NATIONAL CHENG KUNG UNIVERSITY

MASTER THESIS

Classical communication cost of simulating Bell-nonlocal correlations

模擬貝爾不定域相關性之古典通訊成本

Author:
Bo-An TSAI
作者: 蔡柏安

Supervisor:
Prof. Yeong-Cherng LIANG
指導老師: 梁永成 教授

A thesis submitted in fulfillment of the requirements
for the degree of Master

in the

Department of Physics

January 2023

國立成功大學

碩士論文

模擬貝爾不定域相關性之古典通訊成本

Classical communication cost of simulating Bell-nonlocal correlations

研究生：蔡柏安

本論文業經審查及口試合格特此證明

論文考試委員：

陳岳男

周建宏　李政明　謝宏彰

指導教授：梁永成

單位主管：

（單位主管是否簽章授權由各院、系（所、學位學程）自訂）

中 華 民 國 112 年 1 月 9 日

摘要

貝爾不定域性在量子基礎科學與量子資訊中扮演了基本的角色。其中最重要的是即使測量是發生在類空分離的狀況下，量子糾纏所產生的關聯性足以違反貝爾不等式。在操作上，這意味著從測量結果中這些關聯性不能被定域共同原因或共享隨機性所解釋。自然的，我們可以問說需要多少的古典通訊資源去模擬這些貝爾違反的關聯性？

有趣的是，從Toner and Bacon的結果，我們已知僅僅一個比特就足夠從古典上模擬最大糾纏並使用投影測量的兩個量子比特之關聯性。儘管有這些振奮人心的結果，但至少需要多少的古典資源才能夠模擬兩體任意糾纏純態的關聯性猶未可知。舉例來說，一個比特不足以模擬四元量子最大糾纏態的關聯性。

在此，爲了知悉這些模擬貝爾不定域的關聯性，我們採用並研究Pironio的方法——從平均上的古典通訊成本來模擬這些關聯性。其中我們的目標是在不同的貝爾情境下找到最好的平均古典通訊成本下界，並取得了一些研究進展。

關鍵字: 量子關聯性，量子模擬，古典通訊成本。

Abstract

Bell nonlocality plays a fundamental role in quantum foundations and quantum information science. One of the most striking manifestations of quantum entanglement lies in its ability to produce correlations that violate a Bell inequality even when measurements are carried out in a spacelike separated manner. Operationally, this means that such correlations between measurement outcomes cannot be explained by a local common cause, or equivalently by shared randomness alone. When such a Bell-inequality-violating correlation is seen as a resource, it is natural to ask how much classical communication is needed to simulate it. Interestingly, as little as one bit is known to be sufficient for the simulation of a two-qubit maximally entangled state with projective measurements. Despite all these encouraging results, the minimum amount of classical communication required to reproduce all the correlations originating from a general bipartite entangled pure state is still not known. For instance, there are examples involving maximally entangled two-ququart state where one classical bit is known to be insufficient for its classical simulation. Here, to gain insights into the difficulty in simulating Bell-nonlocal correlations, we adopt the approach of Pironio and investigate, instead, the average classical communication cost for simulating such correlations. In particular, we report progress on determining the best lower bound on the largest minimum average communication cost in various Bell scenarios.

Keywords: correlations, simulation, communication cost

致謝

願生命中有足夠的雲翳，來一場不留遺憾的黃昏。

做研究如同墜入兔子洞一般，彷彿若有光。願有一日，我的研究可以造就另外一個人見識到這奇妙世界。首先，讓我摯上我滿滿的謝意給我的導師，梁教授。謝謝帶我領略量子世界與大自然的眞諦——不定域性。強烈的物理直覺總是讓我看到很多新天地。然而我總是相當粗心，容易犯錯，但你仍然對我有耐心。並且，你給予我們這一個很棒的研究環境，可以彼此互動學習。對我來說，這是一個難能可貴的經驗可以互助合作與分享經驗。

在者，獻予我們的Gelo巨巨。我們很常討論很多，不只是我的工作細節，也有很多生活經驗與日常的分享。有時候，我甚至工作到日出時分，但發現你還在。甚至我們還有餘力討論，甚至有了工作進展。這些吉光片羽，我覺得非常有趣。Shiladitya也常跟我分享他的生活經驗或是個人哲學與思考，使我如沐春風，常常我能夠從他身上學習到許多。Swati像是我們的大姊，常常關心我的工作進度，時常耳提面命，有時會分享給我一些有趣的論文，並在白板上，一條一條寫下並解析他們。

岳錠是我進成大第一個認識的朋友。我們一同分享物理的見解，一起修課，並選擇一樣的研究室。甚至當我去練習霹靂舞時，他也參與其中，這些對我來說都是很棒的回憶。願你能在未來得償所望。楷翔，是陪伴我到最後的一個人。我們嘗嘗討論一些證明的細節，或者提點我一些重要的理論基礎。我認爲你在邁向博士的旅程上會有滿滿的收穫。心愉跟萬冠，謝謝你們時常忍受我的爛笑話。希望在你們可以好好享受即將到來的博士生生涯。寫在這，即將邁入尾聲，我依稀記得那些已經從研究室畢業的人們：我以前常常跟士賢和峻禕去做重訓跟吃宵夜；霈昇也常午餐帶我們去吃道地的餐廳；宥峻也常常講一些關於時事的笑話。至碩士多年以來，那些影響我的人們，我抱持著感激。

最後，感謝我的家庭的多年無私支持，感謝我的摯愛，宥儀。

Acknowledgements

"May there be enough clouds in your life to make a beautiful sunset."

Doing the research is jumping into the rabbit hole. Hope my research someday will help someone goes on adventures in another wonderland. First, I want to give a big thanks to my supervisor, professor Liang. Thank you for taking me to explore the quantum world and understand the beauty of nature: Nonlocality. The strong physical intuitions always impress me a lot. However, I'm always the one so careless making mistakes and you always have patience with me. Further, you also create a good environment where we can discuss and work with each other. For me, it's precious to learn how people collaborate and share knowledge with each other.

Second, is our big brother, Gelo. We often discuss very a lot, not just my working details but also generous sharing of knowledge and life. Sometimes, I work or stay until sunrise and you are still around. We even can have work progress after that. I really enjoy each time we have the discussion.

Shiladitya also expresses different life experiences and personal philosophies on many things. Just like a breath of fresh air and I can always learn something from you. Swati is a big sister who always reminds me I haven't finished something yet. Sometimes, share an interesting paper with me. After that, you always have patience writing on the whiteboard so that we can learn something interesting from the papers many times.

Alex is my first friend when I get into NCKU. We always share our understanding of physics, join the course together, and choose in the same group. It is my pleasure that you also come with me when I practiced for breakdance. Wish you will fulfill your goal in the future. Kai-Siang, the last one stayed with me. We are always thinking about some details of the proofs and reminded me a lot of things happened in the office. After many years, you will definitely achieve your goals in the Ph.D. programs. Thank you for always getting my back. Hsin-Yu and Wan-Guan, thank you for always bearing my bad jokes. Hope you will also enjoy for upcoming Ph.D. life in the future.

Last but not least, I still remember those who have left here for a while. I always went work out and had midnight snacks with Jyun-Yi and Shih-Xian. Pei-Sheng always took us to some nice restaurant at launch. Yu-Chun always told us some jokes from the news. I'm grateful to everyone who influenced me in the past master years.

In the end, thank my family and Yo-Yi, my beloved, for many years of unconditional support.

Contents

List of Figures

List of Abbreviations

EPR	Einstein-Podolsky-Rosen
NS	nonsignaling conditions
LHV	local hidden-variable
DI	device-independent
KL	Kullback-Leibler
DAG	directed acyclic graph
CPD	causal parameter-dependent
COD	causal outcome-dependent
MCPD	message causal parameter dependent
POVM	positive-operator valued measure
MES	maximally entangled states
CG	Collins-Gisin
CHSH	Clauser-Horne-Shimony-Holt
NPA	Navascués-Piriono-Acín
PPT	positive partial transposition
SDP	semi-definite program
CCP	communication complexity problem

Chapter 1

Introduction

One of the most striking manifestations of quantum entanglement lies in its ability to produce correlations that violate a Bell inequality [Bel64] even when measurements are carried out space-like separated. A precursor to this discovery was first discussed in the famous paper of Einstein-Podolsky-Rosen in 1935. Later, in 1964, Bell showed that some correlations allowed in quantum mechanics are impossible in local hidden variable theories.

These nonlocal quantum correlations, represented by joint probability distribution of each parties' outcomes conditioned on their choice of measurement, are known to serve as an the indispensable resource for device-independent (DI) quantum information processing [Sca15]. An important goal here is to distinguish them from classical resources. Operationally, such correlations between measurement outcomes cannot be explained by a local common cause [Bel04], or equivalently, by shared randomness alone. In Chapter 2, we will discuss nonlocality briefly and refer the interested readers Appendix A for more details.

In recent years, the development of quantum technology has progressed rapidly. One of the significant breakthroughs is quantum supremacy [AAB$^+$19]. In 2019, Google proudly announced the first demonstration of a quantum computational advantage compared to classical computations. In some sense, the development of quantum information theory can be traced back to Bell's discovery in 1964. As mentioned previously, a Bell experiment can be used to demonstrate that some quantum correlations cannot be reproduced by any local causal model. This leads to new questions concerning the essential features of quantum theory. In particular, people have been trying to understand the ways in which quantum theory is different from classical theory. In this thesis, we focus on the notion of Bell nonlocality.

In this context, Elitzur, Popescu, and Rohrlich (abbreviated as EPR2) asked the interesting question: does each pair of photons in the ensemble behave nonlocally in an experiment? Such questions motivate a resource-theoretic approach to nonlocality. In other words, given a nonlocal correlation, one can measure its nonlocality by the fraction of times a nonlocal resource has to be invoked while shared randomness (any local resource) is assumed to be free. Similarly, when seen as a resource, it is thus natural to ask how much classical communication cost is needed to simulate quantum nonlocal correlations. Interestingly, as little as one bit is known to be sufficient [TB03] for the simulation of a maximally entangled state of two qubits with projective measurements. To be precise, the Shannon entropy corresponding to the communication of the protocol can be compressed to 0.85 bits. On the other hand, for some restricted class of projective measurements, 1.82 bits are sufficient for the simulation of maximally entangled states of any Hilbert space dimension [RT07]. We leave these discussions to the interested readers in Appendix A.2.

Could it be that one-bit communication is sufficient to reproduce all quantum correlations [Gis07]? A counter-example was provided for the case of two-ququart maximally entangled state [VB09]. In our research, we also want to find the counter-example in the lower dimension or minimal setting.

After many protocols are proposed, the limitation of hand-crafted designed protocols is obvious: The protocols treat all correlations equally. In contrast, given the correlation, we can minimize the usage of classical resource. Hence, we can ask what is the minimum amount of communication is needed to reproduce all quantum correlation? Here, to gain some insight on the problem of simulating Bell-nonlocal correlations, we adopt the approach of Pironio [Pir03] and investigate the average communication needed for simulation in each Bell scenario in Chapter 3 and Chapter 4. Also, a brief tangential discussions taking a machine-learning approach is provided in Appendix D.

Chapter 2

Preliminary

First in Section 2.1 and Section 2.2, we will introduce the correlation space in Bell scenario to discuss the general framework for considering quantum and even beyond quantum. Next, we discuss several ways of quantifying nonlocality. To begin with the nonlocal resource, one way is to start from the Bell scenario supplemented with classical communication. We will introduce different categories to quantify communication. Finally, we will focus on the minimum average communication cost and nonlocal content to quantify the nonlocality.

2.1 Bell experiments

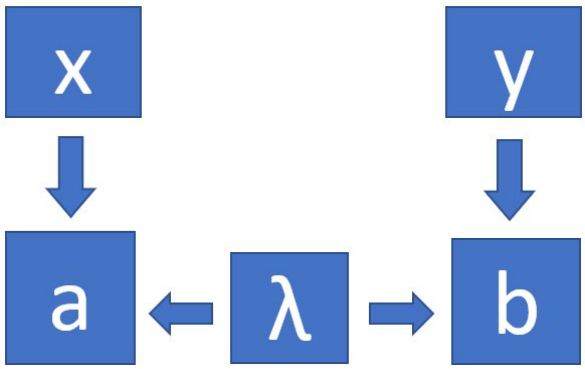


FIGURE 2.1: Bell experiment:
A source λ distributes two physical systems to two spatially separated observers, Alice and Bob. Each observer measures the system and each measurement yields an outcome. Alice's measurement choice and its outcome are labeled by x and a. Similarly, Bob's measurement choice and its outcome are labeled by y and b.

In a Bell experiment, if a nonlocal correlation is found, we can make statements about the devices Alice and Bob use that depend only on the measurement statistics involved. This means that we can use them in quantum technologies in a device-independent way [Sca15]. In a device-independent framework, we are not concerned with the precise details of how the devices function. Rather, we only assume that the devices obey the laws of quantum theory and that the devices do not secretly communicate with each other.

In a typical Bell experiment, two parties, Alice and Bob, share two particles from the same source and stay in space-like separation to ensure that they obey the principle of locality (with respect to causality in special relativity). We assume that Alice and Bob can freely choose to perform the respective measurements x and y and obtain the outcomes a, b, respectively. Locally, Alice and Bob produce their outcomes, and the resulting joint conditional probability distribution $P(a, b|x, y)$ is called the correlation. Here, Alice and Bob are permitted to use some shared randomness, represented by the parameter λ with possible values from a set Λ. Note that λ is independent of parties' inputs alphabets $x, y, z, etc.$ From this perspective, several applications of sharing a correlated random string can be discussed, such as secret key generation, privacy amplification, communication complexity, and the classical simulation of quantum nonlocal statistics.

In the 1920s, there was a series of intense debates between Einstein and Bohr that revolved around the discussion of quantum mechanics and local hidden-variable models. Many decades later, in around 1964 [Bel64], Bell developed a theorem that demonstrated a contradiction for some correlations between the two distant magnetic moments that are possible in quantum theory and any local realistic models that attempt to explain them. After that, various models and assumptions can be discussed. In this thesis, we focus on one small corner of all of these discussions that attempt to understand the precise relation between classical information and quantum theory.

2.2 Correlation space

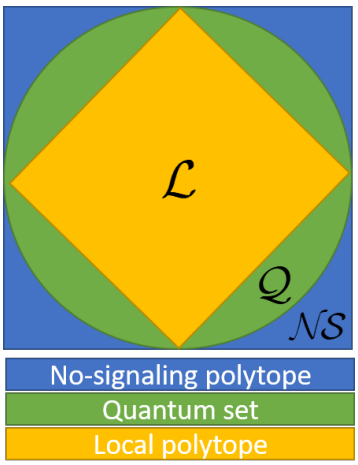


FIGURE 2.2: Schematic view of correlation space: From outside to inside pattern are the no-signaling set $\mathcal{NS}$, the quantum set $\mathcal{Q}$, the local set $\mathcal{L}$, respectively.

In this section, it is essential to introduce the space of correlations in which we are interested. To characterize the different sets of correlations, the following sets are introduced: no signaling sets ($\mathcal{NS}$) in Section 2.2.1, local sets in Section 2.2.2, and quantum sets in Section 2.2.3, respectively. First, we will introduce the corresponding physical idea and the $\mathcal{NS}$ set.

2.2.1 The set of no signaling correlation

Let us continue with the introduction of the Bell scenario. Here Alice and Bob may share some common information obtained from a source but they are not allowed to exchange information. It is natural to look into the joint conditional probability distribution space $P(a, b|x, y)$, but what would be a legitimate distribution from the physical point of view? For this, we may look at the relationship between relativity and quantum theory. To be compatible with both theories, faster-than-light communication cannot be allowed. Thus, in addition to the positivity condition $P(a, b|x, y) \geq 0$ and normalization $\sum_{a,b} P(a, b|x, y) = 1$, the no-signaling conditions are imposed; then each party's outcome distribution is independent of the others' measurement choices.

$$\sum_b P(a, b|x, y) = \sum_b P(a, b|x, y') = P(a|x) \forall y \tag{2.1}$$

$$\sum_a P(a, b|x, y) = \sum_a P(a, b|x', y) = P(b|y) \forall x \tag{2.2}$$

So the $\mathcal{NS}$ sets can be described as the intersection of hyperplanes given by Eq. (2.1) Eq. (2.2) and positivity half-spaces. Hence, the $\mathcal{NS}$ set is a polytope. For further information about a polytope, see Appendix C.1. Summing over from one's inputs x, y and other inputs x', y' results in their marginal distribution $P(a|x)$ and $P(b|y)$. Consider the two-party scenario with n_{in} inputs and n_{out} outcomes as (n_{in}, n_{out}). The normalization constraint and the no-signaling conditions mean that the dimension d_{NS} is reduced to [CG04]

$$d_{NS} = (n_{in} - 1)^2 n_{out}^2 + 2(n_{in} - 1)n_{out} \tag{2.3}$$

2.2.2 The set of Bell-local correlation $\mathcal{L}$

A correlation corresponds to the observed statistics from the experimental data. Apart from that, physicists are familiar with the Hamiltonian formalism to describe the dynamics of systems but sometimes we do care about whether a system behaves like a "quantum" or "classical" one. From an informational point of view, all these classical correlations can be simply described by the local hidden variable model:

$$P(a, b|x, y) = \sum_\lambda P(\lambda) P(a|x, \lambda) P(b|y, \lambda) \tag{2.4}$$

where $P(a|x, \lambda)$, $P(b|y, \lambda)$ are local response functions over all possible local hidden variables λ.

In some sense, the hidden variables λ that Alice and Bob share correspond to certain classical properties. In classical mechanics, we can measure things such as particle number, phase, wavelength, moment, etc., and we say that the outcome reveals a pre-existing value. In some cases, we will have some ignorance of these definite values, and this can be represented by a probability distribution $P(\lambda)$.

A Bell inequality B_{abxy} says that a linear combination of elements of the conditional probability distribution $P(a,b|x,y)$, which we call the Bell value, is bounded by some threshold B_0 called the local bound:

$$\sum_{abxy} B_{abxy} P_{abxy} \overset{\mathcal{L}}{\leq} B_0 \tag{2.5}$$

In practice, the set of local correlation is a convex set obtained from linear combinations of deterministic strategies $\delta_{a,f(x)}$, $\delta_{b,f(y)}$. So the maximum value of the right-hand side of Eq. (2.5) is obtained by maximizing over all these deterministic strategies. Some details are given in Appendix B.1. It is straightforward to show that the number of extremal points will grow exponentially as the setting increases. For the bipartite scenario (n_{in}, n_{out}), some of the full lists are known:

$(n_{in}, 2)$ with $n_{in} \in \{2, 3\ [\text{CG04}], 4\ [\text{CG19}]\}$; $(2, n_{out})$ with $n_{out} \in \{2, 3\ [\text{CG04}]\}$. The partial list $(3, 3)$ is given in [SBSL18] and [CC19]. In Section 4.1, we will employ Bell inequalities to explore the nonlocality.

2.2.3 The set of quantum correlation

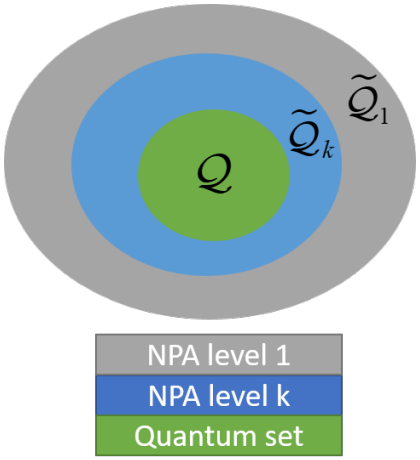


FIGURE 2.3: The NPA hierarchy with k level $\tilde{\mathcal{Q}}_k$ is the outer approximation to the quantum set $\mathcal{Q}$.

To better understand the counterintuitive features of quantum mechanics, it pays to investigate the difference between the set of Bell-local correlations and those allowed in quantum theory. This can be done, for example, by investigating the maximal quantum violation of a Bell inequality. To begin with the characterization of quantum correlations, recall that the outcome probabilities for quantum measurements are governed by Born's rule.

$$P(a,b \mid x,y) = \text{tr}\left(\rho M_{a|x} \otimes M_{b|y}\right) \tag{2.6}$$

The joint conditional probability can be defined for any bipartite density matrix ρ and a set of positive operator-valued measures [NC11] $M_{a|x}$ and $M_{b|y}$ for Alice and Bob, respectively. By that definition, it is straightforward to show that the quantum set $\mathcal{Q}$ strictly belongs to the $\mathcal{NS}$ set.

Also, if density matrix is separable, the corresponding probability distribution must be local. In general, the following relations hold for the sets considered here:

$$\mathcal{L} \subsetneq \mathcal{Q} \subsetneq \mathcal{NS}. \tag{2.7}$$

A full characterization of the set of quantum correlations is an important but difficult issue. The standard approach is to use some hierarchy of SDPs that provides a collection of outer approximations [NPA07] [NPA08] [WDTL08] [MBL$^+$13] to the quantum set.

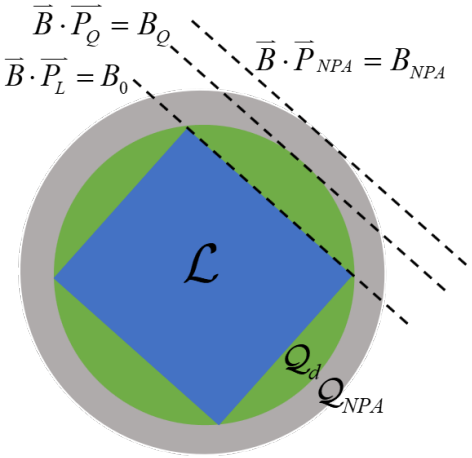


FIGURE 2.4: The local bound B_0, the quantum bound from see-saw method B_Q, and the NPA method B_{NPA} are given from different SDP optimization methods.

The NPA hierarchy, which is a convergent hierarchy with different levels, provides a quantum approximation by replacing the tensor product in Born's rule with commuting operators. These details of the SDP hierarchy can be found in [NPA08]. In practice, since we maximize over an outer approximation, the following SDP originating from NPA gives an upper bound to the maximal quantum value.

$$\begin{aligned} \max \quad & \vec{B}\cdot\vec{P} \\ \text{subject to} \quad & \\ & \chi[\rho] \geq 0 \end{aligned} \tag{2.8}$$

Here we write the Bell value as the inner product of $\vec{B}$, which is the vector form of Bell inequality coefficients, and the correlation $\vec{P}$. The positivity constraint on the moment χ in Eq. (2.8) provides a necessary (but generally insufficient) condition for the correlation to be quantum, thus resulting in an outer approximation.

$$\begin{aligned} \max \quad & \vec{B} \cdot \vec{P} \\ \text{subject to} \quad & \\ & \vec{P} \in Q_d \end{aligned} \tag{2.9}$$

To check whether the bound obtained from the SDP is tight, we also perform heuristic maximizations of the Bell value over fixed d-dimensional Hilbert space. This can be written as Eq. (2.9) where $\mathcal{Q}_d$ denotes the convex hull of the set of quantum correlations arising from up to d-dimensional Hilbert space. One such method goes by the name of the seesaw algorithm [WW01] and [LD07]. Also, the details of seesaw algorithm are left in Appendix C.3.1.

By using both methods, one can compare the lower and the upper bound therefore we have a promise of a tight bound if they coincide, although in practice this only happens up to some numerical precision. So far, we have understood how to approximate the quantum bound. In the following section, we proceed to with the quantification of nonlocality.

2.3 Quantification of nonlocality

As mentioned in the previous section, estimating the maximum violation of a Bell inequality provides a simple and natural way to quantify the amount of nonlocality in a given correlation. Alternatively, one can try to quantify nonlocality by the amount of noise Eq. (C.5) needed so that a given correlation becomes a local one. Several other nonlocality measures are discussed in Appendix A.1. In this section, we will focus on the assumptions needed for manifesting Bell nonlocality.

The property of Bell locality is equivalent to the conjunction of two conditions [Jar84]. (I) Parameter independence: Given the hidden variable, the outcome probabilities for a measurement on any subsystem are independent of what measurement is being performed on any other subsystem. (II) Outcome independence: Given the hidden variable, the outcome probabilities for a measurement on any subsystem are independent of the outcome of a measurement on any other subsystem. In the subsequent section, we discuss several corresponding models which relaxed the above conditions by supplemented with the communication.

2.3.1 Quantification of communication

In fact, this kind of question is not only concerned with foundational perspectives, but also real-life applications like quantum key distribution. For example, the fact that eavesdroppers can compromise security and even hack quantum cryptography systems can be seen from the fact that real devices can behave in ways that do not respect some presupposed causal structure [LWW$^+$10]. Thus, a quantitative study of communication models is also beneficial beyond its relevance to the simulation of nonlocal correlations.

Quantify the communication with the intervention

As mentioned previously, the locality holds for these assumptions. If arbitrarily distributed communication is allowed, any correlation can be reproduced by classical theory. Hence, several structures are proposed with different complementary perspectives. These paradigmatic models can be described by the Bayesian networks and show the relation between several casual nodes [CKBG15]. The directed acyclic graph (DAG) consists of jointly distributed random variables $(x_1, \ldots, x_n)$ and directed edges in order to demonstrate functional dependence. The probability $P(\mathbf{x}) = P(x_1, \ldots, x_n)$ encodes the causal relationships introduced by the DAG can be written in the form

$$P(\mathbf{x}) = \prod_{i=1}^{n} P(x_i \mid \mathrm{pa}_i) \tag{2.10}$$

where pa_i are the parent variables. For example in Fig. 2.1, the $P(x, y, \lambda) = P(x)P(y)P(\lambda)$.

One possible quantification is achieved through the concept of intervention $do(x_i')$ which means the causal nodes x_i are replaced by a new value x_i'. Therefore, x_i' introduces a new mechanism while keeping the other mechanism unperturbed.

$$p(\mathbf{x} \mid do(x_i')) = \begin{cases} \prod_{j \neq i}^{n} p(x_j \mid \mathrm{pa}_j) & \text{if } x_i = x_i' \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

Compared to Eq. (2.10), it removes the part $i = j$.

$$\mathcal{C}_{A \to B} = \max_{b,y,a,a'} \sum_{\lambda} P(\lambda) \mid P(b \mid do(a), y, \lambda) - P(b \mid do(a'), y, \lambda) \mid . \tag{2.12}$$

The direct causal influence of A into B is defined as the maximization of the shift of probability of B caused by interventions in A (denote as do(a) and do(a')) on average of local strategies λ.

Based on this research, different communication DAGs are established with different causal influences and reveal the relationship between direct causal influence and Bell's inequalities [BC17]. The classes can be branched into several DAG representations and decomposition as following.

(I). The CPD (causal parameter-dependent) model:

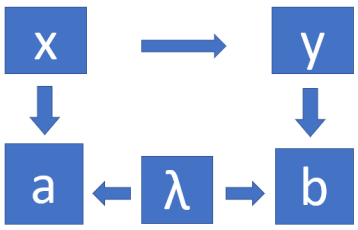


FIGURE 2.5: The causal parameter-dependent model

$$P(a,b \mid x,y) = \sum_{\lambda} P(\lambda)P(a \mid x,\lambda)P(b \mid x,y,\lambda) \tag{2.13}$$

The main idea is that Alice transmits her input information to Bob's side. In this case, the input of one party can directly affect the outcomes of others. One-way communication implies an invalid one-sided no-signaling condition $P(b \mid x,y,\lambda) \neq P(b \mid y,\lambda)$ compared to the original Bell-type LHV model.

(II). The COD (causal outcome-dependent model) model:

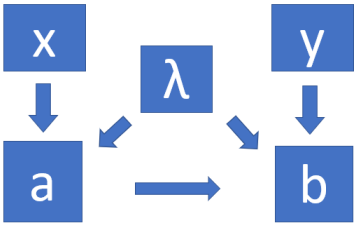


FIGURE 2.6: The causal outcome-dependent model

$$P(a,b \mid x,y) = \sum_{\lambda} P(\lambda)P(a \mid x,\lambda)P(b \mid a,y,\lambda) \tag{2.14}$$

The other model is related to the communication of outcomes, which means that the outcomes of one party have causal influences over the other outcomes. An example in the (3,2) scenario shows that a COD model is unable to reproduce certain nonlocal correlations [CKBG15]. Also, they generalized that COD models are insufficient to reproduce quantum correlations for a larger number of inputs as well. From this example, if one wants to find out which correlation is more nonlocal with respect to the communication cost for simulating correlations, increasing the inputs may be preferable to increase the outcomes. Interestingly, we will show this intuition, in general, is not true in Chapter 4.

(III). The MCPD (message causal parameter dependent model) model:

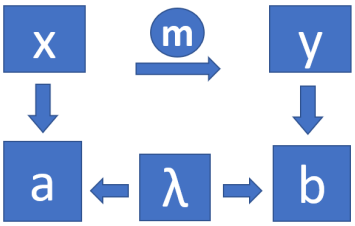


FIGURE 2.7: The message causal parameter dependent model

$$P(a,b \mid x,y) = \sum_{\lambda,m} P(\lambda)P(m \mid x,\lambda)P(a \mid x,\lambda)P(b \mid m,y,\lambda) \tag{2.15}$$

The message can be understood as a function of one's inputs and outcomes. The other party can previously determine the protocol to transmit the information. By the way, the size of

the message matters; otherwise, there would be no cost in reproducing any correlation. One natural quantification is to start with some entropic measure. Given the observed data, one can start from the preparation-and-measure scenario and use entropy witnesses and dimension witnesses [CKBG15]. In that paper, it is shown that a quantum message can outperform any classical message by some entropy witness measure. In this work, rather than having quantum messages, we will discuss the relationship between the classical message and the quantum correlation under the measure of the communication cost on average. To begin with the communication cost, several approaches from the protocols are discussed.

Quantify the communication with the protocol

How can the classical theory of information describe quantum-mechanical statistics? In the past decades, this kind of question can be answered from various information-theoretical perspectives. Some of the motivations for discussing the framework are related to the non-trivial application of quantum technology. One of the celebrated techniques is that of quantum teleportation, which implies that the physical transmission of a qubit can be replaced by the transmission of two classical bits and shared entanglement. This raises several interesting questions: What is the key role and information content between the classical bits and qubits? Why two? What is the minimum amount of communication required to reproduce the statistics of quantum mechanics? For example, the communication of 2.19 bits is on average sufficient for the classical teleportation of the qubit when restricted to Von Neumann measurements [CGM00]. Moreover, for the teleportation protocol, recent research has shown that it is sufficient to transmit one classical bit if the sender has access to multiple copies of the qubits to be teleported [Par22]. As you can see, choosing different restrictions or implementations can be divided into different scenarios. The experimental application like the teleportation will motivate the discussion of the informative framework of communication models. To recapitulate, we are concerned with the probability distribution space and the minimum amount of communication cost to simulate the quantum correlations. In the following, an impressive example is from Toner and Bacon [TB03], who state that the 1-bit communication enables one to simulate the correlations obtained from projective measurements on Bell pairs. Afterward, the problem was asked how many bits of communication are required to simulate the correlation arising from two-outcome projective measurements without restricting the size of the Hilbert space. The protocol proposed in [RT07], states the two-bit communication can simulate any quantum correlation arising from a restricted family of projective measurements. One of the worst-case examples shows that one-bit communication cannot simulate maximally entangled four-dimensional quantum systems [VB09]. Very recently, it's shown that two bits are sufficient to simulate the correlation coming from arbitrary measurements on any two-qubit entangled state [RQ22].

All these results can be understood in terms of the worst-case communication cost. Considering some of the details of the protocols mentioned above may be helpful in understanding the physical picture of simulating quantum states with communication. Beyond that, we want to look into the correlation space $P(a, b|x, y)$ and ask the question: what is the minimum cost on average with respect to Bell scenario?

The minimum average communication cost

Note that we are concerned with the average amount of communication required, as opposed to the worst-case communication that was first considered by Brassard et al. [BCT99]. Here we have two parties that are only allowed to possess classical resources, which are often mentioned to be shared randomness, and a means to communicate classical bits. The goal is to try to reproduce any correlation $\vec{P}$. If the correlation is nonlocal, the use of classical communication is necessary. Some of the technical terms and discrepancies are described below.

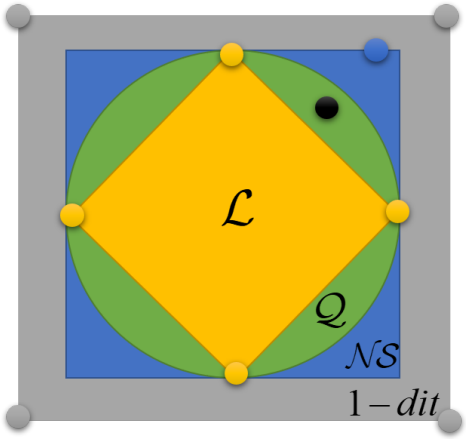


FIGURE 2.8: Given the nonlocal correlation (black point), the notion of MACC involves an optimization over the convex decompositions in terms of local extremal points (yellow points) and 1-dit polytope extremal points (grey points). Similarly, the nonlocal is a notion that involves an optimization over the convex decomposition in terms of local extremal points (yellow points) and an arbitrary no-signaling (boundary) point (the blue point).

1. The minimum average communication cost (abbreviated as MACC): The minimum amount of communication Alice and Bob need to exchange to reproduce a certain correlation. We want to consider the set of correlations that can be produced by exchanging up to 1-dit of classical communication. It is known that such a set forms a convex polytope [Pir03], whose extreme points $\{\vec{D}_i^{(j)}\}_{j=0}^{d}$ are naturally classified according to the number of communication bits needed for the simulation. For example, the i^{th} extreme points $\vec{D}_i^{(j)}$ simply correspond to j-dit deterministic strategies. Any given $\vec{P}$ in a (n_{in}, n_{out}) Bell scenario can be written as a convex decomposition, and

the MACC $\bar{C}(\vec{P})$ is described as follows.

$$\begin{aligned} \min \quad & \sum_i \sum_{j\geq 1} c_i^{(j)} \\ \text{subject to} \quad & \\ & \vec{P} = \sum_{j=0}^{d} \sum_i c_i^{(j)} \vec{D}_i^{(j)}, \\ & c_i^{(j)} \geq 0 \ \forall \ i, \\ & \sum_{i,j} c_i^{(j)} = 1. \end{aligned} \tag{2.16}$$

where i is the index label for the various extreme points. The MACC $\bar{C}(\vec{P})$ for the given correlation $\vec{P}$ can be obtained by minimizing $\sum_i \sum_{j\geq 1} c_i^{(j)}$ on all such decompositions. In this thesis, all of the results for the MACC refers to the 1-bit polytope, unless specified otherwise.

2. The worst case versus the largest case: The worst case means the maximum amount of communication exchanged in certain protocol [BCT99]. In this thesis, to clarify the distinction, the largest case is defined to be the largest possible value of MACC, say, for a given Bell scenario or a given state. For example, in the $(3, 2)$ scenario, the worst case is $log_2 3$ bits [BT03] because each party can tell the other exactly which measurement they have selected. This limit is naturally introduced in the NS scenario, the maximum of MACC in $(3, 2)$ is $log_2 3$ bits. While for the largest case we have found is $3\sqrt{3}/2 - 2 \approx 0.5981$, which sits in the quantum sets, and we will discuss that in the future chapter.

3. Pironio's bound: Given the Bell's value $B(P)$ corresponding to the Bell inequality, we have the simple analytic lower bound of MACC [Pir03].

$$\bar{C}_{P,B} = \frac{B(P) - B_0}{B_{dit} - B_0} \tag{2.17}$$

Here, B_0 is the corresponding local bound, and B_{dit} is the corresponding 1-dit bound. For example, if we consider the Clauser-Horne-Shimony-Holt (CHSH) Bell inequality in $(2, 2)$ scenario, the corresponding local bound $B_0 = 2$ is the maximum of the inner products with 16 local extremal points. Similarly, the corresponding 1-bit bound $B_{dit} = 4$ is the maximum over the inner products with 64 1-bit extremal points. For the maximum CHSH violation $B(P) = 2\sqrt{2}$, we will obtain the Pironio's bound $\sqrt{2} - 1 \approx 0.4141$.

4. NS Pironio's bound: From previous the description of Pironio's bound, means appending the constraint $\vec{B} \cdot \vec{P} = B_v$ to Eq. (4.4) when given Bell's value B_v. In addition, we can also impose the NS constraint. The programming of NS Pironio's bound $C_{P,B,NS}$

is

$$\begin{aligned} \min \quad & \sum_i \sum_{j\geq 1} c_i^{(j)} \\ \text{subject to} \quad & \\ & \vec{P} = \sum_{j=0}^{d} \sum_i c_i^{(j)} \vec{D}_i^{(j)}, \\ & c_i^{(j)} \geq 0 \ \forall \ i, \quad \sum_{i,j} c_i^{(j)} = 1, \\ & \vec{B} \cdot \vec{P} = B_v, \quad \vec{P} \in NS. \end{aligned} \tag{2.18}$$

5. Nonlocal content (EPR2 cost) : The nonlocal content $\bar{C}_{NL}$ of a correlation P is obtained by finding the convex decomposition that maximizes the local weight $\sum_j c_i^{(0)}$ relative to local extremal points [EPR92]. (Conventionally, in this thesis, sometimes we will call it EPR2 cost) The linear program is as follows:

$$\begin{aligned} \min \quad & 1 - \sum_j c_i^{(0)} \\ \text{subject to} \quad & \\ & \vec{P} \geq \sum_i c_i^{(0)} \vec{D}_i^{(0)}, \\ & c_i^{(0)} \geq 0 \ \forall \ i. \end{aligned} \tag{2.19}$$

   Note the distinction from MACC: for MACC, we minimize the nonlocal part of the decomposition with respect to the communication polytope extreme points, but for EPR2, we maximize the local part of the decomposition with respect to the no-signaling extreme points. In a separate work, it was found [Che22] that the nonlocal content of a given correlation lower bounds its MACC.

$$\bar{C}_P \geq \bar{C}_{NL} \tag{2.20}$$

6. Communication advantage: If $\bar{C}_P \neq \bar{C}_{NL}$, then this means that there is room for a smaller value with some other nonlocal extremal points. Specifically, if we have MACC $C_{P,d}$ in the d-dit polytope and $C_{P,d+1} < C_{P,d}$, we will say that for that correlation P has the advantage of communication in d-dit.

   Note that if the communication dimension d is greater than Alice's input number X, the Stirling number will not be defined in Appendix B.2. It means that it leaves no communication advantage for $d = X$ because the extremal points are the same for the X-dit and $X + 1$-dit polytope by construction in Appendix B.2. In other words, Alice can tell Bob more information than which measurements she chosen, but leaves no advantage.

Chapter 3

The MACC from random sampling

In this chapter, we measure the amount of nonlocality by the MACC or by the nonlocal content of randomly sampled correlations in different Bell scenarios. For this, we discuss two sampling methods: Post-selected sampling from the NS space in Section 3.1 and sampling with the unitary operation in Section 3.2.

3.1 Sampling from outer quantum approximations

We briefly discuss how to describe a point in the NS space with a minimal number of parameters. Denote by $\mathcal{P}$ the set of conditional probability distributions $\vec{P} = \{P(a, b \mid x, y)\}_{a,b,x,y}$ where the outcomes are $a, b \in \{1, \ldots, n_{out}\}$ and the inputs are $x, y \in \{1, \ldots, n_{in}\}$, respectively. Due to the no-signaling conditions Eq. (2.1) Eq. (2.2) and normalization $\sum_{a,b} P(a, b \mid x, y) = 1$, the dimension of bipartite correlation is given in Eq. (2.3) $d = (n_{out} - 1)^2 n_{in}^2 + 2n_{in}(n_{out} - 1)$.

Therefore, for minimal parameterization of the elements $\vec{P} \in \mathcal{NS}$, it is convenient to use Collins-Gisin (CG) coordinates [CG04] $\vec{P} = \{P(a \mid x), P(b \mid y), P(a, b \mid x, y)\}_{a,b,x,y}$ where $P(a \mid x)$, $P(b \mid y)$ are marginal probability distributions. For an example $n_{in} = 2, n_{out} = 2$, if $P(a = 1|x = 1, 2), P(b = 1|y = 1, 2)$ and conditional joint probability distribution $P(a = 1, b = 1|x = 1, 2, y = 1, 2)$, respectively.

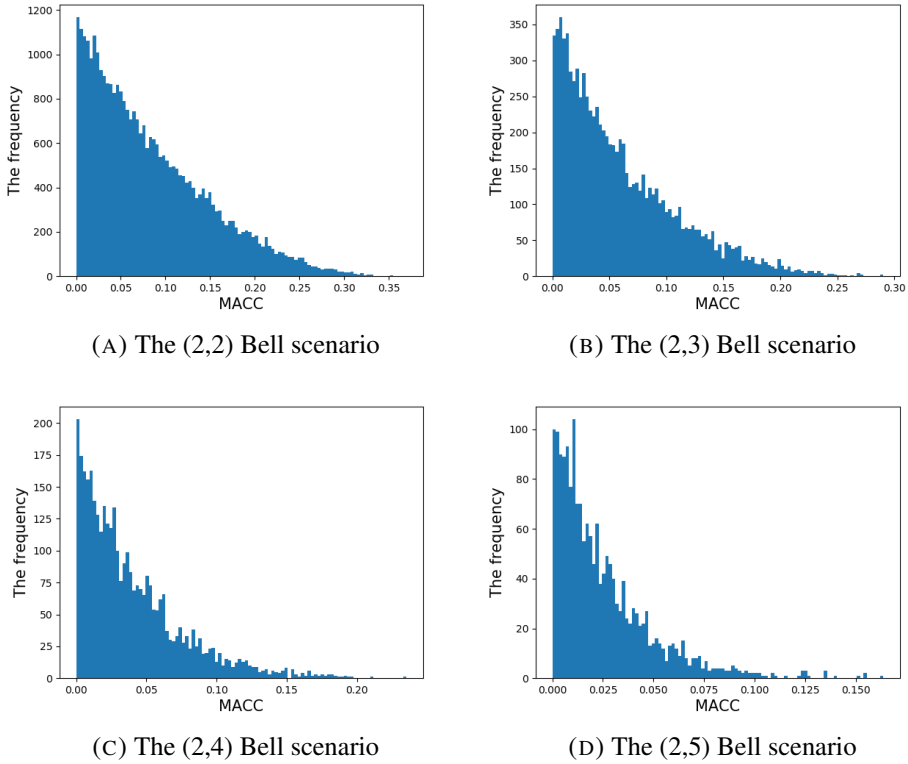
In [LVL21], a large number of correlations were sampled uniformly from the NS space for various Bell scenarios. The authors have further analyzed if each of these samples falls within what they called the tightest quantum approximation to the quantum set and whether they are nonlocal. Here. we compute the MACC for all those correlations from [LVL21] that are nonlocal and are inside their tightest quantum approximations. The number of such (post-selected) samples for each scenario is summarized in Section 3.1 as $N_{\tilde{Q},NL}$.

(n_{in}, n_{out})	$\bar{C}_{P,ave}$	$\bar{C}_{P,max}$	$\bar{C}_{NL,ave}$	$\bar{C}_{NL,max}$	k	$N_{\tilde{Q},NL}(10^4)$
(2,2)	0.0831	0.3699	0.0831	0.3699	4	5
(3,2)	0.1000	0.3863	0.0980	0.3863	3	29
(4,2)	0.1263	0.4190	0.1183	0.3929	2	46
(5,2)	0.1571	0.4226	0.1389	0.3898	2	29
(2,3)	0.0585	0.3077	0.0585	0.3077	3	5.9
(2,4)	0.0394	0.2558	0.0394	0.2558	2	2.58
(2,5)	0.0252	0.1637	0.0252	0.1637	2	0.58

TABLE 3.1: Summary of the numerically estimated MACC $\bar{C}_P$ and the nonlocal content $\bar{C}_{NL}$ obtained from random sampled points $P_{\tilde{Q},NL}$ from NS space in Bell scenario (n_{in}, n_{out}). The goal is to estimate the both measures on average over these points $\bar{C}_{P,ave}$, $\bar{C}_{NL,ave}$ and the maximum $\bar{C}_{P,max}$, $\bar{C}_{NL,max}$ over all post-selected points.

As shown in Table 3.1, the results of the random sampling MACC in the NS sets are given. The analysis is similar to the work of relative volume analysis in [LVL21].

3.1.1 $(2, n_{out})$ case



(A) The (2,2) Bell scenario

(B) The (2,3) Bell scenario

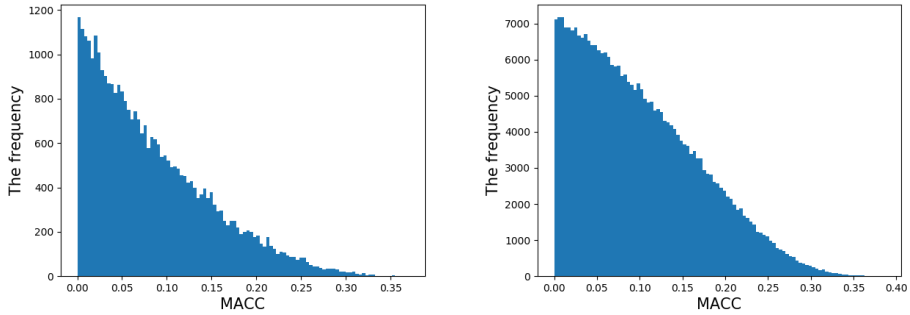
(C) The (2,4) Bell scenario

(D) The (2,5) Bell scenario

FIGURE 3.1: The histogram of MACC of randomly sampled correlations $P_{\tilde{Q},NL}$ from NS space in each Bell scenario $(n_{in}, n_{out}) = (2,3), (2,4), (2,5)$. Each histogram split the Table 3.1 into 100 bars and show the density for each bin of the MACC.

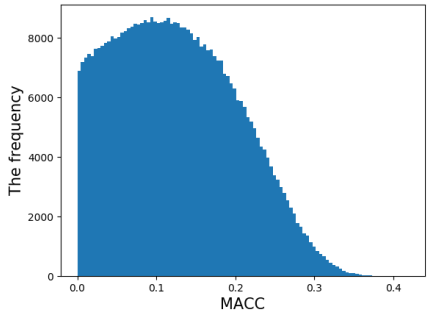
For the case $(2, n_{out})$, we know that one-bit communication can simulate all quantum correlations because one bit is sufficient to simulate a PR box in [CGMP05], which means that one bit is sufficient for the simulation for any 2-input correlation [ZCG19]. In fact, it can be shown [Che22] that there cannot be any communication advantage in these scenarios. In addition, the MACC over the points on average $\bar{C}_{P,ave}$ and nonlocal content $\bar{C}_{NL,ave}$ decrease monotonically as the number of outcomes increases. Furthermore, in the histograms given in Fig. 3.2, we see that in general, the distribution of both measures becomes more local as the outcome number increases. In this case, the histograms can be compared to the relative volume results in Figure 2 of [LVL21], which shows that the fraction of nonlocal points decreases with an increasing number of outcomes. Similarly, for the MACC and EPR2 measures, both decrease strongly from $n_{out} = 2to5$.

3.1.2 $(n_{in}, 2)$ case

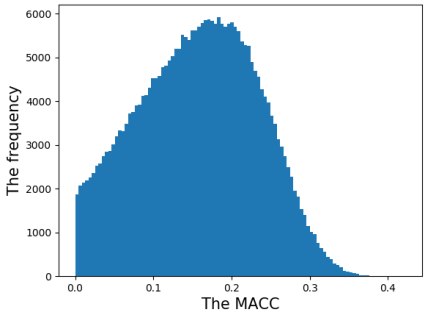


(A) The (2,2) Bell scenario

(B) The (3,2) Bell scenario



(C) The (4,2) Bell scenario



(D) The (5,2) Bell scenario

FIGURE 3.2: The histogram of MACC of randomly sampled correlations $P_{\tilde{Q},NL}$ from NS space in each Bell scenario $(n_{in}, n_{out}) = (3,2), (4,2), (5,2)$. Each histogram split the Table 3.1 into 100 bars and show the density for each bin of the MACC..

Given Table 3.1 when the output number is fixed, the MACC $\bar{C}_{P,ave}$ and the nonlocal content $\bar{C}_{NL,ave}$ increases monotonically with the number of input settings n_{in}. From Figure 2 and Figure 4 of [LVL21], the ratio of nonlocal points increases, but the nonlocal region at certain NPA level decreases in the case $(n_{in}, 2)$. Shown in Table 3.1, the number of nonlocal points at certain NPA level $N_{\tilde{Q},NL}$ increases in $n_{in} = 4$ and then decreases in $n_{in} = 4$. From Fig. 3.2, we qualitatively observe that the overall distribution of MACC increases and concentrates in the middle as the number of input settings n_{in} increases.

The above numerical results tell us something about the MACC and Bell nonlocality in the NS space. In the next section, we discuss what happens specifically in the quantum region, by considering correlations that can be realized with quantum states and measurements. This is the next task: we sample quantum correlations where Alice and Bob's corresponding measurements are related by a unitary transformation.

3.2 Sampling with random unitary operation

In this section, we investigate the quantum region rather than its outer approximation. In the following, we start from a simple bipartite scenario.

Let Alice and Bob share the maximally entangled two-qubit state $\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle)$. Denote the azimuthal angle ϕ and polar angle θ respect to the XZ plane of the Bloch sphere.

1. Alice performs the measurement $A_1 = (0, 0), A_2 = (\pi/2, 0)$ and Bob performs the measurements $UA_1U^\dagger, UA_2U^\dagger$ for a Haar-random unitary U [Mez06]. More specifically, we sample the unitary matrix from the QR decomposition Appendix C.4.1.

2. Repeat the above sampling process to obtain 10^5 nonlocal points.

3. Compute the MACC $\bar{C}_P$ for each sampled point.

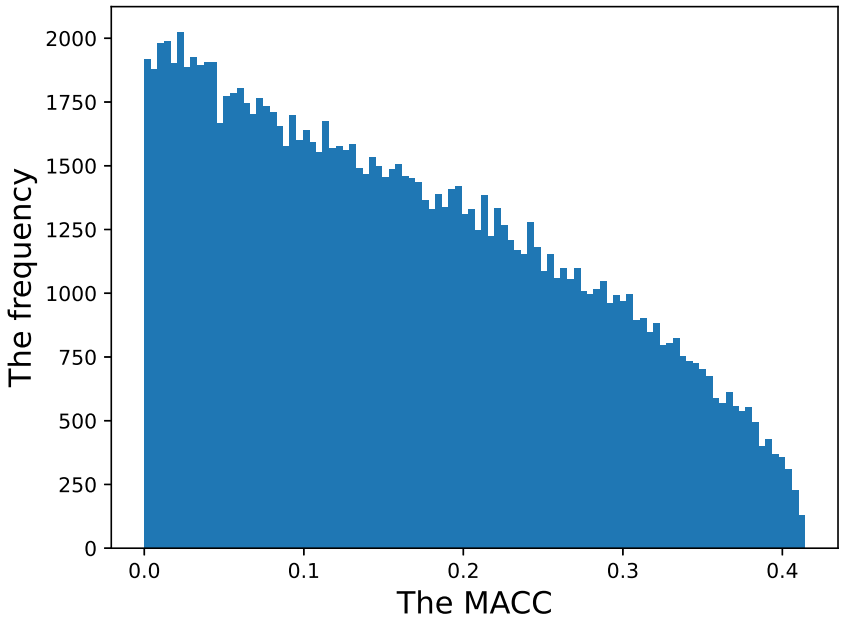


FIGURE 3.3: In (2,2) scenario, follow the unitary operation sampling process. The histogram of MACC split the 124446 nonlocal correlations into 100 bars.

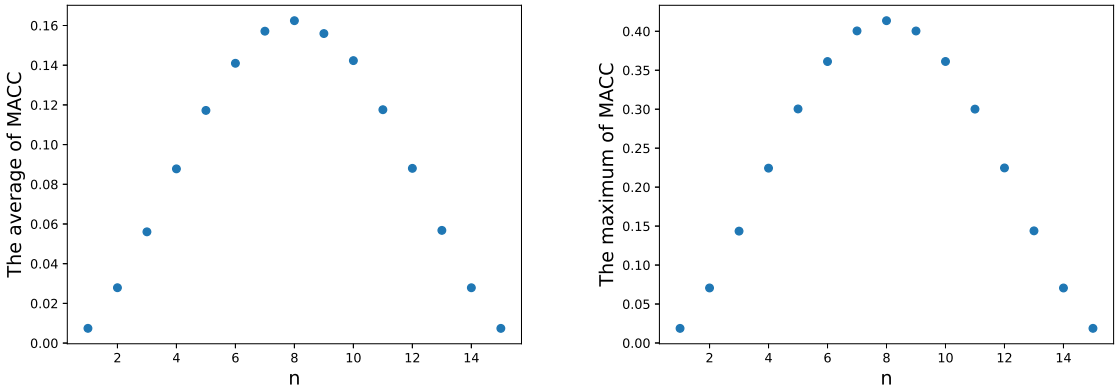
The first goal is to understand the MACC of the quantum correlations after varying the measurements or shared state. In Fig. 3.3, we see that the frequency decreases with increasing MACC. The maximum MACC obtained is 0.4141 bits, which is approximately the same as the known Pironio's bound up to 5 digits in the $(2, 2)$ scenario [Pir03].

Second, we try to find out what is the largest MACC of quantum correlation in each Bell scenario. Compared to Fig. 3.2 (A), this special measurement setting produces a higher average MACC value.

Next, we vary the state or the relative angle δ and discuss the MACC using a similar sampling procedure.

If one wants to find the largest value of MACC in each Bell scenario, we need to do something better than just taking random samples. These other approaches are discussed in the next chapter.

3.2.1 Varying the relative angle of measurements



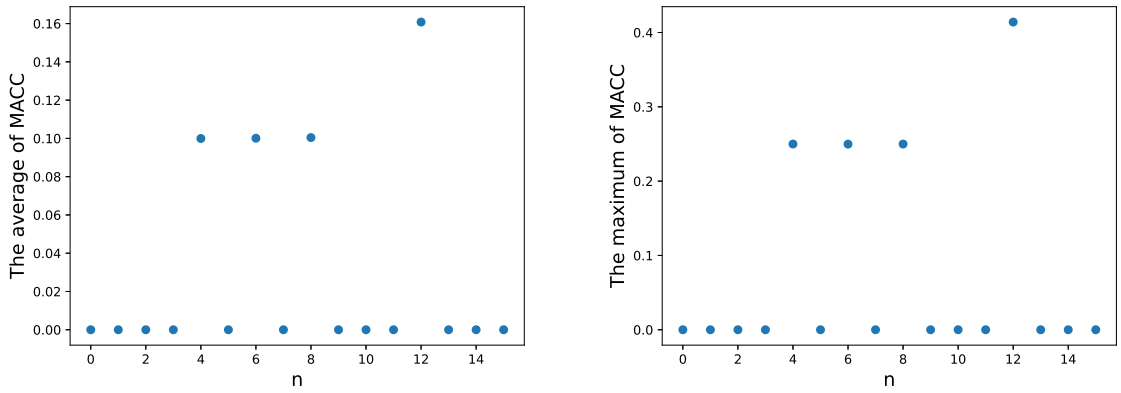
(A) The average of MACC from each δ setting

(B) The maximum of MACC from each δ setting

FIGURE 3.4: The graph of the MACC in averaged (LHS) and maximum (RHS) while varying the relative angle $\delta = n\pi/16$.

One thing we are concerned with is how to choose Alice's initial measurements. Here it is natural to consider a set of mutually unbiased bases (MUBs) [DEBZ10], or measurements that achieve the largest quantum value of a Bell inequality. We begin with the latter.

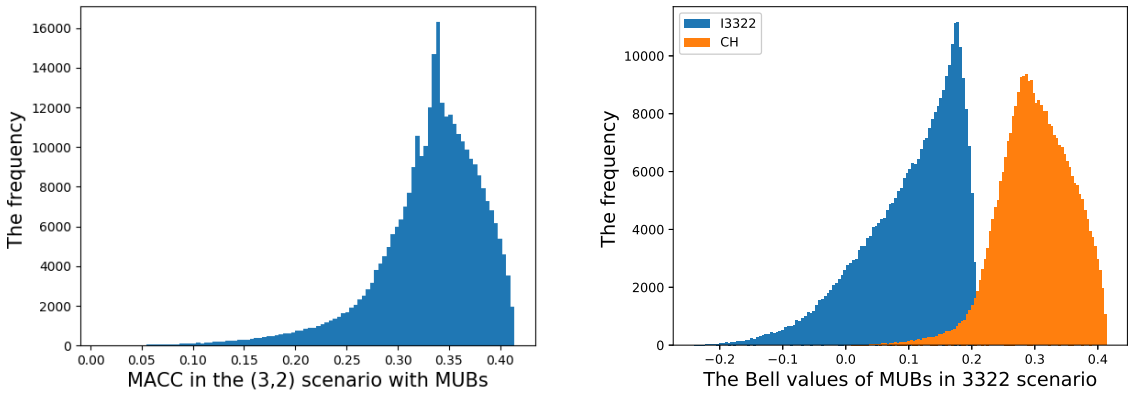
Alice and Bob share the maximally entangled state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and perform the measurement $A_1 = (0, 0), A_2 = (\delta, 0); B_1 = UA_1U^\dagger, B_2 = UA_2U^\dagger$. For each angle δ, we sample 10^5 nonlocal points and compute the MACC. The graph of our results is provided in Fig. 3.7. Our results recover Pironio's bound of 0.4141 bits for the (2,2) scenario. Recall the previous method, we start from the settings of optimal measurements and state of Tsirelson's point Appendix A.1. Next, given the example of (3,2), the preference choice can be selected from the optimal setting of maximum violation.



(A) The average of MACC from each δ setting (B) The maximum of MACC from each δ setting

FIGURE 3.5: The graph of the MACC in averaged (LHS) and maximum (RHS) while varying the opening angle $\delta = n\pi/36$.

In (3,2) scenario, we are inspired by the measurements corresponding to the maximum violation of the I3322 inequality with qubits [CG04]. One possible choice of settings is as follows: Alice chooses the measurements $A_1 = (0, \delta), A_2 = (0, 2\delta), A_3 = (0, 3\delta)$ and Bob chooses measurements that are equivalent up to a local unitary transformation. In this case, we have found that the maximum value comes from $\delta = \pi/3$.



(A) The historgram of MACC (B) The historgram of Bell value

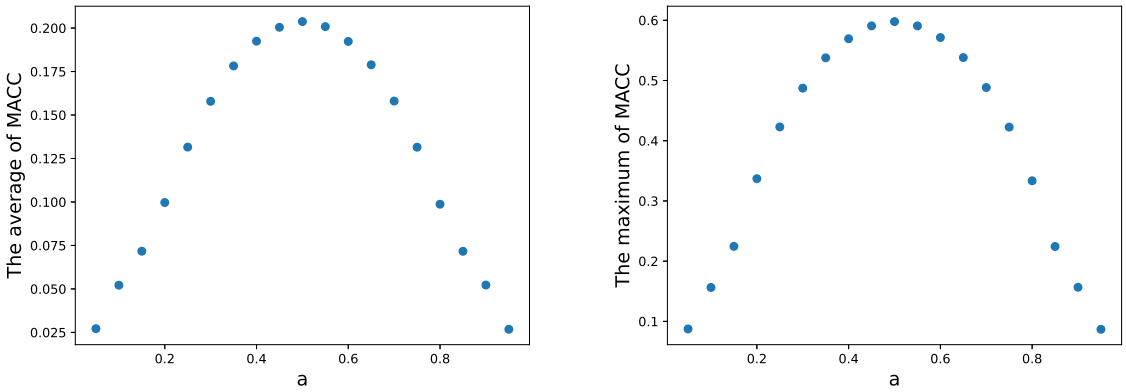
FIGURE 3.6: Both histograms consist of 147840 nonlocal points from MUBs that are divided into 100 bars. The left plot (A) shows the histogram of MACC. The right plot (B) shows the histogram of Bell values violations of I3322 and CH inequalities.

As mentioned above, another choice of measurement settings is a set of MUBs. In the following, we compare the histogram of the violation of the B_{CH} and B_{I3322} Bell inequality with the MACC.

$$B_{I3322} = \max_{\vec{B}\in I_{3322}} \vec{B}\cdot\vec{P}, \quad B_{CH} = \max_{\vec{B}\in I_{CH}} \vec{B}\cdot\vec{P} \tag{3.1}$$

Given a correlation $\vec{P}$, we look for the corresponding maximal violation by considering all possible relabelings. The result in Fig. 3.6 shows the distributions are quite similar, and the maximum of MACC is 0.4141 bits, which is close to MACC of the correlation from CH maximum violation (also from Pironio's bound for CHSH). We also observe all the violation can be estimated by CH inequality in this case. In next subsection, we consider varying the state with a fixed choice of measurements.

3.2.2 Varying the state



(A) The average of MACC from each δ setting (B) The maximum of MACC from each δ setting

FIGURE 3.7: The graph of the MACC in averaged (LHS) and maximum (RHS) while varying the coefficient a of state in $(3, 2)$ scenario.

From $(3, 2)$ case, the largest maximum we found is that relative angle $\delta = \pi/3$ (Also known for from the optimal setting for maximum violation of I3322).

Alice and Bob share the entangled state $(a|00\rangle + b|11\rangle)$ with normalization $a^2 + b^2 = 1$. Alice choose the measurement $A_1 = (0, \delta), A_2 = (0, 2\delta), A_3 = (0, 3\delta)$ while $\delta = \pi/3$. Bob chooses the unitary transformations.

For the example $(2, 2), (3, 2)$ mentioned above, we found that the maximally entangled two-qubit state is the most nonlocal for both the averaged and the maximum sampling value.

To summarize, we consider here specific quantum state and measurements in order to produce a quantum correlation. One of the pros of this is that it ensures that we obtain quantum correlations. However, it becomes computationally demanding and time-consuming to increase the number of settings in Bell scenario. Also, for now we have limited ourselves to looking at a certain preferred set of measurements or state. To develop a more efficient way to find correlations with larger MACC, we utilize the facets of local polytopes and one-dit polytopes.

In the next chapter, different approaches to find the possible largest case will be discussed.

Chapter 4

Lower bounding the largest value of MACC in each Bell scenario

What is the minimum amount of communication (on average) needed to simulate quantum correlations? To address the problem more generally, we introduce a numerical method that explicitly computes a double optimization. Note that all the methods we mentioned below are heuristic and are not guaranteed to find the true optimal point. In other words, to improve the lower bound of MACC, different methods need to be tested. We list all the highest MACC values that we have found in Section 4.5.

4.1 Bounds from the maximal-facet-Bell-inequality violation

We know some connection between the MACC and Bell values, e.g., Pironio's bound [Pir03]. In the following, we look for the largest MACC for quantum correlations by considering some facet inequalities. (See the facet definition in Appendix C.1). In general we adapt the methodology in Fig. 4.1.

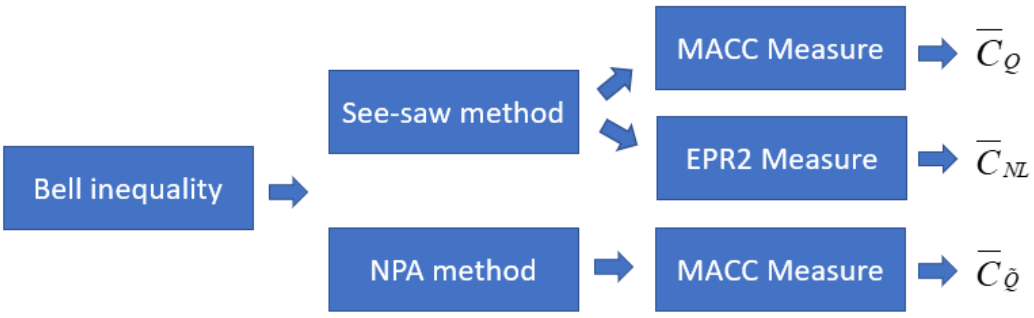


FIGURE 4.1: Methodology: Given a Bell inequality, we obtain the best lower bound and the best upper bound of its maximal quantum violation using, respectively, the seesaw and the NPA method. Then, we compute the MACC of these optimizing correlations and denote them by $\bar{C}_Q$ and $\bar{C}_{\tilde{Q}}$. For the optimizing correlation obtained from the Seesaw method, we also compute its EPR2 cost.

4.1.1 Full and partial facets

$$\bar{C}_Q = \max_P \bar{C}_P, \vec{P} \in Q \tag{4.1}$$

$$\bar{C}_{\tilde{Q}} = \max_P \bar{C}_P, \vec{P} \in \tilde{Q}. \tag{4.2}$$

We know the complete list of facets in some Bell scenarios Section 2.2.2. First, we find the maximum quantum violation of these inequalities using See-saw (Appendix C.3.1) or NPA [NPA08]. Note that these methods refer to a lower bound and an upper bound on the quantum value of the Bell inequalities, respectively. For each candidate optimal point, we compute the MACC $\bar{C}_Q, \bar{C}_{\tilde{Q}}$ and look for the largest value obtained. For these optimization with the facet Bell inequalities, we obtain these inequalities from what is known in the literature Section 2.2.2, or from the *lrs* software [Avi00].

(n_{in}, n_{out})	$\bar{C}_Q$	$\bar{C}_{\tilde{Q}}$	$\|\psi\rangle$	Bell ineq.	$\bar{C}_{NL}$	$\bar{C}_{P,B}$	$\bar{C}_{P,B,NS}$
(2,2)	0.4142	0.4142	MES_2	CHSH [CHSH70]	0.4142	0.4142	0.4142
(2,3)	0.4574	0.4574	$NMES_3$	$CGLMP_3$ [CGL$^+$02]	0.4574	0.4574	0.4574
(2,4)	0.4863	0.4863	$NMES_4$	$CGLMP_4$ [CGL$^+$02]	0.4863	0.4863	0.4863
(3,2)	0.5000	0.4220	MES_2	I3322 [CG04]	0.2509	0.25	0.375
(4,2)	0.6955	0.6955	MES_2	AS1[CG19]	0.6955	0.5857	0.5857
(3,3)	0.5981	0.5981	ME_3	I3333[CC19]	0.5981	0.1495	0.5981

TABLE 4.1: From left to right, we list the results for the known facets given the Bell scenario (n_{in}, n_{out}), the MACC $\bar{C}_Q$ obtained from the see-saw method (explicit quantum strategies), the MACC $\bar{C}_{\tilde{Q}}$ obtained from the NPA hierarchy, the optimal state achieving the bound, the Bell inequality maximally violated by the resulting correlation, the nonlocal content $\bar{C}_{NL}$, the lower bound on the MACC obtained from Pironio's expression $\bar{C}_{P,B}$, and impose the NS condition on Pironio's expression $\bar{C}_{P,B,NS}$. Note that we only have a partial list of Bell inequalities in (2,4),(3,3).

At first, for the binary input cases (2,2) and (2,3), and the binary output cases (3,2) and (4,2), we computed the MACC $\bar{C}_Q$, $\bar{C}_{\tilde{Q}}$ and the nonlocal content with NPA or seesaw optimization method. A partial list of (2,4) inequalities can be obtained using the software *lrs* [Avi00]. We make the following observations:

For the (2,2) and (2,3) cases, the largest MACC value that we have found in this way originates from the maximal quantum violation of the CGLMP inequalities. At the same time, we found that non-local content $\bar{C}_{NL}$ is equal to MACC up to 5 or 6 digits. For the (2,4) case, we have an example of a quantum correlation with MACC larger than the one that maximally violates CGLMP in Section 4.4.

For case (3,2), we observe the MACC gaps $(\bar{C}_Q, \bar{C}_{\tilde{Q}}) = (0.5000, 0.4220)$, for the corresponding I3322 inequality $(B_Q, B_{\tilde{Q}}) = (0.2500, 0.2509)$. This shows that Bell values and MACC represent different measures of nonlocality. In particular, a higher Bell value does not guarantee a higher MACC for any correlation. However, if we consider correlations that produce the same Bell value, Pironio's MACC lower bound $\bar{C}_{P,B}$ can be determined. The largest value of MACC that we found in this Bell scenario is around $0.5981b$its, as we can

see in Section 3.2.2. If this is indeed the true optimal quantum MACC. then it means that the largest case cannot be found using facet Bell inequalities.

Furthermore, the communication advantage can be discussed in this (3,2) example. We have verified that the MACC of one specific NS extremal point (called PR3, see Eq. (C.18)) requires $log_2 3$ bits (1 trit) to simulate. (This coincides with the worst case in the (3,2) scenario [BT03], since it needs to communicate the input information.) For the correlation that yields the maximal qubit violation of I3322, the MACC when Alice and Bob are restricted to 1-bit strategies is $\bar{C}_Q = 0.5$.

$$\begin{aligned} &\min \quad w \\ &\text{subject to} \\ &\qquad \vec{P} = \sum_i c_i^{(0)} \vec{D}_i^{(0)} + w\vec{P}_{PR3} \\ &\qquad \sum_i c_i^{(0)} + w = 1 \end{aligned} \tag{4.3}$$

In addition, we verify that the convex weight w of PR3 is 0.25 in Eq. (4.3). If Alice and Bob share the trit strategies, the MACC with respect to the trit communication polytope will be 0.25 times $log_2 3$, which is approximately $\bar{C}_{Q,trit} = 0.3963$. In this case, it demonstrates the communication advantage.

What this tells us is that if we are concerned about the MACC of some correlation P, then it matters what NS extreme points are included in the NS convex decomposition of P. Beyond that scenario (2,2), more NS extremal points are allowed with different MACC. Based on this reason, one can start from this idea and develop different approaches.

For case (4,2), the largest MACC value that we have found is 0.6955 bits. All the methods we tried approach this value up to 5 or 6 digits of precision. The Pironio's bound $\bar{C}_{P,B}$ we obtained is 0.5857 bits. In another approach described in Section 4.3, we observe that if we employ the "right"inequalities, then the $\bar{C}_{P,B,NS}$ will be the same as the $\bar{C}_P$ and even the same as the nonlocal content $\bar{C}_{NL}$. In other words, we believe that if one can guess the "correct" symmetry for the MACC, one might find the (unique) correlation that gives the largest value.

For the case (3,3), the value found in this case is nothing but the possible largest case in (3,2). We use the facets of the partial list provided in [CC19].

4.1.2 The nonlocal content $\bar{C}_{NL}$ from CGLMP inequalites

n_{out}	$\bar{C}_{NL}$	CGLMP value
3	0.4574	2.915
4	0.4863	2.973
5	0.5079	3.016
6	0.5249	3.050
7	0.5388	3.078
8	0.5506	3.101

TABLE 4.2: For completeness, we list here the nonlocal content found for the correlation maximally violating the CGLMP inequality for $n_{out} \leq 8$ (courtesy of Gelo Tabia [Gel22]).

4.2 Bounds obtained via the visibility program

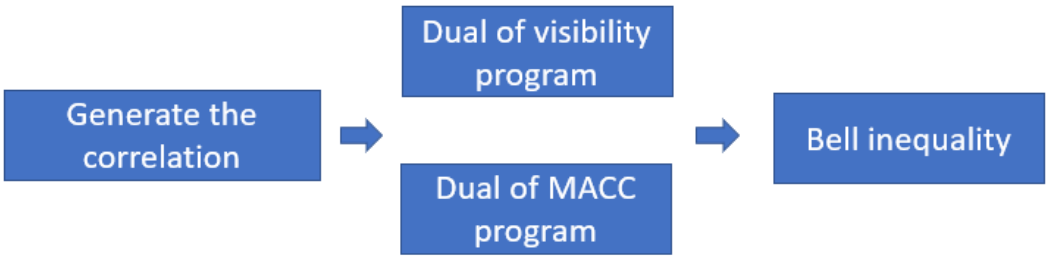


FIGURE 4.2: Methodology: Generate around the 10^5 correlations with random rank-1 projective measurement by the unitarty operations (See Appendix C.4.1) and with normalized Schmidt coefficients. Afterward, the dual of visibility program/MACC program gives the Bell inequality.

To generalize this method of finding Bell inequalities, we can search from the visibility program in Appendix C.2.1. To be precise, first we generate the Bell inequality through Fig. 4.2 and then go via the process of Fig. 4.1.

(n_{in}, n_{out})	$\bar{C}_Q$	$\bar{C}_{\tilde{Q}}$	$\|\psi\rangle$	$\bar{C}_{NL}$	$\bar{C}_P$	$\bar{C}_{P,NS}$
(2,2)	0.4142	0.4142	MES_2	0.4142	0.4142	0.4142
(2,3)	0.4574	0.4574	$NMES_3$	0.4574	0.4574	0.4574
(3,2)	0.5000	0.4220	MES_2	0.2509	0.25	0.375
(3,3)	-	0.5671	-	0.5671	-	-
(4,2)	0.6955	0.6955	MES_2	0.6955	0.5857	0.5857
(4,3)	-	-	-	0.6184	-	-
(5,2)	0.5965	0.5965	MES_2	0.3028	-	-

TABLE 4.3: From left to right, we list the results searching from visibility program and obtain the Bell inequalities from the dual problem: the MACC $\bar{C}_Q$ obtained from see-saw method, the MACC $\bar{C}_{\tilde{Q}}$ obtained from the NPA hierarchy, the optimal state achieving the bound from see-saw, the Bell inequality maximally violated by the resulting correlation, the nonlocal content $\bar{C}_{NL}$, the lower bound on MACC obtained from Pironio's expression $\bar{C}_{P,B}$, and impose the NS condition on Pironio's expression $\bar{C}_{P,B,NS}$.

Compared to Section 4.1.1, it reproduced the results of $(n_{in}, 2)$ and generate the results in other scenario. Starting from the facet inequalities is the natural choice, and we also search from the 1-bit facets in the next subsection.

4.3 Bounds from the maximal quantum violation of one-bit-augmented Bell-type inequalities

4.3.1 Full and partial facets

(n_{in}, n_{out})	$\bar{C}_Q$	$\bar{C}_{\tilde{Q}}$	$\|\psi\rangle$	One-bit ineq.	$\bar{C}_{NL}$	$\bar{C}_P$	$\bar{C}_{P,NS}$
(2,2)	0.4142	0.4142	MES_2	I2222+1[BT03]	0.4142	0.4142	0.4142
(3,2)	0.5981	0.5981	MES_2	I3322,288 [BT03][ZCG19]	0.5981	0.2990	0.5981

TABLE 4.4: From left to right, we list the results searching from the full list of one-bit inequalities, the MACC $\bar{C}_Q$ obtained from the seesaw method, the MACC $\bar{C}_{\tilde{Q}}$ obtained from the NPA hierarchy, the optimal state that achieves the bound from the see-saw, the one-bit inequality maximally violated by the resulting correlation, the non-local content $\bar{C}_{NL}$, the lower bound on MACC obtained from Pironio's expression $\bar{C}_{P,B}$, and impose the NS condition on Pironio's expression $\bar{C}_{P,B,NS}$.

In this section, we consider the facets of the one-bit-augmented Bell polytope (see Appendix B.2 for further details about such polytopes Appendix B.2). The complete list of one-bit facets full-correlation inequalities constructed from the CPD model of Section 2.3.1 is known in the $(2, 2)$ and $(3, 2)$ scenarios in [BT03]. Later, the one-bit facets of the bidirectional

MCPD Section 2.3.1 model in the asymmetrical scenario $(222, 22)$ is found in [MC14] (In asymmetrical scenario, each number represents the number of outcomes for each input on Alice/Bob side). In the (3,2) scenario, 668 one-bit facets are known [ZCG19], and the authors even conjectured that they give a complete description of the one-bit-augmented Bell polytope.

In (2,2), the value we obtain is 0.4141 which is similar to the CHSH case. In (3,2) case, the largest MACC value found originates from I3322+1 inequality of NO.288 [ZCG19]. This case matches numerically with the random sampling results of Section 3.2.2. Also, we found that the lower bounds $\bar{C}_{NL}$, $\bar{C}_{P,B,NS}$ match the MACC $\bar{C}_P$. In other words, we found the right inequality and no communication advantage for the corresponding correlation.

4.3.2 Bounds obtained via the primal-dual program of MACC

Applying vertex enumerations from software is, in fact, computationally demanding. On the other hand, we construct the double optimization approaches with the dual problem from the MACC program. In the following, we will introduce this procedure for the double optimization involving the procedures in Fig. 4.2.

1. Randomly generates a correlation P.
2. Find the MACC solving Eq. (C.7) in Appendix C.2.2. Obtain the MACC $\bar{C}_P$ and the inequalities B.
3. Maximize the quantum violation of the inequality B by see-saw/NPA methods.
4. Feed the optimizing correlation P from the last step to step 2 until $\bar{C}_P$ converges.
5. Repeat step 1- step 4 for over 10^5 correlations P in each Bell scenario.

(n_{in}, n_{out})	$\bar{C}_Q$	$\bar{C}_{\tilde{Q}}$	$\|\psi\rangle$	One-bit ineq.	$\bar{C}_{NL}$	$\bar{C}_{P,B}$	$\bar{C}_{P,B,NS}$
(2,2)	-	0.4142	-	-	-	-	-
(2,3)	-	0.4142	-	-	-	-	-
(2,4)	-	0.4142	-	-	-	-	-
(2,5)	-	0.4142	-	-	-	-	-
(3,2)	0.5981	0.5981	MES_2	I3322+1	0.5981	0.2990	0.5981
(3,3)	-	0.5981	-	-	-	-	-
(4,2)	0.6955	0.6955	MES_2	I4422+1	0.6955	0.1200	0.6955
(4,3)	-	0.6181	-	-	-	-	-
(5,2)	0.7553	0.7553	MES_2	I5522+1	0.7553	0.1259	0.7553

TABLE 4.5: From left to right, we list the results searching from the MACC program and obtain the one-bit inequalities from the dual problems, the MACC $\bar{C}_Q$ obtained from see-saw method, the MACC $\bar{C}_{\tilde{Q}}$ obtained from the NPA hierarchy, the optimal state achieving the bound from see-saw given in the fourth column, the one-bit inequality maximally violated by the resulting correlation, the nonlocal content $\bar{C}_{NL}$, the lower bound on MACC obtained from Pironio's expression $\bar{C}_{P,B}$, and impose the NS condition on Pironio's expression $\bar{C}_{P,B,NS}$.

Compared with the double optimization (where we choose a certain parametrization of both state and measurements) that used the MATLAB fminunc function, the advantage of this approach is that it allows us to explore efficiently a huge number of local extrema of MACC. However, for (2,3), (2,4), (2,5), we only found the trivial case (0.4141 as we mentioned in the previous subsection). For (3,n_{out}) and (4,n_{out}), we also did not find anything larger than that found for $n_{out} = 2$. Despite these discouraging results, this approach did lead us to the largest value of MACC found in binary (n_{in},2) setting. For (3,2), (4,2), (5,2) binary output scenario, all candidates for the largest MACC that we found are due to a maximally entangled two-qubuit state. There are several properties from these largest cases that are worth mentioning. For the largest case found by the one-bit inequality, the lower bound $\bar{C}_{P,B,NS}$ matches the MACC $\bar{C}_Q$, $\bar{C}_{\tilde{Q}}$, even for the nonlocal content $\bar{C}_{NL}$. This suggests that the one-bit-augmented Bell-type inequality is the right inequality to explore when looking for the largest MACC value in these scenarios. Furthermore, in this case it is possible to do it in reverse order, where from the correlation we try to guess the inequalities to reproduce the same quantum bound we have found. The details are discussed in Section 4.4. In general, finding the largest possible case is a double optimization problem. We end this part with a conjecture:

Conjecture 1. *For the (n_{in},2) binary case, if m is an even number, the largest MACC can be obtained from both optimization of the 1-bit and local facets. If m is the odd number, it can be obtained from the 1 bit facets.*

4.3.3 Symmetry arguments

For the search of the largest obtained so far, several symmetry constraints have been imposed, partly motivated by numerical results from the simplest Bell scenarios:

- The symmetry of party exchange: The largest case should be observer-free. This means that the correlation after the two parties exchange should have the same MACC value.

- The symmetry of measurement exchange: The quantum correlation lies within in the NS polytope and therefore satisfies NS conditions for both parties. This means that the order of measurements (inputs) does not matter of the nonlocality measure.

We should note that we have also performed calculations without these constraints, and so far they have not yielded better candidates.

Notice that the correlation giving the largest MACC value found is not unique. But up to relabeling, is it unique? More importantly, it can be decomposed into different NS extreme points, and the communication cost would correspond to the convex decomposition with just the right amount of nonlocal contributions. For example in the (3,2) scenario, a general nonlocal quantum correlation would have a decomposition with nonzero weights for the PR Eq. (C.19) and PR3 Eq. (C.18) NS extremal points. If we compute MACC with respect to the 1-bit polytope, the largest case obtained has 2 different PR components. In some sense, this is because PR is the "cheaper" option compared to PR3 when considering the communication cost. However, this leaves no room for PR3, because if some of the PRs can be replaced by PR3, this may well produce a new decomposition with less cost even thought it is more costly to simulate PR3 itself. Because PR2 require 1 bit of communication to simulate, this results in no communication advantages. But in our numerical observations, the correlation from the largest value case of MACC shares the same value of its EPR2 cost in every Bell scenario that we explored. Therefore, one of our approaches is to search for the correlation with the largest cost of EPR2 and then check its MACC afterward.

However, this might not find the largest quantum MACC because the maximum of EPR2 cost can't exceed than 1. But we know that there exists a correlation from a ququart system with infinite input settings that violates the 1-bit bound [VB09]. If one can know the geometry between the largest case and these NS extremal points, it is helpful to understand the relation of the nonlocal resources. In this manner, in the next section, we will also try to construct the Bell inequalities inspired by the observation of the correlations from the largest case and also by the construction of PR boxes.

4.4 Some analytic bounds

In general, our goal is to find the largest MACC from the for quantum probability distributions $\vec{P}$.

$$\begin{aligned} &\max_{\vec{P}}(\min \quad \sum_i \sum_{j\geq 1} c_i^{(j)}) \\ &\text{subject to} \\ &\qquad \vec{P} = \sum_{j=0}^{d} \sum_i c_i^{(j)} \vec{D}_i^{(j)}, \\ &\qquad c_i^{(j)} \geq 0 \ \forall \ i, \\ &\qquad \sum_{i,j} c_i^{(j)} = 1. \end{aligned} \tag{4.4}$$

First, we develop the double optimization program using the MATLAB fminuc function [Gel22]. See the details of the algorithm in Appendix C.4. For the algorithm, we consider a particular parametrization of states and rank-1 projective measurements. The inner loop is the computation of the MACC for any given $\vec{P}$ while the outer loop optimizes over $\vec{P}$ by varying the state and measurements. Because of the success in employing a symmetrical form in the $(n_{in}, 2)$ Bell scenario (see Section Section 3.2), we assume that Alice chooses the measurement M_A and for Bob's side is $UM_AU^\dagger$. In the following, we give the best value of MACC $\bar{C}_P$ with 10^5 correlations found in this way.

(n_{in}, n_{out})	$\bar{C}_P$	$\|\psi\rangle$	$\bar{C}_{NL}$
(2,2)	0.4141	MES_2	0.4141
(2,3)	0.4574	NMES	0.4574
(2,4)	0.4874	NMES	0.4874
(2,5)	-	NMES	0.5072
(3,3)	0.6456	NMES	0.6456
(3,4)	1	MES_4	1
(3,5)	-	NMES	0.7688
(4,3)	-	NMES	0.7366
(4,4)	-	MES_4	1
(5,3)	-	NMES	0.6456
(5,4)	-	MES_4	1

TABLE 4.6: From left to right, we list the results of the double optimization of nonlocal content in the symmetric measurement setting: MACC $\bar{C}_P$, the optimal state achieved in the optimization, and the nonlocal content $\bar{C}_P$.

From the observation of the symmetrical form of correlation for the possible largest quantum MACC in the binary output scenario (n_{in},2), we determine its corresponding measurements and state [Gel22] to be the following:

- All the state is the maximally entangled state in binary case.
- All measurements are in the same plane.

For example, in the (3,2) scenario, denote the azimuthal angle with respect to the plane.

$$A = (\pi/6, \pi/3, \pi/2) \tag{4.5}$$

$$B = (5\pi/12, 7\pi/12, 3\pi/4) \tag{4.6}$$

Furthermore, for case (4,2)

$$A = (0, \pi/8, \pi/4, 3\pi/8) \tag{4.7}$$

$$B = (3\pi/16, 5\pi/16, 7\pi/16, 9\pi/16) \tag{4.8}$$

Therefore, we make the following conjecture for the $(n_{in}, 2)$ scenario. For this we denote the relative angle between the parties θ and the relative angle between the measurements δ.

Conjecture 2. *In $(n_{in}, 2)$ scenario, the largest case can be obtained from the maximum entangled state $\psi = \frac{1}{\sqrt{2}}(|00> + |11>)$ with projective measurements*

$$\theta = (n_{in} - 1)\tfrac{\pi}{2n_{in}}$$
$$\delta = \tfrac{\pi}{4n_{in}} + 2((n_{in} + 1) \quad mod \quad 2)$$

4.4.1 The construction of Bell inequality: $(n_{in}, 2)$ scenario

Much to our surprise, correlations leading to the largest MACC found in these scenarios show some elegant symmetry. Moreover, their MACC coincide with their nonlocal content. To see the symmetry, we write the correlation $P(a, b|x, y)$ in a matrix form:

$$P = \begin{pmatrix} \mathsf{P}_{0,0} & \mathsf{P}_{0,1} & \cdots & \mathsf{P}_{0,d-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{P}_{d-1,0} & \cdots & \cdots & \mathsf{P}_{d-1,d-1} \end{pmatrix}, \tag{4.9}$$

where $\mathsf{P}_{x,y}$ is a 2×2 matrix collating the entries $P(a, b|x, y)$ for the input pair (x, y).

Then depending on whether d is even or odd, P has zeros at the same positions as that shown in the following block matrices:

$$\beta_{\text{even}} = \begin{pmatrix} D & D & D & \cdots & D \\ A & D & D & \cdots & D \\ A & A & D & \cdots & D \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A & A & A & \cdots & A \end{pmatrix}, \ \beta_{\text{odd}} = \begin{pmatrix} U & D & D & \cdots & D \\ A & U & D & \cdots & D \\ A & A & U & \cdots & D \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A & A & A & \cdots & A \end{pmatrix}. \tag{4.10}$$

where

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \tag{4.11}$$

Interestingly, for $n_{in} \leq 11$, we observed that the optimal decomposition of the correlation P that appears in its EPR2 decomposition involves a NS extreme point that has exactly the same structure shown in Eq. (4.10). Remembering that these correlations originate from a Bell state, our observations naturally relate to the following question from [VB09]:

Whether 2 PR-boxes were enough to simulate two-outcome projective measurements on maximally entangled qudits?

For two-outcome projective measurements on maximally entangled qubits, two PR boxes can simulate the corresponding correlations giving the largest MACC found.

Finally, if we want the form of the Bell inequality that gives directly the MACC as a function of the number of settings n_{in}, we can consider the solution to the dual problem of the linear program for computing the EPR2 decomposition, which yields Bell coefficients that take the form of a block Toeplitz matrix (diagonal-constant matrix):

$$\beta_{\text{even}} = \begin{pmatrix} A & U & \cdots & \cdots & \cdots & U & D \\ A & A & \cdots & \cdots & \cdots & U & U \\ U & A & A & \cdots & \cdots & U & U \\ U & U & A & \ddots & \cdots & \vdots & U \\ \vdots & \vdots & \cdots & \ddots & A & \vdots & \vdots \\ U & U & \cdots & \cdots & A & A & U \\ U & U & \cdots & \cdots & U & A & A \end{pmatrix}, \; \beta_{\text{odd}} = \begin{pmatrix} U & U & \cdots & \cdots & U & D & D \\ A & U & \cdots & \cdots & \cdots & U & D \\ A & A & U & \cdots & \cdots & U & U \\ U & A & A & \cdots & \cdots & \vdots & U \\ \vdots & \vdots & \cdots & \ddots & U & \vdots & \vdots \\ U & U & \cdots & \cdots & A & U & U \\ U & U & \cdots & \cdots & A & A & U \end{pmatrix}, \tag{4.12}$$

The optimal measurement of the qubit system is denoted by the azimuthal angle θ_k and δ_k in same plane, where $k = 1, 2, \ldots, n_{in}$:
If n_{in} is even:

$$\theta_k = (k-1)\pi/2n_{in} \tag{4.13}$$
$$\delta_k = \theta_k + (2n_{in}+1)\pi/4n_{in} \tag{4.14}$$

If n_{in} is odd:

$$\theta_k = k\pi/2n_{in} \tag{4.15}$$
$$\delta_k = \theta_k + (2n_{in}-3)\pi/4n_{in} \tag{4.16}$$

For example, the quantum maximum violation as a function of the number of measurement settings m:

$$B_2 = 2 - \sqrt{2} \approx 0.5858 \tag{4.17}$$
$$B_3 = 3 - \frac{3\sqrt{3}}{2} \approx 0.4019 \tag{4.18}$$
$$B_4 = 4(1 - \cos(\pi/8)) \approx 0.3045 \tag{4.19}$$
$$\tag{4.20}$$

It can be shown that the Bell inequality of tr$(\beta_{even/odd}P)$ from Eq. (4.12) has a local lower bound of 1 and 1-bit lower bound of 0.

The Pironio's bound as defined in Eq. (2.17), we obtain

$$\bar{C}_{n_{in}} = 1 - B_{n_{in}} \tag{4.21}$$

In general, the quantum bound can be written as

$$B_{n_{in}} = n_{in}(1 - cos(\pi/2n_{in})), \tag{4.22}$$

and Pironio's bound on the MACC is given by

$$\bar{C}_{n_{in}} = 1 - B_{n_{in}} = 1 - n_{in}(1 - cos(\pi/2n_{in})). \tag{4.23}$$

When we take n_{in} to infinity, the Pironio's bound is $\bar{C}_{\infty} = 1$.

4.4.2 Explore the symmetrical form in (3,3) scenario

As mentioned above, we can also consider the solution to the dual problem of the linear program for computing the decomposition of EPR2.

In this scenario, we observe that a symmetric Bell inequality, invariant under the exchange of two parties, can be constructed for the largest MACC that we have found for this scenario. Again, as above, we express the Bell coefficients using a block matrix:

$$B_{33} = \begin{pmatrix} A_{11} & A_{21} & U \\ A_{12}^T & U & A_{23} \\ U & A_{23}^T & A_{33} \end{pmatrix} \tag{4.24}$$

where

$$A_{11} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, A_{21} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, A_{23} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}, A_{33} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \tag{4.25}$$

The resulting Bell inequality has a local lower bound of 0.4544. It can be verified that the maximum violation of this inequality satisfies an equation analogous to Eq. (4.21), thus giving a MACC value of 0.6456, which coincides with that presented in Table 4.6 (for double optimization).

4.4.3 Elegant example in (3,4)

In the (3,4) scenario, we have found a quantum correlation whose MACC and non-local content are exactly 1. This means that it is a quantum correlation lying on the no-signaling boundary. Let,

$$M_1 = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}, \tag{4.26}$$

$$M_4 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \quad M_5 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad M_6 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \tag{4.27}$$

$$M_7 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \qquad M_8 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}. \tag{4.28}$$

The said correlation can then written in the block matrix form as

$$\frac{1}{8} \begin{pmatrix} M_1 & M_2 & M_3 \\ M_3 & M_4 & M_5 \\ M_6 & M_7 & M_8 \end{pmatrix}. \tag{4.29}$$

We can verify that this is a quantum correlation since it can be obtained from the state $|\psi\rangle = (|00\rangle + |11\rangle + |22\rangle + |33\rangle)/2$, with rank-one projective measurements summarized by the following matrices from Alice side:

$$A_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \tag{4.30}$$

and similarly for Bob's measurements:

$$B_1 = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad B_3 = \begin{pmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}. \tag{4.31}$$

In the above equations, the j-column of the matrix A_x gives the (unnormalized) eigenvector corresponding to the projector $\Pi^{\mathrm{A}}_{j|x}$ for the j-th outcome of Alice's x-th measurement. As an

explicit example, the projector for the 3rd outcome of Alice's 2nd measurement is

$$\Pi^{\mathrm{A}}_{3|2} = \frac{1}{4}\begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \end{pmatrix}(1\ -1\ 1\ 1) = \frac{1}{4}\begin{pmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{pmatrix} \tag{4.32}$$

Bob's projectors for his y-th measurement can be similarly extracted from B_y.

4.4.4 The largest case in (2,4)

In previous work in Section 4.1.1, we have the conjecture for (2,n_{out}) scenario that the largest case can be found by the CGLMP inequalities. In Section 4.4, we have found the nonlocal content is 0.4874 bits. It is already larger than we found in CGLMP case 0.4863 bits. This shows that the conjecture that the largest case can be obtained from the CGLMP is not true. Moreover, we also optimize over the Bell inequalities by fminuc. One need to choose the good initial Bell inequalities with the NPA method to obtain a correlation with larger nonlocal content. By choosing the inequality given by the dual program of the nonlocal content when using these best points we have. Finally, we have found the best lower bound for largest MACC to be 0.4882 bits.

4.5 Summary of the largest MACC found

In this section, we collect and list the largest value of MACC that we have found so far.

(n_{in}, n_{out})	$\bar{C}_Q$	$\bar{C}_{\tilde{Q}}$	$\|\psi\rangle$	Ineq or methods	$\bar{C}_{NL}$	$\bar{C}_{P,B}$	$\bar{C}_{P,B,NS}$
(2,2)	0.4142	0.4142	MES_2	CHSH, I(2,2)+1	0.4142	0.4142	0.4142
(2,3)	0.4574	0.4574	$NMES_3$	$CGLMP_3$	0.4574	0.4574	0.4574
(2,4)	0.4882	0.4882	$NMES_4$	Dual of EPR2	0.4882	-	-
(2,5)	0.5079	0.5079	$NMES_5$	$CGLMP_5$	0.5079	0.5079	0.5079
(2,6)	0.5249	0.5249	$NMES_6$	$CGLMP_6$	0.5249	0.5249	0.5249
(2,7)	0.5389	0.5389	$NMES_7$	$CGLMP_7$	0.5389	0.5389	0.5389
(2,8)	-	-	$NMES_8$	$CGLMP_8$	0.5506	0.5506	-
(2,∞)	-	-	-	$CGLMP_\infty$	-	1	-
(3,2)	0.5981	0.5981	MES_2	I(3,2)+1	0.5981	0.2990	0.5981
(3,3)	0.6456	-	NMES	Double optimization	0.6456	-	-
(3,4)	1	-	MES_4	Double optimization	1	-	-
(4,2)	0.6955	0.6955	MES_2	AS1, I(4,2)+1	0.6955	0.1200	0.6955
(4,3)	0.7447	0.7447	NMES	Dual of EPR2	0.7447	0.7447	-
(4,4)	-	-	MES_4	Double optimization	1	-	-
(5,2)	0.7553	0.7553	MES_2	I(5,2)+1	0.7553	0.1259	0.7553
(5,3)	0.8293	0.8293	NMES	Dual of EPR2	0.8293	0.8293	-
(5,4)	-	-	MES_4	Double optimization	1	-	-
(∞,2)	-	-	MES	New (n_{in},2) family	-	1	-

TABLE 4.7: From left to right, we list Bell scenario, the MACC $\bar{C}_Q$ for the correlation optimizing (by Seesaw) the violation of the Bell inequality from the 5th column, the MACC $\bar{C}_{\tilde{Q}}$ obtained for the correlation optimizing (by NPA hierarchy) the violation of the Bell inequality from the 5th column, the nature of the optimal state achieving value in the second column, the method used to obtain the correlation, the nonlocal content $\bar{C}_{NL}$ of the correlation optimizing the Bell violation from the 5th column, the lower bound on MACC obtained from Pironio's expression $\bar{C}_{P,B}$, and a tightened lower bound based on the Bell value $\bar{C}_{P,B,NS}$ see also Figure Fig. 4.1. Note that the maximum violation $(2, n_{out})$ can be known from [ZRGW10] when n_{out} goes to infinite, therefore the Pironio's bound is obtained from Eq. (2.17).

Chapter 5

Conclusion

The 'Second quantum revolution' is concerned with the development of quantum technologies, for applications such as quantum computing, quantum cryptography and quantum communications. As a researcher interested in quantum foundations, we use our understanding of essential features of quantum theory to help us develop models for quantum information processing. Someday, I hope that some of our theoretical results will prove useful also in the construction of real and practical quantum computers.

In this thesis, we first considered correlations obtained from measurement directions that lie on the plane of the Bloch sphere. For the case of n_{in} inputs and two outputs, we found that such settings give the best MACC lower bound to date. This coincidence might be related to some observations involving Platonic solids in [PV22]. In that work, the bases leading to maximum violation of Platonic-solid-based Bell inequalities in the two-dimensional case are also located in the plane of the Bloch sphere. In some sense, one of the important goals in this thesis is an attempt to address the following question:

What is the most nonlocal quantum correlation?

In many situations, nature demonstrates its symmetry. The simple construction of Bell inequalities for these largest MACC values known to date is fascinating in its mathematical beauty. To this end, we utilize and optimize the unitary transformation. Further, we can ask the following general question:

What is the minimum classical communication required to simulate all quantum correlations?

Analytically, the maximum quantum violation of certain Bell inequalities can be related to the lower bound of communication cost. For example, the maximum violation of the CGLMP inequality in infinite dimension is discussed in [ZRGW10], therefore, the Pironio bound results in 1. However, in $(2, 4)$ we have found a case larger than CGLMP, so it leads to the question: What is the minimum outcome number such that the bound will be 1?

Very recently, it has been shown in [RQ22] ([RTQ22]) that one trit (two bits) is sufficient for the classical simulation of correlations arising from *any* partially entangled two-qubit state if we consider projective measurements. (general positive-operator-valued measures). Our result for the two-qubit case complements the result of [TB03], showing that one classical bit is not only sufficient but also necessarily required for the classical simulation of some nonlocal quantum correlation arising from a Bell pair $|\Psi_2^+\rangle$ in *every single round* of the simulation.

It is also worth comparing our results for the case of two-ququart maximally entangled state $|\Psi_4^+\rangle$ with those given in [VB09]. There, it was shown that by considering an infinite number of measurements, i.e., with $n_s \to \infty$, one bit is insufficient (in the worst case) to classically simulate the correlations arising from such a state. On the contrary, our result indicates that, already for $n_{in} = 3$, one classical bit is already required on average (in fact, even in every round) of the simulation. Our result therefore provides further circumstantial evidence to the insufficiency of any 1-bit protocol in a Bell scenario involving finite n_{in} measurements.

5.1 Future work

5.1.1 Other approaches

In this thesis, we have found some nontrivial lower bound on largest MACC value in each Bell scenario. Intuitively, one expects to find the largest MACC or EPR2 case by the maximum violation of some Bell inequality. In general, however, we do not know whether these lower bounds summarized in Section 4.5 are tight. A systematic approach to find the upper bound (like how the NPA method can be used to upper bound the maximal quantum violation) is desirable.

In other words, all the numerical approaches we have tried, at best, guarantees a local extrema, so there is no guarantee that the global optimal solution has been found. We have made several conjectures, and a fully analytic, rigorous approach is desirable. For example, discussions of local content are made in [PBG12]. Without loss of generality, a two qubit pure state is given by

$$|\psi(\theta)\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle \tag{5.1}$$

In [PBG12] the authors verify Scarani's conjecture [Sca08] that the upper bound of local content is $\cos 2\theta$ for all pure two-qubit states. Our numerical method provides, for any Bell scenario, a lower bound on the largest value of MACC and nonlocal content. If one can find an upper bound like in [PBG12], it may help to show that the bound is tight, at least, in specific Bell scenarios.

5.1.2 The MACC with random mutual unbiased bases

As mentioned in Section 3.2, one can extend this case to another Bell scenario. A similar analysis is proposed in [TBYL22], where it is shown that there is a significant chance of obtaining a Bell violation from measuring random mutually unbiased bases (MUBs).

5.1.3 Generalization in multipartite Bell scenario

Depending on the purposes, different communication models can be constructed. The correlation can be explained by different communication models and becomes more complicated in the multipartite Bell scenario.

For example, one of the versions of EPR2 decomposition is defined in [ACSA10] to detect genuine multipartite full-nonlocal correlations. Further, this is one example of the inspiration of the Baysian's game related to Bell scenario[BL13], the multistage games in the multipartite Bell scenario can be discussed. Moreover, a communication protocol is related to sampling probability distributions when the separated parties are not allowed to communicate among themselves except with a leader [BDG19]. Very recently, some nontrivial results have revealed that the quantum correlation produced by the graph state shows that the successive edge communication protocol cannot reproduce [MGM22].

In Chapter 4, we explore the various methods to find the largest MACC or largest nonlocal content case. One can generalize these methods and explore them in the multipartite Bell scenario.

Appendix A

Details pertinent to Chapter 1

In this chapter, leave here some technical details of the models to clarify the idea mentioned in the preliminary chapter Chapter 2.

A.1 Different quantifications of nonlocality

How to quantify the nonlocality and distinguish the difference between the quantum and classical systems are key questions. Typically, starting with finding the maximum violation of the Bell inequalities is a preferable choice. In other words, the distinction between the target correlation P_t and the set of local correlations can be derived in various ways. For example: given the correlation, define the nonlocal measure D as the Euclidean distance between the target correlation P_t and all of the local extremal points $P_{L,i}$ with the indexes i in the following.

$$D(P) = \frac{1}{N_L} \sum_{i}^{N_L} |P_t - P_{L,i}| \tag{A.1}$$

where $N = |x||y||a||b|$ counts for the cardinality of random variables and N_L for the number of local extremal points. D_1 is measured between the target correlation P and all local extremal points P_L. However, we will show that this is not a good quantification in this case.

If we consider the no-signaling extremal point (PR box) P_{PR}

		$x=0$		$x=1$	
		0	1	0	1
$y=0$	0	0.5	0	0.5	0
	1	0	0.5	0	0.5
$y=1$	0	0.5	0	0	0.5
	1	0	0.5	0.5	0

Also, given the correlations P_t obtained from the maximum violation of the CHSH inequality [CHSH69] (Tsirelson's point)

		$x=0$		$x=1$	
		0	1	0	1
$y=0$	0	0.0732	0.4267	0.0732	0.4267
	1	0.4267	0.0732	0.4267	0.0732
$y=1$	0	0.0732	0.4267	0.0732	0.4267
	1	0.4267	0.0732	0.4267	0.0732

Next, the white noise P_w is given

		$x=0$		$x=1$	
		0	1	0	1
$y=0$	0	0.25	0.25	0.25	0.25
	1	0.25	0.25	0.25	0.25
$y=1$	0	0.25	0.25	0.25	0.25
	1	0.25	0.25	0.25	0.25

Intuitively, we expect the relation may be $D(P_{PR}) > D(P_T) > D(P_w)$, but it turns out not the case $D(P_{PR}) = D(P_T) = D(P_w)$. Apparently, the measures of the nonlocality are different from the measuring, let alone the Euclidean distance. Presumably, the different measures will have different operational meanings. Hence, the question becomes what kind of definitions are "good" measures. Such the discussion is made in [dV14]. Here we leave the conclusions:

- Relabeling operation is unable to increase the nonlocality.
- Mixing with local behavior like P_w, the measure decreases.
- The measure of the point in the local set is zero.
- The "lifting" operation is unable to increase the nonlocality.

Defining the statistical distance measures between the two sets has several choices. In addition to the Euclidean distance norm in Eq. (A.1), different measures are discussed in [BAC18]. Such measures are monotone through free operations of resource theory. For an example of KL divergence, this quantifier can provide advantages in such computational tasks [BAC18], but it doesn't satisfy the last condition above [dV14].

Apart from the measures of the distinguishability, here we start from a causal relaxation of the local hidden variables model Section 2.3.

A.2 Communication protocols

Address the topic of quantum correlations of the communication complexity problem (CCP). How many bits are the minimum requirements to simulate all quantum correlations? To answer this question, different communication models are proposed. In the following, we review some important models to understand the spirit and to help us construct new protocols.

A.2.1 Toner-Bacon model

The Toner-Bacon model [TB03] is inspired by the original Bell's protocol [Bel64] and considered several simple properties of Bell correlation.
(i) Anticorrelation: if Alice and Bob perform the same direction of measurement $\hat{x} = \hat{y}$, the outcomes will be anticorrelated bits $a = -b$.
(ii) Reversibility: each party can reverse their measurement and then flip their outcomes.
(iii) Bell's expression: the joint probability is only dependent on the inner product of the measurement direction $\hat{x} \cdot \hat{y}$.

Note that the third property (iii) implies that rotationally invariant protocols can be constructed and encoded with two random three-dimensional vectors $\hat{\lambda}_1, \hat{\lambda}_2$ [TB03]. The protocols are the follows:

1. Alice outputs $a = -sgn(\hat{x} \cdot \hat{\lambda}_1)$
2. Alice sends the communication message $c = sgn(\hat{x} \cdot \hat{\lambda}_1)sgn(\hat{x} \cdot \hat{\lambda}_2)$
3. Bob receives the message and output $b = sgn(\hat{y} \cdot (\lambda_1 + c\lambda_2)$

It is shown that the protocol (1) and (3) satisfy properties (ii) and (I), respectively. From the properties (ii), each party's results will change sign under symmetry

$$\hat{\lambda}_1 \leftrightarrow -\hat{\lambda}_1, \hat{\lambda}_2 \leftrightarrow -\hat{\lambda}_2 \tag{A.2}$$

and shows the expectation value $< a >=< b >= 0$. Moreover, in this protocol, the joint expectation value results in Bell expression

$$< ab >= -\hat{x} \cdot \hat{y}. \tag{A.3}$$

The most significant result is that even if a large number of simulations are performed in parallel, the averaged communication measured by Shannon entropy can be compressed to 0.85 bits.

To generalize the previous result, now that the one-bit protocol can simulate the maximally entangled state by projective measurement in two-qubit space. What about the general POVM or any dimension of state?

A.2.2 Regev-Toner model

In this subsection, we will introduce Regev-Toner 's model [RT07] trying to generalize the approach in a d-dimensional quantum state system.

- Alice and Bob share the density matrix ρ in space $\mathbb{C}^d \otimes \mathbb{C}^d$ and the Hermitian matrix $\mathbf{X}$ and $\mathbf{Y}$ as inputs for each side with the eigenvalue ± 1.
- Alice and Bob's outcomes are binary a and $b \in \{-1, 1\}$.
- The correlation satisfies $\mathrm{E}[ab] = \mathrm{Tr}(\mathbf{X} \otimes \mathbf{Y} \cdot \rho)$

Based on Tsirelson's theorem, the simulation of quantum correlation is equivalent to the following CCP:

- Alice and Bob receive a unit vector $\vec{x}, \vec{y} \in \mathbb{R}^n$
- The Alice and Bob outcomes are binary a and $b \in \{-1, 1\}$.
- The correlation satisfies $\mathrm{E}[ab] = \langle \vec{x}, \vec{y} \rangle$

The main goal is to approach the correlation in quantum:

$$\mathrm{E}[ab] = h(\langle \vec{x}, \vec{y} \rangle) \tag{A.4}$$

where h represents the correlation function maps

$$h : [-1, 1] \to [-1, 1] \tag{A.5}$$

The main idea is to find the communication model such that the correlation function is simply $h(x) = x$. The orthant protocol is in following.

- Alice and Bob share a random matrix G with size $(k + 1) \times n$.
- Alice generates the outcomes $a_i = G\vec{x}$ and transmit the bit strings $c_i = a_0 a_i$
- Bob outputs $G\vec{y}$ and bit $1, c_1 .....c_k$

The significance of that protocol can be the approach to the correlation function $h(x) = x$ in 2 bits. Therefore, it is known two bits can simulate the maximally entangled state for any d dimension. Note that one can check their proof in [RT07]; roughly speaking, the correlation function is expanded by the power series, and also they proof the mappings $C(x)$ and $C(y)$ exist.

$$\mathrm{E}[\alpha\beta] = h(\langle C(\vec{a}), C(\vec{b}) \rangle) = h \circ h^{-1}(\langle \vec{a}, \vec{b} \rangle) = \langle \vec{a}, \vec{b} \rangle \tag{A.6}$$

In some sense, the protocol only reproduces some of the correlation with certain projective families in the d-dimension. We can also ask the question further. How to design the protocol by using fewer communication resources is still a difficult problem. Because we have less knowledge about the pattern of quantum correlation to encode them in the classical protocol. Many communication models are designed artificially, it's also a question of whether technology can solve this tricky problem. We are looking for a solution in machine learning approaches Appendix D.

Appendix B

Local and n-dits extremal points

B.1 Local extremal points

Consider the local hidden variable model, Alice and Bob share the hidden variables λ, and the joint probabilities distribution is given by:

$$\sum_{\lambda} P(\lambda) P(a|x,\lambda)\, P(b|y,\lambda)$$

Technologically, we can sort of this model as polytope optimization problems and rewrite them as a convex combination of local deterministic strategies, which are so-called extremal points of the local polytope.

$$\sum_{\lambda} P(\lambda) \delta_{a,f(x,\lambda)} \delta_{b,f(y,\lambda)}$$

To characterize the sets of deterministic strategies and start from a simple example, we will write some examples in the following tables. For example, consider that Alice and Bob are in two input binary settings and they can decide on the strategies beforehand. One of the strategy matrices D_{ax} from Alice's side is given by

	$x=0$	$x=1$
$a=0$	1	0
$a=1$	0	1

Similarly, the strategy matrix for Bob's side D_{bx}

	$y=0$	$y=1$
$a=0$	0	1
$a=1$	1	0

Vectorize D_{ax} and D_{bx} and compute the tensor product of them. The joint probability distribution $P(a,b|x,y)$ can be constructed by

$$vec(D_{xa})^T \otimes vec\,(D_{yb}) = [1\ 0\ 0\ 1] \otimes \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

Result in one of the deterministic strategies $D_i^0(a,b|x,y)$

		$x=0$		$x=1$	
		0	1	0	1
$y=0$	0	0	1	1	0
	1	0	0	0	0
$y=1$	0	0	0	0	0
	1	0	1	1	0

B.2 One-dit extremal points

Consider the relaxation of local hidden variables supplemented by communication. The correlation can be explained in the following formula.

$$\begin{aligned} P(a,b \mid x,y) &= \sum_{c} P(a,c,b \mid x,y) \\ &= \sum_{\lambda',c} p_{\lambda'} P\left(a,c \mid x,\lambda'\right) P\left(b \mid y,\lambda',c\right) \\ &= \sum_{\lambda',c} p_{\lambda'} P\left(c \mid x,\lambda'\right) P\left(a \mid x,\lambda'\right) P\left(b \mid y,\lambda',c\right) \end{aligned}$$

Similarly, the deterministic strategy supplemented with communication can be constructed

$$\sum_{\lambda,c} p_{\lambda} \left(\delta_{a,f_{\mathrm{A}}(x,\lambda)} \delta_{c,f'_{\mathrm{A}}(x,\lambda)}\right) \delta_{b,f_{\mathrm{B}}(y,c,\lambda)}$$

To understand the formula above, the following explanations can be made. The first part $\delta_{a,f_{\mathrm{A}}}$ presents the local extremal points mentioned earlier. Alice can determine the message by the response function $f_A(x,\lambda\prime)$ according to the inputs x and outcomes a. Afterward, Bob receives the message and determines the outcome b by the response function $f_B(y,c,\lambda\prime)$ according to her inputs y. Without loss of generality, take an example for the Bell scenario $(2,3)$ and supplemented by 1-bit communication. Alice executes the communication strategies D_{ycb}.

$x=1$

	$c=0$	$c=1$
$a=0$	1	0
$a=1$	0	0
$a=2$	0	0

$x=2$

	$c=0$	$c=1$
$a=0$	0	0
$a=1$	0	1
$a=2$	0	0

For Bob's side, the communication strategies D_{xcb} matrix is given following :

$y = 1$

	$b = 0$	$b = 1$	$b = 1$
$c = 0$	1	0	0
$c = 1$	0	1	0

$y = 2$

	$b = 0$	$b = 1$	$b = 1$
$c = 0$	0	0	1
$c = 1$	0	1	0

The joint deterministic strategy $D_i^1(a, b|x, y)$ of 1-dit (1-bit extremal point) turns out to be

		$x = 0$			$x = 1$		
		0	1	2	0	1	2
$y = 0$	0	1	0	0	0	0	1
	1	0	0	0	0	0	0
	2	0	0	0	0	0	0
$y = 1$	0	0	0	0	0	0	0
	1	0	1	0	0	1	0
	2	0	0	0	0	0	0

Note that one can construct with the trit strategy but just append the third column full of zeros in the strategies D_{ycb}. It will finally show the trit strategy is the same as the bit strategy case. In general, we only have up to $n - 1$ (inputs-1) dit polytope due to the construction.

By generating for all d-dit strategies in $xyab$ scenario (CG representation), the number of vertices of the 1-dit polytope will be

$$a^x \left(b^y + \left\{ \begin{array}{c} x \\ d \end{array} \right\} \left(b^{dy} - b^y \right) \right) \tag{B.1}$$

The first term $a^x b^y$ is the number of local extremal points in the Bell scenario. In the second terms, $a^x \left\{ \begin{array}{c} x \\ d \end{array} \right\}$ accounts for the strategies for D_{xcb} while $a^{dy} - a^y$ accounts for the strategies for D_{ycb} subtracted by the situation that Bob does not use the communication dit.

Appendix C

Optimization programming

C.1 The Polytope

According to the Farkas-Minkowski-Weyl theorem, any polyhedron can represent from V-representation to H -representation and vice versa. The polytope which is a convex set $\mathcal{P}$ can be described as the intersection of half-spaces $Ax \geqslant c$. In the H-representation, it can be written as

$$\mathcal{P} = \left\{ x \in \mathbb{R}^t | Ax \geqslant c \right\} \tag{C.1}$$

where the matrix of inequalities A is $\in \mathbb{R}^{r \times t}$ and the vector of constants is $\in \mathbb{R}^{r \times t}$. The V representation is written by the extremal points $x_k \in \mathbb{R}^t$

$$\mathcal{P} = \left\{ x = \sum_k \lambda_k x_k \mid \sum_k \lambda_k = 1, \lambda_k \geqslant 0 \right\} \tag{C.2}$$

Given a polytope $\mathcal{P}$ with dimension d, the minimal representation of points x_k is called vertices or extremal points, while the intersection of $\mathcal{P}$ with a bounding hyperplane $Ax \geqslant c$ is called facets if it has dimension $d-1$. The procedure converting between both representations is namely H/V representation enumeration. Note that the software Lrs [Avi00] provides the function to find the convert both representations.

C.2 Linear programming

Linear programming is the optimization of an objective function under linear inequality constraints. The canonical forms are presented: Given fixed coefficients matrix $\mathbf{C} \in \mathbb{R}^n$, the constant vector $\mathbf{q} \in \mathbb{R}^d$ for the matrix $\mathbf{G}$ combined all inequalities. For the standard form:

$$\max_{\mathbf{x}} \mathbf{C}^T \mathbf{x} \quad \text{subject to} \quad \mathbf{G}\mathbf{x} \leq \mathbf{q}, \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \geq \mathbf{0} \tag{C.3}$$

where x represents the variables and the $\mathbf{C}^T$ act on them and result in a scalar that we want to optimize. If the solution is allowed in the domain $\mathbf{x}$ and satisfied all the constraints so-called feasible region, the domain is bound and the problem is feasible. Therefore, it forms a convex polytope. The maximum principle states that the optimum value can be achieved at the extremal point. At the same time, the convex combination of the extremal points can form the convex polytope. Also, one can represent the polytope by the H-representation or

V-representation. The former describes the information about facets and the latter provide the information about vertices. In this regard, solving the linear problem has an equivalent expression called the dual problem.

$$\min_{\mathbf{y}} \mathbf{q}^T \mathbf{y} \quad \text{subject to} \quad G^T \leq \mathbf{C}. \tag{C.4}$$

Referring to the previous version, we have the primal form and the latter is the dual form. To solve linear problems, one can start the idea from the simplex algorithm. The simplex algorithm starts from all the vertices and searches along the edges until it grants the optimal points. Inspired by the simplex algorithm, the following algorithms are concerned with searching the Bell inequalities.

C.2.1 Visibility method

The Bell-facet inequalities always play an important role and provide quantum advantages in theoretical informative tasks. Sometimes the complete list of local facets is not known. The linear program provides a useful exploration. Given the visibility program, the white noise $P_i^w = \frac{1}{n_a n_b}$ that stays within the local polytope is added to the initial point P_i with visibility $1 - v \in [0, 1]$. If the optimal $v^* < 1$, it indicates the initial point P_i outside the local polytope.

The primal problem of the visibility method shows:

$$\begin{aligned} &\max_{c_i^{(0)}} \ v \\ &\text{subject to} \\ &\qquad \sum_i c_i^{(0)} D_i^{(0)} = v P_i + (1 - v) P_i^w \ \ \forall\, i \\ &\qquad \sum_i c_i^{(0)} = 1, \\ &\qquad c_i^{(0)} \geq 0 \ \ \forall\, i. \end{aligned} \tag{C.5}$$

where local extremal points are denoted as $D_i^{(0)}$

The dual problem is written as :

$$\begin{aligned} &\min_{\{B_i\}} \ \sum_i B_i P_i^w - B_0 \\ &\text{subject to} \\ &\qquad \sum_i B_i (P_i^w - P_i) - 1 = 0, \\ &\qquad \sum_i B_i D_i^{(0)} \geq B_0 \ \ \forall\, i, \end{aligned} \tag{C.6}$$

where B_i are the corresponding coefficients of the Bell inequality and B_0 is the corresponding Bell value.

C.2.2 The primal and dual of j-dit MACC program

The primal problem of MACC is given:

$$\begin{aligned} \min \quad & \sum_i \sum_{j\geq 1} c_i^{(j)} \\ \text{subject to} \quad & \\ & \vec{P} = \sum_{j=0}^{d} \sum_i c_i^{(j)} \vec{D}_i^{(j)}, \\ & c_i^{(j)} \geq 0 \ \forall \ i, \\ & \sum_{i,j} c_i^{(j)} = 1. \end{aligned} \tag{C.7}$$

For the dual program, the dual variables $\vec{B}_j$ are given and the vector $\vec{J}$ corresponds to the amount of communication in the communication polytope j-dit. In convention, combine all elements $\vec{D}_i^{(j)}$ as the matrix $\mathbf{D}$.

$$\begin{aligned} \max \quad & \vec{B}_j \cdot \vec{P} \\ \text{subject to} \quad & \\ & \mathbf{D}^T \vec{B} \leq \vec{J} \end{aligned} \tag{C.8}$$

C.2.3 The primal and dual of local content program

The primal program of local content:

$$\begin{aligned} \max \quad & \sum_j c_i^{(0)} \\ \text{subject to} \quad & \\ & \vec{P} \geq \sum_i c_i^{(0)} \vec{D}_i^{(0)}, \\ & c_i^{(0)} \geq 0 \ \forall \ i. \end{aligned} \tag{C.9}$$

The dual program of local content

$$\begin{aligned} \min_B \quad & \vec{B} \cdot \vec{P} \\ \text{subject to} \quad & \\ & \mathbf{D}_0^T \vec{B} \geq \vec{1} \\ & B_i \geq 0 \ \forall \ i. \end{aligned} \tag{C.10}$$

C.3 SDP programming

In semidefinite programming (SDP), the linear objection is subject to constraints related to the combination of positive semidefinite matrices. These constraints may be nonlinear but convex, hence, the SDP is a convex optimization problem. The problem of minimizing the linear function of variables $x \in \mathbb{R}^n$ subject to the matrix inequality can be written as:

$$\begin{aligned} &\min \quad c^T x \\ &\text{subject to} \\ &\qquad F(x) \geq 0 \end{aligned} \tag{C.11}$$

C.3.1 See-saw method

See-saw algorithm is an iterative SDP method proposed initially for maximizing the quantum violation of a Bell correlation inequality for a given state [WW01]. Later, it was generalized to find the maximal Bell violation for a given Hilbert space dimension in [PV09] and also in [LLD09]. The See-saw algorithm iterates two SDPs. First, we consider the formulation from Born's rule where the outcomes a, b and inputs x, y are denoted:

$$P_{ab}^{\vec{xy}} = \mathrm{tr}\left(\rho A_a^x \otimes B_b^y\right) \tag{C.12}$$

The algorithm begins with random initial density matrix ρ, d_a dimensional POVM elements with A_a^x from a th outcome of x th input and similarly for d_b dimensional B_b^y. In practice, the initial choices are projective measurements and pure state. The first part of iterations is to find the maximum violation by varying one side of POVMs and treating the other side as the constant.

$$\begin{aligned} &\max_{B_x^a} \ \vec{B}_j \cdot P_{ab}^{\vec{xy}} \\ &\text{subject to} \\ &\qquad p_{ab}^{xy} = \mathrm{tr}\left(\rho A_a^x \otimes B_b^y\right) \\ &\qquad \sum_{a=1}^{n_a} B_a^x = \mathbf{1}_{d_B} \quad \forall x \\ &\qquad B_x^a \geq 0 \quad \forall x, a \end{aligned} \tag{C.13}$$

where POVM elements are imposed as positive and completeness conditions (sum up to the identity matrix $\mathbf{1}_{d_B}$). Apparently, the same analysis follows if we fix others' measurement settings, and optimize Alice's POVM elements.

The second procedure relates to varying the density matrix ρ but treating all POVM elements as constant.

$$\begin{aligned} &\max_{\rho} \ \vec{B}_j \cdot \vec{P_{ab}^{xy}} \\ &\text{subject to} \\ &\qquad p_{ab}^{xy} = \mathrm{tr}\left(\rho A_x^a \otimes B_y^b\right) \\ &\qquad \rho \geq 0, \ \ \mathrm{tr}\,\rho = 1 \end{aligned} \tag{C.14}$$

where the density matrix ρ is imposed the positive constraint. The two procedures will repeat until the objective function converges. A nontrivial lower bound can be obtained by different measurements and states from sets randomly generated initially.

C.4 Double optimization

We employ the MATLAB function fminunc, a nonlinear multivariable optimization tool, to build a heuristic algorithm. The double optimization consists of an inner loop and an outer loop. The inner loop is the programming of the EPR2 cost Eq. (2.19) while the process of the outer loop is described below : We assume the measurement M_A of Alice's side and $UM_AU^\dagger$ of Bob's side. Using MATLAB function fminunc, we optimize that unitary operation U to find suboptimal point P.

C.4.1 The random unitary matrix from QR decomposition

The unitary operator can be constructed from the QR decomposition. The square matrix A with elements $\in [0, 1]$ can be decomposed into a unitary matrix $Q \in U(n)$ and a triangular matrix R. Note that QR decomposition is not unique for any $\Lambda \in U(n)$,

$$QR = (Q\Lambda)\,(\Lambda^* R) = Q'R' \tag{C.15}$$

However, QR decomposition will not be granted for unique Q. One post-processing will ensure the unique decomposition.

$$\Lambda = \begin{pmatrix} \frac{r_{11}}{|r_{11}|} & & & \\ & \frac{r_{22}}{|r_{22}|} & & \\ & & \ddots & \\ & & & \frac{r_{nn}}{|r_{nn}|} \end{pmatrix} \tag{C.16}$$

where r_{ii} are matrix elements of R. In short, instead of finding the measurements of Bob side, the variables will be replaced by matrix A to construct the unitary matrix Q'.

C.5 The No-signaling extremal points

To have further discussion in Chapter 4, some of the extremal points of NS are mentioned in [BSG06]. For example, PR1 boxes are written of CG form in $(2, 2)$ scenario.

$$PR = \frac{1}{2} \times \begin{array}{c|cc} & 1 & 1 \\ \hline 1 & 1 & 1 \\ 1 & 1 & 0 \end{array} \tag{C.17}$$

Further, we have known the class of no-signaling extremal points which can sort of two classes PR_3, PR in $(3, 2)$ scenario.

$$PR_3 = \frac{1}{2} \times \begin{array}{c|ccc} & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array} \tag{C.18}$$

$$PR = \frac{1}{2} \times \begin{array}{c|ccc} & 1 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{array} \tag{C.19}$$

Appendix D

The machine learning approach

Recapitulate the construction of different communication models, the difficulties are how to design the protocol. Particularly, if the quantum correlations come from a maximum entanglement state, it will guarantee some structures to utilize. Beyond the case, less knowledge we have known. In recent years, many machine learning-based models have been proposed to overcome these intractable problems [KCC$^+$20] such as the non-convex problems. Different models are developed based on machine learning, such as a supervised learning model to detect nonlocality [CBC19] and a reinforcement learning model that acts as an optimization tool in quantum informative tasks [BHVK19]. Now that so many applications are desirable, we want to employ the machine learning technique to construct the communication model. Also, in [MNPC20], there is also a question of whether machine learning can be applied in a similar way to the analysis of equilibrium points or payoff functions. In this thesis, we are inspired by [Kri20] and construct the new machine-learning-based protocols.

D.1 The neural network oracle: local model

In this section, the main idea is to encode the causal structure Section 2.3.1 in neural networks. Given the correlation, we can design a numerical protocol to reproduce. Basically, the neural network can be considered as a non-linear mapping.

$$f : [0, 1] \to [0, 1] \tag{D.1}$$

We can construct the vector $V_{k,m} = f(\vec{x}, \lambda)$ for m th input of k th party. The function $f(\vec{X})$ can be written as the nonlinear function α and W weight matrix act on the inputs $\vec{X}$:

$$f(x) = \alpha W \vec{X} \tag{D.2}$$

Denote the bit strings as $x_j \in \{0, 1\}$ and sharing the randomness $\lambda_j \in [0, 1]$. Given an example in (2,2), the vector $V_{a,j}$ can be constructed:

$$V_{a,1} = \alpha \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ \lambda_1 \end{pmatrix} \tag{D.3}$$

where α is the softmax nonlinear function that results in a normalized vector $V_{a,j}$ and elements W_{kl} are the variables that we want to optimize.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{D.4}$$

Similarly, on Bob's side we have:

$$V_{b,1} = \alpha \begin{pmatrix} W'_{11} & W'_{12} \\ W'_{21} & W'_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ \lambda_1 \end{pmatrix} \tag{D.5}$$

After that, the joint probability distribution $P_j(a, b|1, 1)$ can be constructed from the tensor product of $V_{a,1}$ and $V_{b,1}$ while the others $P_j(a, b|1, 2), P_j(a, b|2, 1), P_j(a, b|2, 2)$ are zeros.

$$P_i(a, b|1, 1) = V_{a,1} \otimes V_{b,1} \quad \forall a, b = 0, 1 \tag{D.6}$$

$$P_i(a, b|x, y) = 0 \quad \forall a, b = 0, 1 \quad x, y \neq 1 \tag{D.7}$$

Note such this idea is related to the construction of local extremal points in Appendix B.1.

$$P(a, b \mid x, y) = \sum_{\lambda} P(\lambda) P(a \mid x, \lambda) P(b \mid x, y, \lambda) \tag{D.8}$$

Similarly, repeat the construction with the random bit string x_j and randomness λ_j by using the matrix W and W' on each side. The final correlation can be obtained from summing up for each λ and divide by each batch $N_{k,l}$ with respect to the $P_i(a, b|k, l)$ have made.

$$P(a, b|x, y) = \sum_{k,l} \sum_{j} P_j / N_{k,l} \tag{D.9}$$

The loss function is determined by the $1 - d$ norm with N entries between the local model $P(a, b|x, y)$ and target distribution $P_t(a, b|x, y)$

$$L_1 = \frac{1}{2N} |P_t(a, b|x, y) - P(a, b|x, y)| \tag{D.10}$$

The neural network process can be activated by minimizing the objective function.

$$\min L_1 \tag{D.11}$$

Different from typical supervised learning task, in this work, the neutral network plays the role in the generation of local strategies. Normally, the construction of LHV is described by the linear combination of local deterministic strategies inAppendix B.1. Instead of optimizing overall possible local deterministic strategies, the correlation generated in the protocol is inside the local polytope. It means that it is able to search different possible mixtures in the

local polytope respective to the target correlation by optimizing the loss function. Inspired by this idea, we construct new neutral-network-based communication protocols.

D.2 The neural network oracle: communication model

In the following protocol, Alice prepares the bit strings x_1, x_2..... and generate the outcome bit strings a_1, a_2..... Afterward, Alice will send the bit $c \in \{0, 1\}$ to Bob's side.
For example in the (2,3) scenario, the vector $V_{1,j}$ consists of nonlinear softmax function $alpha$ and weight matrix W :

$$V_{a,1} = \alpha \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ \lambda_1 \end{pmatrix} = \begin{pmatrix} v_{a,1,1} \\ v_{a,1,2} \end{pmatrix} \tag{D.12}$$

Compared to the previous protocol, we construct the communication transformation matrix. The subscripts of element $V_{n,k,d}$ means transmitting probability for the kth outcome to the d th message dimension of nth party. Similarly, we will have $V_{1,2,2}$ with the elements $v_{1,2,1}$,$v_{1,2,2}$. If Alice's first outcome is 0 and the second is 1, the transformation matrix P_{xac} can be built up as follows: .

$x = 1$

	$c = 0$	$c = 1$
$a = 0$	$v_{1,1,1}$	$v_{1,1,2}$
$a = 1$	0	0
$a = 2$	0	0

$x = 2$

	$c = 0$	$c = 1$
$a = 0$	0	0
$a = 1$	$v_{1,2,1}$	$v_{1,2,2}$
$a = 2$	0	0

Similarly for Bob's side, Bob prepares the bit strings y_1, y_2..... and generates the outcome bit strings b_1, b_2..... We can build up the vector $V_{2,1,1}$ for the input k with elements $v_{2,k,d,l}$ which represent the probability of transition from the message c to the lth outcome of kth input. For example $l = 1, k = 1$:

$$V_{2,1,1} = \alpha \begin{pmatrix} W'_{11} & W'_{12} \\ W'_{21} & W'_{22} \\ W'_{31} & W'_{32} \end{pmatrix} \begin{pmatrix} y_1 \\ \lambda_1 \end{pmatrix} = \begin{pmatrix} v_{2,1,1,1} \\ v_{2,1,1,2} \\ v_{2,1,1,3} \end{pmatrix} \tag{D.13}$$

Therefore, the transformation matrix on Bob's side P_{ycb} can be constructed.

$y = 1$

	$b = 0$	$b = 1$	$b = 1$
$c = 0$	$v_{2,1,1,1}$	$v_{2,1,1,2}$	$v_{2,1,1,3}$
$c = 1$	$v_{2,1,2,1}$	$v_{2,1,2,2}$	$v_{2,1,2,3}$

$y = 2$

	$b = 0$	$b = 1$	$b = 1$
$c = 0$	$v_{2,2,1,1}$	$v_{2,2,1,2}$	$v_{2,2,1,3}$
$c = 1$	$v_{2,2,2,1}$	$v_{2,2,2,2}$	$v_{2,2,2,3}$

The strategy P_j can be obtained by multiplying two transformation matrix P_{xac} and P_{ycb}

$$P_{xac} \times P_{ycb} = P_j \tag{D.14}$$

Summing over these strategies from all randomness λ and divided by all numbers N will result in the correlation $P(a, b|x, y)$.

$$P(a, b|x, y) = \sum_j P_j/N \tag{D.15}$$

Here, we start the discussion of the construction of communication models. We proposed a new model where the main idea is not to use all extremal points on one point. Instead, we provide a way to automatically search in the 1-dit polytope for a given correlation. Moreover, one can improve this supervised learning model implementation. Other machine-learning techniques can also be considered. For example, the generative adversarial net (GAN) [CPZC21], a famous machine learning framework was utilized to detect the entanglement ranging from two-qubit to ten-qubit systems. Such frameworks may help us solve our problems more efficiently.

References

[AAB+19] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, Oct 2019.

[ACSA10] Mafalda L. Almeida, Daniel Cavalcanti, Valerio Scarani, and Antonio Acín. Multipartite fully nonlocal quantum states. *Phys. Rev. A*, 81:052111, May 2010.

[Avi00] David Avis. *A Revised Implementation of the Reverse Search Vertex Enumeration Algorithm*, pages 177–198. Birkhäuser Basel, Basel, 2000.

[BAC18] S. G. A. Brito, B. Amaral, and R. Chaves. Quantifying bell nonlocality with the trace distance. *Phys. Rev. A*, 97:022111, Feb 2018.

[BC17] J B Brask and R Chaves. Bell scenarios with communication. *Journal of Physics A: Mathematical and Theoretical*, 50(9):094001, jan 2017.

[BCT99] Gilles Brassard, Richard Cleve, and Alain Tapp. Cost of exactly simulating quantum entanglement with classical communication. *Phys. Rev. Lett.*, 83:1874–1877, Aug 1999.

[BDG19] Gilles Brassard, Luc Devroye, and Claude Gravel. Remote sampling with applications to general entanglement simulation. *Entropy*, 21(1), 2019.

[Bel64] J. S. Bell. On the einstein podolsky rosen paradox. *Physics*, 1:195–200, Nov 1964.

[Bel04] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. Cambridge University Press, 2 edition, 2004.

[BHVK19] Kishor Bharti, Tobias Haug, Vlatko Vedral, and Leong-Chuan Kwek. How to teach ai to play bell non-local games: Reinforcement learning. 12 2019.

[BL13] Nicolas Brunner and Noah Linden. Connection between bell nonlocality and bayesian game theory. *Nature Communications*, 4(1):2057, Jul 2013.

[BSG06] Nicolas Brunner, Valerio Scarani, and Nicolas Gisin. Bell-type inequalities for nonlocal resources. *Journal of Mathematical Physics*, 47(11):112101, 2006.

[BT03] D. Bacon and B. F. Toner. Bell inequalities with auxiliary communication. *Phys. Rev. Lett.*, 90:157904, Apr 2003.

[CBC19] Askery Canabarro, Samuraí Brito, and Rafael Chaves. Askery:2019. *Phys. Rev. Lett.*, 122:200401, May 2019.

[CC19] Thomas Cope and Roger Colbeck. Bell inequalities from no-signaling distributions. *Phys. Rev. A*, 100:022114, Aug 2019.

[CG04] Daniel Collins and Nicolas Gisin. A relevant two qubit Bell inequality inequivalent to the CHSH inequality. *J. Phys. A: Math. Theo.*, 37(5):1775, 2004.

[CG19] E. Zambrini Cruzeiro and N. Gisin. Complete list of tight bell inequalities for two parties with four binary settings. *Physical Review A*, 99(2), Feb 2019.

[CGL$^+$02] Daniel Collins, Nicolas Gisin, Noah Linden, Serge Massar, and Sandu Popescu. Bell inequalities for arbitrarily high-dimensional systems. *Phys. Rev. Lett.*, 88:040404, Jan 2002.

[CGM00] N. J. Cerf, N. Gisin, and S. Massar. Classical teleportation of a quantum bit. *Phys. Rev. Lett.*, 84:2521–2524, Mar 2000.

[CGMP05] N. J. Cerf, N. Gisin, S. Massar, and S. Popescu. Simulating maximal quantum entanglement without communication. *Phys. Rev. Lett.*, 94:220403, Jun 2005.

[Che22] Kai-Siang Chen. Minimal average communication cost of nonlocal correlations. 2022.

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969.

[CHSH70] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden variable theories. *Phys. Rev. Lett.*, 24:549–549, Mar 1970.

[CKBG15] R. Chaves, R. Kueng, J. B. Brask, and D. Gross. Unifying framework for relaxations of the causal assumptions in bell's theorem. *Phys. Rev. Lett.*, 114:140403, Apr 2015.

[CPZC21] Yiwei Chen, Yu Pan, Guofeng Zhang, and Shuming Cheng. Detecting quantum entanglement with unsupervised learning. *Quantum Science and Technology*, 7(1):015005, nov 2021.

[DEBZ10] THOMAS DURT, BERTHOLD-GEORG ENGLERT, INGEMAR BENGTSSON, and KAROL ŻYCZKOWSKI. On mutually unbiased bases. *International Journal of Quantum Information*, 08(04):535–640, 2010.

[dV14] Julio I de Vicente. On nonlocality as a resource theory and nonlocality measures. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424017, oct 2014.

[EPR92] Avshalom C. Elitzur, Sandu Popescu, and Daniel Rohrlich. Quantum nonlocality for each pair in an ensemble. *Physics Letters A*, 162(1):25–28, 1992.

[Gel22] Gelo. Minimal average communication cost of nonlocal correlations. 10 2022.

[Gis07] Nicolas Gisin. Bell inequalities: many questions, a few answers. 2007.

[Jar84] Jon P. Jarrett. On the physical significance of the locality conditions in the bell arguments. *Noûs*, 18(4):569–589, 1984.

[KCC+20] Tamás Kriváchy, Yu Cai, Daniel Cavalcanti, Arash Tavakoli, Nicolas Gisin, and Nicolas Brunner. A neural network oracle for quantum nonlocality problems in networks. *npj Quantum Information*, 6(1):70, Aug 2020.

[Kri20] A neural network oracle for quantum nonlocality problems in networks. *npj Quant. Inf.*, 6:70, August 2020.

[LD07] Yeong-Cherng Liang and Andrew C. Doherty. Bounds on quantum correlations in bell-inequality experiments. *Phys. Rev. A*, 75:042103, Apr 2007.

[LLD09] Yeong-Cherng Liang, Chu-Wee Lim, and Dong-Ling Deng. Reexamination of a multisetting bell inequality for qudits. *Phys. Rev. A*, 80:052116, Nov 2009.

[LVL21] Pei-Sheng Lin, Tamás Vértesi, and Yeong-Cherng Liang. Naturally restricted subsets of nonsignaling correlations: typicality and convergence. 07 2021.

[LWW+10] Lars Lydersen, Carlos Wiechers, Christoffer Wittmann, Dominique Elser, Johannes Skaar, and Vadim Makarov. Hacking commercial quantum cryptography systems by tailored bright illumination. *Nature Photonics*, 4(10):686–689, Oct 2010.

[MBL+13] Tobias Moroder, Jean-Daniel Bancal, Yeong-Cherng Liang, Martin Hofmann, and Otfried Gühne. Device-independent entanglement quantification and related applications. *Phys. Rev. Lett.*, 111:030501, Jul 2013.

[MC14] Katherine Maxwell and Eric Chitambar. Bell inequalities with communication assistance. *Phys. Rev. A*, 89:042108, Apr 2014.

[Mez06] Francesco Mezzadri. How to generate random matrices from the classical compact groups. *Notices of the American Mathematical Society*, 54, 10 2006.

[MGM22] Uta Isabella Meyer, Frédéric Grosshans, and Damian Markham. Inflated graph states refuting communication-assisted lhv models, 2022.

[MNPC20] George Moreno, Ranieri Nery, Alberto Palhares, and Rafael Chaves. Multistage games and bell scenarios with communication. *Phys. Rev. A*, 102:042412, Oct 2020.

[NC11] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2011.

[NPA07] Miguel Navascués, Stefano Pironio, and Antonio Acín. Bounding the set of quantum correlations. *Phys. Rev. Lett.*, 98:010401, Jan 2007.

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

[Par22] Abhishek Parakh. Quantum teleportation with one classical bit. *Scientific Reports*, 12(1):3392, Mar 2022.

[PBG12] Samuel Portmann, Cyril Branciard, and Nicolas Gisin. Local content of all pure two-qubit states. *Phys. Rev. A*, 86:012104, Jul 2012.

[Pir03] Stefano Pironio. Violations of Bell inequalities as lower bounds on the communication cost of nonlocal correlations. *Phys. Rev. A*, 68:062102, Dec 2003.

[PV09] Károly F. Pál and Tamás Vértesi. Quantum bounds on bell inequalities. *Phys. Rev. A*, 79:022120, Feb 2009.

[PV22] Ká roly F. Pál and Tamás Vértesi. Platonic bell inequalities for all dimensions. *Quantum*, 6:756, jul 2022.

[RQ22] Martin J. Renner and Marco Túlio Quintino. The minimal communication cost for simulating entangled qubits. arXiv:2207.12457, 2022.

[RT07] Oded Regev and Ben Toner. Simulating quantum correlations with finite communication. *SIAM J. Comput.*, 39, 08 2007.

[RTQ22] Martin J. Renner, Armin Tavakoli, and Marco Túlio Quintino. The classical cost of transmitting a qubit. arXiv:2207.02244, 2022.

[SBSL18] Sacha Schwarz, Bänz Bessire, André Stefanov, and Yeong-Cherng Liang. Corrigendum: Bipartite bell inequalities with three ternary-outcome measurements—from theory to experiments (2016 new j. phys. 18 035001). *New Journal of Physics*, 20(4):049502, apr 2018.

[Sca08] Valerio Scarani. Local and nonlocal content of bipartite qubit and qutrit correlations. *Phys. Rev. A*, 77, 04 2008.

[Sca15] Valerio Scarani. The device-independent outlook on quantum physics (lecture notes on the power of bell's theorem). 2015.

[TB03] B. F. Toner and D. Bacon. Communication cost of simulating bell correlations. *Phys. Rev. Lett.*, 91:187904, Oct 2003.

[TBYL22] Gelo Noel M. Tabia, Varun Satya Raj Bavana, Shih-Xian Yang, and Yeong-Cherng Liang. Bell inequality violations with random mutually unbiased bases. *Phys. Rev. A*, 106:012209, Jul 2022.

[VB09] Tamas Vertesi and Erika Bene. Lower bound on the communication cost of simulating bipartite quantum correlations. *Physical Review A*, 80, 04 2009.

[WDTL08] S. Wehner, A. C. Doherty, B. Toner, and Y. Liang. The quantum moment problem and bounds on entangled multi-prover games. In *Twenty-Third Annual IEEE Conference on Computational Complexity - CCC 2008*, pages 199–210, Los Alamitos, CA, USA, jun 2008. IEEE Computer Society.

[WW01] Reinhard F. Werner and Michael M. Wolf. Bell inequalities and entanglement. *Quantum Info. Comput.*, 1(3):1–25, oct 2001.

[ZCG19] Emmanuel Zambrini Cruzeiro and Nicolas Gisin. Bell inequalities with one bit of communication. *Entropy*, 21(2), 2019.

[ZRGW10] S. Zohren, P. Reska, R. D. Gill, and W. Westra. A tight tsirelson inequality for infinitely many outcomes. *Europhysics Letters*, 90(1):10002, apr 2010.