, i.e.,

$$\Pr(Z|Y) = \begin{pmatrix} \frac{1}{2}(1+\xi) & \frac{1}{2}(1-\xi) \\ \frac{1}{2}(1-\xi) & \frac{1}{2}(1+\xi) \end{pmatrix},$$

then the signal decay theorem says

$$\frac{I(X; Z)}{I(X; Y)} \leq \xi^2. \tag{3.8}$$

This theorem has been proven as the tight bound in [50, 51]. Note the equality is held only when propagating the weak signal for noisy the channel $\Pr(Z|Y)$. i.e., $\Pr(Y|X = 0)$ and $\Pr(Y|X = 1)$ are almost indistinguishable. For more detail, please see appendix A.

In [32], we defined $X = a_i$, $Y = a_0 + \vec{x} \cdot \vec{y}$ and $Z = \beta$. On purpose, the bit a_i is encoded as $a_0 + \vec{x} \cdot \vec{y}$ such that $I(a_i; a_0 + \vec{x} \cdot \vec{y}) = 1$. Using the tight bound of (3.8), in this case we can get

$$I(a_i\ ; \beta | b = i) \leq \xi_i^2. \tag{3.9}$$

For our communication protocol, the noise parameter ξ_i is denoted as $\xi_{\vec{y}}$ and can be related to Alice's input marginal and the joint probability of the NS-box as follows:

$$\frac{1+\xi_{\vec{y}}}{2} = \sum_{\{\vec{x}\}} \Pr(\vec{x}) \Pr\left(B_{\vec{y}} - A_{\vec{x}} = \vec{x}\cdot\vec{y} | \vec{x}, \vec{y}\right). \tag{3.10}$$

Assuming that Alice's database is i.i.d., we can then sum over mutual information between β and a_i to arrive

$$\sum_i I(a_i\,;\beta|b=i) \leq \sum_i \xi_i^2. \tag{3.11}$$

Though this object is quadratic, we can linearize it by the Cauchy-Schwarz inequality, i.e., $|\sum_i \xi_i| \leq \sqrt{k\sum_i \xi_i^2}$. For $d=k=2$ case with uniform input marginal probabilities $\Pr(a_i)$, it is easy to show that $\sum_i \xi_i \leq \sqrt{2}$ (or $\sum_i \xi_i^2 \leq 1$) is nothing but the conventional Tsirelson inequality. Moreover, in [32] we use the SDP algorithm in [56] to generalize to $d=2$ and $k>2$ cases and show that the corresponding Tsireslon's inequalities are

$$\sum_i \xi_i \leq \sqrt{k}. \tag{3.12}$$

This is equivalent to say $\sum_i \xi^2 \leq 1$. From the signal decay theorem (3.9) we find that it implies the quantum communication complexity is consistent with the information causality (3.1).

We now generalize the above construction to $d>2$ cases. First, we start with $d=3$ case by considering a cascade of two channels $X \hookrightarrow Y \hookrightarrow Z$ with the second one a 3-input, 3-output symmetric channel. Again, we want to find the upper bound of $\frac{I(X;Z)}{I(X;Y)}$. In the Appendix A we show that the ratio saturates the upper bound whenever three conditional probabilities $\Pr(Y|X=i)$ with $i=0,1,2$ are almost indistinguishable. Moreover, it can be also shown that the upper bound of the ratio is again given by (3.8) for the symmetric channel between Y and Z specified by

$$\Pr(Z|Y) = \begin{pmatrix} \frac{2\xi+1}{3} & \frac{1-\xi}{3} & \frac{1-\xi}{3} \\ \frac{1-\xi}{3} & \frac{2\xi+1}{3} & \frac{1-\xi}{3} \\ \frac{1-\xi}{3} & \frac{1-\xi}{3} & \frac{2\xi+1}{3} \end{pmatrix}. \tag{3.13}$$

One can generalize the above to the higher d cases for the symmetric channel between Y and Z specified as follows: $\Pr(Z=i|Y=i) = \frac{(d-1)\xi+1}{d}$ and $\Pr(Z=s\neq i|Y=i) = \frac{1-\xi}{d}$ with $i \in \{0,1,...,d-1\}$. Again we will arrive (3.8). Based on the signal decay theorem

with $X := a_i$, $Y := a_0 + \vec{x} \cdot \vec{y}$ and $Z := \beta$ and assuming that Alice's input probabilities are i.i.d., we can sum over all the mutual information between each a_i and β,

$$\sum_{i=0}^{k-1} I(\beta; a_i | b = i) \leq \sum_{i=0}^{k-1} \xi_i^2 \log_2(d). \tag{3.14}$$

In our communication protocol, the noise parameter ξ_i is denoted as $\xi_{\vec{y}}$ and can be expressed as

$$\xi_{\vec{y}} = \frac{d \sum_{\vec{x}} \Pr(\vec{x}) \Pr(B_{\vec{y}} - A_{\vec{x}} = \vec{x} \cdot \vec{y} | \vec{x}, \vec{y}) - 1}{d - 1}. \tag{3.15}$$

As for the $d = 2$ case, we assume the upper bound of (3.14) is capped by the information causality to yield a quadratic constraint on the noise parameters. Again, using the Cauchy-Schwarz inequality to linearize the quadratic constraint, we find that the generalized inequality $\sum_{\vec{y}} \xi_{\vec{y}} \leq \sqrt{k}$. This inequality could be the Tsirelson-type inequality and it need to be checked. Especially, if the input marginal probabilities $\Pr(a_i)$ are uniform, the bound on the object $\sum_{\vec{y}} \xi_{\vec{y}}$ yields a constraint on $\Pr(A_{\vec{x}}, B_{\vec{y}} | \vec{x}, \vec{y})$. Therefore, we obtain a proper object of characterizing non-locality: $\sum_{\vec{y}} \xi_{\vec{y}}$ with uniform $\Pr(a_i)$.

Then, it is ready to ask the question: does the joint probabilities $\Pr(A_{\vec{x}}, B_{\vec{y}} | \vec{x}, \vec{y})$ giving the maximal quantum non-locality saturate the upper bound of information causality? Next, we are going to answer this question.

3.3 Convexity and mutual information

3.3.1 Feasibility for maximizing mutual information by convex optimization?

In order to test information causality for difference quantum communication protocols, we have to maximize mutual information I over quantum channel and Alice's input probability. One way is to formulate the problem as the convex optimization programming, so that we may exploit some numerical recipes such as [63] to carry out the task.

Minimizing a function with the equality or inequality constraints is called convex optimization. The object function could be linear or non-linear. For example, SDP is a kind of convex optimization with a linear object function. Regardless of linear or non-linear object functions, the minimization (maximization) problem requires them to be convex (concave). Thus, if we define the mutual information I as the object function for maximization in the context of information causality, we have to check the concavity of it.

A concave function $f(x)$ ($f : \mathbb{R}^n \to \mathbb{R}$) should satisfy the following condition:

$$f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda) f(x_2), \tag{3.16}$$

where x_1 and x_2 are n-dimensional real vectors, and $0 < \lambda < 1$.

Mutual information between input X and output Z can be written as

$$I(X;Z) = H(Z) - H(Z|X) = H(Z) - \sum_i \Pr(X=i) H(Z|X=i), \tag{3.17}$$

where $H(Z) = -\sum_i \Pr(Z=i) \log_2 \Pr(Z=i)$ is the entropy function. We will study the convexity of $I(X;Z)$ by varying over the marginal probability $\Pr(X)$ and the channel probability $\Pr(Z|X)$.

The following theorem is mentioned in [68]. If we fix the channel probability $\Pr(Z|X)$ in (3.17), then $I(X;Z)$ is a concave function with respect to $\Pr(X)$. This is the usual way in obtaining the channel capacity i.e., maximizing mutual information I over input marginal probability for a fixed channel.

However, in the context of information causality, the channel probability is related to both the joint probability of the NS-box and the input marginal probability. This means that the above twos will be correlated if we fix the channel probability. This cannot fit to our setup in which we aim to maximize the mutual information I by varying over the joint probability of NS-box and the input marginal probability. For example, in $d = 2$ and $k = 2$ case, the channel probability is given by

$$\Pr(\beta|a_i, b=i) = \begin{pmatrix} \alpha_i & 1-\alpha_i \\ 1-\lambda_i & \lambda_i \end{pmatrix}.$$

where

$$\begin{aligned}
\alpha_0 :=& \quad \Pr(\beta=0|a_0=0,b=0) = \sum_{\ell=0}^{1} \Pr(B_y - A_x = 0|x=\ell, y=0) \Pr(a_1=\ell), \\
\lambda_0 :=& \quad \Pr(\beta=1|a_0=1,b=0) = \sum_{\ell=0}^{1} \Pr(B_y - A_x = 0|x=\ell, y=0) \Pr(a_1=1-\ell), \\
\alpha_1 :=& \quad \Pr(\beta=0|a_1=0,b=1) = \sum_{\ell=0}^{1} \Pr(B_y - A_x = \ell|x=\ell, y=1) \Pr(a_0=\ell), \\
\lambda_1 :=& \quad \Pr(\beta=1|a_1=1,b=1) = \sum_{\ell=0}^{1} \Pr(B_y - A_x = \ell|x=\ell, y=1) \Pr(a_0=1-\ell).
\end{aligned} \tag{3.18}$$

From the above, we see that the channel probability $\Pr(\beta|a_i, b = i)$ cannot be fixed by varying over $\Pr(B_y - A_x|x, y)$ and $\Pr(a_i)$ independently. Similarly, for higher d and k protocols, we will also have the constraints between the above three probabilities. Thus, maximizing the mutual information I by varying the NS-box in the context of information causality is quite different from the usual way of finding the channel capacity.

To achieve the goal of maximizing the mutual information I over the NS-box, we should check if it is a convex (or concave) optimization problem or not. If it is yes, then we can adopt the numerical recipe as [63] to carry out the task. Otherwise, we can either impose more constraints for our problem or just do it by brutal force. It is known that [67] one can check if maximizing function $f(y_1, \cdots, y_n)$ over y_i's is a concave problem or not by examining its Hessian matrix

$$H(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial y_1^2} & \frac{\partial^2 f}{\partial y_1 y_2} & \cdots & \frac{\partial^2 f}{\partial y_1 y_n} \\ \frac{\partial^2 f}{\partial y_2 y_1} & \frac{\partial^2 f}{\partial y_2^2} & \cdots & \frac{\partial^2 f}{\partial y_2 y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial y_n y_1} & \frac{\partial^2 f}{\partial y_n y_2} & \cdots & \frac{\partial^2 f}{\partial y_n^2} \end{pmatrix}. \tag{3.19}$$

For the maximization to be a concave problem, the Hessian matrix should be negative semidefinite. That is, all the odd order principal minors of $H(f)$ should be negative and all the even order ones should be positive. Note that each first-order principal minor of $H(f)$ is just the second derivative of f, i.e. $\frac{\partial^2 f}{\partial y_i^2}$. So, the problem cannot be concave if $\frac{\partial^2 f}{\partial y_i^2} > 0$ for some i.

With the above criterion, we can now show that the problem of maximizing I over $\Pr(B_{\vec{y}} - A_{\vec{x}}|\vec{x}, \vec{y})$ and $\Pr(a_i)$ cannot be a concave problem. To do this, we rewrite the mutual information I defined in (3.1) as following:

$$I = \sum_{i=0}^{k-1} \sum_{n=0}^{d-1} \sum_{j=0}^{d-1} \Pr(\beta = n, a_i = j|b = i) \log_2 \frac{\Pr(\beta = n, a_i = j|b = i)}{\Pr(\beta = n|b = i) \Pr(a_i = j)}. \tag{3.20}$$

Furthermore, one can express the above in terms of $\Pr(B_{\vec{y}} - A_{\vec{x}}|\vec{x}, \vec{y})$ and $\Pr(a_i)$ by the following relations

$$\begin{aligned} \Pr(\beta = n, a_i = j|b = i) &= \sum_{\{a_{k\neq i}\}} \Pr(B_{\vec{y}} - A_{\vec{x}} = n - a_0|\vec{x}, \vec{y}) \Pr(a_i = j) \, \Pi_{k\neq i} \Pr(a_k), \\ \Pr(\beta = n|b = i) &= \sum_{j=0}^{d-1} \Pr(\beta = n, a_i = j|b = i), \end{aligned} \tag{3.21}$$

where $\vec{x}$ and $\vec{y}$ in the above are given by the encoding of the RAC protocol, namely, $\vec{x} := (x_1, \cdots, x_{k-1})$ with $x_i = a_i - a_0$ and $\vec{y} := (y_1, \cdots, y_{k-1})$ with $y_i = \delta_{b,i}$ for $b \neq 0$ and $\vec{y} = 0$ for $b = 0$.

Moreover, both $\Pr(B_{\vec{y}} - A_{\vec{x}}|\vec{x}, \vec{y})$ and $\Pr(a_i)$ are subjected to the normalization conditions of total probability. Thus we need to solve these conditions such that the mutual information I is expressed as the function of independent probabilities. After that, we can evaluate the corresponding Hessian matrix to examine if the maximization of I over these probabilities is a concave problem or not.

For illustration, we first consider the $d = 2$ and $k = 2$ case. By using the relations (3.21) and the normalization conditions of total probability to implement the chain-rule while taking derivative, we arrive

$$\begin{aligned}
&\frac{\ln 2 \cdot \partial^2 I}{\partial(\Pr(B_y - A_x = 0|x = 0, y = 0))^2} = \\
&-(\frac{1}{\Pr(\beta = 0|b = 0)} + \frac{1}{\Pr(\beta = 1|b = 0)})(\Pr(a_0 = 0)\Pr(a_1 = 0) - \Pr(a_0 = 1)\Pr(a_1 = 1))^2 \\
&+(\Pr(a_0 = 0)\Pr(a_1 = 0))^2(\frac{1}{\Pr(\beta = 0, a_0 = 0|b = 0)} + \frac{1}{\Pr(\beta = 1, a_0 = 0|b = 0)}) \\
&+(\Pr(a_0 = 1)\Pr(a_1 = 1))^2(\frac{1}{\Pr(\beta = 0, a_0 = 1|b = 0)} + \frac{1}{\Pr(\beta = 1, a_0 = 1|b = 0)}). \qquad (3.22)
\end{aligned}$$

Obviously, (3.22) cannot always be negative. This can be seen easily if we set $\Pr(a_0) = 1 - \Pr(a_1)$ so that the first term on the RHS of (3.22) is zero. Then, the remaining terms are non-negative definiteness. This then indicates that maximizing I over the joint probability is not a concave problem.

The check for the higher d and k cases can be done similarly, and the details can be found in the Appendix B. Again, we can set all the $\Pr(a_i)$ to be uniform so that we have

$$\begin{aligned}
&\frac{d^{2k} \ln 2 \cdot \partial^2 I}{\partial(\Pr(B_{\vec{y}} - A_{\vec{x}} = 0|\vec{x} = \vec{0}, \vec{y} = \vec{0}))^2} = \\
&\sum_{n=0}^{d-1}(\frac{1}{\Pr(a_0 = n, \beta = n|b = 0)} + \frac{1}{\Pr(a_0 = n, \beta = n + 1 - d|b = 0)}) > 0. \qquad (3.23)
\end{aligned}$$

3.3.2 Convex optimization for symmetric and isotropic channels with i.i.d. and uniform input marginal probabilities

Recall that we would like to check if the boundaries of the information causality and the quantum non-locality agree or not. To achieve this, we may either maximizing the mutual

information I with the quantum constraint, or maximizing the quantum non-locality and then evaluate the corresponding mutual information I which can be compared with the bound of information causality. These two tasks are not equivalent but complementary. However, unlike the first task, the second task will be concave problem as known in [56, 45]. The only question in this case is if the corresponding mutual information I is monotonically related to the quantum non-locality or not. If yes, then maximizing quantum non-locality is equivalent to maximizing the mutual information I. The answer is partially yes as we will show because this monotonic relation holds only for symmetric and isotropic channels with i.i.d. and uniform input marginal probabilities.

Assuming Alice's input is i.i.d., we have $H(\beta|b=i) = \log_2(d)$. Also, once the channel is symmetric, we have $\Pr(\beta = t|a_i = j, b = i) = \frac{(d-1)\xi_i+1}{d}$ for $t = j$, and $\Pr(\beta = t|a_i = j, b = i) = \frac{1-\xi_i}{d}$ for $t \neq j$. Thus, the mutual information I becomes

$$I = k\log_2 d + \sum_{i=0}^{k-1}[\frac{(d-1)\xi_i+1}{d}\log_2(\frac{(d-1)\xi_i+1}{d}) + (1-\frac{(d-1)\xi_i}{d})\log_2(\frac{1-\xi_i}{d})] \quad (3.24)$$

If we also assume the channels are isotropic i.e., $\xi_{\vec{y}} = \xi$, then for such a case the mutual information I can be further simplified to

$$I = k[\log_2 d + \frac{(d-1)\xi+1}{d}\log_2(\frac{(d-1)\xi+1}{d}) + (1-\frac{(d-1)\xi}{d})\log_2(\frac{1-\xi}{d})]. \quad (3.25)$$

The value of ξ is in the interval $[0, 1]$ with $\xi = 0$ for the completely random channel, and $\xi = 1$ for the noiseless one.

We can show that the mutual information I is the monotonic increasing function of the quantum non-locality parameterized by the noise parameter ξ. To do this, we calculate the first and second derivative of I with respect to ξ and obtain

$$\frac{dI}{d\xi} = \frac{d-1}{d}\log\frac{(d-1)\xi+1}{1-\xi},$$
$$\frac{d^2I}{d\xi^2} = \frac{d-1}{d}(\frac{d-1}{(d-1)\xi+1} + \frac{1}{1-\xi}).$$

From the above, we see that $\frac{dI}{d\xi}$ is always positive for $\xi \in [0, 1]$. Moreover, it is easy to see that I is minimal at $\xi = 0$ since $\frac{d^2I}{d\xi^2} = d - 1 > 0$. Thus, the mutual information I is a monotonically increasing function of ξ for the symmetric and isotropic quantum channel with i.i.d. and uniform input marginal probabilities.

3.4 Finding the bound of Bell-type inequality from the hierarchical semidefinite programming

We now will prepare for numerically evaluating the maximum of (3.5) which is monotonic increasing with mutual information I under some assumptions. In order to ensure that the maximum of (3.5) can be obtained by quantum resource, we have to use the same method as in [44, 45]. In [44, 45], they checked if a given set of probabilities can be reproduced from quantum mechanics or not. The test can be formulated as a hierarchy of semidefinite programming (SDP). This is a very important issue in quantum information. With the test, we can know the limitation for transmitting quantum information and the non-local nature of the quantum correlation.

3.4.1 Projection operators with quantum behaviors

We will briefly review the basic ideas in [44, 45] and then explain how to use it for our program. In [44, 45] they use the projection operators under following measurement scenario. Two distant partite Alice and Bob share a NS-box. Alice and Bob input X and Y to the NS-box, respectively, and obtain the corresponding outcomes $a \in A$ and $b \in B$. Here A and B are used to denote the set of all possible Alice's and Bob's measurement outcomes, respectively. We use $X(a)$ and $Y(b)$ to denote corresponding inputs. These outcomes can be associated with some sets of projection operators $\{E_a : a \in A\}$ and $\{E_b : b \in B\}$. The joint probability of the NS-box can then be determined by the quantum state ρ of the NS-box and the projection operators as following:

$$\Pr(a, b) = \mathrm{Tr}(E_a E_b \rho). \tag{3.26}$$

Note that $\Pr(a, b)$ is the abbreviation of $\Pr(A_{\vec{x}}, B_{\vec{y}}|\vec{x}, \vec{y}) = \mathrm{Tr}(E_{A_{\vec{x}}} E_{B_{\vec{y}}} \rho)$ defined in the previous sections.

If E_a and E_b are the genuine quantum operators, then they shall satisfy (i) hermiticity: $E_a^\dagger = E_a$ and $E_b^\dagger = E_b$; (ii) orthogonality: $E_a E_{a'} = \delta_{aa'}$ if $X(a) = X(a')$ and $E_b E_{b'} = \delta_{b,b'}$ if $Y(b) = Y(b')$; (iii) completeness: $\Sigma_{a \in X} E_a = \mathbb{I}$ and $\Sigma_{b \in Y} E_b = \mathbb{I}$; and (iv) commutativity: $[E_a, E_b] = 0$.

In our measurement scenario, the distant partite Alice and Bob perform local measurements so that property (iv) holds. On the other hand, the property (iii) implies no-signaling as it leads to (3.2) via (3.26). Furthermore, this property also implies that

there is redundancy in specifying Alice's operators E_a's with the same input since one of them can be expressed by the others. Thus, we can eliminate one of the outcomes per setting and denote the corresponding sets of the remaining outcomes for the input X by $\tilde{A}_X$ (or $\tilde{B}_Y$ for Bob's outcomes with input Y). The collection of such measurement outcomes $\sum_X \tilde{A}_X$ is denoted as $\tilde{A}$. Similarly, we denote the collection of Bob's independent outcomes as $\tilde{B}$.

Using the reduced set of projection operators $\{E_a : a \in \tilde{A}\}$ and $\{E_b : b \in \tilde{B}\}$, we can construct a set of operators $O = \{O_1, O_2, ..., O_i, ...\}$. Here O_i is some linear function of products of operators in $\{\mathbb{I} \cup \{E_a : a \in \tilde{A}\} \cup \{E_b : b \in \tilde{B}\}\}$. The set O is characterized by a matrix Γ given by

$$\Gamma_{ij} = \mathrm{Tr}(O_i^\dagger O_j \rho). \tag{3.27}$$

By construction, Γ is non-negative definite, i.e.,

$$\Gamma \succeq 0. \tag{3.28}$$

This can be easily proved as follows. For any vector $v \in \mathbb{C}^n$ (assuming Γ is a n by n matrix), one can have

$$v^\dagger \Gamma v = \Sigma_{s,t} v_s^* \mathrm{Tr}(O_s^\dagger O_t \rho) v_t = \mathrm{Tr}(V^\dagger V \rho) \geq 0. \tag{3.29}$$

Recall that our goal is to judge if a given set of joint probabilities such as (3.26) can be reproduced by quantum mechanics or not. In this prescription, the joint probabilities is then encoded in the matrix Γ satisfying the quantum constraints (3.26) and (3.28). However, Γ contains more information than just joint probabilities (3.26). For examples, the terms appearing in the elements of Γ such as $\mathrm{Tr}(E_a E_{a'} \rho)$, $\mathrm{Tr}(E_b E_{b'} \rho)$ for $X(a) \neq X(a')$ and $Y(b) \neq Y(b')$ can not be expressed in terms of the joint probabilities of the NS-box. This is because these measurements are performed on the same partite (either Alice or Bob) and are not commutative. Therefore, to relate the joint probabilities of the NS-box to the matrix Γ, we need to find the proper combinations of Γ_{ij} so that the final object can be expressed in terms of only the joint probabilities. Therefore, given the joint probabilities, there shall exist some matrix functions F_q's such that the matrix Γ is constrained as follows:

$$\Sigma_{s,t} (F_q)_{s,t} \Gamma_{s,t} = g_q \tag{3.30}$$

where g_q's are the linear functions of joint probabilities $\Pr(a, b)$'s.

We then call the matrix Γ a certificate if it satisfies (3.28) and (3.30) for a given set of joint probabilities of NS-box. The existence of the certificate will then be examined numerically by SDP. If the certificate does not exist, the joint probabilities cannot be reproduced by quantum mechanics.

Examples on how to construct F_q and g_q for some specific NS-box protocols can be found in [44, 45]. For illustration, here we will explicitly demonstrate the case not considered in [44, 45], that is the $k = 2$, $d = 3$ RAC protocol. We will use the notation which we defined in the previous sections. We start by defining the set of operators $\mathcal{E} = \{\mathcal{E}_i\} := \mathbb{I} \cup \{E_{A_x} : A_x \in \{0,1\}, x \in \{0,1,2\}\} \cup \{E_{B_y} : B_y \in \{0,1\}, y \in \{0,1\}\}$ with the operator label $i \in \{0, 1, 2, ..., m_a, ..., m_a + m_b\}$. The operator $\mathcal{E}_{i=0}$ is the identity operator $\mathbb{I}$, and $\mathcal{E}_{1<i\leq m_a} \in E_{A_x}$, $\mathcal{E}_{m_a<i\leq m_a+m_b} \in E_{B_y}$.

The associated quantum constraints can be understood as the relations between joint probability $\Pr(a, b)$ and $\mathrm{Tr}(\mathcal{E}_a^\dagger \mathcal{E}_b \rho)$ (or marginal probability $\Pr(a)$ and $\mathrm{Tr}(\mathbb{I}\mathcal{E}_a\rho)$). That is,

$$\begin{aligned}
&\mathrm{Tr}(\rho) = 1, \qquad \mathrm{Tr}(\mathbb{I}E_{A_x}\rho) = \Pr(A_x|x), \qquad \mathrm{Tr}(\mathbb{I}E_{B_y}\rho) = \Pr(B_y|y),\\
&\mathrm{Tr}(E_{A_x}E_{A'_x}\rho) = \delta_{A_x,A'_x}\Pr(A_x|x), \mathrm{Tr}(E_{B_y}E_{B'_y}\rho) = \delta_{B_y,B'_y}\Pr(B_y|y),\\
&\mathrm{Tr}(E_{A_x}E_{B_y}\rho) = \Pr(A_x, B_y|x, y). \qquad (3.31)
\end{aligned}$$

Note that these equations also hold when permuting the operators as $\mathrm{Tr}(E_{A_x}E_{B_y}\rho) = \mathrm{Tr}(E_{B_y}E_{A_x}\rho)$.

Moreover, we can make the matrix Γ to be real and symmetric by redefining it as $\Gamma = (\Gamma^* + \Gamma)/2$. Thus, in the following we will only display the upper triangular part of Γ. We then use the quantum constraints (3.31) to construct F_q and g_q by comparing them with (3.30). We then see that every constraint in (3.31) yields a matrix function F_q which has only one non-zero element, and also yields a function g_q which is either zero or contains only a single term of a marginal or joint probability. These constraints can be further divided into four subsets labeled by $q = (q_1, q_2, q_3, q_4)$ as follows:

1. The labels $q_1, q_2 \in \{0, 1, ..., m_a + m_b\}$ are used to specify the marginal probabilities $\mathrm{Tr}(\mathbb{I}\mathcal{E}_{q_1}\rho)$ and $\mathrm{Tr}(\mathcal{E}_{q_2}^\dagger\mathcal{E}_{q_2}\rho)$. The corresponding matrix functions F_q are given by $(F_{q_1})_{s,t} = \delta_{s,1}\delta_{t,q_1+1}$ and $(F_{q_2})_{s,t} = \delta_{s,q_2+1}\delta_{t,q_2+1}$, and the g_{q_1} and g_{q_2} are the corresponding marginal probabilities.

2. The label $q_3 \in \{1, ..., d^{k-1}+k\}$ is used to specify the probabilities associated with the

orthogonal operator pairs, $\mathrm{Tr}(\mathcal{E}_{2q_3-1}\mathcal{E}_{2q_3}\rho)$. The matrix element $(F_{q_3})_{s,t} = \delta_{s,2q_3}\delta_{t,2q_3+1}$, and $g_{q_3} = 0$.

3. The label $q_4 \in \{1, ..., m_a m_b\} = 4(2x + A_x) + (2y + B_y + 1)$ is used to specify the joint probabilities of the NS-box. The corresponding F_q and g_q are given by $(F_{q_4})_{s,t} = \delta_{s,2x+A_x+2}\delta_{t,m_a+2y+B_y+2}$, and $g_{q_4} = \Pr(A_x, B_y|x, y)$.

Considering the above set of quantum constraint, we can define the associated Γ matrix

$$\Gamma = \begin{pmatrix}
1 & \Pr(0|0)_A & \Pr(1|0)_A & \Pr(0|1)_A & \Pr(1|1)_A & \Pr(0|2)_A & \Pr(1|2)_A & \Pr(0|0)_B & \Pr(1|0)_B & \Pr(0|1)_B & \Pr(1|1)_B \\
 & \Pr(0|0)_A & 0 & \chi_0 & \chi_1 & \chi_2 & \chi_3 & \Pr(00|00) & \Pr(01|00) & \Pr(00|01) & \Pr(01|01) \\
 & & \Pr(1|0)_A & \chi_4 & \chi_5 & \chi_6 & \chi_7 & \Pr(10|00) & \Pr(11|00) & \Pr(10|01) & \Pr(11|01) \\
 & & & \Pr(0|1)_A & 0 & \chi_8 & \chi_9 & \Pr(00|10) & \Pr(01|10) & \Pr(00|11) & \Pr(01|11) \\
 & & & & \Pr(1|1)_A & \chi_{10} & \chi_{11} & \Pr(10|10) & \Pr(11|10) & \Pr(10|11) & \Pr(11|11) \\
 & & & & & \Pr(0|2)_A & 0 & \Pr(00|20) & \Pr(01|20) & \Pr(00|21) & \Pr(01|21) \\
 & & & & & & \Pr(1|2)_A & \Pr(10|20) & \Pr(11|20) & \Pr(10|21) & \Pr(11|21) \\
 & & & & & & & \Pr(0|0)_B & 0 & \chi_{12} & \chi_{13} \\
 & & & & & & & & \Pr(1|0)_B & \chi_{14} & \chi_{15} \\
 & & & & & & & & & \Pr(0|1)_B & 0 \\
 & & & & & & & & & 0 & \Pr(1|1)_B
\end{pmatrix} \tag{3.32}$$

, where $\Pr(A_x|x)_A$'s and $\Pr(B_y|y)_B$'s are the marginal probabilities for Alice and Bob, respectively, and $\Pr(A_x, B_y|x, y)$'s are the joint probability of the NS-box. The elements χ_i's in the above cannot be defined by the given marginal and joint probabilities because they corresponds to the probability of different measurement settings for only one party. Thus, they cannot appear in the constraints (3.30) but are still constrained by the non-negative definiteness of Γ.

Testing the existence of the certificate— The task of testing the existence of the certificate can be formulated as a SDP by defining the standard primal and the associated dual problems. The details can be found in Appendix C. The primal problem of SDP is subjected to certain conditions associated with a positive semidefinite matrix, which can be either linear equalities or inequalities. Each primal problem has an equivalent dual problem. Therefore, when the optimal value of the primal problem is the same as the optimal value of the dual problem, the feasible solution of the problem is obtained.

For our case the primal problem of SDP is as follows:

$$\begin{aligned}
&\textit{maximize} && \lambda && && \text{(3.33a)} \\
&\textit{subject to} && \mathrm{Tr}(F_q^T\Gamma) = g_q, && q = 1, ..., m, && \text{(3.33b)} \\
& && \Gamma - \lambda\mathbb{I} \succeq 0. && && \text{(3.33c)}
\end{aligned}$$

Obviously, if the maximal value $\lambda \geq 0$ is obtained, the non-negative definiteness of Γ is guaranteed under the quantum constraints (3.28).

On the other hand, the associated dual problem is given by

$$\begin{aligned} \textit{maximize} \quad & \sum_q y_q g_q, && \text{(3.34a)} \\ \textit{subject to} \quad & \sum_q y_q F_q^T \succeq 0, && \text{(3.34b)} \\ & \sum_q y_q \mathrm{Tr}(F_q^T) = 1. && \text{(3.34c)} \end{aligned}$$

Note that the quantity $\sum_q y_q g_q$ is the object in characterized the quantum non-locality since g_q's are mainly the two-point correlation function. Therefore, if maximizing this quantity is equivalent to finding the generalized Tsireslon bound. Therefore, if the solution of this SDP is feasible, then the associated certificate exists and there yields the generalized Tsireslon bound.

3.4.2 Hierarchy of the semidefinite programming

Different operator sets O's yield different quantum constrains (3.26) and (3.28). It seems no guideline in choosing the set O and examining the existence of the corresponding certificate. However, it is easy to see that the certificates associated with different operator sets are equivalent. This can be seen as follows. Let us assume O and O' are two linearly equivalent set of operators such that $O_i \in O$ can be expressed by a linear combination of the elements in O', i.e., $O_i = \sum_j C_{i,j} O_j^{'}$. If there exists a matrix Γ' satisfying (3.28) and(3.30) for the corresponding operator set O', then there will exist another matrix Γ whose elements are $\Gamma_{s,t} = \sum_{q,l} C_{q,s}^* \Gamma_{q,l}^{'} C_{l,t}$ also satisfying (3.28) and (3.30) for the set O. Therefore, we only need to stick to one set of operators in this linear equivalence class when examining the existence of the corresponding certificate.

Besides, a systematic way of constructing O is proposed in [44, 45] so that the task of finding the certificate can be formulated as a hierarchy of SDP. This is constructed as follows. The length of the operator O_i, denoted by $|O_i|$, is defined as the minimal number of projectors used to construct it. We can then divide the set O into different subsets labeled by the maximal length of the operators in the corresponding subset. Thus, we decompose the operator set O into a sequence of hierarchical operator sets denoted by S_n

where n is the maximal length of the operators in S_n. That is,

$$
\begin{aligned}
&S_0 = \{\mathbb{I}\} \\
&S_1 = \{S_0\} \cup \{E_a : a \in \tilde{A}\} \cup \{E_b : b \in \tilde{B}\} \\
&S_2 = \{S_0\} \cup \{S_1\} \cup \{E_a E_{a'} : a, a' \in \tilde{A}\} \cup \{E_b E_{b'} : b, b' \in \tilde{B}\} \cup \\
&\{E_a E_b : a \in \tilde{A}, b \in \tilde{B}\} \\
&\ldots
\end{aligned} \tag{3.35}
$$

Furthermore, to save the computer memory space used in the numerical SDP algorithm, in the above sequence we can add an intermediate set between S_n and S_{n+1}, which is given by $S_{n+AB} := \{S_n\} \cup \{S \in S_{n+1} | S = E_a E_b S' : a \in \tilde{A}, b \in \tilde{B}\}$. For example, when $n = 1$ we have $S_{1+AB} = \{S_1\} \cup \{E_a E_b : a \in \tilde{A}, b \in \tilde{B}\}$ such that $S_1 \subseteq S_{1+AB} \subseteq S_2$. Note that S_{1+AB} does not have the product of the marginal projection operators in the form of $\{E_a E_{a'} : a, a' \in \tilde{A}\}$ and $\{E_b E_{b'} : b, b' \in \tilde{B}\}$. It is clear that $S_{1+AB} \subseteq S_2$. All the operators in O can be expressed in terms of the linear combination of the operators in S_n for large enough n.

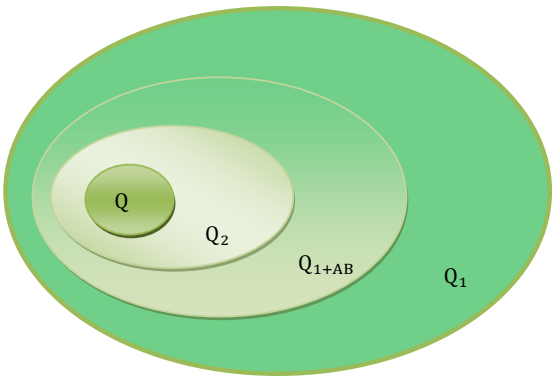

Figure 3.3: The geometric interpretation of collection Q_n

Since we know $S_n \subseteq S_{n+AB} \subseteq S_{n+1}$, the associated constraints produced by S_{n+1} is stronger than S_{n+AB} and S_n. We can start the task from S_1 then S_{1+AB}, S_2 and so on. Let the certificate matrix associated with the set S_n be denoted as $\Gamma^{(n)}$. Finding the certificate associated with this sequence can be formulated as a hierarchical SDP. Once

the given joint probabilities satisfy the quantum constraints (3.28) so that the associated certificate $\Gamma^{(n)}$ exists, we then denote the collection of these joint probabilities as Q_n. Since we know that the associated constraints are stronger than the previous steps of the hierarchical sequence, the collection Q_n will become smaller for the higher n. That is, the non-quantum correlations will definitely fail the test at some step in the hierarchical SDP. The geometrical interpretation of the above fact is depicted in Fig 3.3.

It was shown in [44, 45] that the probability is ensured to be quantum only when the certificate associated with $S_{n\to\infty}$ exists, i.e., for the joint probabilities in the collection Q of Fig 3.3. In this sense, it seems that we have to check infinite steps. To cure this, a stopping criterion is proposed in [44, 45] to terminate the check process at some step of the hierarchical SDP. This can ensure that the given joint probability is quantum at finite n if it is.

The stopping criterion is satisfied when the rank of sub matrix of $\Gamma^{(n)}$ is equal to the rank of $\Gamma^{(n)}$, i.e.,

$$rank(\Gamma^{(n)}_{X,Y}) = rank(\Gamma^{(n)}). \tag{3.36}$$

The element of $\Gamma^{(n)}_{X,Y}$ is constructed by the operators in the set $S_{X,Y} := \{S_{n-1}\} \bigcup \{S = E_a E_b S' : a \in \tilde{A}_X, b \in \tilde{B}_Y, |S| \leq n\}$.

The above stopping criterion is for integer n. However, it was also generalized in [45] for the intermediate certificate $\Gamma^{(n+AB)}$: the stopping criterion is satisfied if the following equation is satisfied for all the measurement settings X and Y,

$$rank(\Gamma^{(n+XY)}) = rank(\Gamma^{(n+AB)}), \tag{3.37}$$

so that the certificate $\Gamma^{(n+AB)}$ has a rank loop. Here $\Gamma^{(n+XY)}$ is the certificate associated with $S_{n+XY} := \{S_n\} \cup \{S \in S_{n+1} | S = E_a E_b S' : a \in \tilde{A}_X, b \in \tilde{B}_Y\}$.

Now we are ready to implement the above criterion to numerically examine the quantum behaviors of the given joint probabilities for our RAC protocols with higher k and d.

3.4.3 The bound of Bell-type inequality and the corresponding mutual information in the hierarchical semidefinite programming

Any Bell-type inequality including (3.5) can be written as the linear combination of joint probabilities, then the hierarchical SDP can be used to approach the quantum bound

of Bell-type inequality (the Tsirelson bound). Recall that the quantum non-locality and the mutual information I are monotonically related for symmetric and isotropic channels with i.i.d. and uniform input marginal probabilities. After obtaining the maximum of Bell-type function at each step of the aforementioned hierarchical SDP, we can calculate the corresponding mutual information I and compare with the information causality. Since the quantum constraint is stronger in the hierarchical SDP and the collection of Q_n will become smaller while n is increasing. We then know that the bound of Bell-type inequality and the associated mutual information I will become tighter for larger n and it will converge to the quantum bound for large enough n. Once the bound of mutual information I at some step of hierarchy does not saturate the information causality, we can then infer that the quantum bound of mutual information will not saturate the information causality, too.

First, let us discuss how to find the bound of Bell-type inequality. As discussed before, the problem of finding the Tsirelson bound can be reformulated as a SDP. The primal problem of this SDP is defined as

$$\begin{aligned}
&\textit{maximize} && \mathrm{Tr}(C^T\Gamma^{(n)}) && \text{(3.38a)}\\
&\textit{subject to} && \mathrm{Tr}(F_q^T\Gamma^{(n)}) = g_q(p), \quad q = 1, \cdots, m; && \text{(3.38b)}\\
& && \Gamma^{(n)} \succeq 0. && \text{(3.38c)}\\
& && \mathrm{Tr}(H_w^T\Gamma^{(1)}) \geq 0, \quad w = 1, \cdots, s; && \text{(3.38d)}
\end{aligned}$$

The matrix C is given to make $\mathrm{Tr}(C^T\Gamma^{(n)})$ the Bell-type function which we would like to maximize. Eq. (3.38b) and (3.38c) are the quantum constraints discussed in the previous subsections so that the quantum behaviors are ensured during the SDP procedure. Moreover, with proper choice of the matrix H_w [5], the condition (3.38d) is introduced to ensure the non-negativity of the joint probabilities which are the off-diagonal elements of $\Gamma^{(1)}$.

In the following we define the matrix C for our case. Eq. (3.5), which can be expressed as the linear combination of the joint probabilities, i.e., $\sum_{\vec{x},\vec{y}} \mathrm{Pr}(B_{\vec{y}} - A_{\vec{x}} = \vec{x}\cdot\vec{y}|\vec{x},\vec{y})$, is the object for our SDP (3.38). Since we only consider $d-1$ marginal probabilities per measurement setting, we should further rewrite our object according to the completeness

[5] Since we only consider $a \in \tilde{A}$ and $b \in \tilde{B}$ to save the computer memory space, we need to choose H_w to ensure the non-negative definiteness of not only the $(d-1)^2$ terms of $\Gamma^{(1)}$ but also the other $d^2-(d-1)^2$ terms which are the linear combinations of the elements of $\Gamma^{(1)}$.

conditions, i.e., $\Sigma_{a\in X}E_a = \mathbb{I}$ and $\Sigma_{b\in Y}E_b = \mathbb{I}$. After rewriting, we can write down the matrix C in (3.38). We take $d = 3$, $k = 2$ RACs protocol for example. For $\Gamma^{(1)}$,

$$C = \frac{1}{2}\begin{pmatrix}
1. & 2. & 0. & 0. & -1. & 1. & 1. & 0. & 3. & 0. & 0. \\
2. & 0. & 0. & 0. & 0. & 0. & 0. & -1. & -2. & -1. & -2. \\
0. & 0. & 0. & 0. & 0. & 0. & 0. & 1. & -1. & 1. & -1. \\
0. & 0. & 0. & 0. & 0. & 0. & 0. & -1. & -2. & 2. & 1. \\
-1. & 0. & 0. & 0. & 0. & 0. & 0. & 1. & -1. & 1. & 2. \\
1. & 0. & 0. & 0. & 0. & 0. & 0. & -1. & -2. & -1. & 1. \\
1. & 0. & 0. & 0. & 0. & 0. & 0. & 1. & -1. & -2. & -1. \\
0. & -1. & 1. & -1. & 1. & -1. & 1. & 0. & 0. & 0. & 0. \\
3. & -2. & -1. & -2. & -1. & -2. & -1. & 0. & 0. & 0. & 0. \\
0. & -1. & 1. & 2. & 1. & -1. & -2. & 0. & 0. & 0. & 0. \\
0. & -2. & -1. & 1. & 2. & 1. & -1. & 0. & 0. & 0. & 0.
\end{pmatrix}. \tag{3.39}$$

The size of (3.39) is equal to the size of $\Gamma^{(1)}$ (the first step in our hierarchical SDP). If $n \neq 1$, the size of matrix C will be bigger, we could define (3.39) as the sub-matrix of matrix C and the other elements of C are zero such that the object functions $\mathrm{Tr}(C^T\Gamma^{(n)})$ are all equal for different steps of our hierarchical SDP.

For higher d and k, we write down the quantum constraints (3.28) for $\Gamma^{(1)}$ and $\Gamma^{(1+AB)}$ and estimate its number in Appendix E. However, due to the limitation of the computer memory (we have $128GB$), we cannot finish all the tests of our hierarchical SDP but stop at level of $1 + AB$. In our calculation, we take the $\sum_{\vec{x},\vec{y}}\Pr(B_{\vec{y}} - A_{\vec{x}} = \vec{x}\cdot\vec{y}|\vec{x},\vec{y})$ as the object of SDP, which is monotonically related to the object of communication complexity $\sum_{\vec{y}}\xi_{\vec{y}}$ in a straightforward way via (3.15).

At the $n = 1$ level the numerical results of our SDP object $\sum_{\vec{x},\vec{y}}\Pr(B_{\vec{y}}-A_{\vec{x}} = \vec{x}\cdot\vec{y}|\vec{x},\vec{y})$ for various k and d are listed below:

k	d=2	d=3	d=4	d=5
2	3.4142	4.8284	6.2426	7.6569
3	9.4641	19.3923	32.7846	49.6410
4	24.0000	72.0000	160.0000	
5	57.8885	255.7477		

The entries are in the table are the values of $\sum_{\vec{x},\vec{y}}\Pr(B_{\vec{y}} - A_{\vec{x}} = \vec{x}\cdot\vec{y}|\vec{x},\vec{y})$.

Similarly, at the $n = 1+AB$ level the results for the same SDP object are listed below:

k	d=2	d=3	d=4	d=5
2	3.4142	4.6667	5.9530	7.1789
3	9.4641	18.6633		
4	24.0000			
5	57.8885			

The stopping criterion is checked at the same time. Unfortunately, it is not satisfied for $\Gamma^{(1+AB)}$, this means that the bound associated with Γ^{1+AB} is not the Tsirelson bound. However, our numerical computational capacity cannot afford the higher level calculations.

Few more remarks are in order: (i) Even we do not require our channel to be isotropic, i.e., uniform $\xi_{\vec{y}}$ for our SDP, we find that the channel for maximizing the SDP object to be isotropic for our level $n = 1$ and $n = 1+AB$ check. (ii) We find the bound at the $n = 1$ level is the same as the bound derived from signal decay theorem in section 3.2. (iii) For $d = 2$ case, the bound for the SDP object at the $n = 1$ and $n = 1 + AB$ level are equal, which is also the same as the Tsirelson bound as can be proved by Tsirelson's theorem [56]. Since the bound is the Tsirelson bound, it will not change for the further steps of the hierarchical SDP. (iv) For $d > 2$, the bound of the SDP object at the $n = 1 + AB$ level becomes tighter than the one at the $n = 1$ level, as expected. However, it needs more numerical efforts to arrive the true tight bound for the quantum non-locality, i.e., the generalized Tsireslon bound.

Since the optimal channel for the above SDP procedure is isotropic, we can then obtain the value of the noise parameter ξ and use (3.25) to evaluate the corresponding mutual information I:

At the $n = 1$ level,

	d=2	d=3	d=4	d=5
Information causality	1.0000	1.5850	2.0000	2.3220
k=2	0.7982	1.3547	1.7845	2.1357
k=3	0.7680	1.3360	1.7895	2.1680
k=4	0.7549	1.3333	1.8048	
k=5	0.7476	1.3345		

The entries are the corresponding mutual information I given by (3.25).

At the $n = 1 + AB$ level,

	d=2	d=3	d=4	d=5
Information causality	1.0000	1.5850	2.0000	2.3220
k=2	0.7982	1.1972	1.5478	1.7788
k=3	0.7680	1.1531		
k=4	0.7549			
k=5	0.7476			

Note that our results support the information causality. This is because the maximal mutual information I evaluated from the joint probabilities constrained by the $n = 1$ certificates is already smaller than the bound from the information causality. Thus, as implied by the geometric picture of Fig. 3.3, the the quantum bound on the mutual information I obtained in the large n limit will also satisfy the information causality, at least for the symmetric and isotropic channels with i.i.d. and uniform input marginal probabilities. Moreover, for a given d the maximal mutual information I from the certificates decrease as k increases. However, it is hard to find the quantum bound of the mutual information I exactly because the stopping criterion fails at the $n = 1 + AB$ level. It needs more checks for higher n certificate to arrive the quantum bound of the mutual information I. However, we will not carry out this task due to the limitation of the computational power.

3.5 Maximizing mutual information for general quantum communication channels

Most of the RACs protocols discussed so far and in the literatures are under some assumptions such as i.i.d., uniform input marginal probabilities for the symmetric and isotropic channels. If we want to test the information causality, we should maximize the mutual information I for the more general quantum communication channels.

Recall that from the proof of section 3.3, we cannot formulate the problem of maximizing the mutual information I over the joint and the input marginal probabilities of the NS-box as a convex optimization programming. Thus, for more general quantum communication channels, we are forced to solve the problem by brutal force. The procedure is as follows. Firstly, we divide the defining domains of the joint and input marginal probabil-

ities into many fine points. We then pick up the points satisfying the consistent relations for a given communication channel(protocol). Secondly, we test if these joint probabilities can be reproduced by quantum mechanics or not. If they can, we then evaluate the corresponding mutual information I. Thirdly, by comparing these mutual information I, we can obtain the maximal one and then check if the information causality is satisfied or not. By this brutal force method, we can then obtain the distribution of mutual information I over the joint and the input marginal probabilities produced by quantum mechanics. This yields far more than just the maximal mutual information consistent with quantum mechanics. The price to pay is the cost for the longer computing time. Due to the restriction of the computer power, we can only work for $d = 2$ and $k = 2$ case.

We start the discussion of more general communication channels by fixing either the joint probabilities $\Pr(B_y - A_x|x, y)$ or the input marginal probabilities $\Pr(a_i)$. Firstly, we assume the input probabilities are i.i.d. and uniform such that we could take the CHSH function as the object of quantum communication complexity. Therefore we could study the relation between the mutual information and the quantum communication complexity. Note that, when requiring our communication channel (3.18) to have the i.i.d. and uniform input marginal probabilities, the channel between a_i and β then becomes symmetric automatically. Secondly, in order to study the influence of the input marginal probabilities $\Pr(a_i)$ on the mutual information, we pick up three sets of the joint probabilities $\Pr(B_y - A_x|x, y)$ constrained by quantum mechanics and then evaluate the corresponding mutual information with different input marginal probabilities $\Pr(a_i)$. Besides these communication channels, in order to test if the information causality is always satisfied, we will consider the most general quantum communication channel, namely, we do not impose any condition on the communication channel except the quantum constraints for joint probabilities.

Before evaluating the corresponding mutual information, the chosen joint probabilities $\Pr(B_y - A_x|x, y)$ should pass a test. For $d = 2$ and $k = 2$ RAC protocol, the quantum constraint is as follows:

$$G = \begin{pmatrix} 1 & \theta_1 & C_{00} & C_{01} \\ & 1 & C_{10} & C_{11} \\ & & 1 & \theta_2 \\ & & & 1 \end{pmatrix} \succeq 0, \tag{3.40}$$

where $C_{x,y} := (-1)^{xy}[2\Pr(B_y - A_x = xy|x,y) - 1]$ is the correlation function of the measurement setting x, y for Alice and Bob, respectively. The condition was pointed out in [56, 42, 44, 45] and can be derived as the necessary and sufficient condition for the quantum correlation functions $C_{x,y}$ (or equivalently the joint probabilities $\Pr(B_y - A_x|x,y)$) by Tsirelson's theorem [55], in which the marginal probabilities $\Pr(A_x|x)$ and $\Pr(B_y|y)$ are unbiased. Actually, G is the sub matrix of the $n = 1$ certificate $\Gamma^{(1)}$. Due to the positivity, (3.40) is satisfied once $\Gamma^{(1)} \succeq 0$.

Since the condition (3.40) is related to a positive semidefinite matrix, we need to use the numerical recipe to solve it. Once the joint probabilities are not fixed in the communication channel(protocol), we have to pick up many sets of joint probabilities from their defining domains. This seems not efficient enough to test all possible sets of joint probabilities by SDP. Therefore, instead of condition (3.40) we use a set of linear inequalities to test if the joint probabilities can be produced by quantum mechanics or not. In this way, the test will become simpler and more efficient. The linear inequalities are [43, 54]

$$|arcsin(C_{00}) + arcsin(C_{01}) + arcsin(C_{10}) - arcsin(C_{11})| \leq \pi, \tag{3.41a}$$

$$|arcsin(C_{00}) + arcsin(C_{01}) - arcsin(C_{10}) + arcsin(C_{11})| \leq \pi, \tag{3.41b}$$

$$|arcsin(C_{00}) - arcsin(C_{01}) + arcsin(C_{10}) + arcsin(C_{11})| \leq \pi, \tag{3.41c}$$

$$|-arcsin(C_{00}) + arcsin(C_{01}) + arcsin(C_{10}) + arcsin(C_{11})| \leq \pi. \tag{3.41d}$$

Actually, the condition (3.41) is equivalent to (3.40). If the linear inequalities (3.41) are satisfied, we then can find valid θ_1 and θ_2 to make condition (3.40) satisfied, and vise versa [42, 44, 45].

Once the corresponding correlation functions $C_{x,y}$ satisfy (3.41), we will know that these joint probabilities $\Pr(B_y - A_x|x,y)$ can be reproduced by quantum system. But we have to notice that some of them could also be expressed by the local hidden variable model. This means the shared correlation is local. Since the bound of communication complexity for local correlations is different from the quantum non-local ones, we could use the communication complexity to divide them.

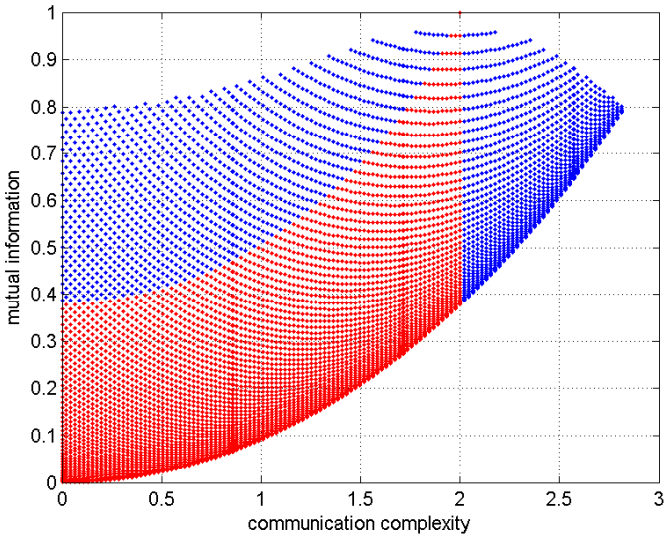

Figure 3.4: Mutual information v.s. (quantum) communication complexity for $d = 2$, $k = 2$ RAC protocol with i.i.d. and uniform input marginal probabilities. Here, the (quantum) communication complexity is characterized by the CHSH function. The red part can be achieved also by sharing the local correlation.

3.5.1 Symmetric channels with i.i.d. and uniform input marginal probabilities

We start with the most simple case: the $d = 2$, $k = 2$ RAC protocol with the symmetric channels and i.i.d., uniform input marginal probabilities. In this case, the successful probability for Bob to guess Alice's database right is equivalent to the CHSH function i.e., $|C_{0,0} + C_{0,1} + C_{1,0} - C_{1,1}|$. Thus, we could take the CHSH function as the object of communication complexity. Moreover, using the CHSH function and its three symmetric partners by shifting the minus sign, we could ensure that the shared correlations can be described by the local hidden variable model. Once the corresponding values of all these functions are less than 2, the shared correlation is local. Otherwise, the shared correlation could be quantum non-local or beyond. The latter happens when some of these values are larger than $2\sqrt{2}$ which is nothing but the Tsirelson bound. When the Tsirelson bound is reached, the quantum non-locality between two partite (Alice and Bob) is the maximum.

In our numerical calculations, we divide the defining domain of the joint probabilities $\Pr(B_y - A_x|x, y)$ into 100 points. Follow the procedure of our brutal force method, we obtain the distribution of the mutual information I over the quantum communication complexity as shown in Fig 3.4. for symmetric channels with i.i.d. and uniform input

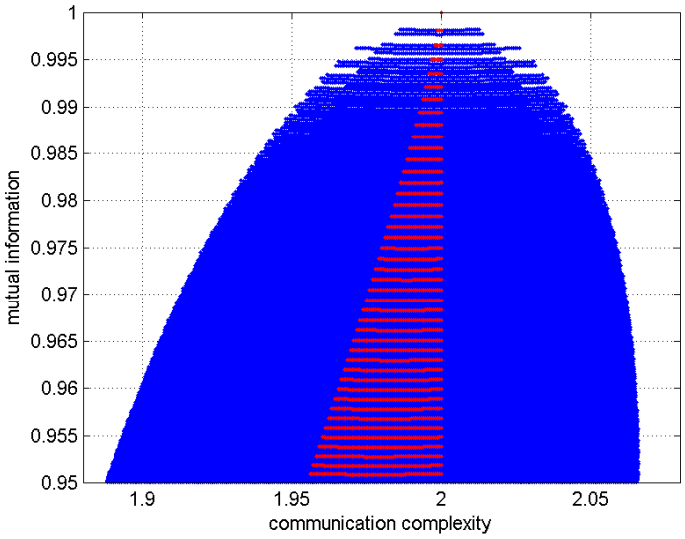

Figure 3.5: Some points near the top region in Fig. 3.4.

marginal probabilities. Note that, the quantum communication complexity here (x-axis of Fig 3.4) is characterized by the value of the CHSH function, $|C_{0,0} + C_{0,1} + C_{1,0} - C_{1,1}|$. In Fig 3.4, all the points satisfy quantum constraint (3.41). We particularly use the red color to denote the points which also can be obtained by the local correlations, i.e., the bound of CHSH inequality and its three symmetric partners are all less than 2. Moreover, it seems that the distribution of the mutual information over the quantum communication complexity as shown in Fig 3.4 is not continuous. This is not the case but because we did not partition the defining domain of the joint probability fine enough. In Fig 3.5 we partition more finely on the defining domain of the joint probability in the top region of Fig 3.4 and show that the empty region in Fig 3.4 is now filled. Similarly, the empty region on the top of Fig 3.5 could be filled again by the more fine partitioning.

The results in Fig 3.4 is consistent with the information causality since the maximal mutual information for the local or quantum correlations is bound by 1, the bound suggested by information causality. However, the peculiar part of Fig 3.4 is that some of the local correlations can achieve the larger mutual information than $I \simeq 0.8$, which is achieved by the correlations with the maximal quantum communication complexity. This peculiar part is the red region above $I \simeq 0.8$ in Fig 3.4. Especially, the maximal mutual information $I = 1$ is reached when the shared correlation is marginally non-local, i.e., the communication complexity is equal to 2. This indicates that the mutual information

is not monotonically related to the quantum communication complexity. Or put this in the other way, the more quantum non-locality may not always yield the more mutual information. We think it is interesting to understand this phenomenon in the future works.

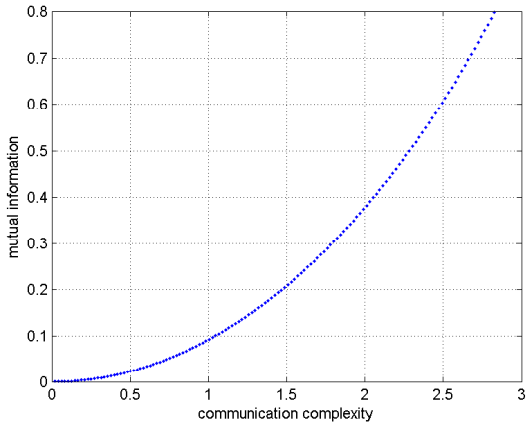

Figure 3.6: Mutual information vs (quantum) communication complexity for isotropic channels with i.i.d. and uniform input marginal probabilities.

Form these symmetric quantum channels with i.i.d. and uniform input marginal probabilities, we pick up the isotropic ones ($\xi_0 = \xi_1$) and obtain Fig 3.6. It shows that the mutual information I and the quantum communication complexity are monotonically related. This explicitly demonstrate what we have discussed in the previous section.

3.5.2 Channels with non-uniform input marginal probabilities

In the above communication channels, the input marginal probabilities are fixed to be i.i.d. and uniform. Now we would like to demonstrate the effect of non-uniform input marginal probabilities. In this case, we would like to vary the input marginal but keep the joint probabilities fixed. To see this effect for different channels, we consider three different sets of the joint probabilities corresponding to (i) symmetric, (ii) symmetric and isotropic and (iii) asymmetric channel.

To be more specific, for the case (i) the joint probabilities should be constrained by $\Pr(B_y - A_x = 0|x, y = 0) = 1$ and $\Pr(B_y - A_x = xy|x, y = 1) = \frac{1}{2}$ for $x = 0, 1$ such that the noise parameters are given by $\xi_0 = 1$ and $\xi_1 = 0$. For the case (ii) all the joint probabilities $\Pr(B_y - A_x = xy|x, y)$ are equal to $\frac{1}{2}(1 + \frac{1}{\sqrt{2}})$ such that $\xi_0 = \xi_1 = \frac{1}{\sqrt{2}}$. For

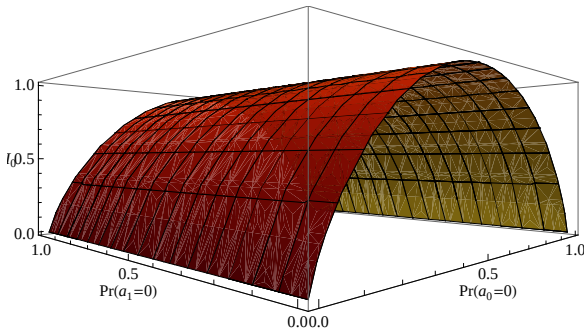

Figure 3.7: $I = I_0$ vs $\Pr(a_{0,1} = 0)$ for case (i).

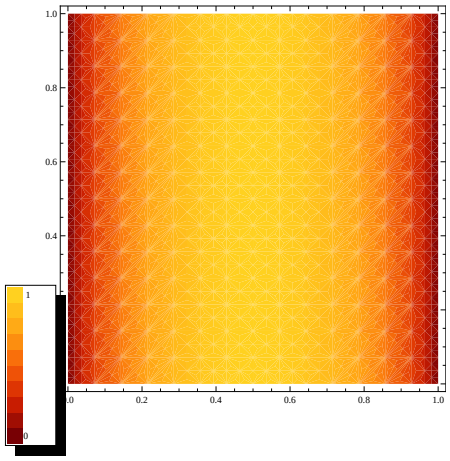

Figure 3.8: Density plot of the Left figure.

the case (iii) the joint probabilities are given by $\Pr(B_y - A_x = 0|x = 0, y) = \frac{1}{2}(1 + \frac{1}{\sqrt{2}})$ and $\Pr(B_y - A_x = xy|x = 1, y) = \frac{1}{2}$ for $y = 0, 1$. Obviously, it is asymmetric for general input marginal probabilities.

In the following discussion, we denote the mutual information $I(a_0; \beta|b = 0)$ as I_0 and $I(a_1; \beta|b = 1)$ as I_1, which are functions of two input marginal probabilities, namely, $\Pr(a_0 = 0)$ and $\Pr(a_1 = 0)$. Here I_i can be thought as the mutual information for the sub-channel between a_i and β, and the corresponding noise parameter is ξ_i. The mutual information I is just $I = I_0 + I_1$. Note that, I_0 does not depend on $\Pr(B_y - A_x = xy|x, y = 1)$ and I_1 not on $\Pr(B_y - A_x = 0|x, y = 0)$. Thus, the sub-channel for I_0 can be made symmetric by just requiring $\Pr(B_y - A_x = xy|x, y = 0)$'s for $x = 0, 1$ are equal, and similarly for the sub-channel for I_1 to be symmetric. An important feature for these symmetric channels is that I_i will depend only on $\Pr(a_i)$ not on $\Pr(a_{(i+1 \mod 2)})$.

For case (i), both sub-channels are symmetric. Moreover, since $\xi_0 = 1$ and $\xi_1 = 0$ so that the sub-channel for ξ_0 is noiseless and the one for ξ_1 is completely noisy. This then leads to $I_1 = 0$ and $I = I_0$. The dependence of $I = I_0$ on the input marginal probability, i.e., $\Pr(a_0 = 0)$ only, is shown in Fig 3.7-3.8. Note that I reaches its maximal value, 1 at $\Pr(a_0 = 0) = \frac{1}{2}$ as expected for the noiseless symmetric channel. This point is nothing but the point of maximal I in Fig 3.4. Note that this maximum saturates the bound by information causality. This implies that we can reach the causally allowed mutual information bound by sacrificing one of the sub-channel without any comprise. This is a bit surprising.

For case (ii), the channel is both symmetric and isotropic, we then expect that the

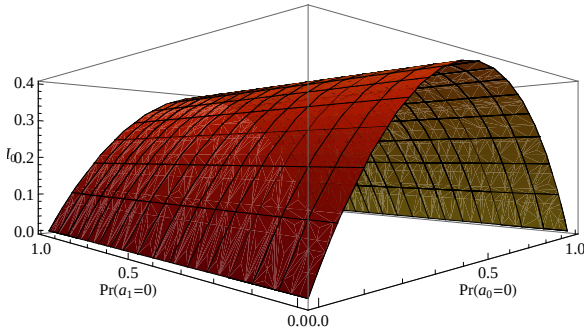

Figure 3.9: I_0 vs $\Pr(a_{0,1} = 0)$ for case (ii).

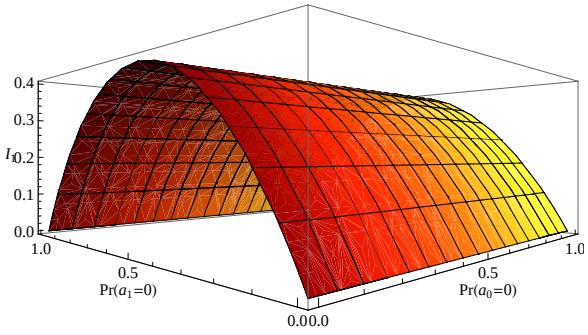

Figure 3.10: I_1 vs $\Pr(a_{0,1} = 0)$ for case (ii).

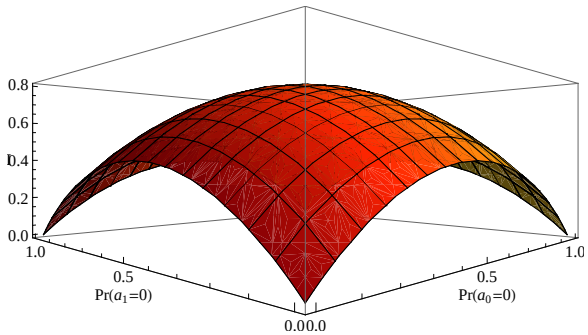

Figure 3.11: I vs $\Pr(a_{0,1} = 0)$ for case (ii).

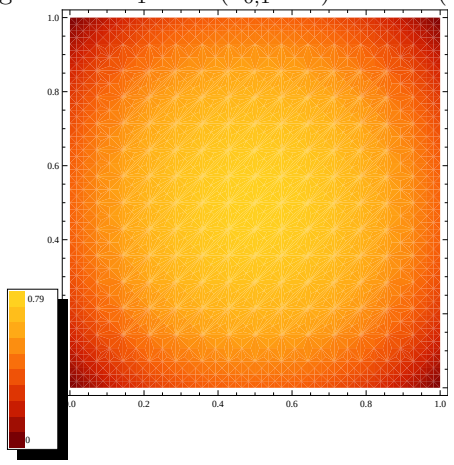

Figure 3.12: Density plot of the Left figure.

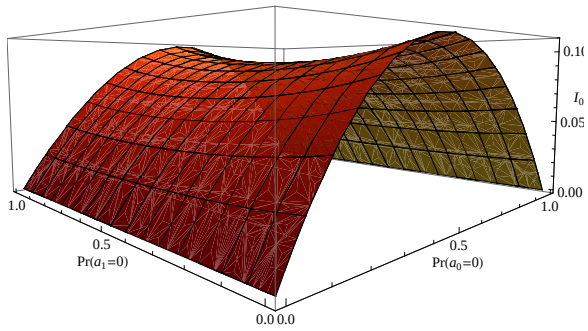

Figure 3.13: I_0 vs $\Pr(a_{0,1} = 0)$ for case (iii).

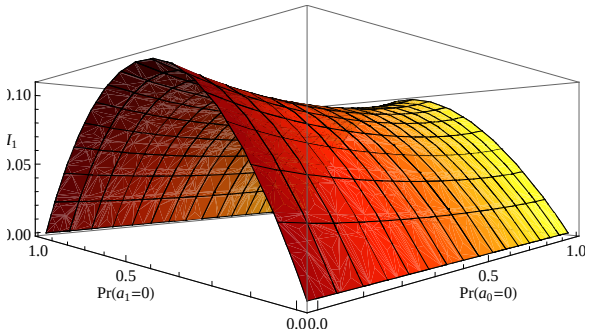

Figure 3.14: I_1 vs $\Pr(a_{0,1} = 0)$ for case (iii).

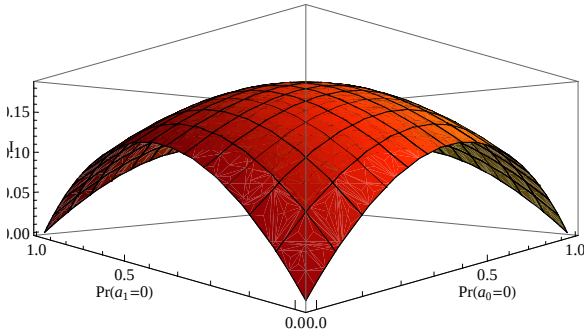

Figure 3.15: I vs $\Pr(a_{0,1} = 0)$ for case (iii).

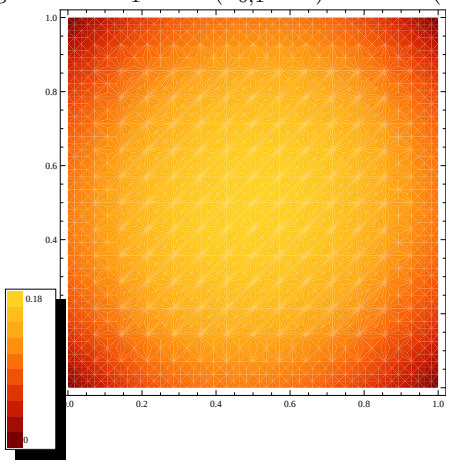

Figure 3.16: Density plot of the Left figure.

isotropy will also appear in the plot for I vs the input marginal probabilities, and that I_0 and I_1 will have the same shape. This is indeed the case as shown in Fig 3.9-3.12. Note that I_i only depends on $\Pr(a_i)$ though $I = I_0 + I_1$ depends on both. We see that the maximal value of I occurs at the symmetric point, i.e., all the $\Pr(a_i)$ equal to $\frac{1}{2}$. However, the maximal value is 0.7983 which is less than 1 of the information causality but is the same value for the case of maximal quantum communication complexity.

Finally, for case (iii), i.e., the particular asymmetric channel, I_i's are now dependent on both $\Pr(a_i)$'s unlike in the previous two cases. However, the mutual information I has the isotropic form as in the case (ii) but with a far smaller maximal value at the symmetric point. The results are shown in Fig 3.13-3.16.

Our above results implies that the the closer $\Pr(B_y - A_x = xy|x, y)$ to 1, the larger of the mutual information I. This is consistent with our RAC protocol as Bob can perfectly guess Alice's inputs by using the PR box [21]. Of course, the information causality ensures that the NS-box constrained by quantum mechanics can not be the PR box. Also, note that the maximum of I occurs at the symmetric point of the input marginal probabilities

for case (ii) and (iii) but it is not the case for case (i). Therefore, the uniform input marginal probabilities do not always lead to the maximal I.

3.5.3 Information causality for the most general channels

After testing the information causality for the more general channels as discussed in the previous sections, we would wonder if the information causality holds for the most general channels or not, i.e., the channels without any additional constraint on the joint probabilities and the input marginal probabilities except the necessary quantum and no-signaling constraints. For our $d = 2$, $k = 2$ RAC protocol, we check this by partitioning the defining domains of the probabilities into 100 points and then using the brutal force methods to do the numerical check. We find that the information causality is always satisfied. This yields a more general support for the information causality.

Furthermore, we find that the information causality is saturated, i.e., $I = 1$ when one of the sub-channel is noiseless and the other one is completely noisy. This is similar to the case (i) discussed in the previous subsection.

3.6 Summary

Information causality was proposed as a new physical principle and gives an intuitive picture on the meaning of causality from the information point of view. Therefore, to test its validity for different settings will help to establish it as a physical principle. Motivated by this, in this work we try our best to extend the framework of the original proposal to the more general communication protocols, such as the multi-level and multi-setting or removing the conditions of symmetric channel or uniform input marginal probabilities. We then test the information causality for these general protocols by either adopting the SDP for numerical check, or using the brutal force method for the more general communication channels. With all these efforts, our results are rewarding: we see that the information causality are preserved in all the protocols discussed in this work. This reinforce the validity of the information causality further than before. Though more checks for more general protocols should be always welcome. We also find that the information causality is saturated not by sharing the correlations with the maximal quantum non-locality, but by the ones which are marginally non-local. This then raises the issues on the intimate relation between the shared mutual information and the quantum non-locality. Especially,

this result challenges our intuition that a channel can transfer more information with more quantum non-local resources. We think our findings in this work will shed some light on the related topics.

Chapter 4

Conclusion

Information causality was proposed as the physical principle to single out quantum mechanics. In order to have more understanding about it, we study the constraints on the information causality and test the information causality for the different communication protocols.

In Chapter 2, we combine the information causality and the signal decay theorem and then obtain a series of Tsirelson-type inequalities for two-level and bi-partite quantum systems. Moreover, according to the theory of noisy computation, the constraint on the information causality leads to a large computational noise and therefore the reliable computation cannot be achieved.

In Chapter 3, in order to have more confidence of treating the information causality as the physical principle, we test the information causality for more general quantum communication protocols such as multi-level and multi-setting ones or the asymmetric communication channels. Specifically, we calculate the maximal mutual information shared between the sender and the receiver for more these general quantum communication protocols, and compare it to the bound from the information causality. For the quantum communication protocols discussed in this work, the information causality is never violated. Thus, the information causality is supported and could be treated as a physical principle to single out quantum mechanics.

Moreover, for the two-inputs/two-outputs case, we also find that the information causality is saturated not for the channels with the maximal quantum non-locality associated with the Tsirelson bound but for the marginal cases saturating local bound of the CHSH inequality. This means that, sharing more non-local correlation does not imply the better performance in the communication. Thus, one may ask what is the essential ingredient for the efficient quantum communication protocols.

So far, the multi-partite information causality is still unclear. In the above works, we use the semidefinite programming to characterize the bi-partite quantum correlations.

This work can be generalized to the multi-partite cases. Assuming the information causality can single out the quantum correlations, the maximal mutual information bound by the information causality is then equal to the one over all quantum communication protocols. One may then use the semidefinite programming to bound the multi-partite quantum correlations and find the associated constraint for the multi-partite information causality.

Appendix A

Signal decay and data processing inequality for multi-nary channels

In this appendix, we will first sketch the key steps of [50] in obtaining the maximal bound on the signal decay for the binary noisy channels, and then generalize this derivation to the one for the multi-nary channels.

Our setup is to consider a cascade of two communication channels: $X \to Y \to Z$. The decay of the signal is implied by the data processing inequality, i.e.,

$$I(X;Z) \leq I(X;Y). \tag{A.1}$$

The mutual information $I(X;Y) = H(Y) - \sum_i \Pr(X=i) H(Y|X=i)$, where $H(Y)$ and $H(Y|X)$ are the Shannon entropies for the probability $\Pr(Y)$ and the conditional probability $\Pr(Y|X)$, respectively.

Furthermore, for the binary symmetric channel A characterized by

$$A = \begin{pmatrix} \frac{1+\xi}{2} & \frac{1-\xi}{2} \\ \frac{1-\xi}{2} & \frac{1+\xi}{2} \end{pmatrix}, \tag{A.2}$$

it was shown in [50] that the bound on the signal decay is characterized by the following bound

$$\frac{I(X;Z)}{I(X;Y)} \leq \xi^2. \tag{A.3}$$

Note that this bound is tighter than the one obtained in [48], which is $\frac{I(X;Z)}{I(X;Y)} \leq \xi$.

In this appendix, we will generalize the above result to the one for the dinary channel characterized by $\Pr(Z=i|Y=i) = \xi$ and $\Pr(Z=s\neq i|Y=i) = \frac{1-\xi}{d-1}$ with $i \in \{0,1,...,d-1\}$, so that the signal decay is bound by

$$\frac{I(X;Z)}{I(X;Y)} \leq (\xi - \frac{1-\xi}{d-1})^2. \tag{A.4}$$

A.1 Sketch of the proof in [50]

The derivation in [50] consists of two key steps. The first one is to show the following theorem for weak signal:

Theorem I: The ratio $\frac{I(X;Z)}{I(X;Y)}$ *reaches its maximum if the conditional probabilities* $\Pr(Y|X=0)$ *and* $\Pr(Y|X=1)$ *are almost indistinguishable, i.e.,* $|\Pr(Y=0|X=0)-\Pr(Y=0|X=1)| \to 0$.

To prove this theorem we need the following lemma:

Lemma I: For any strictly concave function f *and* g *on the interval* $[0,1]$*, and any* $p \in [0,1]$*, the ratio*

$$r(x,y) = g_2(x,y,p)/f_2(x,y,p) \tag{A.5}$$

reaches its maximum in the limit $|x-y| \to 0$*. Here* $f_2(x,y,p) = f(px+(1-p)y) - pf(x) - (1-p)f(y)$ *denotes the second order difference of the function* f *with the weight* p*, and similarly for the* $g_2(x,y,p)$*.*

We sketch the proof of this lemma, which will be useful when generalizing to the multinary channel. We assume that the ratio r reaches its maximum at $x = x^*$ and $y = y^*$, and for concreteness assuming $x^* < y^*$. Note that $0 < r < \infty$ due to the concave f and g. We can perform affine transformation to scale this maximal value of $r(x^*, y^*, p)$ to be 1, and also to make $f(x^*) = g(x^*)$ and $f(y^*) = g(y^*)$. This immediately leads to $f(px^* + (1-p)y^*) = g(px^* + (1-p)y^*)$. That is, there is a point $z^* = px^* + (1-p)y^*$ inside the interval $[x^*, y^*]$ at which f also equals to g. Use this fact, it is easy to convince oneself that either $r(z^*, y^*) \geq r(x^*, y^*)$ or $r(x^*, z^*) \geq r(x^*, y^*)$. For more subtle details, please see [50]. By repeating this procedure we prove the lemma.

Observe that $I(X;Y)$ and $I(X;Z)$ are the second order difference of the (concave) entropy functions $H(Y)$ and $H(Z)$, respectively with the weight $p = \Pr(X=0)$. We can then prove the Theorem I by the above lemma.

The second step is first to rewrite the ratio $\frac{I(X;Z)}{I(X;Y)}$ in terms of relative entropy $D(p\|q) := \sum_x \Pr(p=x) log \frac{\Pr(p=x)}{\Pr(q=x)}$, that is,

$$\frac{I(X;Z)}{I(X;Y)} = \frac{\sum_{i=0}^{1} \Pr(X=i) D(\Pr(Y|X=i)\cdot A \| \Pr(Y)\cdot A)}{\sum_{i=0}^{1} \Pr(X=i) D(\Pr(Y|X=i) \| \Pr(Y))}. \tag{A.6}$$

Then, based on the above theorem we can parameterize the conditional probability $\Pr(Y|X=0) = \vec{p} + \vec{\epsilon}$ where $\vec{p} = \sum_{i=0}^{1} \Pr(X=i)\Pr(Y|X=i)$ and $\vec{\epsilon} = (\epsilon, -\epsilon)$ with ϵ being sufficiently small. With this condition, (A.6) can be simplified to

$$\frac{I(X;Z)}{I(X;Y)} \approx \frac{D((\vec{p}+\vec{\epsilon})\cdot A \parallel \vec{p}\cdot A)}{D(\vec{p}+\vec{\epsilon} \parallel \vec{p})}. \tag{A.7}$$

Note that the ratio now does not depend on $\Pr(X)$.

Finally, given the binary channel (A.2) we can expand the relative entropy in terms of $\epsilon/\Pr(Y)$, so for the ratio $\frac{I(X;Z)}{I(X;Y)}$. Then, fixing ϵ and then varying the first order term of the ratio $\frac{I(X;Z)}{I(X;Y)}$ in the above expansion over $\vec{p}$, we obtain the bound in (A.3).

A.2 Generalizing to the multi-nary channels

We now generalize the above derivation to the trinary noisy channels, then the generalization to the dinary channel will just follows. The key steps are similar to the binary ones. The first step is to use the same method to prove the following theorem:

Theorem II: The ratio $\frac{I(X;Z)}{I(X;Y)}$ *reaches its maximum only when all the three conditional probabilities* $\Pr(Y|X=i)$ *with* $i=0,1,2$ *are almost indistinguishable.*

The strategy to prove this theorem is to observe that we can treat the pair $(\Pr(Y=0|X=i), \Pr(Y=1|X=i))$ for each i (note that $\Pr(Y=2|X=i)$ is not independent of this pair) as a point inside the unit square $([0,1],[0,1])$. Then the three points $\Pr(Y|X=i)$ for $i=0,1,2$ form a triangle. We can then follow the same way of proving the Lemma I in the previous subsection for the trinary case. First, we assume the maximal value of r occurs at all three vertices of some triangle. We then perform the affine transformation to rescale this maximal value to 1, and to make $f=g$ (or more specifically $H(Y|X=i) = H(Z|X=i)$) at the three vertices of the above triangle. This then immediately leads to that there exists some point inside the triangle such that $f=g$. We can use this point to construct a smaller triangle with any two of the vertices of the original triangle and show that the ratio r for this new triangle is greater than the one for the original larger triangle. Repeating this procedure we can prove the above theorem. It is also clear that we can generalize the theorem for the multi-nary channels by generalizing the triangle to the concave body of the higher dimensional space.

Here, we should point out that one can always reduce the concave body to the linear interval one, so that we can reduce to the situation for the binary case. That is, we

set all the conditional probabilities except one to be equal, and then study the closeness condition of the remaining two distinct conditional probabilities for the maximal ratio of $\frac{I(X;Z)}{I(X;Y)}$. In the following, we will always restrict to such a situation.

We then go to the second step as for the binary channel, that is to use Theorem II to reduce the problem of maximizing $\frac{I(X;Z)}{I(X;Y)}$ to the one of maximizing the ratio of relative entropies. We rewrite the ratio of two mutual information as following,

$$\frac{I(X;Z)}{I(X;Y)} = \frac{\sum_{i=0}^{2} \Pr(X=i) D(\Pr(Y|X=i)\cdot A \| \Pr(Y)\cdot A)}{\sum_{i=0}^{2} \Pr(X=i) D(\Pr(Y|X=i) \| \Pr(Y))}. \tag{A.8}$$

To simplify the expression for further manipulations, we denote the average probability of Y as $\vec{p} = \sum_{i=0}^{2} \Pr(X=i) \Pr(Y|X=i)$, and parameterize the probability $\Pr(Y|X=0) = \vec{p} + \vec{\epsilon}_0$ and $\Pr(Y|X=1) = \vec{p} + \vec{\epsilon}_1$. Thus, the probability $\Pr(Y|X=2)$ is forced to be $\vec{p} - \frac{\Pr(X=0)}{\Pr(X=2)}\vec{\epsilon}_0 - \frac{\Pr(X=1)}{\Pr(X=2)}\vec{\epsilon}_1$. The parameter vectors $\vec{\epsilon}_0$ and $\vec{\epsilon}_1$ should be sufficiently small as required by Theorem II to have maximal ratio $\frac{I(X;Z)}{I(X;Y)}$. Furthermore, we will further reduce the triangle to the linear interval case by assuming $\vec{\epsilon}_0 = \vec{\epsilon}_1$, i.e., $\Pr(Y|X=0) = \Pr(Y|X=1)$.

The ratio (A.8) then becomes

$$\frac{I(X;Z)}{I(X;Y)} \approx \frac{D((\vec{p}+\vec{\epsilon_0})\cdot A \parallel \vec{p}\cdot A)}{D(\vec{p}+\vec{\epsilon_0} \parallel \vec{p})}. \tag{A.9}$$

Note again the ratio now does not depend on $\Pr(X)$.

Before serious expansion of (A.9) in the power of $\vec{\epsilon}_0$, we need to specify $\vec{p} = \left(\Pr(Y=0),\Pr(Y=1),\Pr(Y=2)\right)$ and $\vec{\epsilon}_0 = (v_0,v_1,v_2)$. Note that, $v_0+v_1+v_2 = 0$. As for the bi-nary channel, we expand the relative entropy in terms of $\frac{v_i}{\Pr(Y=i)}$. The leading term of the expansion for the denominator of (A.9) is found to be

$$D(\vec{p}+\epsilon_0 \parallel \vec{p}) = \frac{1}{2ln2} \sum_{i=0}^{1} \frac{v_i^2}{\Pr(Y=i)}. \tag{A.10}$$

To find the expansion of the numerator, we need to specify the channel A between Y and Z. The generic trinary channel is given by

$$A = \Pr(Z|Y) = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}, \tag{A.11}$$

where the elements of the channel should satisfy $a_1+a_2+a_3 = 1$, $b_1+b_2+b_3 = 1$, and $c_1+c_2+c_3 = 1$. Then, the leading term in the expansion of the numerator of (A.9) is

found to be

$$D((\vec{p}+\vec{\epsilon_0})\cdot A \parallel \vec{p}\cdot A) = \frac{1}{2ln2}(\frac{v_0a_1+v_1b_1+v_2c_1}{p(Z=0)} + \frac{v_0a_2+v_1b_2+v_2c_2}{p(Z=1)} + \frac{v_0a_3+v_1b_3+v_2c_3}{p(Z=2)}). \quad \text{(A.12)}$$

For simplicity, we only consider the symmetry trinary channel as follows

$$A = \Pr(Z|Y) = \begin{pmatrix} \xi & \frac{1-\xi}{2} & \frac{1-\xi}{2} \\ \frac{1-\xi}{2} & \xi & \frac{1-\xi}{2} \\ \frac{1-\xi}{2} & \frac{1-\xi}{2} & \xi \end{pmatrix}. \quad \text{(A.13)}$$

Then, (A.12) then becomes

$$D((\vec{p}+\vec{\epsilon_0})\cdot A \parallel \vec{p}\cdot A) = (\frac{3\xi-1}{2})^2 \frac{1}{2ln2}\sum_{i=0}^{2}\frac{v_i^2}{\Pr(Z=i)}. \quad \text{(A.14)}$$

Since we know that for symmetric channel, the maximal mutual information is achieved for uniform input probabilities. Thus, we assume uniform $\Pr(Y)$ and $\Pr(Z)$ so that (A.9) depends only on variable ξ. We then obtain

$$\frac{I(X;Z)}{I(X;Y)} \leq (\frac{3\xi-1}{2})^2. \quad \text{(A.15)}$$

This is the generalization of (A.3) for binary channel to the trinary one.

Similarly, we can generalize the above derivation to the dinary channels. If the channel between Y and Z is a dinary and symmetry channel specified as follows: $\Pr(Z=i|Y=i)=\xi$ and $\Pr(Z=s\neq i|Y=i)=\frac{1-\xi}{d-1}$ with $i\in\{0,1,...,d-1\}$, then the bound of the ratio $\frac{I(X;Z)}{I(X;Y)}$ is given by (A.4).

Appendix B

The concavity of mutual information

In this appendix, we want to prove the mutual information I is not a concave function to joint probabilities $\Pr(B_{\vec{y}} - A_{\vec{x}}|\vec{x},\vec{y})$ and input marginal probabilities $\Pr(a_i)$. Thus, we could not formulate the problem (maximizing mutual information I) as a convex optimization programming.

First, we reexpress mutual information I by $\Pr(B_{\vec{y}} - A_{\vec{x}}|\vec{x},\vec{y})$ and $\Pr(a_i)$. If maximizing mutual information is a concave function to these probabilities, the second order partial derivative of mutual information respecting to each probability should be negative. Here, we find a violation when calculating $\frac{\partial^2 I}{\partial(\Pr(B_{\vec{y}} - A_{\vec{x}}=0|\vec{x}=0,\vec{y}=0))^2}$. In following paragraphs, we denote the joint probability $\Pr(B_{\vec{y}} - A_{\vec{x}} = 0|\vec{x} = 0, \vec{y} = 0)$ as V.

The mutual information can be rewritten as

$$I = \sum_{i=0}^{k-1} I_{b=i}, \tag{B.1}$$

where $I_{b=i}$ is equal to $I(a_i;\beta|b=i)$. Since the joint probability V only contribute to $I_{b=0}$, we only need to calculate $\frac{\partial^2 I_{b=0}}{\partial V^2}$. The reexpression of $I_{b=0}$ is

$$I_{b=0} = \sum_{n=0}^{d-1}\sum_{j=0}^{d-1} \Pr(\beta = n, a_0 = j|b=0) log_2 \frac{\Pr(\beta = n, a_0 = j|b=0)}{\Pr(\beta = n|b=0)\Pr(a_0 = j|b=0)}. \tag{B.2}$$

Therefore, the first order partial derivative respecting to $\Pr(B_{\vec{y}} - A_{\vec{x}} = 0|\vec{x} = 0, \vec{y} = 0)$ is

$$\begin{aligned}
&\frac{\partial I_{b=0}}{\partial V} = \\
&\sum_{n=0}^{d-1}\sum_{j=0}^{d-1} \frac{\partial \Pr(a_0 = j, \beta = n|b=0)}{\partial V} log_2 \frac{\Pr(a_0 = j, \beta = n|b=0)}{\Pr(\beta = n|b=0)\Pr(a_0 = j|b=0)} \\
&+\frac{1}{ln2}\left(\frac{\partial \Pr(a_0 = j, \beta = n|b=0)}{\partial V} - \frac{\Pr(a_0 = j, \beta = n|b=0)}{\Pr(\beta = n|b=0)} \frac{\partial \Pr(\beta = n|b=0)}{\partial V}\right)
\end{aligned} \tag{B.3}$$

We can express $\Pr(a_0 = j, \beta = n|b = 0)$ as the combination of joint probabilities $\Pr(B_{\vec{y}} - A_{\vec{x}}|\vec{x},\vec{y})$ and input marginal probabilities $\Pr(a_i)$ to obtain $\frac{\partial \Pr(a_0=j,\beta=n|b=0)}{\partial V}$. Since joint probabilities $\Pr(B_{\vec{y}} - A_{\vec{x}}|\vec{x},\vec{y})$ are subjected to the normalization conditions

of total probability, if $n - j \neq (d-1)$,

$$\Pr(a_0{=}j,\beta{=}n|b{=}0){=}\textstyle\sum_{a_{k\neq 0}} \Pr(B_{\vec{y}} - A_{\vec{x}}{=}n-j|\vec{x},\vec{y}{=}0)\Pr(a_0{=}j)\,\Pi_{k\neq 0}\Pr(a_k); \tag{B.4}$$

if $n - j = (d-1)$,

$$\Pr(a_0{=}j,\beta{=}n|b{=}0){=}\textstyle\sum_{a_{k\neq 0}}(1-\sum_{t=0}^{d-2}\Pr(B_{\vec{y}} - A_{\vec{x}}{=}t|\vec{x},\vec{y}{=}0))\Pr(a_0{=}j)\,\Pi_{k\neq 0}\Pr(a_k), \tag{B.5}$$

where $\vec{x}$ in the above functions is given by the encoding of the RAC protocol, namely, $\vec{x} := (x_1, \cdots, x_{k-1})$ with $x_i = a_i - a_0$

Now, we can calculate the derivatives. The patrial derivative

$$\frac{\partial \Pr(a_0 = j, \beta = n|b = 0)}{\partial V} \tag{B.6}$$

is not equal to zero for two cases, the first one is $j = n$, we can obtain $\Pi_k \Pr(a_k = n)$ for (B.6). The second case is $n - j = (d-1)$, we can obtain $-\Pi_k \Pr(a_k = n - (d-1))$. Therefore, since $\Pr(\beta = n|b = 0) = \sum_j \Pr(a_0 = j, \beta = n|b = 0)$, we can obtain

$$\frac{\partial \Pr(\beta = n|b = 0)}{\partial V} \quad = \Pi_k \Pr(a_k = n) - \Pi_k \Pr(a_k = n - (d-1)). \tag{B.7}$$

Put above result to (B.3), for fixed j, we can find that $\sum_{n=0}^{d-1} \frac{\partial \Pr(a_0{=}j,\beta{=}n|b{=}0)}{\partial V} = 0$, thus the second term of (B.3) will vanish.

We then can calculate the second order derivative

$$\begin{aligned}\tfrac{\partial^2 I_{b=0}}{\partial V^2} &= \tfrac{1}{ln2}\textstyle\sum_{n=0}^{d-1}\sum_{j=0}^{d-1}(\frac{\partial \Pr(a_0{=}j,\beta{=}n|b{=}0)}{\partial V})^2 \frac{1}{\Pr(a_0{=}j,\beta{=}n|b{=}0)} \\ -\tfrac{2}{\Pr(\beta{=}n|b{=}0)}&\tfrac{\partial \Pr(a_0{=}j,\beta{=}n|b{=}0)}{\partial V}\tfrac{\partial \Pr(\beta{=}n|b{=}0)}{\partial V} + (\tfrac{\partial \Pr(\beta{=}n|b{=}0)}{\partial V})^2 \tfrac{\Pr(a_0{=}j,\beta{=}n|b{=}0)}{(\Pr(\beta{=}n|b{=}0))^2}\end{aligned} \tag{B.8}$$

For $d = 2$ and $k = 2$, (B.8) becomes

$$\begin{aligned}\frac{\partial^2 I}{\partial V^2} &= \frac{1}{ln2}[(\Pr(a_0 = 0)\Pr(a_1 = 0))^2(\frac{1}{\Pr(a_0 = 0, \beta = 0|b = 0)} + \frac{1}{\Pr(a_0 = 0, \beta = 1|b = 0)}) \\ &+(\Pr(a_0 = 1)\Pr(a_1 = 1))^2(\frac{1}{\Pr(a_0 = 1, \beta = 0|b = 0)} + \frac{1}{\Pr(a_0 = 1, \beta = 1|b = 0)}) \\ &-(\frac{1}{\Pr(\beta = 0|b = 0)} + \frac{1}{\Pr(\beta = 1|b = 0)})(\Pr(a_0 = 0)\Pr(a_1 = 0) - \Pr(a_0 = 1)\Pr(a_1 = 1))^2]\end{aligned} \tag{B.9}$$

Once $\Pr(a_0 = 0) = 1 - \Pr(a_1 = 0)$, the above function is non-negative.

For higher d and k, once the input marginal probabilities $\Pr(a_i)$ are uniform. We then

can obtain

$$\frac{\partial^2 I}{\partial V^2} = \frac{\partial^2 I_{b=0}}{\partial V^2} =$$
$$\frac{1}{ln2}\sum_{n=0}^{d-1}\frac{1}{d^{2k}}\left(\frac{1}{\Pr(a_0 = n, \beta = n|b=0)} + \frac{1}{\Pr(a_0 = n, \beta = n-(d-1)|b=0)}\right)$$
$$> 0 \tag{B.10}$$

It is clear that mutual information I is not a concave function to joint probabilities $\Pr(B_{\vec{y}} - A_{\vec{x}}|\vec{x}, \vec{y})$ and input marginal probabilities $\Pr(a_i)$.

Appendix C

Semidefinite programming

In this appendix, we briefly introduce the semidefinite programming (SDP) [61]. SDP is the problem of optimizing a linear function subjected to certain conditions associated with a positive semidefinite matrix X, i.e., $v^\dagger X v \geq 0$, for $v \in \mathbb{C}^n$, and is denoted by $X \succeq 0$. It can be formulated as the standard primal problem as follows. Given the $n \times n$ symmetric matrices C and D_q's with $q = 1, \cdots, m$, we like to optimize the $n \times n$ positive semidefinite matrix $X \succeq 0$ such that we can achieve the following:

$$\begin{aligned} &minimize && Trace(C^T X) && \text{(C.1a)} \\ &subject \quad to && Trace(D_q^T X) = b_q, \quad q = 1, \cdots, m\ . && \text{(C.1b)} \end{aligned}$$

Corresponding to the above primal problem, we can obtain a dual problem via a Lagrange approach [64]. The Lagrange duality can be understood as the following. If the primal problem is

$$\begin{aligned} &minimize && f_0(x) && \text{(C.2a)} \\ &s.t. && f_q(x) \leq 0, \quad q \in 1...m. && \text{(C.2b)} \\ & && h_q(x) = 0, \quad q \in 1...p, && \text{(C.2c)} \end{aligned}$$

the Lagrange function can be defined as

$$L(x, \lambda, \nu) = f_0(x) + \Sigma_{q=1}^m \lambda_q f_q(x) + \Sigma_{q=1}^p \nu_q h_q(x), \tag{C.3}$$

where $\lambda_1, \ldots, \lambda_m$, and $\nu_1, \ldots, \nu_p$ are Lagrange multipliers respectively. Due to the problem and (C.3), the minima of f_0 is bounded by (C.3) under the constraints when $\lambda_1, \ldots, \lambda_m \geq 0$.

$$\inf_x f_0 \geq \inf_x L(x, \lambda, \nu).$$

Then the Lagrange dual function is obtained.

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu).$$

$g(\lambda, \nu) \leq p$ (p is the optimal solution of $f_0(x)$), for $\lambda_1, \ldots, \lambda_m \geq 0$ and arbitrary $\nu_1, \ldots, \nu_p$. The dual problem is defined.

$$\begin{aligned} maximize \quad & g(\lambda, \nu) && \text{(C.4a)} \\ s.t. \quad & \lambda_q \geq 0. \quad (q \in \{1...m\}) && \text{(C.4b)} \end{aligned}$$

We can use the same method to define the dual problem for SDP. From the primal problem of SDP (C.1), we can write down the dual function by using minimax inequality [65].

$$\begin{aligned} \inf_{X \succeq 0} Trace(C^T X) \quad & = \inf_{X \succeq 0} Trace(C^T X) + \textstyle\sum_{q=1}^{m} y_q (b_q - Trace(D_q^T X)) \\ & = \inf_{X \succeq 0} \sup_y \textstyle\sum_{q=1}^{m} y_q (b_q) + Trace((C^T - \sum_{q=1}^{m} y_q D_q^T) X) \\ & \geq \sup_y \inf_{X \succeq 0} \textstyle\sum_{q=1}^{m} y_q (b_q) + Trace((C^T - \sum_{q=1}^{m} y_q D_q^T) X) \\ & = \sup_y \inf_{X \succeq 0} \textstyle\sum_{q=1}^{m} y_q (b_q) + Trace((C - \sum_{q=1}^{m} y_q D_q)^T X). \end{aligned} \quad \text{(C.5)}$$

The optimal solution of dual function is bounded under some vector y.

$$\sup_y \inf_{X \succeq 0} \sum_{q=1}^{m} y_q (b_q) + \mathrm{Tr}((C - \sum_{q=1}^{m} y_q D_q)^T X) = \begin{cases} \sup_y \sum_{q=1}^{m} y_q (b_q) & ; when \quad C - \sum_{q=1}^{m} y_q D_q \succeq 0 \\ -\infty & ; otherwise. \end{cases}$$

The correspond dual problem is

$$\begin{aligned} maximize \quad & \sum_{q=1}^{m} y_q (b_q) && \text{(C.6a)} \\ s.t. \quad & S = C - \sum_{q=1}^{m} y_q D_q \succeq 0. && \text{(C.6b)} \end{aligned}$$

If the feasible solutions for the primal problem and the dual problem attain their minimal and maximal values denoted as p' and d' respectively, then $p' \geq d'$, which is called the duality gap. This implies that the optimal solution of primal problem is bounded by dual problem. This then leads to the following: Both the primal and the dual problems attain their optimal solutions when the duality gap vanishes, i.e., $d' = p'$.

Appendix D

The Tsirelson-type inequality derived from the information causality

In this appendix, we write down the detail of getting the Tsirelson-type inequality derived from IC. We review the the RAC protocol as follows. Alice has a database of k bits $a_0, a_1, , , a_{k-1}$ where $a_i \in \{0, 1\}$ is the random variable $\forall i \in (0, \cdots, k-1)$. The distant Bob is given a random variable $b \in (0, ...k-1)$ and a bit α sent by Alice. Bob's task is to guess a_b. Here we will consider the RAC protocol with different settings. Case (a) is proposed in the main text. In case (b), Alice's and Bob's settings are modified. In the following, Alice's input is denoted by an N-bit string $\vec{x} = x_1 \dots x_N$. Let $x = 1 + \sum_{i=1}^{N} 2^{i-1} x_i$, $1 \leq x \leq 2^N$. Bob's input is denoted by N- bit string $\vec{y} = y_1 \dots y_N$.

Case (a)

Here $N = k-1$, and $x_i = a_0 + a_i \ \forall i \in \{1, ..., k-1\}$. $y_i = \delta_{i,b} \ \forall i \in \{1, ..., k-1\}$, if $b \neq 0$. $\vec{y} = \vec{0}$ if $b = 0$. Let $y = 1 + \sum_{i=1}^{N} i y_i$, $1 \leq y \leq k$. In this case, the Tsirelson-type inequality derived from information causality following the procedure in the chapter 2. is

$$| \sum_{\{\vec{x}\},\{\vec{y}\}} (-1)^{\vec{x}\cdot\vec{y}} C_{\vec{x},\vec{y}}| \leq 2^{k-1}\sqrt{k}. \tag{D.1}$$

Case (b)

Here $N = k$, $x_i = a_{i-1}$, and $y_i = \delta_{i,b+1} \ \forall i \in \{1, ..., k\}$. Let $y = \sum_{i=1}^{N} i y_i$, $1 \leq y \leq k$. Then, the Tsirelson-type inequality from information causality is

$$| \sum_{\{\vec{x}\},\{\vec{y}\}} (-1)^{\vec{x}\cdot\vec{y}} C_{\vec{x},\vec{y}}| \leq 2^{k}\sqrt{k}. \tag{D.2}$$

D.1 Checking the Tsirelson-type bound by semidefinite programming

We now use SDP to check the Tsirelson-type bound. To cast the above problem of finding the Tsirelson's bound in the context of quantum mechanics, we need to use Tsirelson's theorem [55]. It says that for any quantum state $|\Psi\rangle \in \mathbb{A} \bigotimes \mathbb{B}$ shared by two observers Alice and Bob with their measurement outcomes being $A_x \in [-1, 1]$ and $B_y \in [-1, 1]$, respectively. The correlation function can be expressed by the inner product of two real unit vectors α_x, $\beta_y \in \mathbb{R}^{t+v}$. Therein, t and v are the numbers of Alice's and Bob's measurement settings, respectively. In detail, $C_{\vec{x},\vec{y}}$ used in (D.1) or (D.2), the Tsirelson's theorem guarantees that we have $C_{\vec{x},\vec{y}} = \alpha_x \cdot \beta_y$. Then, we can cast the problem of finding the Tsirelson bound in (D.1) or (D.2) into the following form of optimal problem for SDP,

$$maximize \qquad | \sum_{\{\vec{x}\},\{\vec{y}\}} (-1)^{\vec{x}\cdot\vec{y}} \alpha_x \cdot \beta_y | \tag{D.3a}$$

$$s.t. \qquad \|\alpha_x \| = \|\beta_y \| = 1 \, , \qquad \forall \, x, y. \tag{D.3b}$$

Then, the associated dual problem is

$$minimize \qquad \sum_{q=1}^{m} y_q \tag{D.4a}$$

$$s.t. \qquad S = \sum_{q=1}^{m} y_q D_q - C \succeq 0. \tag{D.4b}$$

We now will turn the problem (D.3) into the primal problem (C.1) by constructing the matrices X, C and A_i's from the unit vectors α_x and β_y. Following the way in [56], the mapping is as follows. Define the matrix P whose columns are vectors $(\alpha_1, ..., \alpha_t, \beta_1, ...\beta_v)$. Then the SDSP matrix X is given by $P^T P$, which can be put into the following block form

$$X = \begin{pmatrix} E & F \\ G & H \end{pmatrix}$$

where the matrix elements of each block are $E_{ij} = \alpha_i \cdot \alpha_j$, $F_{ib} = \alpha_i \cdot \beta_b$, $G_{aj} = \beta_a \cdot \alpha_j$ and $H_{ab} = \beta_a \cdot \beta_b$ with $i, j = 1, \cdots, t$ $(t = 2^N)$ and $a, b = 1, \cdots, v$ $(v = k)$. Note that

F and G are used in (D.3), and instead E and H are used in (D.3b). Therefore, we can write down the matrices C and D_q's accordingly so that the problem (D.3) is equivalent to the problem (C.1). It is easy to see that C is a matrix with only non-vanishing off-diagonal block of matrix elements given by $(-1)^{\vec{x}\cdot\vec{y}}$, and D_q's are the diagonal matrices with $(D_q)_{st} = \delta_{s,q}\delta_{t,q}$. We omit their detailed forms here.

We take $k = 2$ and $k = 3$ in case(a) for example.

k=2

Here $\vec{x} = x_1$ and $\vec{y} = y_1$. According Eq. (D.1), we want to maximize $|C_{0,0} + C_{0,1} + C_{1,0} - C_{1,1}|$. Using the Tsirelson theorem, it is equivlaent to maximizing $\alpha_1 \cdot \beta_1 + \alpha_1 \cdot \beta_2 + \alpha_2 \cdot \beta_1 - \alpha_2 \cdot \beta_2$. Such Tsirelson bound has been showed by Wehner [56] using SDP. We just show the numerical result. For more details, please see [56]. After using SeDuMi program [62] to solve SDP, the optimal for both primal and dual problem is 2.8284. It is consistent with the Tsirelson bound [53] ($2\sqrt{2}$) for the case two settings per site.

k=3 Here $\vec{x} = x_1x_2$ and $\vec{y} = y_1y_2$. Notably, $\vec{y} \in \{00, 10, 01\}$. The problem which we want to solve is

$$\begin{aligned}
&maximize \quad |C_{00,00} + C_{00,10} + C_{00,01} + C_{01,00} + C_{01,10} - C_{01,01} \\
&+C_{10,00} - C_{10,10} + C_{10,01} + C_{11,00} - C_{11,10} - C_{11,01}| \\
&= maximize \quad \alpha_1 \cdot \beta_1 + \alpha_1 \cdot \beta_2 + \alpha_1 \cdot \beta_3 + \alpha_3 \cdot \beta_1 + \alpha_3 \cdot \beta_2 - \alpha_3 \cdot \beta_3 \\
&+\alpha_2 \cdot \beta_1 - \alpha_2 \cdot \beta_2 + \alpha_2 \cdot \beta_3 + \alpha_4 \cdot \beta_1 - \alpha_4 \cdot \beta_2 - \alpha_4 \cdot \beta_3. \qquad \text{(D.5)}
\end{aligned}$$

The X matrix for primal problem is $X = S^T S$ where the columns of S correspond the unit vectors $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \beta_3)$.

$$X = \begin{pmatrix}
\alpha_1 \cdot \alpha_1 & \alpha_1 \cdot \alpha_2 & \alpha_1 \cdot \alpha_3 & \alpha_1 \cdot \alpha_4 & \alpha_1 \cdot \beta_1 & \alpha_1 \cdot \beta_2 & \alpha_1 \cdot \beta_3 \\
\alpha_2 \cdot \alpha_1 & \alpha_2 \cdot \alpha_2 & \alpha_2 \cdot \alpha_3 & \alpha_2 \cdot \alpha_4 & \alpha_2 \cdot \beta_1 & \alpha_2 \cdot \beta_2 & \alpha_2 \cdot \beta_3 \\
\alpha_3 \cdot \alpha_1 & \alpha_3 \cdot \alpha_2 & \alpha_3 \cdot \alpha_3 & \alpha_3 \cdot \alpha_4 & \alpha_3 \cdot \beta_1 & \alpha_3 \cdot \beta_2 & \alpha_3 \cdot \beta_3 \\
\alpha_4 \cdot \alpha_1 & \alpha_4 \cdot \alpha_2 & \alpha_4 \cdot \alpha_3 & \alpha_4 \cdot \alpha_4 & \alpha_4 \cdot \beta_1 & \alpha_4 \cdot \beta_2 & \alpha_4 \cdot \beta_3 \\
\beta_1 \cdot \alpha_1 & \beta_1 \cdot \alpha_2 & \beta_1 \cdot \alpha_3 & \beta_1 \cdot \alpha_4 & \beta_1 \cdot \beta_1 & \beta_1 \cdot \beta_2 & \beta_1 \cdot \beta_3 \\
\beta_2 \cdot \alpha_1 & \beta_2 \cdot \alpha_2 & \beta_2 \cdot \alpha_3 & \beta_2 \cdot \alpha_4 & \beta_2 \cdot \beta_1 & \beta_2 \cdot \beta_2 & \beta_2 \cdot \beta_3 \\
\beta_3 \cdot \alpha_1 & \beta_3 \cdot \alpha_2 & \beta_3 \cdot \alpha_3 & \beta_3 \cdot \alpha_4 & \beta_3 \cdot \beta_1 & \beta_3 \cdot \beta_2 & \beta_3 \cdot \beta_3
\end{pmatrix} \qquad \text{(D.6)}$$

According to (D.5), the matrix C is defined

$$C = \frac{-1}{2} \times \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 \end{pmatrix}.$$

The norm of the vectors $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \beta_3)$ must be one is the source of the constrain. Each of the matrix D_q $(q = 1...7)$ is a 7×7 diagonal matrix with the q-th diagonal element being one and zero others. The value b_q $(q = 1...7)$ is one. The numerical result shows that the tight bound is 6.9282, which essentially agrees with (D.1). When we get the optimal solution, the correlation function matrix is

$$X = \begin{pmatrix} 1.0000 & 0.3333 & 0.3333 & -0.3333 & 0.5774 & 0.5774 & 0.5774 \\ 0.3333 & 1.0000 & -0.3333 & 0.3333 & 0.5774 & -0.5774 & 0.5774 \\ 0.3333 & -0.3333 & 1.0000 & 0.3333 & 0.5774 & 0.5774 & -0.5774 \\ -0.3333 & 0.3333 & 0.3333 & 1.0000 & 0.5774 & -0.5774 & -0.5774 \\ 0.5774 & 0.5774 & 0.5774 & 0.5774 & 1.0000 & 0.0000 & 0.0000 \\ 0.5774 & -0.5774 & 0.5774 & -0.5774 & 0.0000 & 1.0000 & -0.0000 \\ 0.5774 & 0.5774 & -0.5774 & -0.5774 & 0.0000 & -0.0000 & 1.0000 \end{pmatrix}. \quad \text{(D.7)}$$

X satisfying the constraint that X is SDSP with non-negative eigenvalues [66].

For the case $k = 3$ to $k = 8$

After setting up the SDP for finding the Tsirelson bound, we still use the package named SeDuMi to solve it for both case (a) and (b) with any value of k. The result agrees extremely well with the bound obtained from information causality up to $\mathcal{O}(10^{-4})$. To be more concrete, the numerical results are shown below: for case (a) up to $k = 8$, we have

k	3	4	5	6	7	8
SDP	6.9282	16.0000	35.7771	78.3837	169.3281	362.0387

This agrees extremely well with the RHS of (D.1). Similarly, for case (b) up to $k = 8$, we have

k	3	4	5	6	7	8
SDP	13.8564	32.0000	71.5542	156.7673	338.6562	724.0773

It again agrees extremely well with (D.2). Therefore, based on our numerical simulation, information causality indeed singles out the Tsirelson bound of a physical theory such as quantum mechanics.

Appendix E

The quantum constraints for $n = 1$ and $n = 1 + AB$ certificate

We divide this appendix into two parts. In the first part, we will write down the associated quantum constraints for $\Gamma^{(1)}$ and $\Gamma^{(1+AB)}$ when finding the bound of Bell-type inequality. In the second part, we will estimate the number of these constraints and find a efficient way to write down these constraints.

E.1 The quantum constraints for $n = 1$ and $n = 1+AB$ certificate

When maximizing the Bell-type inequality under some quantum constraints, the joint probabilities are not given, they are variables. Therefore, when writing down quantum constraints (3.38b), we only need to consider the elements with the specific value (0 and 1) and the relation between different elements such as some elements are the same. For convenience, instead of $A_{\vec{x}}$ and $B_{\vec{y}}$, we use $a : a \in \tilde{A}$ and $b : b \in \tilde{B}$ to denote Alice's and Bob's outcomes and $X(a)$ and $Y(b)$ are the associated measurement setting. The indexes s, t of Γ denote associated operators, i.e., $\Gamma_{a,b} = \mathrm{Tr}(E_a E_b \rho)$.

For $\Gamma^{(1)}$, the associated quantum constraints are

- $\Gamma^{(1)}_{1,1} = \mathrm{Tr}(\rho) = 1$.
- $\Gamma^{(1)}_{a,a'} = \delta_{aa'}\Gamma^{(1)}_{1,a}$ if $X(a) = X(a')$.
- $\Gamma^{(1)}_{b,b'} = \delta_{bb'}\Gamma^{(1)}_{1,b}$ if $Y(a) = Y(a')$.
- $\Gamma^{(1)}_{s,t} = \Gamma^{(1)}_{t,s}$.

We reexpress $\Gamma^{(1+AB)}$ by 4 sub-matrixes,

$$\begin{pmatrix} v_{1,1} & v_{1,2} \\ v_{2,1} & v_{2,2} \end{pmatrix} \tag{E.1}$$

Since $\Gamma^{(1+AB)}$ is symmetric matrix, the sub-matrix $v_{2,1}$ is equal to the transpose of $v_{1,2}$,

and both sub-matrix $v_{1,1}$ and $v_{2,2}$ are symmetric matrixes. Note that, $v_{1,1} = \Gamma^{(1)}$. The elements of matrices $v_{1,2}$ and $v_{2,2}$ are constrained by following quantum constrains:

- $\Gamma^{(1+AB)}_{1,ab} = \Gamma^{(1+AB)}_{a,ab} = \Gamma^{(1+AB)}_{a,b} = \Gamma^{(1+AB)}_{b,ab}$.
- $\Gamma^{(1+AB)}_{ab,a'b} = \Gamma^{(1+AB)}_{a,a'b} = \Gamma^{(1+AB)}_{a',ab}$.
- $\Gamma^{(1+AB)}_{ab,ab'} = \Gamma^{(1+AB)}_{b,ab'} = \Gamma^{(1+AB)}_{b',ab}$.
- $\Gamma^{(1+AB)}_{a,a'} = 0$, $\Gamma^{(1+AB)}_{a,a'b} = 0$, and $\Gamma^{(1+AB)}_{ab,a'b'} = 0$ if $X(a') = X(a)$.
- $\Gamma^{(1+AB)}_{b,b'} = 0$, $\Gamma^{(1+AB)}_{b,ab'} = 0$, and $\Gamma^{(1+AB)}_{ab,a'b'} = 0$ if $Y(b) = Y(b')$.
- $\Gamma^{(1+AB)}_{s,t} = \Gamma^{(1+AB)}_{t,s}$.

E.2 Estimating the number of constrains for $n = 1$ and $n = 1+AB$ certificates

Due to the limitation of computer memory, we need to estimate the number of these quantum constraints for different k and d RAC protocols. The dimension of $\Gamma^{(1)}$ is $1 + (d-1)(d^{k-1} + k)$, we denote it as dim. The number of conditions corresponding to different quantum behaviors is as follows.

n=1	symmetric matrix	$\mathrm{Tr}(\rho) = \Gamma^{(1)}_{1,1} = 1$	orthogonality	$E_aE_a = E_a, E_bE_b = E_b$
number	$\frac{dim(dim-1)}{2}$	1	$\frac{(d-1)(d-2)}{2}(d^{k-1} + k)$	$dim - 1$

The dimension of $\Gamma^{(1+AB)}$ is $1 + (d-1)(d^{k-1} + k) + (d-1)(d^{k-1}k)$, we denote it as dim_{1+AB}. The number of conditions corresponding to different quantum behaviors is as follows:

n=1+AB	symmetric matrix	$Tr(\rho) = \Gamma^{1}_{1,1} = 1$	orthogonality	$E_aE_a = E_a, E_bE_b = E_b$	same
number	$\frac{dim_{1+AB}(dim_{1+AB}-1)}{2}$	1	$otha + othb + othc$	$dim_{1+AB} - 1$	$\sum_{i=1}^{7} same_i$

The quantum constraints orthogonality and commutativity make some elements of certificate to be 0 or to be the same. We will specify to estimate the number of these special elements in $n = 1 + AB$ certificate. First, we estimate the number of elements whose value is zero.

- The variable $otha = \frac{(d-1)(d-2)}{2}(d^{k-1} + k)$ is used to specify the number of zero elements for right upper matrix of $v_{1,1}$.

- The variable $othb = 2(d-1)^2(d-2)kd^{k-1}$ is used to specify the number of zero elements for sub-matrix $v_{1,2}$.
- The variable $othc = \frac{kd^{k-1}(d-1)^2}{2}((d-2)(d-1)(d^{k-1}+k-2)+(d-1)^2-1)$ is used to specify the number of zero elements for right upper matrix of $v_{2,2}$.

We estimate the variable $same_i$ which is used to denote the number of equal pairs.

- $\Gamma^{(1+AB)}_{1,ab} = \Gamma^{(1+AB)}_{a,ab}$, $same_1 = (d-1)^2(d^{k-1}k)$.
- $\Gamma^{(1+AB)}_{a,ab} = \Gamma^{(1+AB)}_{a,b}$, $same_2 = (d-1)^2(d^{k-1}k)$.
- $\Gamma^{(1+AB)}_{b,ab} = \Gamma^{(1+AB)}_{a,b}$, $same_3 = (d-1)^2(d^{k-1}k)$.
- $\Gamma^{(1+AB)}_{ab,a'b} = \Gamma^{(1+AB)}_{a,a'b}$, $same_4 = (d-1)^3d^{k-1}k(d^{k-1}-1)/2$.
- $\Gamma^{(1+AB)}_{a,a'b} = \Gamma^{(1+AB)}_{a',ab}$, $same_5 = (d-1)^3d^{k-1}k(d^{k-1}-1)$.
- $\Gamma^{(1+AB)}_{ab,ab'} = \Gamma^{(1+AB)}_{b,ab'}$, $same_6 = (d-1)^3d^{k-1}k(k-1)/2$.
- $\Gamma^{(1+AB)}_{b,ab'} = \Gamma^{(1+AB)}_{b',ab}$, $same_7 = (d-1)^3d^{k-1}k(k-1)$.

After estimating the number of conditions, we can think how to write down these conditions with minimal computer memory. Here, we use the numerical package named CVXOPT [63] to calculate the bound of Bell-type inequality. The primal problem of the cone programming defined in CVXOPT is

$$\begin{aligned} minimize \quad & c \cdot x && \text{(E.2a)} \\ subject \quad to \quad & Ax - b = 0 && \text{(E.2b)} \\ & h - Gx \geq 0 && \text{(E.2c)} \end{aligned}$$

Given c, h which are the vectors and A, G which are matrixes, we can optimize the linear combination $c \cdot x$. Here matrix G is used to specify the positive definiteness constraint. Writing down the positive definiteness constraint of a matrix Z whose size is $s \times s$, we need the matrix G with size $s^2 \times n$ to define the condition (where n is the number of variables x). That means, if we reduce the number of variables, we can save the computer memory. To do this, we define the same variable for two elements instead of constraining two variables with the same value. On the other hand, if the value of some elements are zero, it could also reduce the number of variables.

After using the conditions to reduce the number of variables, we can estimate the number of variables in the certificate.

The number of variables in $\Gamma^{(1)}$ for different RAC protocols:

n=1	d=2	d=3	d=4	d=5
k=2	10	50	153	364
k=3	28	288	1596	6160
k=4	78	1922	20706	132612

The number of variables in $\Gamma^{(1+AB)}$ for different RAC protocols:

n=1+AB	d=2	d=3	d=4	d=5
k=2	15	182	1287	5964
k=3	82	4068	61860	474160
k=4	486	71258	1995810	24012612

Due to the constraint of the computer memory (128GB), we could not find the bound of Bell-type inequality for arbitrary RAC communication protocols. We find the bound what we can do and show the result in the main text.

References

[1] P. W. Shor., "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer", SIAM Journal on Computing, 26(5):1484V1509, (1997).

[2] C. H. Bennett and G. Brassard., "Quantum cryptography: Public key distribution and coin tossing", In Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, pages 175V179, 1984.

[3] A. Einstein, B. Podolsky, and N. Rosen, "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?", Phys. Rev. **47**, 777V780 (1935).

[4] Bohm, D.,1951, Quantum Theory, New York: Prentice Hall.

[5] J. S. Bell, "On the Einstein Podolsky Rosen Paradox", Physics 1, 3, 195-200 (1964).

[6] J. Clauser, M. Horne, A. Shimony, and R. Holt, "Proposed experiment to test local hidden-variable theories", Phys. Rev. Lett. **23**, 880 (1969).

[7] B. S. Cirel'son, "Quantum Generalizations of Bell's Inequality", Lett. Math. Phys. **4**, 93 (1980).

[8] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, "Quantum correlations for arbitrarily high-dimensional Bell inequality", Phys. Rev. Lett. **88**, 040404 (2002).

[9] N. Gisin, "Bell inequality for arbitrary many settings of the analyzers", Phys. Lett. A **260**, 1 (1999).

[10] G. C. Ghirardi, A. Rimini, and T. Weber, "A general argument against superluminal transmission through the quantum mechanical measurement process", Lett. Nuov. Cim. **27**, 293 (1980).

[11] L. Masanes, A. Acin, and N. Gisin, "General properties of nonsignaling theories", Phys. Rev. A **73**, 012112 (2006).

[12] J. Barrett, "Information processing in generalized probabilistic theories", Phys. Rev. A **75**, 032304 (2007).

[13] H. Barnum, J. Barrett, M. Leifer, and A.Wilce, "Generalized No-Broadcasting Theorem", Phys. Rev. Lett. **99**, 240501 (2007).

[14] J. Barrett, L. Hardy, and A. Kent, "No Signalling and Quantum Key Distribution", Phys. Rev. Lett. **95**, 010503 (2005).

[15] A. Acin, N. Gisin, and L.Masanes, "From Bell's Theorem to Secure Quantum Key Distribution", Phys. Rev. Lett. **97**, 120405 (2006).

[16] Masanes, Ll., Acin, A., Gisin, N., "General properties of nonsignaling theories", Phys. Rev. A. **73**, 012112 (2006).

[17] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, "Non-local correlations as an information theoretic resource", Phys. Rev. A **71**, 022101 (2005).

[18] V. Scarani, N. Gisin, N. Brunner, L. Masanes, S. Pino, and A. Acín, "Secrecy extraction from no-signalling correlations", Phys. Rev. A **74**, 042339 (2006).

[19] N. Linden, S. Popescu, A. J. Short, and Andreas Winter, "Quantum nonlocality and beyond: Limits from nonlocal computation", Phys. Rev. Lett. **99**, 180502 (2007).

[20] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu and D. Roberts, "Information processing in generalized probabilistic theories", Phys. Rev. A **71**, 022101 (2005).

[21] S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom", Foundations of Physics, **24**(3):379V385, 1994.

[22] A. Fine, "Hidden Variables, Joint Probability, and the Bell Inequalities", Phys. Rev. Lett. **48**, 291 (1982).

[23] Yao, A. C. (1979), "Some Complexity Questions Related to Distributed Computing", Proc. of 11th STOC 14: 209V213

[24] van Dam W., "Implausible consequences of superstrong nonlocality", arXiv:quant-ph/0501159v1.

[25] H. Buhrman, R. Cleve, S. Massar and R. de Wolf, "Nonlocality and communication complexity", Rev. Mod. Phys., vol. **82**, 665 (2010).

[26] R.Cleve and H.Buhrman, "Substituting quantum entanglement for communication", Physical Review A, 56(2):1201V1204, 1997. G.

[27] Brassard et al., "Limit on Nonlocality in Any World in Which Communication Complexity Is Not Trivial", Phys. Rev. Lett. **96**, 250401 (2006).

[28] N. Brunner and P. Skrzypczyk, "Non-locality distillation and post-quantum theories with trivial communication complexity", Phys. Rev. Lett. 102, 160403 (2009).

[29] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, "Information causality as a physical principle", Nature (London) **461**, 1101 (2009).

[30] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, "Dense Quantum Coding and Quantum Finite Automata", Journal of the ACM **49**, 496 (2002).

[31] M. Pawlowski and M. Zukowski, "Entanglement-assisted random access codes", Phys. Rev. A **81**, 042326 (2010).

[32] L. -Y. Hsu, I-C. Yu and F. -L. Lin, "Information Causality and Noisy Computations", Phys. Rev. A **84**, 042319 (2011).

[33] I-C. Yu and F. -L. Lin, "Testing Information Causality for General Quantum Communication Protocols", arXiv:1301.1448.

[34] Navascues M. and Wunderlich H., "A glance beyond the quantum model",(2009) Proc. R. Soc. A **466**, 881.

[35] J. Allcock, N. Brunner, M. Paw lowski, and V. Scarani, "Recovering part of the boundary between quantum and nonquantum correlations from information causality", Phys. Rev. A **80**, 040103 (2009).

[36] D. Cavalcanti, A. Salles, and V. Scarani, "Macroscopically local correlations can violate information causality", Nat. Comm. **1**, 136 (2010).

[37] T. H. Yang, D. Cavalcanti, M. L. Almeida, C. Teo, and V. Scarani, "Information-Causality and Extremal Tripartite Correlation", New J. Phys. **14**, 013061 (2012).

[38] Y. Xiang, W. Ren, "Bound on genuine multipartite correlations from the principle of information causality", Quantum Inf. Comput. **11**, 948 (2011).

[39] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués, "Quantum correlations require multipartite information principles", Phys. Rev. Lett. **107**, 210403 (2011).

[40] Sabri W. Al-Safi and Anthony J. Short, "Information causality from an entropic and a probabilistic perspective", Phys. Rev. A **84**, 042323 (2011).

[41] J. Uffink, "Quadratic bell inequalities as tests for multipartite entanglement", Phys. Rev. Lett **88**, 230406 (2002).

[42] L. Landau, "Empirical Two-Point Correlation Functions", Found. Phys. **18**, 449 (1988).

[43] L. Masanes, "Necessary and sufficient condition for quantum-generated correlations", quant-ph/0309137.

[44] M. Navascues, S. Pironio, and A. Acin, "Bounding the set of quantum correlations", Phys. Rev. Lett. **98**, 010401 (2007).

[45] M. Navascues, S. Pironio, A. Acin, "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations", New J. Phys. **10**, 073013 (2008).

[46] C. E. Shannon, "A mathematical theory of communication", Bell System Tech. J., 27:379V423; 623V656, 1948.

[47] J. von Neumann, "Probabilistic logics and the synthesis of reliable organisms from unreliable components", In C. E. Shannon and J. McCarthy, editors, Automata Studies, pages 43V98. Princeton University Press, 1956.

[48] N. Pippenger, "Reliable computation by formulas in the presence of noise", IEEE Transactions on Information Theory, 34(2):194V197, March 1988.

[49] W. Evans' PhD thesis, "Information Theory and Noisy Computation", http://www.cs.ubc.ca/ will/papers/thesis.pdf

[50] W. Evans and L. J. Schulman, "Signal Propagation, with Application to a Lower Bound on the Depth of Noisy Formulas", Proceedings of the 34th Annual Symposium on Foundations of Computer Science, **594** (1993).

[51] W. Evans and L. J. Schulman, "Signal Propagation and Noisy Circuits", IEEE Trans. Inf. Theory, **45**, 2367 (1999).

[52] Oscar C. O. Dahlsten, D. Lercher and R. Renner. "Tsirelson's bound from a generalized data processing inequality", New J. Phys. **14**, 063024 (2012).

[53] B. Cirelson (Tsirelson), "Quantum generalizations of Bell's inequality", Lett. Math. Phys. **4**, 93 (1980).

[54] B. Tsirelson, "Quantum analogues of the Bell inequalities", J. Sov. Math. **36**, 557 (1987).

[55] B. Tsirelson, "Some results and problems on quantum Bell-type inequalities", Hadronic J. Suppl. **8**, 329 (1993).

[56] S. Wehner, "Tsirelson bounds for generalized Clauser-Horne-Hold inequalities", Phys. Rev. A **73**, 022110 (2006).

[57] T. Vértesi and K. Pál, "Generalized Clauser-Horne-Shimony-Holt inequalities maximally violated by higher-dimensional systems", Phys. Rev. A **77**, 042106 (2008)

[58] A. Acín, T. Durt, N. Gisin and J. I. Latorre "Quantum nonlocality in two three-level systems", Phys. Rev. A **65**, 052325 (2002)

[59] http://www.theory.caltech.edu/people/preskill/ph229/

[60] M. A. Nielsen, I. L. Chuang, Quantum Computation and Quantum Information (Cambridge University Press, Cambridge, 2000).

[61] L. Vandenberghe and S. Boyd, SIAM Review 38, 1 (1996).

[62] J. Sturm and AdvOL, http://sedumi.mcmaster.ca.

[63] http://abel.ee.ucla.edu/cvxopt/ .

[64] Boyd, Stephen and Vandenberghe, Lieven (2004). Convex Optimization. Cambridge University Press.

[65] $http://homepages.cwi.nl/\ monique/ow-seminar-sdp/files/ow_intro_sdp.pdf$

[66] C. Helmberg. Semidefinite programming for combinatorial optimization. Technical Report ZIB-Report ZR-00-34, Konrad-Zuse-Zentrum Berlin, 2000.

[67] Singiresu S. Rao, Engineering Optimization: Theory and Practice, Fourth Edition, John Wiley , Sons, Inc, 2009.

[68] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York: Wiley, 1991.