Universidade Federal de Minas Gerais
Departamento de Física
Instituto de Ciências Exatas — ICEx

Classical and Quantum Memory in Contextuality Scenarios

Gabriel Fagundes

PhD Thesis

Presented to the Graduate Program in Physics of Federal University of Minas Gerais, as a partial requirement to obtain the title of Doctor in Physics.

Abstract

Quantum theory can be described as a framework for calculating probabilities of measurement outcomes. A great part of its deep foundational questions comes from the fact that these probabilities may disagree with classical calculations, under reasonable premises. The classical notion in which the observables are frequently assumed as predefined before their measurement motivates the assumption of *noncontextuality*, i.e. that all the observables have preassigned values before the interaction with the experimental apparatus, independently on which other observables are jointly measured with it. It is known that this classical view is inconsistent with quantum predictions. The main question of this thesis can be phrased as: can we use memory to classically obtain results in agreement with quantum theory applied to sequential measurements? If so, how to quantify the amount of memory needed? These questions are addressed in a specific contextuality scenario: the Peres-Mermin square. Previous results are extended by using a comprehensive scheme, which shows that the same bound of a three-internal-state automaton is sufficient, even when all probabilistic predictions are considered. Trying to use a lower dimensional quantum resource, i.e. the qubit, to reduce the memory cost in this scenario led us to another question of whether or not there is contextuality for this type of system. We find that sequences of compatible and repeatable quantum measurements on a qubit cannot reveal contextuality, even when the measurements are not assumed projective beforehand.

Resumo

A Teoria Quântica pode ser interpretada como um arcabouço teórico para se calcular probabilidades de resultados de medições. Grande parte de suas questões fundamentais vem do fato que estas probabilidades podem entrar em desacordo com cálculos clássicos. A noção clássica de que observáveis têm valores predefinidos antes de suas medições motiva a premissa de *não-contextualidade*(NC), i.e. de que todas as observáveis têm valores pré-atribuídos antes da interação com o aparato de medição, independente de quais outras observáveis estão sendo medidas conjuntamente. É sabido que tal visão clássica é inconsistente com as previsões da teoria quântica. O problema principal desta tese pode ser formulado da seguinte forma: podemos usar memória para obter classicamente resultados de acordo com a Teoria Quântica, para medições sequenciais? E como quantificar quanta memória seria necessária? Tais questões são abordadas em um cenário de contextualidade específico: o quadrado de Peres-Mermin. Resultados anteriores são ampliados usando uma abordagem abrangente, demonstrando que autômatos com três estados internos são suficientes, mesmo quando as previsões probabilísticas são incluídas. A tentativa de usar um recurso quântico de menor dimensão, isto é, um qubit, para reduzir o custo de memória neste cenário nos levou a outra questão: se há contextualidade neste tipo de sistema. Nós descobrimos que uma sequência de medições compatíveis e repetitíveis em um qubit não pode revelar nenhuma contextualidade, mesmo que não sejam assumidas medições projetivas de antemão.

Acknowledgements

Este trabalho não seria feito sem muita gente, naturalmente, dos quais destaco alguns. Não em termos de importância, imensurável, mas por serem mais relacionados com esse trabalho. A obra se confunde com o percurso, e vice-versa.

À meus pais, José Luiz e Sandra, sempre no suporte incondicional e bons conselhos, amo vocs. À minhas irmãs, Ana Luiza, Isabela e Daniele, pedaços de mim. Aos velhos e novos amigos, amigas e companheiras, sem os quais pouca coisa faria sentido.

To the supervisors, Matthias Kleinmann and Marcelo Terra Cunha, for all the work together, dedication and wit, physics became clearer through them. Ao Terra, pela orientação, pelos ensinamentos, pela compreensão e pela liberdade. To Matthias, who accepted me as his student in UPV/EHU, Basque country, for guindance in physics and in foreign lands, and for the pintxos and txuletones. To Géza Tóth, who accepted me as part of his group in UPV. To Otfried Guhne and his group, who hosted me in Siegen, Germany, for a strangely snowing and fruitfull period.

Aos companheiros de profissão, com os quais aprendi muito. Aos Terráqueos, pelo ambiente acolhedor e ao mesmo tempo estimulante; uma pena que a maioria seja sonserina. Aos habitantes da salinha, pelos momentos de trabalho duro e pelos momentos que me impediram de trabalhar. Tomara que a ciência brasileira encontre espaços e suporte cada vez melhores para se desenvolver.

À UFMG, obrigado, y UPV/EHU, eskerrik asko.

Ao Mizerani, grande pareceiro ao longo dessa jornada, pelas horas de cerveja, conversa fiada, cachaça de pistache, ao nosso projeto PicadeiRRÓ (que valia um capítulo nessa tese). À grande amiga Naty, você foi essencial. Ao Lucas Mussnich, grande amigo que compartilha minha sina, muchas gracias pela visita. Ao Leo (El Lobito), pelos momentos que valem mil charutos. Ao Davi, por ser esse amigo numa salinha quase abandonada, e por entender o que eu falo. À Julia, Ana, Tassius, Mari, Denise, Raul, Aloísio, Cobra, Alex, Lígia, Rafael, Andreij, Luísa, Duty. Obrigado Alexandra Elba. To Costantino, Marco Túlio , Mateus, for their friendship and interesting physics discussions. À Shirley e o pessoal da biblioteca, pela graciosidade. À casa lama. À Família Tai Lau Chuan. Aos funcionários e funcionárias do

Departamento de Física, do ICEx e da UFMG.

À trupe do Picadeiro. Aos Pé-de-Porco. À turma do Biziu. Ao grande amigo Ghandi, use com moderação. Erinaldo, monge urbano. Gabriel Arruda, pela inestimvel amizade, e À Bárbara, pela recente. Ao grande Jojo da Bahia. Ao David. À Ana, Larissa, Luciana, Laura, Rafamé Juliana.

Ao companheiros na Espanha que los quiero mucho: Lucas, que sua cachaa nunca seque, Lu e David, vocês sâo demais!, Thaiara, pela buenísima onda de todos os dias, Elê, grande guerreira, Angel, grande amigo, Nerea, pelos belos momentos. Los amigos de UPV: Giuseppe, a good friend, Nuria, Iagoba, João, Maria, Laura. A los compis de piso: Julia, Itzel, Raiane. A la trupe de Sokole, Elena, Lourdes, Giuseppe, Sheila, y todos los amigos de la percu, gracias por este mundo que me dieran la oportunidad de adentrar.

Aos meus queridos forrozeiros, por aceitarem entrar nessa comigo, Nathy, Luis, Eliel do Pandeiro, Maré.

Ao povo brasileiro, pela confiança, ainda quase cega, de que a ciência vale de ser apoiada. À Beagá, por enfim me revelar alguns de seus caminhos.

"...não quero faca nem queijo.
Quero a fome."
Adelia Prado

Contents

Introduction

During the last century, the development of quantum theory (QT) exposed the physicists to theoretical puzzles and conceptual challenges which we are still trying to unravel. When studying the foundations of QT, there exists a great interest about how it differs from classical theory (CT), which advantages it might offer to society and how well can we *understand* the basic new concepts arising from its intricate mathematical formalism. Notions such as entanglement [1], contextuality [2], nonlocality [3, 4], complementarity [5] etc. are in front of us as obstacles to be surpassed but also as tools to be explored [6–8]. Research on these topics serves as much as for constructing the conceptual and theoretical framework as for several technological applications, ranging from the classical simulation of quantum systems [9–17] to better cryptographic strategies [18].

In CT, the physical properties of a system are described via deterministic functions on the set of classical states. Associated with an apple falling from a three, for example, there is a set of questions we might ask about it: what is its color, its position in a specific time, energy, momentum etc. If f is a function which associates each question X, for a physical system s, to the corresponding answer $f_X(s)$, then $f_{\text{color}}(\text{apple}) = \text{red}$, and at the level of probabilities we write $p_{\text{apple}}(\text{red} \mid \text{color}) = 1$, while $p_{\text{apple}}(\text{blue} \mid \text{color}) = 0$ and so on. This description is said to be *dispersion-free*, i.e. it is given in terms of probability distributions which assume only two values, 1 or 0, meaning the system having or not having the property. On the other hand, actual experiments exhibit a degree of imprecision, and ignorance on the actual state of affairs leads to a probabilistic description in terms of *mixtures* of dispersion-free distributions. In subsection 1.4.3, we will see that *no dispersion-free probability distribution can be assigned to*

the outcomes of the set of all quantum mechanical observables. This is an interesting discussion arising with the birth of quantum theory, and exploited by the so-called EPR paper, in 1935, written by A. Einstein, B. Podolsky, and N. Rosen [19]. In this paper, the authors argue about why should QT be considered incomplete, if it is assumed a local-realistic definition of physical models. Their point may be explained in a simpler version, created by D. Bohm [20]. Consider a singlet state $|\psi\rangle = 1/\sqrt{2}(|01\rangle - |10\rangle)$, written in the σ_z basis, distributed over two parties, Alice and Bob. They will make the same measurement σ_z on its part of the system, Alice first and then Bob. When Alice makes her measurement and gets its outcome, she is able to predict which is the measurement outcome of Bob, because they are anticorrelated. As she does not need to interact with Bob's system, this classically means that the measurement outcome is predefined to Bob. Now, the singlet state maintain its form when rotated, and writing it in the σ_x basis it becomes $|\psi\rangle = 1/\sqrt{2}(|+-\rangle - |-+\rangle)$, in which $|\pm\rangle = 1/\sqrt{2}(|0\rangle \pm |1\rangle)$. If instead of measuring σ_z, they measure σ_x, the same reasoning applies. This leads to the conclusion that the outcomes for the measurements in B must be predefined, according to EPR argument, and therefore quantum mechanics is incomplete, given it does not provide means to predict those quantities.

An example of a deterministic extension for QT appeared in 1952, proposed by D. Bohm [21]. It brings a formalism for QT based on the position of the particles and a global wave function which respects Schrödinger's quantum dynamics. Therefore, it is complete, in the sense proposed by EPR, because it allows for deterministic predictions of the properties of the system. The ignorance on the initial position of the system leads to the probabilistic predictions, as it commonly happens in classical theory. On the other hand, it has the strange feature of being *nonlocal*, i.e. actions of distant observers influence the measurement outcomes of others.

In the course of this line of research, the differences between QT and CT became more evident. In 1960, E. Specker [22, 23] found that if the dimension of the quantum state space is three or larger, there is no way to embed quantum logic, as formulated by J. von Neumann using the rules of QT, into classical logic, given a set of jointly decidable propositions, its conjunctions and implications.

In 1964, motivated by the EPR paper and by the de Broglie-Bohm pilot wave theory, mainly by its nonlocal character, Bell introduced the concept of *local hidden variables* [24] to study the differences between classical and quantum probability distributions for a bipartite scenario. A random variable λ, unknown to the experimenters, would contain the information needed for the outcomes of the measurements to become predefined, and then the ignorance on this variable is responsible for the indeterminacy of the outcomes. Given the bipartite scenario specified above, we denote by $A(\hat{a}, \lambda)$ the outcome of a spin measurement by Alice in the direction of the unitary vector $\hat{a}$, and $B(\hat{b}, \lambda)$ the analogous for Bob. If the λ has a probability distribution $\rho(\lambda)$, the expectation values for the product of the two observables is

$$E_L(\hat{a}, \hat{b}) = \int d\lambda \rho(\lambda) A(\hat{a}, \lambda) B(\hat{b}, \lambda). \tag{1}$$

The quantum correlations achieved by the singlet state in the same scenario is

$$E_Q(\hat{a}, \hat{b}) = \langle \sigma_{\hat{a}} \otimes \sigma_{\hat{b}} \rangle_{|\psi\rangle} = -\hat{a} \cdot \hat{b}. \tag{2}$$

Bell showed in this paper that the behaviours of E_L and E_Q are distinct, thus demonstrating theoretically the impossibility of explaining QT through the use of local hidden variables. Thus, these works started to show that, although EPR paper present reasonable assumptions, there is always some price to pay — in Bohm's case, for example, it's nonlocality — when trying to fit QT in classical interpretations, what frustrates the expectations of the paper written in 1935.

In 1966, Bell also showed [25] that when the dimension of the quantum state space is three or larger, there is no way to assign dispersion free distributions to measurement outcomes which satisfy the requirements of QT. He perceived that Gleason's theorem, which will be explained in subsection 1.4.3, implied this conclusion already. Nevertheless, in 1967 S. Kochen and E. Specker proved a similar proposal [26]; they solved the issue by analysing the equivalent problem of embedding quantum logic into classical logic, which is equivalent to assigning consistent definite values to quantum observables. The authors proved the impossibility of this assignment, with the power of attributing values simultaneously only to jointly measurable

observables, due to the impossibility of measuring two or more noncommuting observables in the same experimental apparatus.

Any set of jointly measurable observables defines a *context*, and the *hypothesis of noncontextuality* is defined by the demand that observables have predefined values independent of which other compatible observables are measured together. Thus, if $\{A, B, C, D, \ldots\}$ is a context and $\{A, B', C', D', \ldots\}$ is also a context, a variable λ implies an assignment of the same value to A in both contexts, i.e. there exists a well defined $A(\lambda)$. Thus, the papers discussed above proved that QT is *contextual*, in the sense that QT cannot be modeled by a noncontextual model.

There is a straightforward mathematical relation between contextuality and nonlocality. Observables measured in distinct locations at the same time are compatible by construction, thus the impossibility to model QT with a local model, i.e. the nonlocal character of the theory, may be interpreted as a particular manifestation of the contextual character of the theory, when the subsystems are space-like separated.

Different and simpler arguments proving that QT is contextual appeared later, e.g. Refs. [2, 27–31]. Two particularly important scenarios are discussed in this thesis, the Clauser–Horne–Shimony–Holt (CHSH) scenario [32] and the Peres-Mermin square [33, 34]. The first scenario, in a modern adaptation, is an experimental proposal of the scenario proposed by Bell in 1964, with the generalization consisting of probabilistic outcomes. It attributes probability distributions $p(a \mid A, \lambda)$ and $p(b \mid B, \lambda)$ to measurement outcomes a and b, when measuring A and B, depending on the variable λ, such that the probabilities achieved are of the form

$$p(a, b \mid A, B) = \int d\lambda \rho(\lambda) p(a \mid A, \lambda) p(b \mid B, \lambda). \tag{3}$$

These distributions respect the CHSH *Bell inequality*, in reference to Bell's original paper. Each party has two dichotomic observables, which can assume the values $+1$ or -1. The inequality

is written as

$$\langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2, \tag{4}$$

in which $\langle A_i B_j \rangle$ is the expectation value for the product of the outcomes of measuring A_i and B_j. It is known [35] that QT allows a violation of this inequality up to $2\sqrt{2}$, which is another proof that it cannot be modeled with local hidden variables.

The second scenario, the Peres-Mermin square, is one of the simplest scenarios in which contextuality manifests itself. It consist of one party measuring all the observables of one row or one column of the following square of observables

$$\begin{bmatrix} A & B & C \\ a & b & c \\ \alpha & \beta & \gamma \end{bmatrix}. \tag{5}$$

A noncontextual modeling assigns ± 1 value to each observable depending on λ, i.e $\{A(\lambda), B(\lambda), C(\lambda), a(\lambda), b(\lambda), c(\lambda), \alpha(\lambda), \beta(\lambda), \gamma(\lambda)\}$. An ensemble of this strategies respects the following inequality

$$\langle ABC \rangle + \langle abc \rangle + \langle \alpha\beta\gamma \rangle + \langle Aa\alpha \rangle + \langle Bb\beta \rangle - \langle Cc\gamma \rangle \leq 4. \tag{6}$$

On the other hand, using the following operators

$$\begin{bmatrix} \sigma_z \otimes \mathbb{1} & \mathbb{1} \otimes \sigma_z & \sigma_z \otimes \sigma_z \\ \mathbb{1} \otimes \sigma_x & \sigma_x \otimes \mathbb{1} & \sigma_x \otimes \sigma_x \\ \sigma_z \otimes \sigma_x & \sigma_x \otimes \sigma_z & \sigma_y \otimes \sigma_y \end{bmatrix}, \tag{7}$$

QT reaches the value of 6, which is the algebraic maximum of the expression, independently of which quantum system is being measured. This latter fact also demonstrates that QT cannot be modeled in terms of a noncontextual model, and that this is a feature of the measurements, since this violation is valid for all quantum states.

In the course of foundational research, it has been found that the idiosyncrasies of QT might in fact help in improving computational strategies. This is intriguing, since if we consider a system described by a Hilbert space of dimension d, only d states can be perfectly discriminated in a single-shot experiment, and its information-carring capacity is equal or less than $\log_2 d$ bits, according to Holevo bound [36]. Therefore, one could think that there is no advantage in using quantum systems as substrate to perform computations. These natural considerations are disproved by scientific works since the decade of 1980, when the field of quantum information started being shaped. Since then, there has been a great effort to understand the role of probabilities arising from quantum theory in order to offer more effective resources to computational problems with respect to classical programs [6, 37]. For example, distributing a quantum system over several parties might reduce the communication complexity over what is possible with classical systems alone [7, 12, 38]. The relation between this type of problems and Bell's nonlocality, and the identification of the latter as a resource is studied in [4, 39, 40]. One of the most clear scenarios in which this advantage in communication occurs is in the *superdense coding* scenario. In this situation, Alice and Bob share a state $|\psi\rangle = 1/\sqrt{2}(|00\rangle + |11\rangle)$. By manipulating her part in a specific way and sending the qubit to Bob, she can transmit *two bits* of information, as explained now. She wants to send to Bob one of the strings $00, 01, 10$ or 11, after applying only local measurements. Alice can transform locally the global state to one of the four orthogonal states:

$$\begin{aligned}
00 &: \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right) \qquad \text{(do nothing)},\\
01 &: \frac{1}{\sqrt{2}}\left(|00\rangle - |11\rangle\right) \qquad \text{(apply } \sigma_z\text{)},\\
10 &: \frac{1}{\sqrt{2}}\left(|10\rangle + |01\rangle\right) \qquad \text{(apply } \sigma_x\text{)},\\
11 &: \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right) \qquad \text{(apply } \sigma_x \text{ than } \sigma_z\text{)}.
\end{aligned} \tag{8}$$

In the first situation, to send 00, she does nothing. In the second, sending 01, she applies the σ_z gate in her part. To send 10, she applies σ_x. Finally, to send 11, she applies σ_x and than σ_z. As the resulting states are orthogonal, they can be discriminated in a single experiment. Alice then sends her qubit to Bob, and now in the possession of the two qubits, he can make

the measurement in the Bell basis and find out which is the two bit message Alice sent him.

One natural question regarding quantum nonlocality is quantifying the communication needed between parts in order to classically simulate its behaviours[1]. Generalizing this question to contextuality brings the idea of quantifying the resource necessary to simulate the results of quantum sequential measurements. The latter are found be simulated with the use of some memory, leading to the notion of *memory cost* [15, 16, 41], i.e. the classical memory needed to simulate the correlations occurring in sequences of quantum measurements by means of a classical automaton, defined below. It is impossible to accurately record a quantum state, given that we have to record complex numbers, and, even worse, the number of real parameters needed grows exponentially with the size of the system. One alternative is to simulate the probabilistic behavior of quantum systems in terms of generating outputs with identical probability distributions. During a sequence of measurements $X_1, \ldots, X_n$ with outputs $x_1, \ldots, x_n$, a quantum system updates its state in every step

$$\rho_0 \xrightarrow[x_1]{X_1} \rho_1 \xrightarrow[x_2]{X_2} \rho_2 \ldots \xrightarrow[x_n]{X_n} \rho_n. \tag{9}$$

The classical counterpart is modeled by a classical automaton with access to k distinct states

$$s_0 \xrightarrow[x_1]{X_1} s_1 \xrightarrow[x_2]{X_2} s_2 \ldots \xrightarrow[x_n]{X_n} s_n, \tag{10}$$

in which $s_i \in \{1, \ldots, k\}$. Those states can be seen as an internal memory of the machine; if we represent them as binary strings, we need to have access to a string of length of $\log_2 k$. The aim is to reproduce the distribution $p(x_1, \ldots, x_n \mid X_1, \ldots, X_n)$. It has been found that the memory cost can exceed the amount of information which can be stored in a quantum system defined for a Hilbert space of dimension d, i.e. $k \geq d$, and this implies that using a quantum system can reduce the memory needed for certain input-output processes, yielding a quantum memory advantage [11, 15, 16, 42–44]. In the seminal paper, which we will explain in chapter 2, Galvão and Hardy [11] showed that the simulation of a qubit in a unitary evolution requires

[1] Pragmatically, these behaviours are sets of probability distributions $p(ab\ldots \mid AB\ldots)$ which can be obtained in a given scenario.

an arbitrary large number of classical bits, depending on the number of steps. In Ref. [15], Kleinmann *et al.* introduced the concept of *memory cost to quantum contextuality*, and derived some bounds for different scenarios, including the Peres-Mermin scenario, which is great part of the subject of this thesis. They found a memory advantage while trying to simulate some deterministic predictions for outcomes of quantum measurements. For example, in a sequence of measurements $ABaA$ in a ququart, given that A commutes with every operator, but $[B, a] \neq 0$, QT predicts that the outcome for A must repeat in both measurements. For a certain choice of operators, we need k strict larger than four to be possible to reproduce this sequence, thus defining an instance of quantum memory advantage. In a recent paper, Cabello *et al.* [42] found memory advantages in important scenarios, in terms of the entropy of the set of strings of outputs generated by a quantum system.

Once classical resources are still much more accessible than quantum resources [45], the analysis of which scenarios would yield a quantum advantage is crucial. We are here interested in the analysis of the memory cost with respect to quantum contextuality, which means to determine the classical memory needed when the measurements in a sequence only embraces mutually compatible measurements [15], and compare this cost to the information carriage capacity of the system. In this strict form, the question of whether there exists a quantum memory advantage due to contextuality is still open. Finding a scenario with contextual memory advantage demonstrates a feature of quantum mechanics which can be used to improve computational strategies. On the other hand, if there is not any advantage in the contextuality paradigm, besides being fundamentally intriguing why the bound of d states exists, the incompatibility of quantum measurements would be a useful resource, since in this case there are situations which show memory advantage.

This thesis is organized as follows. Chapter 1 is dedicated to explain the basic probabilistic framework for quantum and classical measurements, and some experimental differences between these two approaches. Both CT and QT can be put in the common ground of *effects* and *instruments*, in which the probabilities extracted from the corresponding states are derived. For QT, the notion of post-measurement state is discussed, with later implications to quantum sequential measurements; in this situation the update rule for quantum states is crucial.

Furthermore, the notion of nonlocality and contextuality are discussed, and the Peres-Mermin scenario and CHSH scenario are explained in details.

Chapter 2 is devoted to extend the notion of classical models, to include other resources such as communication and memory, and how to simulate quantum scenarios with them, quantifying the resource to do this task. Results for communication are better explored and studied, and some lower bounds on communication cost and its relation with the violation of Bell inequalities are discussed. Then, sequential scenarios and its memory cost are considered. First, the memory cost to simulate a specific decision problem relating to unitary evolution of quantum system is addressed. At last, the Peres-Mermin scenario is studied, and a memory cost to simulate certain deterministic predictions is derived, giving a lower bound to the main question considered in this work.

Chapter 3 contains the main results of this thesis, solving two questions. First, the memory cost for simulating the Peres-Mermin scenario, using a proposal for modeling the sequence of measurements; these results have been published in Ref. [46]. We found that three internal states are sufficient to explain quantum correlations for all states and arbitrary sequences of compatible observables. And this raised the second question: the possibility of reproducing the same scenario, which uses two qubits, with a lower dimensional quantum system. As we prove, this is possible even with a classical system with three states of memory, and so it also possible with a qutrit, which leads us to reduce even more the dimension of the resource. The impossibility of reproducing the Peres-Mermin scenario with a qubit led us to rewrite the second question as a more general proposal: is there contextuality in a qubit? We address this issue considering repeatable and compatible instruments.

Chapter 1

Aspects of Classical and Quantum Theory

In this chapter we begin to expose the mathematical background and the main points to be explored in this work, relating to the confrontation between classical and quantum theory. The focus in the thesis is the cost of classical resources to simulate probabilities obtained from quantum states and measurements, therefore we briefly explain the probabilistic background in classical theory (CT), and then a modern formulation in quantum theory (QT). While in CT the outcome of measurements can be interpreted as the value of a property of the system, in which the error about this knowledge can be as close to zero as the experimental set up allows, in QT the notion of incompatible observables appears, and the measurement cannot be interpreted as a noninvasive procedure any more.

1.1 Classical Events

In classical theories, like Newtonian, Lagrangian and Hamiltonian Mechanics, the features of physical systems are usually defined as functions on the elements of the state space. Thus, position, momentum, energy etc. are functions of a classical state to the set of possible outcomes. This means that the events obtained from measuring those observable quantities are represented

by deterministic, or dispersion free, probability distributions; and the act of measuring only reveals the value the system holds for that observable even before the measurement. Usually, it is hard to know the exact classical state we prepare, or even impossible some times. Thus, a degree of imprecision in the description of the exact state leads to a probabilistic description of the measurement outcomes. Besides, the empirical distinction between all classical states is not perfect, because the measurement apparatus has an error associated to the measurement outcome. A third factor which could impact the discrimination of a given state is not knowing the exact evolution of the system through time, before measuring it. Then, when obtaining nondeterministic distributions for measurement outcomes, classical theories allow them to be interpreted as a mixture of the deterministic distributions, i.e., those distributions in which we know everything about the system and dynamics. In principle, it is possible to obtain more precise preparations of each state and more precise measurements so that the statistics obtained approaches a dispersion free distribution, as we will see in Sec. 1.3.1.

In order to make the above reasoning more clear, we will introduce a notation for the measurement scenario, which can be translated into an input-output process. Let the measurement choice be the input represented by the elements $X \in \mathcal{I}$, the outcome of a measurement by $x \in \mathcal{O}$, with $\mathcal{I}$ being the finite set of allowed inputs and $\mathcal{O}$ being the finite set of outputs[1]. An output is a function of the input, the state of the system and the strategy adopted. The latter may refer to direct questions of the properties of the system, and may include other general attributes, such as classical manipulation af available data before the output. It will be a general label that represents all the possibilities which can be achieved for available states and inputs. The output functions for every measurement are denoted by $X(s, \lambda)$, s being the state and λ the strategy. In the scope of classical reasoning, we assume we are testing one specific state for each strategy, i.e. the measurements do not change the state of the system, and thus the state is implied by the strategy, and the output function can be written only as $X(\lambda)$. The probabilities for the measurement outcome are represented by $l_\lambda(x_1 x_2 \dots x_N \mid X_1 X_2 \dots X_N)$, in which the string $x_1 x_2 \dots x_N$ is the string of outputs, $X_1 X_2 \dots X_N$ is the string of inputs and l_λ is the probability distribution for this string of outputs, given the choice of those inputs and

[1] Actually, there can be distinct sets of outputs for each input, but for simplicity we can use only one set, with the required adaptations.

the strategy λ. We are going to restrict ourselves to a finite number of inputs and outputs, because these cases are the only important ones in this thesis, and it simplifies the approach.

If the outcomes are functions of the input, the probabilities are *dispersion free*, which means that they attain the values 0 or 1, reproducing the deterministic behavior. Consequently, we have

$$l_\lambda(x_1 x_2 \dots x_N \mid X_1 X_2 \dots X_N) = \delta^{x_1}_{X_1(\lambda)} \delta^{x_2}_{X_2(\lambda)} \dots \delta^{x_N}_{X_N(\lambda)}. \tag{1.1}$$

In this equation, X_i and x_i, for $i = 1, \dots, N$, are arbitrary members of $\mathcal{I}$ and $\mathcal{O}$, respectively, and δ^a_b is the usual Kronecker delta. Therefore, this expression is 1 if all the outputs x are in accordance with the output functions $X(\lambda)$, and zero otherwise.

There is a larger class of deterministic behaviours, those in which an output is a function of all the inputs. This can happen for instance when the scenario is composed by several parties, in which communication is allowed after the input choice, and thus every party can know what the inputs of the other parties are and change its output accordingly. They are defined as

$$d_\lambda(x_1 x_2 \dots x_N \mid X_1 X_2 \dots X_N) := \delta^{x_1}_{f_1(\lambda, X_1, \dots, X_N)} \delta^{x_2}_{f_2(\lambda, X_1, \dots, X_N)} \dots \delta^{x_N}_{f_N(\lambda, X_1, \dots, X_N)}. \tag{1.2}$$

In the expression above, the f's are the response functions for each strategy λ and the whole string of inputs.

A general strategy generates a random variable λ according to some probability distribution $\rho(\lambda)$. Therefore, the probability distributions achieved are mixtures of the deterministic strategies d_λ, i.e.

$$p_\rho(x_1 x_2 \dots x_N \mid X_1 X_2 \dots X_N) = \sum_\lambda \rho(\lambda) d_\lambda(x_1 x_2 \dots x_N \mid X_1 X_2 \dots X_N). \tag{1.3}$$

1.2 Deterministic Strategies as Vertices of Polytopes

Eq. (1.3) may be interpreted as a convex mixture of vectors in a suitable vector space. Given a specific *scenario* — which is defined by its inputs, outputs and parties involved, together with the conditions they must respect — we write the possible strings of inputs as $E_i = X_1^{(i)} \dots X_{N_i}^{(i)}$, defined as the string i of inputs which are entered in the same run. Define also $e_{ij} = x_1^{(ij)} \dots x_{N_i}^{(ij)}$, given E_i for input, as all the possibilities for the outcomes, ordered by the j index. As an example, let us assume three dichotomic observables $\mathcal{I} = \{A, B, C\}$, in which $\mathcal{O} = \{1, 2\}$. The scenario consist of measuring $\vec{E} := (E_i)_{i=1}^4 = (A, B, AB, C)$, meaning that we can enter each of the inputs separately, but only A and B together. The output events are gathered in an n-tuple $\vec{e}_i = (e_{ij})_j$ such that, for example, $\vec{e}_1 = (1, 2)$ and $\vec{e}_3 = (11, 12, 21, 22)$, the possible outputs for A and AB, respectively. Defining the vectors $\vec{d}_\lambda$, such that $(d_\lambda)_{ij} := d_\lambda(e_{ij} \mid E_i)$, we see that a general convex combination

$$\sum_{\lambda=1}^{L} \rho(\lambda) \vec{d}_\lambda, \tag{1.4}$$

in which L is the total number of strategies, defines a convex polytope [47–49], because it is the convex hull of a finite number of points, which are the extremal vectors $\vec{d}_\lambda$. To be clear, this constitutes a polytope P defined as

$$\begin{aligned} P &:= \{\, \vec{q} \mid q_{ij} = \sum_\lambda d_\lambda(e_{ij} \mid E_i)\rho(\lambda), \quad \text{for some } \rho(\lambda) \,\} \\ &:= \{\, \vec{q} \mid \vec{q} = D\vec{\rho}, \quad \text{for some } \vec{\rho} \} \end{aligned} \tag{1.5}$$

in which $\vec{\rho} = (\rho(\lambda = 1), \rho(\lambda = 2), \dots, \rho(\lambda = L))$, with $\rho(\lambda) \geq 0$ and $\sum_{\lambda=1}^{L} \rho(\lambda) = 1$. The matrix D has its columns formed with the vectors $\vec{d}_\lambda$, i.e. $D = [\vec{d}_1 \ \vec{d}_2 \ \dots \vec{d}_L]$. Therefore, defining $e_{ij} \mid E_i$ as the event of obtaining the outcomes e_{ij} when making the measurements in E_i, q_{ij} is the probability for the event $e_{ij} \mid E_i$ given the ensemble characterized by $\rho(\lambda)$.

The above way of writing the convex polytope, a convex hull of a finite number of vertices, is called *vertex representation* (or V-representation), see Fig.1.1(a); more details about

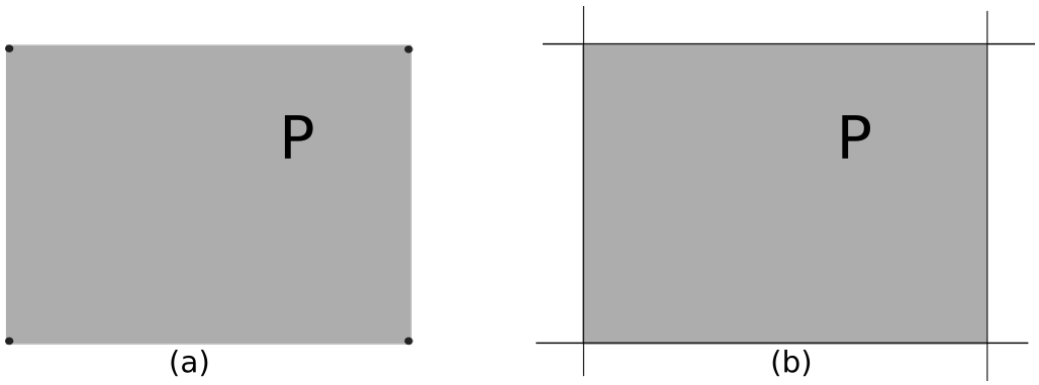

Figure 1.1: (a) Convex set formed as convex mixture of a finite number of vertices. (b) Convex set formed by the points which are located in the intersection of the half-spaces defined by the inequalities.

general representations can be found in [47–49]. Another useful way of characterizing the polytopes, which will be a major tool of the present thesis, is the half-space representation (H-representation), see Fig. 1.1(b). It consists of constructing the polytope as an intersection of half spaces, defined by linear inequalities, i.e.

$$P = \{\vec{q} \mid \vec{h}_l.\vec{q} \leq \alpha_l \ \ \forall\ l\}. \tag{1.6}$$

In this expression, the $\vec{h}_l$ are vectors of coefficients for the linear inequalities and, together with α_l, they determine a half space region in the original vector space. The intersection of these half-spaces defines the same polytope.

1.3 States and Effects

Usually in an experiment, it is possible to define the process in two steps: preparation and measurement. The line separating the two regions is quite arbitrary, and in these procedures might be included transformations, such as unitary gates, on a state that comes out of a source, for example.

The notion of *effects* allows us to embrace both the classical and the quantum probabilistic framework. From this section until Sec. 1.4.8, the content is formulated and disposed according to the book of T. Heinosaari and M. Ziman [50], with the required adaptations. This

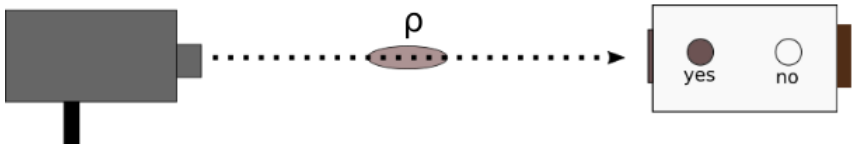

Figure 1.2: A physical system is prepared and sent to a measurement apparatus, which outputs "yes" or "no", based on the received system λ and internal features of the apparatus.

book provides a modern compilation of important results for the mathematical formalism of QT, and the effects are one of the starting points for the construction that follows. The interested reader might also want to read Ref. [51], a earlier publication with a similar approach.

According to the authors, an effect "is a measurement apparatus that produces either 'yes' or 'no' as an outcome". In this general framework, given a state ρ we associate an effect

$$E : \rho \mapsto [0, 1] \tag{1.7}$$

as the probability of an 'yes' answer. By definition, the *identity effect* I assigns probability one to every state, i.e. $I(\rho) = 1$, and the *zero effect* O assigns a null probability, i.e. $O(\rho) = 0$.

If we prepare two states ρ_1 and ρ_2, and randomly select ρ_1 with probability p, this results in a new preparation procedure, and thus the effect must act in the convex mixture of the states. This means that the resultant effect is also a convex mixture of the original effects, i.e.

$$E(p\rho_1 + (1-p)\rho_2) = pE(\rho_1) + (1-p)E(\rho_2). \tag{1.8}$$

Another starting point would be if we turn our attention to the states. Then, it is possible to define them with respect to the allowed effects for each theory. They can be seen as what comes out from the preparation procedure. In this way, the meaning of states would be to represent the physical objects which behave identically under all experiments, i.e. a *state* is the set of preparation procedures that identically leads to the same effects. In other words, we might be able to recognize the set of effects demanded by the experiments, and, labeling by

the states, we can define the functions

$$f_\rho(E) := E(\rho), \tag{1.9}$$

and for two states ρ_1 and ρ_2, we have $\rho_1 = \rho_2$ iff both states define the same function for all effects, i.e.

$$f_{\rho_1}(E) = f_{\rho_2}(E), \quad \forall \ E. \tag{1.10}$$

Finally, convex combination of the states must respect

$$(\lambda f + (1-\lambda)f')(E) = \lambda f(E) + (1-\lambda)f'(E). \tag{1.11}$$

We will take as the starting point the states and derive the effects associated to them. As an effect can be seen as the probability of a "yes" answer, in a measurement with more than two outcomes, the probability of each outcome is represented by an effect.

1.3.1 Classical Effects

The tools described above can be used both in classical and quantum formalisms. The application in the classical realm is done through the identification of classical preparation procedures, i.e. classical states, as members of a state space Ω, and effects as functions from Ω to the interval $[0, 1]$. Usually Ω is infinite, like the usual position vs. momentum phase space, which for N degrees of freedom is associated with $\mathbb{R}^{2N}$. In this thesis, though, we are going to deal only discrete state space. Thus, in the scenario in which Ω contains d elements, a classical preparation procedure is represented by a probability vector

$$\vec{p} = (p_1, \ldots, p_d). \tag{1.12}$$

The equation above is interpreted as preparing the state i with a probability p_i, and it can be seen as an ensemble of the d distinguishable states. On the other hand, the effects are represented by

$$\vec{e} = (e_1, \ldots, e_d), \tag{1.13}$$

satisfying $0 \leq e_i \leq 1$. The probability for a 'yes' answer for this effect is $\vec{p} \cdot \vec{e} = \sum_i p_i e_i$. The identity effect is $I = (1, 1, \ldots, 1)$ and the zero effect is $O = (0, 0, \ldots, 0)$.

There is no state for which all the effects are *dispersion free*, i.e. produces only 0 or 1, but we can restrict the analysis to extremal states and to extremal effects. An element of a convex set is *extremal* iff it cannot be written as a convex combination of any other element. In this situation, both p_i and e_i belong to $\{0, 1\}$, and this corresponds to the deterministic scenario in this formalism. Therefore, there are d extremal states and 2^d extremal effects. If the state i has the property associated to an effect then $e_i = 1$, otherwise it is zero. Consequently, it is possible to interpret the classical effects as a mixture of deterministic effects on a discrete set of states.

1.4 Quantum Theory

Quantum theory (QT) is a consistent probabilistic theory developed through the last century, and its foundations raises interesting confrontations with the foundations of classical theory. One of its most intriguing features is *contextuality*, in which the possibility that quantum systems have predefined values as outcomes of measurements, and therefore dispersion free probability measures for measurement outcomes, implies that the outcome of each measurement has to depend on which other observables are measured alongside. In other words, the measurement outcomes would depend on the *context* of the measurement. We explain now the framework for QT, and later the fundamental aspects and non-usual behaviors with respect to classical reasoning.

1.4.1 Mathematical Formalism

The state space of QT is a subset of the space of linear operators in the Hilbert space $\mathcal{L}(\mathcal{H})$. Considering the mathematical description, the only specific characteristic of the state space is the dimension of the underlying Hilbert space $\mathcal{H}$. The set $\mathcal{S}(\mathcal{H})$ of physical states associated to a Hilbert space $\mathcal{H}$ is

$$\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{L}(\mathcal{H}) \mid \rho \geq 0, \mathrm{tr}(\rho) = 1\}. \tag{1.14}$$

The members of this space are usually called *density matrices*. Consequently, effects are linear mappings $\mathcal{S}(\mathcal{H}) \to [0, 1]$. These properties guarantee (c.f. [50], pg. 68) that for every effect there is a selfadjoint operator $\hat{E}$, such that

$$E(\rho) = \mathrm{tr}(\rho\hat{E}). \tag{1.15}$$

The equation above is related to the *Born rule* in this approach. It associates every measurement outcome to an effect, and the probability for that outcome is given by Eq. (1.15). Every $\hat{E}$ has to respect $0 \leq \hat{E} \leq I$. As we are going to deal only with quantum effects, from now on we will write the selfadjoint operators without the hats. In this way, the set of effects is denoted by $\mathcal{E}(\mathcal{H})$:

$$\mathcal{E}(\mathcal{H}) = \{\hat{E} \in \mathcal{L}_S(\mathcal{H}) \mid 0 \leq E \leq I\}. \tag{1.16}$$

In the equation above, $\mathcal{L}_S(\mathcal{H})$ is the set of selfadjoint linear operators on the Hilbert space $\mathcal{H}$.

Note that projections are effects, but not all effects are projections. For example, assuming the reader familiar with Dirac notation, an operator $t\,|\psi\rangle\langle\psi|$, for positive $t < 1$, is an effect but not a projection. The collection of projections is denoted by $\mathcal{P}(\mathcal{H})$.

1.4.2 Observables and POVMs

Effects are used to formalize the notion of probabilistic "yes" or "no" outcomes inside the Hilbert space formalism of quantum theory. Usually, though, an experiment has several outcomes for a given preparation, each one occuring with a certain probability. The term *observable* defines the mathematical description of this idea.

Given an experiment with n outcomes on a quantum state ρ , each of the outcomes is represented by an effect $E_i, i = 1, \ldots, n$. Assuming that there is always an outcome in every run, the normalization of probabilities implies that

$$\mathrm{tr}(\rho E_1) + \mathrm{tr}(\rho E_2) + \ldots + \mathrm{tr}(\rho E_n) = 1. \tag{1.17}$$

in which $0 \leq \mathrm{tr}(\rho E_j) \leq 1$. As this is required to be valid in all quantum states ρ, the effects have to obey

$$\sum_{j=1}^{n} E_j = I, \tag{1.18}$$

Concluding, we can say that a discrete observable is a collection of effects which obey condition (1.18).

In the general case, we need to define the objects more carefully, with the notion of generalized probability measures on the set of effects. If Ω is a non-empty set, a collection $\mathcal{F}$ of subsets of Ω is a *σ-algebra* if:

1. Ω is in $\mathcal{F}$;
2. if $X_1, X_2, \ldots$ are in $\mathcal{F}$, then $\cup_i X_i$ is also in $\mathcal{F}$;
3. if X is in $\mathcal{F}$, $\Omega \backslash X$ is also in $\mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a measurable space. An element of $X \in \mathcal{F}$ is called an *event*. Thus a σ-algebra allows to assign probabilities to events in a consistent way. A *probability*

measure is an assignment $p : \mathcal{F} \mapsto [0, 1]$ such that

1. $p(\Omega) = 1$;
2. $p(\cup_i X_i) = \sum_i p(X_i)$ for sequences $\{X_i\}$ of pairwise disjoint sets in $\mathcal{F}$.

Then, it is possible to define *positive operator-valued measure* (POVM) as a mapping $A : \mathcal{F} \to \mathcal{E}(\mathcal{H})$ such that

1. $A(\Omega) = I$;
2. $A(\cup_i X_i) = \sum_i A(X_i)$, for sequences $\{X_i\}$ of disjoint sets $X_i \in \mathcal{F}$.

The conditions above guarantee that this mapping A defines a probability measure for every set of outcomes X for all quantum states ρ via $\mathrm{tr}[\rho A(X)]$. The possible observables in quantum theory are now defined with the set of POVMs. For observables which have a countable set of outcomes x_i, this set is identified with Ω, i.e. $\Omega = \{x_1, x_2 \ldots\}$, and $\mathcal{F}$ is chosen to be the power set 2^Ω, which is the set constructed with all possible subsets of Ω and therefore has $2^{|\Omega|}$ elements. By the conditions for POVMs applied to a countable Ω, an observable A of this type is totally characterized by the assignments

$$x_j \mapsto A(x_j), \tag{1.19}$$

with $\sum_i A(x_i) = I$.

An observable is called *sharp*, or a projection, if for every $X \in \mathcal{F}$, $A(X)$ is a projection, i.e. $A(X) = [A(X)]^2$.

Proposition 1.1. *The following statements are equivalent (c.f. [50], Prop. 3.29)*

1. *A is sharp;*
2. *$A(X)A(Y) = A(X \cap Y)$, for every $X, Y \in \mathcal{F}$;*

3. $A(X)A(\Omega/X) = O$, *for every* X.

Proposition 1.2. *The range of a sharp observable* A *consists of mutually commuting projections.*

Proof. It follows from proposition 1.1. If $A(X)$ and $A(Y)$ are in the range of A,

$$A(X)A(Y) = A(X \cap Y) = A(Y \cap X) = A(Y)A(X). \tag{1.20}$$

□

Proposition 1.3. *If* A *is a sharp observable on a Hilbert space of dimension* d*, then it consists of at most* d *outcomes with nonzero probability represented by orthogonal projections.*

Proof. Proposition 1.1 tells us that when $X \cap Y = \emptyset$ then $A(X)A(Y) = 0$. And there can be at most d mutually orthogonal vectors, each one representing one of the projectors, therefore associated to d outcomes.

□

In the usual textbooks of quantum mechanics [52, 53], the observables are defined in terms of selfadjoint operators. How does the formalism explained in this section relates to the usual one? The answer comes from the *spectral decomposition theorem.*

Theorem 1.1. *Spectral Decomposition*

If T *is a Hermitian operator, there exists an orthonormal basis* $\{\phi_i\}$ *and a sequence* $\{\lambda_j\}$ *of real numbers such that*

$$T = \sum_i \lambda_i \left|\phi_i\right\rangle \left\langle\phi_i\right|. \tag{1.21}$$

Given a Hilbert space of dimension d, an experiment with d distinct outcomes $\{a_1, \ldots, a_d\}$,

and an associated orthonormal basis $\{|a_i\rangle\}_{i=1}^{d}$, the following operator is selfadjoint

$$A = \sum_i a_i P_{a_i}, \tag{1.22}$$

in which P_{a_i} is the projector into the subspace associated to the outcome a_i. Besides, starting with some selfadjoint A, its eigenvalues are denoted by a_i, representing the outcomes, with the corresponding eigenvectors spanning the associated subspace. The projector onto those subspaces are represented by P_{a_i}, and the probability for each outcome is given by $\mathrm{tr}[\rho P_{a_i}]$. Consequently, the expectation value of a sharp observable A might be expressed as

$$\begin{aligned} \langle A \rangle_\rho &= \sum_i a_i p(a_i) \\ &= \sum_i a_i \,\mathrm{tr}[\rho P_{a_i}] \\ &= \mathrm{tr}(\rho A). \end{aligned} \tag{1.23}$$

1.4.3 Extremal Effects

We already saw that the extremal states on classical mechanics are dispersion free when restricted to extremal effects. Below we see that trying to interpret the probabilities obtained in QT in terms of dispersion-free assignments is impossible, for $\dim(\mathcal{H}) \geq 3$.

As already mentioned, an element of a convex set is called *extremal* if it cannot be written as a convex sum of other elements. The extremal states from $\mathcal{S}(\mathcal{H})$ are called *pure states*. If a state is not pure, it is called a *mixed state*.

The theorem 1.1 guarantees there is a *canonical convex decomposition* for quantum states, expressed in the following theorem:

Theorem 1.2. *Canonical Convex Decomposition*

If $\rho \in \mathcal{S}(\mathcal{H})$, it can be expressed as a convex sum of projectors

$$\rho = \sum_i \lambda_i P_i. \tag{1.24}$$

As ρ is Hermitian, this is proved using the trace condition $\mathrm{tr}[\rho] = 1$ and positivity. The existence of a canonical convex decomposition implies that *the extremal elements of the state space are the projectors P_i*. It is a good moment to point that the notion of extremality is, in general, different from that of boundary states. A state belongs to the boundary if there is another arbitrarily close linear selfadjoint operator with trace one that is not a state. In other words, ρ is on the boundary if, for each $\epsilon > 0$, there exists a selfadjoint operator ξ_ϵ, with $\mathrm{tr}\,[\xi_\epsilon] = 1$, such that $||\rho - \xi_\epsilon||_{\mathrm{tr}} < \epsilon$, but $\xi_\epsilon \notin \mathcal{S}(\mathcal{H})$. An extremal state is on the boundary, but usually the boundary is formed also by mixed states. The example of the cube illustrates that. The extremal elements are its vertices, and its faces belong to its boundary, although they are formed by convex combinations of its vertices. For QT, the only example in which the boundary coincides with the extremal states, considering a fixed dimension, is in the case of qubits. It is interesting to note that all states which have eigenvalue 0 belong to the boundary.

The results for the extremal effects have similarities with those for the extremal states, as seen below.

Proposition 1.4. *The extremal elements of the set of effects $\mathcal{E}(\mathcal{H})$ are the projections.*

Proof. To see that every projection is an extremal element, try to write it as a convex sum of two other effects

$$P = \lambda E_1 + (1 - \lambda) E_2. \tag{1.25}$$

Find some $|\psi\rangle \in \mathcal{H}$ such that $\langle\psi|P\psi\rangle = 0$, i.e.

$$\lambda \langle\psi|E_1\psi\rangle + (1 - \lambda) \langle\psi|E_2\psi\rangle = 0 \tag{1.26}$$

This implies that $\langle\psi|E_1\psi\rangle = 0$. Therefore, we see that $P|\psi\rangle = 0 \implies E_1|\psi\rangle = 0$. Analogously,

using the projection $I - P$ and another vector $|\phi\rangle$, such that $(I - P)|\phi\rangle = 0$, we see that $P|\phi\rangle = |\phi\rangle \implies E_1|\phi\rangle = |\phi\rangle$. As a projection can be expanded in terms of its eigenvectors, one should conclude that $E_1 = P$, therefore E_2 must also be equal to P and only trivial expansions of a projector are consistent. Therefore, the projectors are extremal.

To see why only the projections are extremal, assume that there is an effect A which is not a projection but is extremal. The derived operator $E_1 = A^2 \neq A$ is also an effect, and $E_2 = 2A - A^2$ is also an effect. The fact that the latter is an effect is easily seen in the expression $I - E_2 = I - 2A + A^2 = (I - A)^2$, which is an effect. Therefore,

$$A = \frac{1}{2}(E_1 + E_2), \tag{1.27}$$

and consequently A is not extremal. □

1.4.4 Gleason's Theorem

Trying to assign a classical view to quantum theory is the same as trying to find dispersion free assignments to the set of extremal POVMs acting on pure states. The projective measurements are extremal, and it is enough to restrict the analysis to them. Below we show the impossibility of dispersion-free assignments to projections, if $\dim \mathcal{H} \geq 3$.

We want to find a probability measure on the projections $P_i \in \mathcal{P}(\mathcal{H})$ such that $f(P_i) \in \{0, 1\}$, and f has to obey

1. $f(I) = 1$ and
2. $f(\sum_i |\varphi_i\rangle\langle\varphi_i|) = \sum_i f(|\varphi_i\rangle\langle\varphi_i|)$,

in which $\{\varphi_i\}$ forms an orthonormal set. The two assumptions are independent; while the first one was already explained, the second is a desirable behaviour of the probability assignments. It is formalizing the notion that if two outcomes x_1 and x_2 are associated to two projections and the latter belong to the range of the same observable — and therefore they are orthogonal

— the observation of an event which is the outcome x_1 *or* x_2 is the sum of each probability. This corresponds to summing the projections.

If two states $|\eta\rangle$ and $|\phi\rangle$ are orthogonal and $f(|\eta\rangle\langle\eta|) = 1$, then $f(|\phi\rangle\langle\phi|) = 0$. If both $f(|\eta\rangle\langle\eta|) = f(|\phi\rangle\langle\phi|) = 0$, then every linear combination of η and ϕ, i.e. $\psi = a|\eta\rangle + b|\phi\rangle$, with $|a|^2 + |b|^2 = 1$, must lead to $f(|\psi\rangle\langle\psi|) = 0$. This comes from noticing that

$$f\left(|a\eta + b\phi\rangle\langle a\eta + b\phi| + |b^*\eta - a^*\phi\rangle\langle b^*\eta - a^*\phi|\right) = f\left(|\eta\rangle\langle\eta| + |\phi\rangle\langle\phi|\right) = 0, \tag{1.28}$$

and, therefore $f(|a\eta + b\phi\rangle\langle a\eta + b\phi|) = f(|b^*\eta - a^*\phi\rangle\langle b^*\eta - a^*\phi|) = 0$, since $a|\eta\rangle + b|\phi\rangle$ and $b^*|\eta\rangle - a^*|\phi\rangle$ are orthogonal.

Theorem 1.3. *If* $\dim\mathcal{H} \geq 3$*, there is no dispersion free probability measure on* $\mathcal{P}(\mathcal{H})$*.*

Proof. Let us suppose that exists a dispersion free probability measure on $\mathcal{P}(\mathcal{H})$ when $\dim\mathcal{H} \geq 3$.

Choose some vector $|\varphi\rangle$ for which $f(|\varphi\rangle\langle\varphi|) = 1$. Suppose there is another vector $|\phi\rangle$ such that $f(|\phi\rangle\langle\phi|) = 0$ and $||\,|\varphi\rangle - |\phi\rangle\,|| \leq 1/3$. We define

$$|\phi\rangle = \frac{|\varphi\rangle + \epsilon|\varphi'\rangle}{\sqrt{1+\epsilon^2}}, \tag{1.29}$$

in which $\epsilon \in \mathbb{R}, \epsilon > 0$. Furthermore, we choose $|\varphi'\rangle$ orthogonal to $|\varphi\rangle$, thus we see that ϵ must be less than $1/2$. Now, choose another vector $|\varphi''\rangle$, orthogonal to both $|\varphi\rangle$ and $|\varphi'\rangle$, and define, for any nonzero real number γ, other two orthogonal vectors

$$\begin{aligned} |\psi\rangle_\gamma &= \frac{1}{N}(-|\varphi'\rangle + \gamma|\varphi''\rangle), \\ |\psi'\rangle_\gamma &= \frac{1}{N'}\left(|\phi\rangle + \frac{\epsilon}{\gamma\sqrt{1+\epsilon^2}}|\varphi''\rangle\right), \end{aligned} \tag{1.30}$$

in which N and N' are normalization factors. By the definitions above, $f(|\phi\rangle\langle\phi|) = f(|\varphi'\rangle\langle\varphi'|) = f(|\varphi''\rangle\langle\varphi''|) = f(|\psi_\gamma\rangle\langle\psi_\gamma|) = f(|\psi'_\gamma\rangle\langle\psi'_\gamma|) = 0$.

Since $\epsilon < 1/2$, we define $\gamma_\pm$ by

$$\epsilon(\gamma_\pm + \gamma_\pm^{-1}) = \pm 1. \tag{1.31}$$

The expression above comes from the inequality $|\gamma + \gamma^{-1}| \geq 2$, for any γ. Finally, defining another pair of vectors

$$\begin{aligned} |\eta_\pm\rangle =& \frac{\epsilon\sqrt{1+\gamma_\pm^2}\,|\psi_{\gamma_\pm}\rangle + \sqrt{1+\epsilon^2(1+\gamma_\pm^{-2})}\,|\psi'_{\gamma_\pm}\rangle}{\sqrt{1+\epsilon^2(\gamma_\pm+\gamma_\pm^{-1})^2}} \\ =& \frac{|\varphi\rangle + \epsilon(\gamma_\pm+\gamma_\pm^{-1})\,|\varphi''\rangle}{\sqrt{1+\epsilon^2(\gamma_\pm+\gamma_\pm^{-1})^2}} = \frac{1}{\sqrt{2}}(|\varphi\rangle \pm |\varphi''\rangle). \end{aligned} \tag{1.32}$$

Now, the $\eta_\pm$ are linear combinations of ψ_γ, thus $f(|\eta_\pm\rangle\langle\eta_\pm|) = 0$. On the other hand,

$$|\varphi\rangle = \frac{1}{\sqrt{2}}\left(|\eta_+\rangle + |\eta_-\rangle\right), \tag{1.33}$$

which leads to $f(|\varphi\rangle\langle\varphi|) = 0$. This is in contradiction with our original assumption, and therefore, examining the logical steps, the assumption $f(|\phi\rangle\langle\phi|) = 0$ is inconsistent. Concluding, as the values must be distributed according to the normalization constraint, there is no way to assign a dispersion free probability distribution for the continuous set of projections. □

A profound theorem which describes the unique generalized probability measure for the set of projections through the state space is the *Gleason's theorem*:

Theorem 1.4. (Gleason's theorem)

If $\dim\mathcal{H} \geq 3$*, for every probability measure* $f : \mathcal{P}(\mathcal{H}) \to [0,1]$ *there is a unique operator* $\rho_f \in \mathcal{S}(\mathcal{H})$ *such that*

$$f(P) = \mathrm{tr}(\rho_f P). \tag{1.34}$$

The above relation means that the probabilities for the extremal effects one might realize

in a lab on a physical system, in the framework of QT, are derived from a specific quantum state through the above formula. As it is never dispersion free, it is impossible for QT to be interpreted like a mixture of deterministic behaviors. It is a good point to note that two of the most important fundamental questions left — and they are reduntant, i.e. answering one automatically answers the other, nonetheless it is important to choose — is why is the state space of QT the set of density matrices associated to the Hilbert space formalism, or why the effect space is $\mathcal{E}(\mathcal{H})$. Gleason's theorem is for systems with dimension greater than two, while for dimension two there are explicit dispersion free models, the interested reader can check in [50], pg. 80.

1.4.5 Composite Systems

Suppose there is a physical system with two distinguishable parts, A and B, described by two quantum states $\rho_A \in \mathcal{S}(\mathcal{H}_A)$ and $\rho_B \in \mathcal{S}(\mathcal{H}_B)$. One should look for a way to consistently describe the joint system. Denoting the effects which act on each part respectively by E_A and E_B, if they are uncorrelated, there should be a joint effect $\gamma(E_A, E_B) \in \mathcal{E}(\mathcal{H}_{AB})$, and a joint quantum system $\bar{\gamma}(\rho_A, \rho_B) \in \mathcal{S}(\mathcal{H}_{AB})$ such that

$$\operatorname{tr}(\bar{\gamma}(\rho_A, \rho_B)\gamma(E_A, E_B)) = \operatorname{tr}(\rho_A E_A)\operatorname{tr}(\rho_B E_B). \tag{1.35}$$

We identify $\mathcal{H}_{AB}$ with the tensor product of the separated spaces $\mathcal{H}_{AB} = \mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}}$. Then, the joint effects are tensor products of each effect $\gamma(E_A, E_B) = E_A \otimes E_B$ and analogously for the state, $\bar{\gamma}(\rho_A, \rho_B) = \rho_A \otimes \rho_B$.

Note that the state space consists not only of the composition of uncorrelated systems and effects, but of more exotic elements ρ_{AB}, which gives rise to many foundational questions. These states are the so-called *entangled states*, which are those which cannot be written as a

convex sum of product states

$$\rho_{AB} \neq \sum_i \gamma_i \rho_A^{(i)} \otimes \rho_B^{(i)}, \tag{1.36}$$

for any $\rho_A^{(i)}$ and $\rho_B^{(i)}$, and $\gamma_i \in \mathbb{R}_+, \sum_i \gamma_i = 1$.

For a composite system, it is possible to obtain information about its parts, by the *partial trace*.

Definition 1.1. *The partial trace for the system A, over B, is a mapping $\mathrm{tr}_B : \mathcal{S}(\mathcal{H}_{AB}) \to \mathcal{S}(\mathcal{H}_A)$ such that*

$$\mathrm{tr}[\mathrm{tr}_B[\rho_{AB}]E] = \mathrm{tr}[\rho_{AB}(E \otimes I)], \tag{1.37}$$

for all effects E. The partial trace over A is defined in analogous way.

The statistics for the system A alone is obtained through the quantum state resulting from the partial trace over the system B, i.e. $\rho_A := \mathrm{tr}_B[\rho_{AB}]$.

1.4.6 Quantum Channels

Physical states are able to undergo transformations, and in QT this is done through quantum channels and quantum operations. The authors of the book in Ref. [50] define a *channel* as a transformation which takes as input a quantum state and maps it to another state. An *operation* is a more general idea which maps quantum states to subnormalized quantum states, thus allowing for example loss of systems during a transformation. *Subnormalized* states are defined as

$$\tilde{\mathcal{S}} = \{\rho \in \mathcal{L}(\mathcal{H}) \mid \rho \geq O, \mathrm{tr}(\rho) \leq 1\}. \tag{1.38}$$

The operations $\mathcal{N}$ which act on quantum states must be linear, i.e.

$$\mathcal{N}(\sum_i \lambda_i \rho_i) = \sum_i \lambda_i \mathcal{N}(\rho_i), \tag{1.39}$$

in order to provide the same action on two different convex decompositions of the same quantum state. Therefore, these operations are considered to be linear mappings on $\bar{\mathcal{S}}(\mathcal{H})$. They have to satisfy

- $\mathrm{tr}[\mathcal{N}(\rho)] \leq 1$,
- $\mathcal{N}(\rho) \geq O$.

The operation $\mathcal{N}$ which respects $\mathrm{tr}[\mathcal{N}(T)] \leq \mathrm{tr}[T]$, for all positive operators $T \in \mathcal{T}(\mathcal{H})$, is called *trace nonincreasing* operation. And, if for all T, it respects $\mathrm{tr}[\mathcal{N}(T)] = \mathrm{tr}[T]$, it is called *trace preserving*.

As the composition of systems is done through tensor product, an operation acting locally on a quantum system must transform the whole composite system into a positive operator. This is guaranteed through demanding *complete positiveness*. In composite system AB, an operation acting on A will be represented by

$$\mathcal{N}_A(\rho_A) \otimes \rho_B = (\mathcal{N}_A \otimes I)(\rho_A \otimes \rho_B). \tag{1.40}$$

The positiveness of $\mathcal{N}_A$ does not always imply positivity of any extension $\mathcal{N}_A \otimes I$, as it happens for example with *transposition*. Choosing an orthogonal basis $\{|\phi\rangle_i\}_{i-1}^d$, the linear operation of *transposition* τ permutes the ket and bra elements, i.e.

$$\tau(|\phi_i\rangle \langle\phi_j|) = |\phi_j\rangle \langle\phi_i| . \tag{1.41}$$

As the eigenvalues are not modified by the transposition, it is a positive operation. Now consider

the operation $\tau_A \otimes I$ applied on the state $|\psi_+\rangle\langle\psi_+|$, in which

$$|\psi_+\rangle = \frac{1}{\sqrt{d}} \sum_j |\phi_j\rangle \otimes |\phi_j\rangle .$$
$$\implies |\psi_+\rangle\langle\psi_+| = \sum_{i,j} \frac{1}{d} |\phi_i\rangle\langle\phi_j| \otimes |\phi_i\rangle\langle\phi_j| . \tag{1.42}$$

Thus, $\tau_A \otimes I(|\psi_+\rangle\langle\psi_+|) = \frac{1}{d}\sum_{i,j} |\phi_j\rangle\langle\phi_i| \otimes |\phi_i\rangle\langle\phi_j|$. This operator is not positive semidefinite, since

$$(\sum_{i,j} |\phi_j\rangle\langle\phi_i| \otimes |\phi_i\rangle\langle\phi_j|) \cdot (|\phi_1\rangle \otimes |\phi_2\rangle - |\phi_2\rangle \otimes |\phi_1\rangle) = -(|\phi_1\rangle \otimes |\phi_2\rangle - |\phi_2\rangle \otimes |\phi_1\rangle), \tag{1.43}$$

and therefore, it has negative eigenvalues.

Definition 1.2. *A linear mapping $\mathcal{N}_A$ is* completely positive *if every mapping $\mathcal{N}_A \otimes I \in \mathcal{T}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is positive, for all finite dimensional extensions $\mathcal{H}_B$.*

Summarizing, a quantum channel/operation obeys the following properties:

1. linearity,
2. complete positiveness,
3. trace preserving/nonincreasing.

The channels have a very useful representation in terms of a sum of operators, which we use to prove a result in Sec. 3.2. First, note that a state of the form $\mathcal{N}_S(\rho) = S\rho S^\dagger$ is valid subnormalized state, for $S^\dagger S \leq I$ and $\mathrm{tr}\,[S] < \infty$. The properties of linearity and trace nonincreasingness are implied by these facts, and the complete positivity comes from

$$\langle\psi \mid (S\rho S^\dagger \otimes \mathcal{I}) \mid \psi\rangle = \langle\psi \mid (S \otimes I)(\rho \otimes I)(S^\dagger \otimes I) \mid \psi\rangle = \langle\tilde{\psi} \mid (\rho \otimes \mathcal{I}) \mid \tilde{\psi}\rangle \geq 0, \tag{1.44}$$

where $|\tilde{\psi}\rangle = (S^\dagger \otimes I)|\psi\rangle$. If $S^\dagger S = I$, then $\mathcal{N}_S$ is a channel.

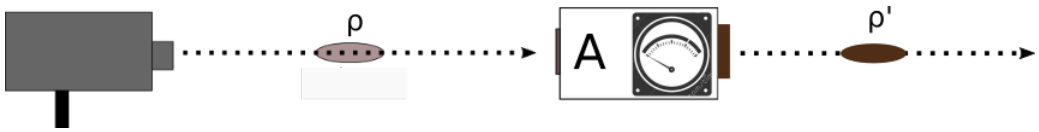

Figure 1.3: A system is prepared in the state ρ; after the measurement it changes to state ρ'.

Also, by linearity, the sum

$$\mathcal{N}_S = \mathcal{N}_{S_1} + \mathcal{N}_{S_2} + \ldots \mathcal{N}_{S_N} \tag{1.45}$$

is an operation or channel, if

$$S_1^\dagger S_1 + S_2^\dagger S_2 + \ldots + S_N^\dagger S_N \leq I. \tag{1.46}$$

By the proposition below, we see that the converse is also true.

Theorem 1.5. *A linear mapping $\mathcal{E} : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})$ is a channel if and only if there exists a sequence of operators $\{A_k\}$ such that (c.f. [50], Prop. 4.21)*

$$\mathcal{E}(T) = \sum_k A_k T A_k^\dagger, \qquad \sum_k A_k^\dagger A_k = I. \tag{1.47}$$

If $\dim \mathcal{H} = d < \infty$, then it is possible to choose d^2 or fewer operators A_k.

The form in Eq. (1.47) is called *operator-sum form*, or *Kraus-form*, of a channel $\mathcal{E}$. The operators A_k are called Kraus operators.

1.4.7 Measurement Models and Instruments

In the process of measuring an observable A, we firstly prepare a quantum system on a state $\rho \in \mathcal{S}(\mathcal{H})$ and send it to the measurement apparatus. An outcome is registered in the latter, and the system may or may not be destroyed in the process. If the quantum state is not always destroyed, as in Fig. 1.4.7, the definition of post-measurement state to be used in subsequent measurements is of significant role. The system interacts with a *probe* initially prepared in a

state $\xi \in \mathcal{K}$, by its own a physical system associated to some Hilbert space dimension. After this interaction the state of the measured system changes, in general, and the probe becomes correlated with the system; finally we make the measurement on the probe. Therefore, in a measurement model we should specify

- $\mathcal{K}$, the Hilbert space associated to the probe system;
- ξ, the initial state of the probe;
- $\mathcal{V}$, a channel from $\bar{\mathcal{S}}(\mathcal{H} \otimes \mathcal{K})$ to $\bar{\mathcal{S}}(\mathcal{H} \otimes \mathcal{K})$;
- the *pointer observable* F, associated to the probe effect space, and assumed to have the same outcome space $(\Omega, \mathcal{F})$ than A.

The quadruple $\mathcal{M} = (\mathcal{K}, \xi, \mathcal{V}, F)$ is a *measurement model* of A if, for all quantum states ρ,

$$\mathrm{tr}[\rho A(X)] = \mathrm{tr}[\mathcal{V}(\rho \otimes \xi)(I \otimes F(X))]. \tag{1.48}$$

This means that the probabilities obtained by measuring the probe are the same as those associated to the effects for the observable A. After the first interaction, and assuming that the system is not destroyed, we will measure another observable B. As we are dealing only with finite number of outcomes, we will change notation slightly. The effects associated to the outcome x and an observable A are going to be denoted by A_x, and for the second observable B with outcome y the effects are B_y. Denoting by $p_\rho(x, y)$ the probability of measurement of A and B to return an outcomes x and y, the joint probability obtained in this scenario is

$$p_\rho(x, y) = \mathrm{tr}[\rho A_x B_y] = \mathrm{tr}[\mathcal{V}(\rho \otimes \xi)(B_y \otimes F_x)]. \tag{1.49}$$

comparing with Eq. (1.37) this is the same as

$$\mathrm{tr}[\mathcal{V}(\rho \otimes \xi)(B_y \otimes F_x)] = \mathrm{tr}[B_y \,\mathrm{tr}_{\mathcal{K}}(\mathcal{V}(\rho \otimes \xi)(I \otimes F_x))]. \tag{1.50}$$

Therefore, it is useful to define the operator

$$\mathcal{I}_x^{\mathcal{M}}(\rho) = \mathrm{tr}_{\mathcal{K}}(\mathcal{V}(\rho \otimes \xi)(I \otimes F_x)) \tag{1.51}$$

such that

$$p_\rho(x, y) = \mathrm{tr}[\mathcal{I}_x^{\mathcal{M}}(\rho) B_y]. \tag{1.52}$$

This mapping from $\rho \to \mathcal{I}_x^{\mathcal{M}}(\rho)$ has the following properties

(i) for each X, $\mathcal{I}_x^{\mathcal{M}}$ is an operation;

(ii) $\mathrm{tr}[I_\Omega^{\mathcal{M}}(\rho)] = 1$ and $\mathrm{tr}[I_\emptyset^{\mathcal{M}}(\rho)] = 0, \quad \forall \ \ \rho$;

(iii) for a sequence $\{x_j\}$ of pairwise disjoint sets

$$\mathrm{tr}[\mathcal{I}_{\cup x_j}^{\mathcal{M}}(\rho)] = \sum_j \mathrm{tr}[\mathcal{I}_{x_j}^{\mathcal{M}}(\rho)]; \tag{1.53}$$

and, by definition

$$p_\rho(x) = \mathrm{tr}[\mathcal{I}_x^{\mathcal{M}}(\rho)]. \tag{1.54}$$

A mapping $\mathcal{I}$, which satisfies the conditions $(i) - (iii)$ above is called an *instrument*; they are used here to define the post-measurement state. Our results in Sec. 3.2 are proven inside this formalism.

1.4.8 Quantum Contextuality

Nonlocality: CHSH scenario

Consider two parties, Alice and Bob, spatially separated, each of which holding a box with a physical system inside, c.f. Fig. 1.4. Each box has a set of input buttons and produces an

output after each button is pressed. The inputs for Alice and Bob are $\{A_1, A_2\}$ and $\{B_1, B_2\}$, respectively. For every input, the output is always -1 or $+1$. These buttons represents measurement choices — "questions" — we can perform in a physical system, plus a classical treatment of the data acquired, producing the outputs, or measurement outcomes.

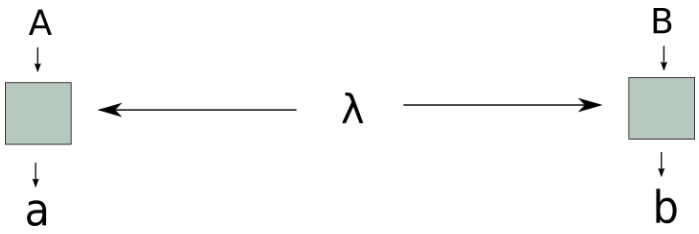

Figure 1.4: Alice and Bob receive a classical system λ, input A and B and output a and b, respectively.

We will start by analysing the scenario in which they share a random variable, carried by a classical system, after pressing the input button. The shared random variable will be denoted by λ. This variable together with the input will influence the output of each box in such a way that the output is a probability distribution depending on the input. Thus, for Alice, the output a has a probability $p(a \mid A, \lambda)$, and analogously for Bob. If λ has a distribution $p(\lambda)$, the joint distribution of the output which can be obtained by this model is

$$p(a, b \mid A, B) = \int_\lambda p(\lambda) p(a \mid A, \lambda) p(b \mid B, \lambda) d\lambda. \tag{1.55}$$

We can restrict the probabilistic character of these distributions to the randomness of the variable λ, following the work of A. Fine [54]. To do that, first rewrite

$$\begin{aligned} p(a \mid A, \lambda) &= \int_\mu p(\mu) \delta^a_{\alpha(\mu, A, \lambda)} d\mu, \\ p(b \mid B, \lambda) &= \int_\nu p(\nu) \delta^b_{\beta(\nu, B, \lambda)} d\nu, \end{aligned} \tag{1.56}$$

in which α and β are the output functions, depending also on the intrinsic random variables μ and ν. In this way, Eq. (1.55) is written as

$$p(a, b \mid A, B) = \int_\lambda \int_\mu \int_\nu [p(\lambda) p(\mu) p(\nu)] \delta^a_{\alpha(\mu, A, \lambda)} \delta^b_{\beta(\nu, B, \lambda)} d\mu d\nu d\lambda. \tag{1.57}$$

Now, we can define some $\lambda'(\lambda, \mu, \nu)$ such that $p(\lambda') = p(\lambda)p(\mu)p(\nu)$ and

$$p(a, b \mid A, B) = \int_{\lambda'} p(\lambda')\delta^a_{\alpha(A,\lambda')}\delta^b_{\beta(B,\lambda')}d\lambda'. \tag{1.58}$$

Therefore, we can make mixtures of deterministic strategies for achieving every classical probability distributions over the joint outputs.

The *nonclassicallity witness* appears when confronting specific expectation values obtained from the experiments, which can be achieved by QT, but not by classical theory, with the modeling represented in Eq. 1.55. This witness is the average of the CHSH expression

$$S := a_1b_1 + a_1b_2 + a_2b_1 - a_2b_2. \tag{1.59}$$

Now, if the outputs are written in a table as

$$\begin{bmatrix} a_1(\lambda') & b_2(\lambda') \\ b_1(\lambda') & a_2(\lambda') \end{bmatrix} := \begin{bmatrix} \alpha(A_1, \lambda') & \beta(B_2, \lambda') \\ \beta(B_1, \lambda') & \alpha(A_2, \lambda') \end{bmatrix} \tag{1.60}$$

and, for each λ', the outputs of A_i and B_j, respectively, are denoted by $a_i(\lambda')$ and $b_j(\lambda')$. There are 2^4 tables of this kind, so we can compute S for each of them, to see the maximum value it attains. On the other hand there is a shortcut to see the maximum value for this and other similar expressions. Instead, defining $c_1(\lambda') = a_1(\lambda')b_1(\lambda')$, $r_1(\lambda') = a_1(\lambda')b_2(\lambda')$, $r_2(\lambda') = a_2(\lambda')b_1(\lambda')$ and $c_2(\lambda') = a_2(\lambda')b_2(\lambda')$, the expression S now becomes

$$S = c_1 + r_1 + r_2 - c_2. \tag{1.61}$$

On the other hand, the product of the elements can be made by the rows or the columns, equivalently. Therefore

$$r_1(\lambda')r_2(\lambda') = c_1(\lambda')c_2(\lambda'). \tag{1.62}$$

In this way, S could be rewritten as

$$S = c_1 + r_1 + r_2 - \frac{r_1 r_2}{c_1}. \tag{1.63}$$

In this form, instead of checking the values of the observables, it is possible to analyze the value of the expression by the individual terms of the sum. There are eight values for the triple $(c_1(\lambda'), r_1(\lambda'), r_2(\lambda'))$, and $S(\lambda')$ is maximal when $(c_1(\lambda'), r_1(\lambda'), r_2(\lambda'))$ is equal to one of the triples $(+1,+1,+1), (-1,+1,+1), (+1,-1,+1), (+1,+1,-1)$, resulting in $S = 2$. Consequently,

$$\langle S \rangle \leq 2. \tag{1.64}$$

On the other hand, in the realm of QT, the following set of selfadjoint operators

$$\begin{bmatrix} A_1 & B_2 \\ B_1 & A_2 \end{bmatrix} = \begin{bmatrix} \sigma_x \otimes I & I \otimes \sigma_z \\ I \otimes \sigma_x & \sigma_z \otimes I \end{bmatrix} \tag{1.65}$$

violates the classical bound. Just note that the operator S^Q has the form

$$S^Q = A_1 B_1 + A_1 B_2 + A_2 B_1 - A_2 B_2 = \sigma_x \otimes \sigma_x + \sigma_x \otimes \sigma_z + \sigma_z \otimes \sigma_x - \sigma_z \otimes \sigma_z, \tag{1.66}$$

and that it has eigenvalues $\pm 2\sqrt{2}$, therefore the corresponding quantum state for the higher value violates (1.64). Indeed, this is shown to be the maximum value achieved in QT [35]. We have the following bounds

$$\langle S \rangle \overset{C}{\leq} 2 \overset{Q}{\leq} 2\sqrt{2}, \tag{1.67}$$

in which the superscript $C(Q)$ means the classical (quantum) bound.

Therefore, we see that the impossibility to assign dispersion free probability distributions for the quantum set has experimental implications. Inequalities like expression (1.64) are commonly called *Bell inequalities*, and they are related to inequalities obtained when trying to

model experiments in terms of local variables, i.e. modeling in the context which leads to Eq. (1.55). Recent experiments confirmed the classical bound is violated [55–57]. Thus, we say QT is *nonlocal*, in the sense that no local modeling of the form (1.55) can explain experiments in which the parties are spatially separated.

Contextuality: the Peres-Mermin Scenario

In the CHSH scenario, it was possible to distribute the observables to different parties. But this is not needed; there is a more general concept to deal with the possibility of non existence of dispersion free probability distributions, and this is related to the phenomenon of *contextuality*. A *context* is any set of joint measurable observables, and the hypothesis of *noncontextuality* is that *a measurement outcome of any observable does not depend on which other observables are jointly measured with it*. Spatially separated observables are jointly measurable by construction, thus they form a context. But there are other scenarios, and one of the most intriguing collection of contexts is the Peres-Mermin scenario [33, 34]. It consists of 9 dichotomic observables, with $+1$ or -1 outputs arranged as in the matrix below:

$$\begin{bmatrix} A & B & C \\ a & b & c \\ \alpha & \beta & \gamma \end{bmatrix}. \tag{1.68}$$

The Peres-Mermin (PM) scenario is made such that the observables on the same row or column are jointly measurable and, therefore, belong to the same context. We have then six contexts, $\{A, B, C\}, \{a, b, c\}, \{\alpha, \beta, \gamma\}, \{A, a, \alpha\}, \{B, b, \beta\}$ and $\{C, c, \gamma\}$.

Noncontextual Model A classical noncontextual model explaining the statistics of nine dichotomic measurements assigns a mixture of matrices with predefined values, identified by

the random variable λ,

$$\begin{bmatrix} A(\lambda) & B(\lambda) & C(\lambda) \\ a(\lambda) & b(\lambda) & c(\lambda) \\ \alpha(\lambda) & \beta(\lambda) & \gamma(\lambda) \end{bmatrix} := \begin{bmatrix} A & B & C \\ a & b & c \\ \alpha & \beta & \gamma \end{bmatrix}_{\lambda}. \tag{1.69}$$

The outputs are $+1$ or -1, thus the total number of distinct deterministic tables is $2^9 = 512$. For the noncontextual model, there is a noncontextuality (NC) inequality associated. It was proposed by Cabello [29] and it can be written as

$$\langle ABC\rangle + \langle abc\rangle + \langle \alpha\beta\gamma\rangle + \langle Aa\alpha\rangle + \langle Bb\beta\rangle - \langle Cc\gamma\rangle \leq 4. \tag{1.70}$$

This inequality can be easily checked by noticing that each one of the 512 tables does not exceed the value of 4 for (1.70) nor the mixture of them, consequently. Another way to see the impossibility of classical violation is to note that, in analogy to the CHSH scenario, it is possible to write product of the nine observables following its columns or its rows. Consequently the product of the outputs taken from the last column context is related to the product of the five other contexts. Defining R_i as the product of outcomes for the observables on the row i, and analogously C_i for the columns, C_3 might be written as

$$C_3 = Cc\gamma = \frac{(ABC)(abc)(\alpha\beta\gamma)}{(Aa\alpha)(Bb\beta)} = \frac{R_1R_2R_3}{C_1C_2}. \tag{1.71}$$

Therefore, the expression (1.70) is maximal when all contexts are $+1$, or when only one among the observables C, c or γ is -1, resulting in the above bound for the classical reasoning.

The Quantum Scenario In the original references [33,34], the authors proposed the following set of observables to fill the matrix:

$$\begin{bmatrix} A & B & C \\ a & b & c \\ \alpha & \beta & \gamma \end{bmatrix} = \begin{bmatrix} \sigma_z \otimes \mathbb{1} & \mathbb{1} \otimes \sigma_z & \sigma_z \otimes \sigma_z \\ \mathbb{1} \otimes \sigma_x & \sigma_x \otimes \mathbb{1} & \sigma_x \otimes \sigma_x \\ \sigma_z \otimes \sigma_x & \sigma_x \otimes \sigma_z & \sigma_y \otimes \sigma_y \end{bmatrix}. \tag{1.72}$$

Every operator commutes with the others in the same row or column. They have the interesting property that

$$ABC = abc = \alpha\beta\gamma = Aa\alpha = Bb\beta = -Cc\gamma = \mathbb{1}, \tag{1.73}$$

and, therefore,

$$\langle ABC\rangle + \langle abc\rangle + \langle \alpha\beta\gamma\rangle + \langle Aa\alpha\rangle + \langle Bb\beta\rangle - \langle Cc\gamma\rangle = 6. \tag{1.74}$$

This is an instance of state independent violation of the NC inequality (1.70). Therefore, its violation comes *only* from the relations among the observables, a feature that cannot happen in a Bell inequality, as it is always possible to find separable states, with local statistics, while no global measurements can be applied.

Taking a different point of view, as a logical conclusion, Eq. (1.73) could never be satisfied by the outcomes of a noncontextual theory, as we can see by the analysis of the condition (1.71). In QT, the outcomes of all observables within a context can be obtained in a joint measurement. For the three dichotomic observables in each context of the Peres-Mermin square, the joint measurement on two qubits has four distinct outcomes, taken from the set of the 8 possible combinations of outcomes $\{(+1,+1,+1),(+1,+1,-1),\ldots,(-1,-1,-1)\}$. Alternatively, the outcomes can be obtained by measuring the observables in a context in a sequential way. This approach has been preferred in recent experiments on quantum contextuality [58–64]. When measuring an observable X from the Peres-Mermin square, the quantum state ρ changes according to the usual state update rule

$$\rho \mapsto \frac{\Pi_{x|X}\rho\Pi_{x|X}}{\mathrm{tr}(\rho\Pi_{x|X})}, \tag{1.75}$$

with $\Pi_{x|X} = \frac{1}{2}(\mathbb{1} + xX)$ depending on the measurement outcome $x = \pm 1$ of X. In a sense, sequential measurements with this Lüders transformation [65] are a special way to implement a joint measurement. Since the quantum state changes according to the choice of the observable and the measurement outcome, one could argue that the quantum state serves as a memory

and the contextual behavior is achieved due to the very presence of this memory.

1.5 Chapter Conclusions

We detailed here how probabilities for events can be modeled in classical and quantum theory, through the notion of effects associated to measurement outcomes. Differently from classical theories, quantum theory does not allow general dispersion free probability distributions when restricted to extremal effects and states. This is interpreted as the manifestations of the intrinsic probabilistic character of QT, which have experimental implications. Using the local and noncontextual classical models to explain the scenarios, we were able to derive some inequalities these models respect and, in contrast, QT violates. The premises of those models, although being extremely reasonable, can be modified to embrace more general classical models, and this comes with a cost, which is the subject of the next chapter.

Chapter 2

Simulations and Costs

In this chapter we discuss some concepts and results about classically simulating a specific set of events and its probabilities. We analyze the costs by searching for lower bounds on the classical resources needed to reproduce the desired correlations. We are mainly concerned with the comparison between classical and quantum resources, but this kind of approach is useful for more general scenarios, once the important resources are defined. The classical resource is quantified in bits, in terms of memory or communication, while the quantum resource by the number of necessary qubits. Those are the simplest information storage systems of each theory.

2.1 Simulating Nonlocality with Communication

One way to quantify nonlocality is by asking how much communication is necessary between the parties such that they are able to classically simulate a nonlocal behaviour [4].

The most basic scenario of nonlocality is between two parties, Alice and Bob, each of which receives an input and has to return an output. They can share a random variable λ — discrete or continuous, with a distribution $p(\lambda)$ — prior to the input, and they can communicate freely before receiving the inputs, and only freely after the input. The input of Alice will be denoted by A, the output by a, Bob's input by B, output by b and the total message

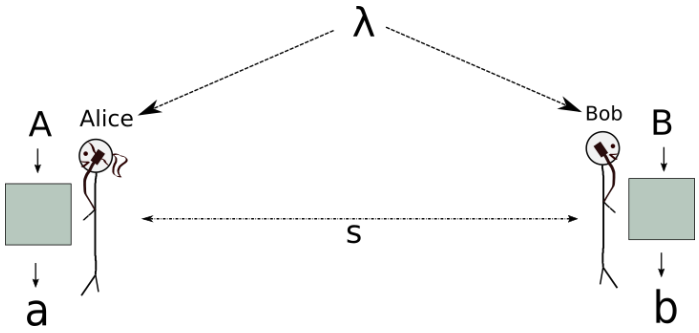

Figure 2.1: If Alice and Bob can communicate before the outputs, how much communication is needed to simulate CHSH correlations?

communicated before the outputs by s. The main goal is to reproduce a given probability distribution $p(a, b \mid A, B)$. The process is illustrated by the figure 2.1.

Using this scheme, the probabilities achieved by the model are

$$p(a, b \mid A, B) = \sum_{\lambda, s} p(\lambda) p(s \mid A, B, \lambda) p(a \mid A, s, \lambda) p(b \mid B, s, \lambda), \tag{2.1}$$

with the sum over λ replaced by an integral when this variable belongs to a continuous set. The communication is always treated as discrete in this thesis.

The set of probabilities which can be reproduced by this model increases with the amount of bits communicated. The minimum amount of communication per run is one bit, while if every part communicates its input to the other, it is trivial to reproduce any probability distribution. We can see this by noting that, if both know A and B, they can use joint deterministic strategies sampled from the original probability distribution $p(a, b \mid A, B)$. For example, they can decompose it into a mixture of joint deterministic distributions, c.f. Eq. (1.2), such that they only use terms which respects the joint probability; a convenient $p(\lambda)$ allows them to do the task.

2.1.1 Communication Cost for Simulating Singlet Correlations

The seminal Bell paper [24] on a similar bipartite scenario tells us that if there is no communication between the parties, or no interaction in any form after distributing the joint quantum

state, the singlet state, implies that the quantum correlations obtained behave differently than what is expected in classical theory. Clauser *et al.* [32] translated this conclusion into an experiment realized in its most strong form in recent years [55–57][1]. The experiments tells us that the CHSH inequality (1.64) is indeed violated. Thus, to achieve what quantum theory gives using only classical resources, how much communication is needed? The question might also be formulated in the other way around, by how much can classical communication be reduced if one uses quantum systems?

We must allow an influence from Alice's side to Bob's side and, as already said, the minimal amount of communication they can exchange is one bit. In the paper of Toner and Bacon [12], the authors proved that one bit of communication is enough to reproduce *all bipartite correlations* arising from the singlet state, and thus it also covers the CHSH scenario. The strategy they used will be explained below.

First, Alice and Bob share a pair of two independent random variables $\hat{\lambda}_1$ and $\hat{\lambda}_2$, taken from a uniform distribution over the unit sphere. Alice then selects her measurement setting, represented by another direction $\hat{A}$ on the unit sphere. Alice outputs a according to

$$a = -\operatorname{sgn}(\hat{A} \cdot \vec{\lambda}_1), \tag{2.2}$$

in which $\operatorname{sgn}(x) = 1$, for $x \geq 0$ and $\operatorname{sgn}(x) = -1$, for $x < 0$. She communicates a bit s to Bob,

$$s = \operatorname{sgn}(\hat{A} \cdot \hat{\lambda}_1) \operatorname{sgn}(\hat{A} \cdot \hat{\lambda}_2), \tag{2.3}$$

and Bob analogously selects another measurement vector $\hat{B}$, and outputs b,

$$b = \operatorname{sgn}\left[\hat{B} \cdot (\hat{\lambda}_1 + s\hat{\lambda}_2)\right]. \tag{2.4}$$

[1] It is strong in the sense that the loopholes of *freedom-of-choice*, *fair-sampling* and *no-signaling* [66] were simultaneously closed in each experiment. The first loophole is engendered when we cannot guarantee that the measurement choices are free or random. The second loophole is related to the fact that some runs of an experiment do not produce experimental data. Thus, improving the detection rate closes this loophole. The third loophole appears when the parties are close enough so that communication can be exchanged between them after the input is chosen. All these three loopholes allow violations of Bell inequalities which can be explained classically.

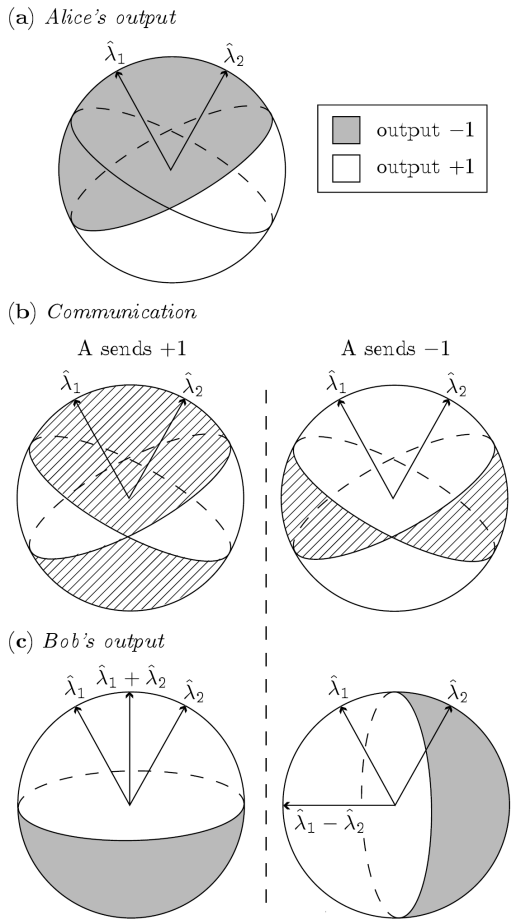

Figure 2.2: Diagram for the 1 bit strategy. Figure taken from reference [12]. In (a), if Alice chooses a measurement direction in the gray region, she outputs -1, and if she chooses a direction in the other hemisphere she outputs $+1$; (b) and (c) are read in analogous way.

The protocol is schematized in the Fig. 2.2, and the calculations are done in the referred paper. Averaging over the random variables gives us

$$\langle ab \rangle = -\hat{A} \cdot \hat{B}. \tag{2.5}$$

Therefore, the authors show that with one bit of classical communication it is possible to reproduce quantum correlations for the singlet. One should note that we are not dealing with a countable number of measurement settings; indeed, any direction on the unit sphere might in principle be realized by a qubit measurement. If you consider only two settings per party, then it is possible to reduce the communication in average, as shown in the next section.

2.1.2 Lower Bounds on Communication Cost

It is usually very difficult to find optimal strategies to simulate specific correlations. Frequently one starts by constructing which outputs are obtained for each amount of communication available. So, for example, in the CHSH scenario with one bit of communication we can construct 64 deterministic output functions. The set of inputs for Alice is $\{A_0, A_1\}$ and for Bob it is $\{B_0, B_1\}$, the output can change depending on the communication received, represented by the bit s. Suppose Alice makes the measurement, she has two inputs to map to two outputs, thus four distinct output functions $a(A_i, \lambda_j)$, $i = 0, 1$ and $j = 1, \ldots, 4$. Then she communicates s to Bob. As his outputs also depends on s, he has 16 output functions $\beta(B_{i'}, s, \lambda_{j'})$, with $i' = 0, 1$ and $j' = 1, \ldots, 16$. Note here that each λ corresponds to a strategy. As the strategies for Alice and Bob are independent, i.e. they can choose any output function a and b, we arrive in 4×16 values for the joint strategy.

As we increase the number of inputs, outcomes and the amount of communication, the number of strategies grows exponentially, making it difficult to calculate the communication cost. In another crucial work, Pironio [13] showed that the violation of Bell inequalities using classical resources imposes lower bounds on the communication cost, and that this lower bound increases with the violation of those inequalities. Below we adapt to more parties the scenario analyzed by Pironio.

Consider N parties performing one among k measurements available with m outputs each. The notation is going to change slightly to accommodate more parties. The scenario is completely characterized by the probabilities $p(x_1, \ldots, x_N \mid X_1 \ldots X_N)$, in which $x_i \in \{0, \ldots, m-1\}$ is the output and $X_j = \{0, \ldots, k-1\}$ is the input. We can arrange

this probabilities in a vector $\vec{p}$, ordering by the input first:

$$\vec{p} = \begin{bmatrix} p(0\ldots0 \mid 0\ldots0) \\ p(0\ldots1 \mid 0\ldots0) \\ \vdots \\ p(m-1\ldots m-1 \mid 0\ldots0) \\ p(0\ldots0 \mid 0\ldots1) \\ \vdots \\ \vdots \\ p(m-1\ldots m-1 \mid k-1\ldots k-1) \end{bmatrix}. \tag{2.6}$$

The normalization constraint applies to every measurement setting,

$$\sum_{x_1,\ldots,x_N} p(x_1,\ldots,x_N \mid X_1\ldots X_N) = 1, \qquad \text{for } X_i = 0,\ldots,k-1, \tag{2.7}$$

and for $i = 1,\ldots,N$. Consequently, the sum of its entries is the number of distinct measurement settings, i.e.,

$$\sum_j p_j = k^N. \tag{2.8}$$

In the paper, Bell type inequalities are written as a linear inequality on the vector of probabilities, i.e.

$$B(\vec{p}) = \vec{b}\cdot\vec{p} \leq B_0, \tag{2.9}$$

in which B_0 is the maximal value it takes for local strategies, i.e. strategies which use no communication.

In classical reasoning, each party uses local strategies to reproduce the probabilities for the outputs, based on their choice for the input and the communication received by them. The

deterministic strategies are $\vec{d^{\lambda}}$, in which each of its entries are

$$d^{\lambda}(x_1,\ldots,x_N \mid X_1,\ldots,X_N) = \delta^{x_1}_{f_1(X_1,\ldots,X_N,\lambda)}\ldots\delta^{x_N}_{f_N(X_1,\ldots,X_N,\lambda)}. \tag{2.10}$$

The functions f_i are the deterministic outputs for every joint input. Since there are a finite number of deterministic functions for this kind of strategy, we will label them with a countable variable λ. Now, every probabilistic strategy can be decomposed in terms of deterministic ones:

$$\vec{p} = \sum_{\lambda} q_{\lambda}\vec{d^{\lambda}}. \tag{2.11}$$

Every deterministic strategy has a communication cost $\mathcal{C}(\vec{d^{\lambda}})$, and we will group them according to this cost. Thus, if a strategy has a communication cost c_i, we will label them with λ_i, i.e., $\mathcal{C}(\vec{d^{\lambda_i}}) = c_i$. In other words, the label "i" tag the communication cost, while λ now can order the strategies with the same cost. In this way, the probabilistic strategies become

$$\vec{p} = \sum_{i}\sum_{\lambda_i} q_{\lambda_i}\vec{d^{\lambda_i}}. \tag{2.12}$$

The average communication cost is, then,

$$\bar{\mathcal{C}} = \sum_{i}\sum_{\lambda_i} q_{\lambda_i}c_i \equiv \sum_{i} q_i c_i, \tag{2.13}$$

in which $q_i := \sum_{\lambda_i} q_{\lambda_i}$ is the probability of selecting a strategy with cost c_i.

It is interesting to see that, grouping the strategies this way, there is a maximum violation of a linear inequality for every communication cost bounded by the maximum over the strategies that have this cost, i.e.

$$B_i = \max_{\lambda_i}\{\vec{b}\cdot\vec{d^{\lambda_i}}\}. \tag{2.14}$$

We can now interpret B_0 as

$$B_0 = \max_{\lambda_0}\{\vec{b}\cdot d^{\vec{\lambda}_0}\}, \tag{2.15}$$

and we read it as the maximum value achieved for strategies with no cost. Now we are in position to state a central result, in the proposition below, taken from [13] with a minor modification.

Proposition 2.1. *The communication cost associated with a particular value for $B(\vec{p})$ by the distribution $\vec{p}$ satisfies the following bound*

$$\mathcal{C}(\vec{p}) \geq \frac{B(\vec{p}) - B_0}{B_{j^*} - B_0} c_{j^*}, \tag{2.16}$$

in which the index j^* is taken from the most efficient strategy which violates the inequality (2.9), i.e., j^* is such that

$$j^* = \arg\max_{j\neq 0} \frac{B_j - B_0}{c_j}. \tag{2.17}$$

Proof. By definition,

$$B(\vec{p}) = \vec{b}\cdot\vec{p} = \sum_{i,\lambda_i} q_{\lambda_i}\vec{b}\cdot d^{\vec{\lambda}_i} \leq \sum_{i,\lambda_i} q_{\lambda_i} B_i = \sum_i q_i B_i. \tag{2.18}$$

For the average communication cost we have

$$\begin{aligned}
\overline{\mathcal{C}(\vec{p})} &= \sum_i q_i c_i \\
&= \sum_i q_i c_i - \left(\frac{B(\vec{p}) - B_0}{B_{j^*} - B_0} c_{j^*}\right) + \frac{B(\vec{p}) - B_0}{B_{j^*} - B_0} c_{j^*} \\
&\geq \sum_i q_i c_i - \left(\frac{\sum_i q_i B_i - (\sum_i q_i) B_0}{B_{j^*} - B_0} c_{j^*}\right) + \frac{B(\vec{p}) - B_0}{B_{j^*} - B_0} c_{j^*} \\
&= \frac{B(\vec{p}) - B_0}{B_{j^*} - B_0} c_{j^*} + \sum_i q_i \left(c_i - \frac{B_i - B_0}{B_{j^*} - B_0} c_{j^*}\right) \\
&\geq \frac{B(\vec{p}) - B_0}{B_{j^*} - B_0} c_{j^*},
\end{aligned} \tag{2.19}$$

in which we used Eq. (2.17) and (2.18) to see that $c_i \geq c_{j^*}(B_i - B_0)/(B_{j^*} - B_0)$. □

Therefore, there is a lower bound which increases with the amount of violation of the Bell inequality.

The author also shows that for certain situations this bound is tight, which includes the CHSH scenario. In the latter, with one bit of communication it is possible to reach the maximum algebraic value for CHSH expression, which is 4, and therefore $c_{j^*} = 1$. Consequently, the communication cost to simulate the singlet state is $\mathcal{C}(\vec{p}) = \frac{2\sqrt{2}-2}{4-2}.1 = \sqrt{2} - 1 \approx 0.4$. The authors also give an explicit strategy which accomplishes this. It consists of selecting properly among these specific local strategies, which reach the value of 2 for $B(\vec{p})$,

	d^{0_0}	d^{1_0}	d^{2_0}	d^{3_0}	d^{4_0}	d^{5_0}	d^{6_0}	d^{7_0}
$d(00 \mid 00)$	1	1	0	0	1	0	0	0
$d(10 \mid 00)$	0	0	0	0	0	1	0	0
$d(01 \mid 00)$	0	0	1	0	0	0	0	0
$d(11 \mid 00)$	0	0	0	1	0	0	1	1
$d(00 \mid 10)$	1	0	0	0	0	0	0	0
$d(10 \mid 10)$	0	1	0	0	1	1	0	0
$d(01 \mid 10)$	0	0	1	1	0	0	1	0
$d(11 \mid 10)$	0	0	0	0	0	0	0	1
$d(00 \mid 01)$	1	1	1	0	0	0	0	0
$d(10 \mid 01)$	0	0	0	1	0	0	0	0
$d(01 \mid 01)$	0	0	0	0	1	0	0	0
$d(11 \mid 01)$	0	0	0	0	0	1	1	1
$d(00 \mid 11)$	1	0	1	1	0	0	0	0
$d(10 \mid 11)$	0	1	0	0	0	0	0	0
$d(01 \mid 11)$	0	0	0	0	0	0	1	0
$d(11 \mid 11)$	0	0	0	0	1	1	0	1

and these 1-bit strategies, which reach the value of 4 for $B(\vec{p})$,

	d^{0_1}	d^{1_1}	d^{2_1}	d^{3_1}	d^{4_1}	d^{5_1}	d^{6_1}	d^{7_1}
$d(00 \mid 00)$	1	0	1	0	1	0	1	0
$d(10 \mid 00)$	0	0	0	0	0	0	0	0
$d(01 \mid 00)$	0	0	0	0	0	0	0	0
$d(11 \mid 00)$	0	1	0	1	0	1	0	1
$d(00 \mid 10)$	0	0	0	0	0	0	0	0
$d(10 \mid 10)$	0	0	1	1	1	0	1	0
$d(01 \mid 10)$	1	1	0	0	0	1	0	1
$d(11 \mid 10)$	0	0	0	0	0	0	0	0
$d(00 \mid 01)$	1	0	1	0	1	1	0	0
$d(10 \mid 01)$	0	0	0	0	0	0	0	0
$d(01 \mid 01)$	0	0	0	0	0	0	0	0
$d(11 \mid 01)$	0	1	0	1	0	0	1	1
$d(00 \mid 11)$	1	1	0	0	1	1	0	0
$d(10 \mid 11)$	0	0	0	0	0	0	0	0
$d(01 \mid 11)$	0	0	0	0	0	0	0	0
$d(11 \mid 11)$	0	0	1	1	0	0	1	1

.

It is an interesting conclusion to relate the communication cost to the violation of Bell type inequalities. On the other hand, the analysis above still needs to compute the most efficient strategy in terms of communication and violation — i.e., to find the j^* index by computing the violations for every value of c_i possible. A great advance in this terms comes from Montina and Wolf [67], in a paper which they compute lower bounds for the communication cost for the CHSH scenario based solely on the probability distributions.

Extending the discussion, Ref. [68] shows that, using two bits of communication, maximally entangled states are reproduced, in a bipartite scenario and an arbitrary number of measurements with two outcomes. Ref. [14] shows that three bits are sufficient to simulate equatorial measurements on the GHZ scenario.

2.2 Simulating Contextuality with Memory

Apart from communication problems, there are other confrontations between classical strategies and quantum strategies. In the past decade, memory as a classical resource was associated to simulating sequences of quantum measurements. It is the case that, in a sequential process, a physical system attains different states during the process, and therefore we can associate a *memory cost* to the scenario. Classically, the information content of d distinguishable states produced with probability p_i, $i = 1, \ldots, d$, is usually quantified in terms of the *Shannon entropy* $H(\vec{p}) = -\sum_i p_i \log p_i$, which is measured in bits, if the log is taken in base 2. The maximum value it attains is $\log d$. Quantically, *Holevo's bound* [36] expresses that the maximum storage capacity of a quantum system associated to a Hilbert space of dimension d is also $\log d$. Nielsen and Chuang provide a simpler argument to prove it in Ref. [69], chapter 12. Superficially, this bound may lead to the idea that we cannot take advantage of the peculiarities of QT while using quantum states to store information or to communicate; an idea which turns out to be wrong. It is interesting to point out that using an orthogonal basis in a d-level state from QT one can simulate classical transformations on d classical states, as long as one can operate in the quantum states analogously, for example mapping from basis vectors to basis vectors. On the other hand, there are probability distributions achievable by quantum operations on a d level quantum state which are not achievable by classical operations on d distinguishable classical states. This raises the possibility of a *memory advantage* of QT over CT.

The first instance of memory advantage of quantum theory was discovered by E. Galvão and L. Hardy [11], detailed in Subsec. 2.2.2, analyzing a quantum unitary dynamics. The application to contextuality was developed by Kleinmann *et al.* [15], explained in Subsec. 2.2.1. In the latter, first the authors studied the application of the concept of memory to scenarios in which *only compatible measurements* are allowed. Then they studied sequences which are related to contextuality, but also allowed incompatible observables in the same sequence. In their model, they considered only deterministic automata and did not find a memory advantage from QT over CT when considering only compatible measurements in the same sequence. There are good reasons to extend the model to a probabilistic one, as explained in 2.2.3. Furthermore,

besides these two models, there were other proposals to define the memory cost [16, 70]. In all of them, the memory is related to the number of distinct classical states a physical system can assume during sequential measurements.

2.2.1 Memory Cost of Quantum Contextuality: Lower Bound on the Peres-Mermin Scenario

As a first approach, the authors of [15] modeled a sequence of quantum measurements as a classical Mealy machine [71]. These machines, or *automata*, are used to model sequential input-output processes, for which the general idea is to simulate a string of outputs $x_1 x_2 \dots x_n$ produced sequentially when entering the inputs $X_1, X_2, \dots, X_n$, also sequentially. The process represents a classical system, in some state s, receiving an interaction with the environment, i.e. the input, then this interaction transforms the state s to another state s' and some part of the environment, represented by the output. The automaton can be defined with a quintuple $\mathcal{M} = (Q, \Sigma, \Theta, F, G)$. In this notation, Q is the set of accessible states of the machine, Σ the set of inputs, Θ the set of outputs, $F : Q \times \Sigma \to Q$ the transition function taking $s \mapsto s'$, conditioned on the environment input, and $G : Q \times \Sigma \to \Theta$ the output function. This can be drawn in a diagram such as in Fig. 2.3.

This thesis is related to the computation of the memory cost of contextuality; specifically for the Peres-Mermin scenario, which has as inputs projective measurements. In a sequential scenario for QT with projective measurements, we see that the measurement outputs of a contextuality scenario respects *repeatability*, i.e. that in a sequence of the same measurement $XX\dots$, the same output is returned, i.e. $xx\dots$. Furthermore, the *compatibility* condition on the measurements is also respected. Following Ref. [72], this condition is defined by the demand that, in a sequence of compatible measurements, the outcomes of each observable cannot be changed after measuring other observable and that the correlations obtained while measuring two compatible observables do not depend on their position on the sequence of measurements. In order to specify it better, we define $\mathcal{S}_{AB} = \{A_1, B_1, A_1A_1, A_1B_2, B_1B_2, B_1A_2, \dots\}$ as the infinite set representing the sequences of measurements which we can perform only with two

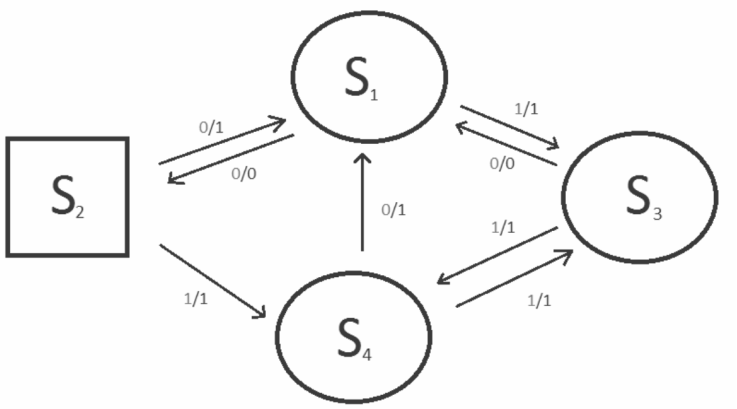

Figure 2.3: A four-state machine, with inputs and outputs belonging to the set {0,1}. Every arrow corresponds to the transition from one state to another, given the i/o (input/output) respective to the process written next to it. The state depicted in the square represents the initial state.

observables A and B. The symbol x_l^S denotes the outcome of the observable X measured on the position l in the sequence $S \in \mathcal{S}_{AB}$. If the observables A and B are compatible, the following conditions are satisfied:

1. Their outcomes a and b cannot be changed during the sequences, i.e.

$$\begin{aligned} a_k^S &= a_l^S, \\ b_m^S &= b_n^S, \end{aligned} \tag{2.20}$$

   for all k, l, m, n and any $S \in \mathcal{S}_{AB}$.

2. For any S_1 and S_2 belonging to $\mathcal{S}_{AB}$, the average of the outcomes for A and B are the same in both sequences and in any position, i.e

$$\begin{aligned} \langle a_k^{S_1} \rangle &= \langle a_l^{S_2} \rangle, \\ \langle b_m^{S_1} \rangle &= \langle b_n^{S_2} \rangle, \end{aligned} \tag{2.21}$$

   for all k, l, m, n.

When modeling this scenario as proposed here, these two conditions must be observed.

The memory cost of quantum contextuality is the cost associated to a set of sequences of compatible observables. If the number of classical states is greater than the Holevo bound associated to the quantum state used to obtain the probabilities, we have an instance of memory advantage. It should be noted that there is still no contextuality scenario in which the memory cost is known to be necessarily greater than the Holevo bound. Answering whether there is a memory advantage in the contextuality paradigm is important even in the nonlocality scenario, in which the observables of the experiment are compatible, due to space-like separation. We are concerned in this thesis with the Peres-Mermin scenario, studied in section 1.4.8.

In reference [15], the authors derived a lower bound for the memory cost of this scenario, which we detail below. As it is a contextuality scenario, at least two classical states are required to simulate it. The authors then prove that two states are not enough, and show a model with three states which respects the contextuality conditions above.

An automaton with access to L states can be represented by L *state matrices* $\{m_s\}_{s=1}^{L}$ with elements $\{f_s(X), s'_s(X)\}$, in which f_s is the output function and s'_s is the state after the output, both depending on the input X and the current state s. As an example, a toy model with three inputs, two outputs and two states can be written as

$$m_1 = \begin{bmatrix} +,1 & +,1 & +,2 \end{bmatrix}, \quad m_2 = \begin{bmatrix} -,2 & +,2 & +,2 \end{bmatrix}, \tag{2.22}$$

in which the inputs A, B and C correspond to the first, second and third entries of each matrix, respectively, and the outputs are $\{+,-\}$. Starting the automaton in $s = 1$, a sequence CA returns $(+,-)$, and the final state is $s = 2$. This toy model respects repeatability, and violates the *triangle inequality* [73]

$$\langle AB\rangle + \langle BC\rangle - \langle CA\rangle \leq 1, \tag{2.23}$$

which is a noncontextuality inequality. The model does not respect compatibility, as a sequence ACA returns different values for A.

For the Peres-Mermin scenario, a state matrix for the state i is a 3×3 matrix with the

corresponding outputs $f_i(X) \equiv X(i)$ and state transition $s'_X(i)$, i.e.

$$m_i = \begin{bmatrix} A(i), s'_A(i) & B(i), s'_B(i) & C(i), s'_C(i) \\ a(i), s'_a(i) & b(i), s'_b(i) & c(i), s'_c(i) \\ \alpha(i), s'_\alpha(i) & \beta(i), s'_\beta(i) & \gamma(i), s'_\gamma(i) \end{bmatrix}. \tag{2.24}$$

Considering this specific scenario, one should note that the fact that the conditions (1.73) cannot be respected in a noncontextual theory means that every state matrix must have at least one contradiction to these quantum conditions in one of its rows or columns. This means that the product of the outputs in the *same* state matrix is not 1 (-1 for the last column) and there is the need for a state change.

A two state automaton has the form

$$m_1 = \begin{bmatrix} A(1), s'_A(1) & B(1), s'_B(1) & C(1), s'_C(1) \\ a(1), s'_a(1) & b(1), s'_b(1) & c(1), s'_c(1) \\ \alpha(1), s'_\alpha(1) & \beta(1), s'_\beta(1) & \gamma(1), s'_\gamma(1) \end{bmatrix}, m_2 = \begin{bmatrix} A(2), s'_A(2) & B(2), s'_B(2) & C(2), s'_C(2) \\ a(2), s'_a(2) & b(2), s'_b(2) & c(2), s'_c(2) \\ \alpha(2), s'_\alpha(2) & \beta(2), s'_\beta(2) & \gamma(2), s'_\gamma(2) \end{bmatrix}. \tag{2.25}$$

Now the authors show that no two-state automaton can simulate QT, in the sense of giving valid outputs for a sequence of inputs. Without loss of generality, we will start the automaton on state 1. Suppose there is a contradiction on the last column context, i.e. $C(1)c(1)\gamma(1) = 1$. In order to fix the contradiction we will assume the state changes at $X = C$, i.e. $s'_{C,1} = 2$. Then, because of repeatability $C(2) = C(1)$,

$$m_1 = \begin{bmatrix} - & - & C(1), 2 \\ - & - & - \\ - & - & - \end{bmatrix}, m_2 = \begin{bmatrix} - & - & C(1), s'_C(2) \\ - & - & - \\ - & - & - \end{bmatrix}. \tag{2.26}$$

But there cannot be only one state change in the last column context. If this is the case, when we measure C in the last position — γcC or $c\gamma C$ — we still would have the contradiction.

Therefore we need to have another state change. We assign it to c:

$$m_1 = \begin{bmatrix} - & - & C(1),2 \\ - & - & c(1),2 \\ - & - & - \end{bmatrix}, m_2 = \begin{bmatrix} - & - & C(1),s'_C(2) \\ - & - & c(1),s'_c(2) \\ - & - & - \end{bmatrix}. \tag{2.27}$$

Note that again $c(2) = c(1)$. Even in this scenario, we still have contradictions in sequences like γcC. But if we have three state changes, repeatability forces the contradiction in both state matrices. Consequently, a two state automaton cannot reproduce the PM scenario.

The authors on [15] found a three state automaton which give valid output sequences, for inputs which belong to the same context, i.e. for sequences of inputs in the same row or column. It can be written as

$$m_1 = \begin{bmatrix} (+1,1) & (+1,1) & (+1,\mathbf{2}) \\ (+1,1) & (+1,1) & (+,\mathbf{3}) \\ (+1,1) & (+1,1) & (+1,1) \end{bmatrix}, m_2 = \begin{bmatrix} (+1,2) & (+1,\mathbf{1}) & (+1,2) \\ (-1,2) & (+1,2) & (-1,2) \\ (-1,2) & (-1,\mathbf{3}) & (+1,2) \end{bmatrix},$$

$$m_3 = \begin{bmatrix} (+1,3) & (-1,3) & (-1,3) \\ (+1,\mathbf{1}) & (+1,3) & (+1,3) \\ (-1,\mathbf{2}) & (-1,3) & (+1,3) \end{bmatrix}. \tag{2.28}$$

This automaton respects repeatability, compatibility and the quantum conditions, as one can check. As they are dispersion free, and the PM square contains incompatible observables, e.g. those in the diagonal, no quantum state can be represented by an automaton like that. This means that when trying to fully simulate a quantum state, including its correlations, and marginal probabilities, the memory cost could increase, but still we have a lower bound on the memory cost. The main purpose of this thesis is to find the memory cost to simulate *any* quantum state in the PM scenario, which is the subject of chapter 3.

2.2.2 How Many Bits to a Qubit?

In 2003, E. Galvão and L. Hardy proposed a scenario [11] in which they compare the memory cost of classical and quantum theories to solve the following problem. Consider a field $\phi(x)$ with the property that

$$\int_A^B \phi(x)dx = \alpha m, \tag{2.29}$$

in which m is an integer and α is a fixed constant, and we consider some interval $x \in [A, B]$. The problem is to tell if m is even or odd, sending a physical system through the field. An analog classical system with continuous degree of freedom can perform the task by the right coupling to the field. For instance, a rod with a fixed end and free to rotate in the yz plane and the coupling defined such as the angle θ from the rod and its position x is implicit in the relation

$$d\theta = \eta\phi(x)dx. \tag{2.30}$$

If $\eta = \pi/\alpha$ the rod will rotate by $m/2$ turns in the end. Then, if m is even, the rod will be pointing in the same direction of the beginning, and if m is odd, it will be pointing to the opposite side.

The quantum strategy uses a two level system, for example a spin 1/2 particle prepared interacting with a field in the $\hat{x}$ direction such that it rotates in the yz plane while it travels. Then we adjust η such that the outcome of a measurement in B will tell the parity of m. The proposal presented in the paper for such idealization uses a transparent rod inside of which is a classical field $\phi(x) = B_x$, produced by a solenoid wrapped around it. If a photon with vertical polarization passes through the rod, the direction of the polarization rotate proportionally to the integral of the field, according to the Faraday effect. Adjusting the physical parameters of the rod, the parity of m can be told by the vertical or horizontal orientation of the polarization at point B.

A spin-half particle has access to two orthogonal states which can be distinguished with certainty in a one shot experiment, and therefore realizes the concept of a qubit. Due to the fact that the classical strategy above has a tolerance to error — we only need to identify the hemisphere at which the rod ends — maybe a discrete classical system might solve the computational task too. What the authors showed is that, when using a discrete number of classical states to simulate the strategies above, there is always a chance to fail.

To prove that a classical system with a finite number of states fail at some point, let us divide the interval AB into N equal intervals, in each of which there is a party which can classically manipulate the system. We now define ϕ_n , $n = 1, \ldots, N$, to be the integrated value of the field in its corresponding interval. Equation (2.29) now turns into

$$\sum_n \phi_n = \alpha m. \tag{2.31}$$

We will choose, for practical purposes, the field such that it takes only a discrete set of values $\phi_n = \alpha k_n/K$, in which $k_n \in \{0, 1, \ldots, 2K-1\}$, where K is a power of 2. Equation (2.30) now implies

$$\sum_n k_n = mK. \tag{2.32}$$

The classical strategy succeeds if it correctly answers the question of the parity of m at point B. In the middle of the process, the nth party receives a classical state l_{n-1} from the previous party, makes the classical manipulation based on the number ϕ_n, and sends l_n to the next party, trying to transmit some information about the partial sum of the integrated values of the field. Each l_n is one of the classical states $l_n \in \{1, \ldots, L\}$. The state output function is then denoted by $l_n = f^n_{l_{n-1}}(k_n)$. The first party possesses $l_0 = 1$ to start. The last party must decide whether m is even or odd based on k_N and on the information transmitted by l_{N-1}. The authors then proved that, given an arbitrary integer $L = 2N - 1$, there is always a scenario which demands more than L states to succeed, proving that the memory must be arbitrarily large.

The proof is by contradiction. Assume that the number of states in which the classical system can be is $L = 2K - 1$, i.e. less than what is needed to transmit k_n. Consequently, for some l_1 there must be $a \neq b$ such that $f_0^1(a) = f_0^1(b) = l_1$. Furthermore, the important function to be evaluated at the end is $S_N = (\sum_{i=1}^{N} k_i) \mod 2K$. If m is even, then $S_N = 0$, otherwise $S_N = K$. Each party n has to keep track of the partial sum $S_n = (\sum_{i}^{n} k_i) \mod 2K$ to get in the end without error. We will see this is not possible with $2N - 1$ states. Its already apparent that in the first step one already has the possibility to fail, and this error will accumulate until there is certainly a message which carries the wrong information.

Let us denote by A_n the set composed with the partial sums S_n which are consistent with the state l_{n-1}. We can see that if A_{N-1} has elements which differ by K, summing k_N can lead both to m odd or even, therefore the Nth party will have the chance to fail, for some k_n.

Associated to an arbitrary l_n, there exists a and b such that $f_{l_{n-1}}^n(a) = f_{l_{n-1}}^n(b) = l_n$. If the $(n+1)$th party receives this l_n, then

$$|A_{n+1}| \geq |A_n \oplus \{a, b\}|. \tag{2.33}$$

The expression $X = Y \oplus Z$ means the set X is formed through the set of distinct sums mod $2K$ of one element from Y and one element from Z.

Now, we will prove that we must find elements which are distant by K. We assume, to arrive later in a contradiction, that a successful protocol exists and that the following relation holds:

$$|A_n \oplus \{a, b\}| = |A_n|. \tag{2.34}$$

We can see actually that

$$A_n \oplus \{a\} = A_n \oplus \{b\}, \tag{2.35}$$

because $A_n \oplus \{a, b\} = (A_n \oplus \{a\}) \cup (A_n \oplus \{b\})$ and unions have the same elements as their

parts if and only if they are equal.

Looking to Eq. (2.35), it also implies

$$A_n = A_n \oplus \{b - a\}. \tag{2.36}$$

By denoting $\Delta_1 = b - a$, we have $A_n = A_n \oplus i\Delta_1$, for any $i \in \mathbb{Z}$. Either this Δ_1 divides $2K$, or there is a smaller Δ_m which does. It will be defined by the following procedure. If Δ_1 does not divide $2K$, there is an i such that $0 < i\Delta_1 \mod 2K < \Delta_1$, and then $\Delta_2 = i\Delta_1 \mod 2K$. Note that $A_n = A_n \oplus j\Delta_2$, for $j \in \mathbb{Z}$. After that one checks if this Δ_2 is the divisor. If not, the procedure must be repeated interactively until a divisor is found. The sequence must terminate, since $0 < \Delta_{n+1} < \Delta_n$. The last member of the sequence must be a divisor of $2K$. We will call it ν.

Since it is a divisor of K, it must be a power of 2. As the sums are taken $\mod 2K$, and K is also a power of 2, this implies that are elements from A_n which differ by K, leading to an error. Therefore we should revise our premise (2.34). It must be false, then

$$|A_{n+1}| \geq |A_n \oplus \{a, b\}| \geq |A_n| + 1 \tag{2.37}$$

is true for any successful protocol. Consequently, the last party might receive a message with $|A_N| = N$. If this number is larger than K, it would mean that there are elements which differ by K in the possible sums, and therefore it will be impossible for the Nth party give the right answer with certainty. Thus, we need $K \geq N$. But we have chosen $L = 2K - 1$. This reasoning implies $L \geq 2N - 1$. As N and K are freely chosen, the number of different internal classical states must be arbitrarily large. Concluding, no finite classical system reproduces the quantum qubit protocol faithfully.

2.2.3 Probabilistic Automata

In the literature, deterministic automata [11, 15, 16, 71] and $\epsilon-$transducers [70, 74] are used to model the sequential scenario. Probabilistic automata [75], those which allow probabilistic answers and state transitions, are more powerful than deterministic ones; the simplest way to see this is in the adapted description of the "Absent Minded Driver" problem [76] described now.

Absent Minded Driver In this situation, illustrated in figure 2.4, a driver wants to get home. The path consists of two roundabouts in the first of which the driver has to pass straightforward and in the second the driver has to take the first exit, to the right. But, when confronted with a roundabout, the driver cannot remember if it is the first or the second. The driver, then, only follow direct instructions such as "if a roundabout found: turn right" or "go straight ahead" . Deterministic strategies for the memoryless driver always leads to wrong exits. These strategies corresponds to always take the first exit or to go straight ahead when confronted with a roudabout, no matter if it is the first or the second one. Then, the driver could never arrive home without the help of friends. On the other hand, there is a better strategy for the memoryless driver: to randomly choose between the two exits. In this way, the driver always has a positive chance to get home. This is a clear instance in which the random strategy is better than deterministic ones. Note also that with one bit of memory, the problem can be solved with certainty. The driver just have to memorize the position of the roundabout, and follow the instructions for each one. In terms of probabilities, denoting by R the instance of confronting the roundabout(input), by 1 the output of going straight and by 0 the output of turning right, the deterministic memoryless strategies always impose $p(01 \mid RR) = 0$, while the probabilistic ones allows for $0 < p(01 \mid RR) < 1$ and the 1-bit strategy allows $p(01 \mid RR) = 1$.

2.2.4 Memory Cost for Probabilistic Behaviour

In our model, which we propose now, this kind of memory saving strategy corresponds to intrinsic randomness for the output, i.e a probabilistic output even when the strategy λ is

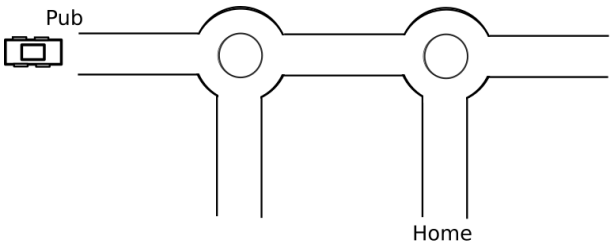

Figure 2.4: Without memory of which roundabout the driver is, the best strategy is to choose randomly.

already considered — $p(x \mid X, \lambda)$ is different from 0 or 1. In order to simulate measurements on quantum systems, intrinsic randomness can help to reduce the memory needed in sequences of noncompatible observables. The sequence of $\sigma_x\sigma_z\sigma_x$ measurements on a qubit in the state $|0\rangle$ needs no internal memory to be implemented by a probabilistic strategy. Just take a coin, and every measurement corresponds to a coin flip. But if there is no intrinsic randomness, having $p(011|\sigma_x\sigma_z\sigma_x) > 0$ is not trivial, at least one bit of memory is required to reproduce the opposite outcome in the last measurement, with respect to the first. This approach has been observed in Ref. [70]. Now consider sequences of only one measurement on $|0\rangle$, like $\sigma_x\sigma_x$ or $\sigma_y\sigma_y$. Using only intrinsic randomness, the automata will need an amount of internal states with size of at least the number of possible outcomes. This is because it has to be random in the first output, but repeatable in the rest of the sequence. Therefore, at this point it is easy to see the role of the external random variable. To reproduce these sequences, we need no memory if the outcome depend on the value of an external random variable λ, and we assume this variable is produced with no cost. At every run, consisting of a complete sequence of measurements, a different λ is selected, and consequently this kind of strategy deals with the diversity of the outcomes, without having to add more internal states. Thus, if we represent a strategy λ by a vector of outputs $(\sigma_x(\lambda), \sigma_y(\lambda))$, we can randomly select two distinct strategies, $(\sigma_x(\lambda_1), \sigma_y(\lambda_1)) = (+1, +1)$ and $(\sigma_x(\lambda_2), \sigma_y(\lambda_2)) = (-1, -1)$. Now, the memory is represented by the number of different internal classical states the automaton uses, which is related to the necessity of state transitions if ones wants to simulate a given set of sequences of measurements.

Concluding, the model we propose is probabilistic, with internal and external randomness and with access to a set of classical states. At this point, we note that the output and state

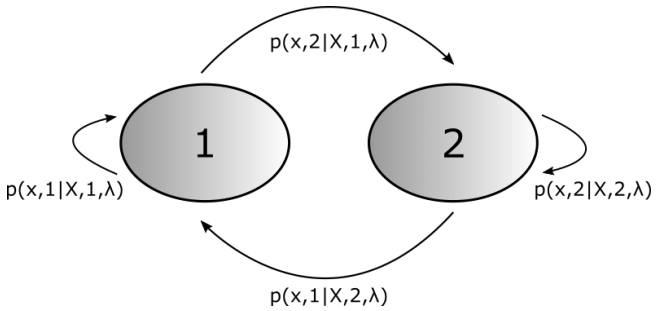

Figure 2.5: Example of an ϵ-transducer with two states. The arrows indicate the transitions, and attached to the arrows are the events with the corresponding probability.

transitions depend on which internal state the machine is, the input and the strategy selected. The minimal number of states needed to reproduce the given probability distributions gives the memory cost for that specific situation.

The framework for general sequences of measurements is the following. Consider a sequential input-output process of length N reproduced with a set of inputs $\mathcal{I} = \{0, \ldots, I-1\}$ and set of outputs $\mathcal{O} = \{0, \ldots, 1-O\}$. By sequential we mean every time an input $X_i \in \mathcal{I}$ is entered an output $x_i \in \mathcal{O}$ is returned, with $i = 1, \ldots, N$. The process is characterized by the probabilities $P(x_1 x_2 \ldots x_N \mid X_1 X_2 \ldots X_N)$. The classical counterpart is modelled with a mixture of probabilistic automata with access to a set M of classical internal states $M = \{1, \ldots, k\}$, representing the memory, and to a random variable λ, with probability distribution $p(\lambda)$. At the beginning of every sequence of measurements, the automaton selects a strategy λ according to $p(\lambda)$, with an starting state $s_0 \in M$ selected with probability $p(s_0 \mid \lambda)$. In any moment of the sequence, before every input, the internal state of the automaton is some $s \in M$. The automaton receives, then, an input X, outputs x and updates its state to s' according to $p(x, s' \mid X, s, \lambda)$. Therefore, the probabilities achieved by our model are

$$P(x_1, \ldots, x_N \mid X_1, \ldots, X_N) = \\ = \sum_{\lambda, \vec{s}} p(\lambda) p(s_0 \mid \lambda) p(x_1, s_1 \mid X_1, s_0, \lambda) \ldots p(x_N, s_N \mid X_N, s_{N-1}, \lambda), \tag{2.38}$$

in which $\vec{s} = (s_1, \ldots, s_N)$ and the sum is over all the possible values of its entries.

Figure 2.5 represents one possible way of visualizing the process. It represents an example of 2 state automaton, for a fixed λ. The automaton starts at an specific strategy λ and selects the starting state s_0. The arrows represent the state transition after the corresponding input, labeled by the probability of this event.

By convention, the memory cost is counted in terms of bits, which is the logarithm in base two of the number of classical states used by the automaton. One should note that a noncontextual probability distribution is the one which uses only one state, i.e., no memory, to be simulated. And this is exactly the point in which we relate contextuality to memory. In addition to being the costly resource, the memory cost can be interpreted as a quantifier for contextual behavior.

2.3 Chapter Conclusions

The relation between the simulation of some quantum scenarios and the classical resources to simulate them was studied. The communication cost was studied for singlet correlations, and it was found that with one bit it is possible to reconstruct the correlations obtained from measurements on this quantum system. The memory cost was studied for the Peres-Mermin scenario, and we saw that at least three internal states are necessary to reproduce the correlations obtained in QT.

Chapter 3

Memory Cost for the Peres-Mermin Scenario

3.1 Memory Cost to Simulate PM Scenario

In this section we present one of the main results of this thesis. The memory cost for the PM scenario is found, based on the probabilistic model we propose.

3.1.1 Probabilistic Models for PM scenario

The paper from Kleinmann *et al.* [15] shows us that no machine with less than three internal states allows us to reproduce conditions (1.73), if compatibility and repeatability are also demanded — see section 2.2.1. The analysis was deterministic, but it is easy to see that a probabilistic model cannot improve the results, as shown now.

To assume the model is probabilistic means that $p(x, s' \mid X, s, \lambda)$ is different from 0 or 1 for at least one configuration of the arguments. Fixing λ and s, we note that, to respect repeatability for a given X, if the output is probabilistic then there should be a transition to two distinct states, one for each output. Thus, this situation already demands three states.

Next, we study when this is not the case, i.e. the output is deterministic, and the model only counts on probabilistic transitions and fixed outputs. The argument goes very similar to the deterministic case. Thus, let us assume we have two states, a starting state s_1 and other general state s_2, and we are in a specific context with the observables, say, A, B and C. Due to the fact that the model is noncontextual, at least for one context the product of the outputs in the state matrix is in contradiction with (1.73). Without loss of generality, we can choose this contradiction to be the ABC context. Therefore we need a state transition to avoid the contradiction, and it will be placed at the A measurement. The repeatability condition demands that the value for A is the same on s_1 and s_2. But if we have only this transition the value of A is already defined by B and C, as we can make the reverse measurement BCA, the product of which is also equal to the identity. Therefore we need at least two state transitions, from s_1 to s_2. Let those transitions be in A and B. Yet, we see that the value of C determines the product of the other output: $C(s_1) = A(s_1)B(s_2) = B(s_1)A(s_2)$, because the product of the three outputs for the sequence ABC is fixed by the context. Together with repeatability, these assumptions tells us that the output of these measurements should be the same in both states, leading to the same contradiction. The last case is when there is three state transitions, but then repeatability already imposes the same output for the three observables.

Concluding, no machine — deterministic or probabilistic — with less than three states can reproduce sequences of quantum measurements simulating the Peres-Mermin scenario.

3.1.2 Other Valid Automata

As we concluded, an automaton needs at least three internal states to be able to reproduce the contextuality and repeatability conditions in the Peres-Mermin scenario. Also, a quantum state cannot be reproduced by any single automaton. Therefore, we will use the automaton proposed in reference [15] and make some operations on it to find other automata which respect the quantum conditions (1.73), and then make convex combinations of them to obtain the desired correlations. As we will see in the analysis below, there is no need for intrinsic randomness in this situation.

Every automata that respects the quantum conditions are valid automata. There are some operations when performed on the automata in Eq. (2.28) produce other valid automata:

i Flip the signs of the outputs in the same way in the three state matrices in such a way that there are either two sign flips or no sign flips in every row and column. As an example, it is shown below the flipping pattern in the first matrix, with the flips represented by diamonds, and then the corresponding valid automaton:

$$\begin{bmatrix} \diamond & \diamond & . \\ \diamond & . & \diamond \\ . & \diamond & \diamond \end{bmatrix} \tag{3.1}$$

$$m_1 = \begin{bmatrix} (-1, m_1) & (-1, m_1) & (+1, \mathbf{m_2}) \\ (-1, m_1) & (+1, m_1) & (-, \mathbf{m_3}) \\ (+1, m_1) & (-1, m_1) & (-1, m_1) \end{bmatrix}, m_2 = \begin{bmatrix} (-1, m_2) & (-1, \mathbf{m_1}) & (+1, m_2) \\ (+1, m_2) & (+1, m_2) & (+1, m_2) \\ (-1, m_2) & (+1, \mathbf{m_3}) & (-1, m_2) \end{bmatrix},$$
$$m_3 = \begin{bmatrix} (-1, m_3) & (+1, m_3) & (-1, m_3) \\ (-1, \mathbf{m_1}) & (+1, m_3) & (-1, m_3) \\ (-1, \mathbf{m_2}) & (+1, m_3) & (-1, m_3) \end{bmatrix}. \tag{3.2}$$

ii Any permutation of the rows applied in the same way in all of the three state matrices; the transition rule must also change accordingly. Or a permutation between the first and the second column.

All automata obtained in this way might be viewed as a deterministic strategy in the set of three state automata. According to the proposed model in Eq. (2.38), the mixture of three state automata is also a valid strategy. Therefore, we shall use them to try to simulate the quantum correlations in the PM scenario.

Applying the rule i arbitrarily results in 15 new tables of the form (2.28). For each one of these 16 tables, we also made the permutations of the rows $1 \leftrightarrow 2, 1 \leftrightarrow 3, 2 \leftrightarrow 3$,

and the column permutation $1 \leftrightarrow 2$, together with the original table. There are other valid automata, but for computational reasons we want to keep the minimum of strategies needed. Consequently, until now we have $16 \times 5 = 80$ automata, and they proved to be sufficient. Every state in each automaton can be chosen to be the initial state, thus we arrive in 240 different values for λ.

3.1.3 Testing Sufficient Sequences

The model should reproduce every correlation obtained from the sequence of measurements of the form $x_1 x_2 \ldots \mid X_1 X_2 \ldots$, given that $X_1, X_2 \ldots$ are pairwise compatible; thus, we would have to test an infinite number of sequences in order to see if the model is right. However, there are some conditions on the probabilities which allow us to reduce the number of sequences tested. Furthermore, practically, the contexts tested in experiments are finite in number and sequence length.

In the PM scenario, when we measure two distinct observables the third one is already determined, because of the conditions (1.73). Therefore,

$$P(xyz \mid XYZ) = P(xy \mid XY)\delta_{x\cdot y\cdot z,\mathrm{sgn}(C)}, \tag{3.3}$$

for the inputs $X \neq Y \neq Z$. The function sgn is defined as $\mathrm{sgn}(C) = 1$, for $C = R_1, R_2, R_3, C_1, C_2$, and $\mathrm{sgn}(C) = -1$, if $C = C_3$. Remeber that C_i and R_j are the column and row contexts, respectively. Then, we should look for infinite sequences of two distinct measurements. Note also that in quantum mechanics and in the model, repeatability is always respected. Consequently, when one measures repeated observables, e.g., a sequence $X \ldots XYXY$, the same probability distribution for the outcomes as those of the $X \ldots XY$ sequence are obtained. This is the case because when the second distinct observable is measured, all the sequences that can result from this scenario have their probability already defined. Furthermore, the outcome for Y is only defined when it is measured, and differences in the length of the string of X might result in different values for y. The only case missing, then, is when one observable repeats itself until

other observable is measured. For quantum mechanics, we have

$$\begin{aligned} P(x,x,\ldots,x,y\,|X,X\ldots,X,Y) = \\ = \mathrm{tr}(\Pi_y^Y\Pi_x^X\ldots\Pi_x^X\Pi_x^X\rho\Pi_x^X\Pi_x^X\ldots]\Pi_y^Y) = \mathrm{tr}(\Pi_y^Y\Pi_x^X\rho\Pi_x^X\Pi_y^Y) = \\ = P(xy\mid XY). \end{aligned} \tag{3.4}$$

However, for the automata, the value of y might not be fixed until we actually measure it, and sequences like XXY and XY might produce different outcomes for y. Therefore we have to look how our specific model behaves in this situation. Analysing the automaton in (2.28), it is possible to conclude that the output for a measurement sequence $XX\ldots XY$ is $xx\ldots xy$ if and only if the output for XY is xy. This is due to the fact that we have only one state transition for column contexts, and for row contexts we always change to a row with identical outputs. Applying transformations i and ii to the original automaton does not change these properties. Therefore it is enough to test probabilities for sequences of size two of distinct measurements.

Usually it is better to construct the deterministic strategies by its outputs, i.e. how many outputs we can obtain given we have access to a fixed amount of memory. For deterministic automata, we can count the number of distinct strategies by determining the number of different automata we can construct. We denote by $\mathcal{I} = \{0,\ldots,I-1\}, \mathcal{O} = \{0,\ldots,O-1\}$ and k the input set, output set and number of internal states of the model, respectively. The deterministic automaton can be represented, as in Eq. (2.28), by a vector of pairs $(X_l(s), s'_l(s))_{l=1}^I$ for every state s. The number of different vectors we can construct is, then, $(O\cdot k)^I$ per state. As every state can be seen as an independent vector, we have a total of $(O\cdot k)^{I\cdot k}$ different automata. In this way, we have to test which kind of probability vectors we can produce, starting from the outputs. The case of probabilistic automata is different, as there exists an infinite number of different behaviors, and we need to parametrize a general automaton. Given the probability distribution $p(s',x\mid s,X,\lambda)$, we will analyze the number of parameters for a fixed state s and strategy λ. Then, every s', x and X defines one parameter, therefore we have $k\cdot O\cdot I$ parameters. There is one normalization to every input, consequently to every state is associated $(k\cdot O-1)I$ parameters. As each state is independent, the total number of parameters is $(k\cdot O-1)I\cdot k$.

The λ labels, then, every distinct selection of these parameters. As it is supposed to be, the set of deterministic probabilities is inside the probabilistic set. Just note that the number of ways to arrange the parameters associated to every state and input such that there is only one 1 and the rest 0's in the probability distribution is $O \cdot k$. Therefore, there are $(O \cdot k)^I$ ways per state to distribute this assignments, leading to $(O \cdot k)^{I \cdot k}$ different deterministic behaviors.

For practical reasons, we are not going to test the set of probabilities $p(xy \mid XY)$. Instead, we deal with the equivalent set of expectation values

$$\begin{aligned} \langle X \rangle &= \sum_{x,y} xP(xy \mid XY), \\ \langle XY \rangle &= \sum_{x,y} xyP(xy \mid XY), \\ \langle XYX \rangle &= \sum_{x,y} yP(xy \mid XY). \end{aligned} \tag{3.5}$$

Considering $\langle X \rangle$, there are 9 possibilities for the PM scenario, one for each observable. For $\langle XY \rangle$, there are in principle 36 — 6 combinations for each context — but, for quantum mechanics, $\langle XY \rangle = \langle YX \rangle$ and for the automata the product xy is always the same as yx. This latter fact happens because there is at least one input for which there is no state change, for every context in the state matrices. As the product of the three observables in the same line is fixed, if one outcome is fixed, the product of the other two is also fixed. Therefore we have 18 different expectation values. Finally, there are 36 expectation values of the form $\langle XYX \rangle$. Thus, if the model can reproduce these 63 expectation values with the 3 state automata, by means of Eq. (3.5), the probabilities for *any* sequence would also be in accordance with the desired probabilities. In the analysis below, we use only deterministic behaviors and we see they are sufficient. If this were not the case, we should include probabilistic behaviors before testing for an increased-memory model.

In this approach, we need to construct the correlation vector for every automaton. The correlation obtained by a quantum state for a sequential measurement is $\langle XY \rangle = \mathrm{tr}(\rho XY)$. On

the other hand, every automaton corresponds to a deterministic correlation vector. To obtain the desired correlations, the automata should be selected randomly according to the probability of the external random variable λ. Consequently, we arrive at a *feasibility program* for $p(\lambda)$:

$$\begin{aligned}
\mathrm{tr}(\rho X) &= \sum_{\lambda} p(\lambda) x_1^X(\lambda), \\
\mathrm{tr}(\rho XY) &= \sum_{\lambda} p(\lambda) x_1^{XY}(\lambda) y_2^{XY}(\lambda) = \sum_{\lambda} p(\lambda) x_1^{X}(\lambda) y_2^{XY}(\lambda), \\
\mathrm{tr}(\rho XYX) &= \sum_{\lambda} p(\lambda) x_1^{XYX}(\lambda) y_2^{XYX}(\lambda) x_3^{XYX}(\lambda) = \sum_{\lambda} p(\lambda) y_2^{XY}(\lambda),
\end{aligned} \tag{3.6}$$

in which $x_l^{X_1 \ldots X_l \ldots X_n}(\lambda)$ is the output in the position l for $X_l = X$ in the sequence $X_1 \ldots X_n$, and $y_m^{X_1 \ldots X_m \ldots X_n}(\lambda)$ is the output in the position m for $X_m = Y$ in the sequence $X_1 \ldots X_n$, given the strategy λ. In the second line of Eq. (3.6) it was considered that $x_1^{XY} = x_1^X$ because there is no influence on future sequences on past outcomes, and in the third line $x_1^{XYX} y_2^{XYX} x_3^{XYX} = y_2^{XY}$ because in addition the value of x repeats in all its occurrences.

It is possible to find which distribution $p(\lambda)$ reproduces specific states by linear programming. As a practical example, reproducing the singlet state $|\psi\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ demands four automata $\lambda = 1, 2, 3, 4$, chosen with equal probability. The automata can be written as:

$$\lambda_1 :$$

$$m_1^{(1)} = \begin{pmatrix} - & + & -, m_2 \\ - & - & +, m_3 \\ + & - & - \end{pmatrix}; m_2^{(1)} = \begin{bmatrix} - & +, m_1 & - \\ + & - & - \\ - & +, m_3 & - \end{bmatrix}; \quad m_3^{(1)} = \begin{pmatrix} - & - & + \\ -, m_1 & - & + \\ -, m_2 & + & - \end{pmatrix}; \tag{3.7}$$

$$\lambda_2 :$$

$$m_1^{(2)} = \begin{pmatrix} - & + & -, m_2 \\ + & + & +, m_3 \\ - & + & - \end{pmatrix}; m_2^{(2)} = \begin{bmatrix} - & +, m_1 & - \\ - & + & - \\ + & -, m_3 & - \end{bmatrix}; \quad m_3^{(2)} = \begin{pmatrix} - & - & + \\ +, m_1 & + & + \\ +, m_2 & - & - \end{pmatrix}; \quad (3.8)$$

λ_3 :

$$m_1^{(3)} = \begin{pmatrix} + & - & -, m_2 \\ - & - & +, m_3 \\ - & + & - \end{pmatrix}; m_2^{(3)} = \begin{bmatrix} + & -, m_1 & - \\ + & - & - \\ + & -, m_3 & - \end{bmatrix}; \quad m_3^{(3)} = \begin{pmatrix} + & + & + \\ -, m_1 & - & + \\ +, m_2 & - & - \end{pmatrix}; \quad (3.9)$$

λ_4 :

$$m_1^{(4)} = \begin{pmatrix} + & - & -, m_2 \\ + & + & +, m_3 \\ + & - & - \end{pmatrix}; m_2^{(4)} = \begin{bmatrix} + & -, m_1 & - \\ - & + & - \\ - & +, m_3 & - \end{bmatrix}; \quad m_3^{(4)} = \begin{pmatrix} + & + & + \\ +, m_1 & + & + \\ -, m_2 & + & - \end{pmatrix}. \quad (3.10)$$

The starting state is now fixed and it is highlighted by the square brackets, i.e. $s_0 = 2$. In this form, one clearly sees that the above strategy represents an eigenstate of the observables of the last column context, as it is supposed to be the case, because there is no state changes in the last columns in the starting state matrices, and the corresponding outputs have the same value in all strategies.

3.1.4 Probing All States

The feasibility program (3.6) is useful when dealing with specific states. In the present case, the query if all quantum states can be classically simulated by a three state automaton must be solved with a different approach. Our method can be seen geometrically: we will show that the set of probabilities obtained from quantum theory is a subset of the set of probabilities given by three-state automata.

The set of correlation vectors produced by the automata is convex and, as we are dealing with deterministic strategies, each correlation vector obtained from every automaton is a vertex. Therefore, the set of strategies is a polytope, as already mentioned, c.f. (1.5) and (1.6). The V-representation for this case is

$$\begin{aligned} P &:= \{\vec{q} \mid \vec{q} = \sum_{\lambda} p(\lambda)\vec{f}_{\lambda}, \qquad \sum_{\lambda} p(\lambda) = 1\}, \\ &\equiv \{\vec{q} \mid q_j = \vec{v}_j \cdot \vec{p}, \sum_{\lambda} p_{\lambda} = 1, p_{\lambda} \geq 0\} \end{aligned} \tag{3.11}$$

in which λ labels the vertices. We also define $D = \{D_j\}_{j=1}^{63} = \{A, B, C, \alpha, \ldots, AB, AC, BC, \ldots, c\gamma c\}$ as the sequence of possible contexts, as defined in Sec. 3.1.3, and $\vec{f}_{\lambda}$ is the vector in which each entry j is the product of the outputs for the context D_j, for a given λ. The vector $\vec{v}_j$ has in its entries the prediction for the same sequence of measurements, for every automata, i.e., $\vec{v}_j = ((f_{\lambda=1})_j, (f_2)_j, \ldots, (f_{240})_j)$, in which the λ's are defined in Sec. 3.1.2. The H-representation is obtained from the V-representation, and it is assumed in the general form

$$P := \{\vec{q} \mid \vec{a}_l \cdot \vec{q} \leq \alpha_l, \text{ for every } l = 1, \ldots\}. \tag{3.12}$$

On the other hand, the quantum set is characterized by

$$Q := \{\vec{q} \mid q_j = \mathrm{tr}(\rho Z_j), \text{for some } \rho\}, \qquad j = 1, \ldots, 63. \tag{3.13}$$

Each operator Z_j is one of the 63 operators obtained by multiplying the observables of the sequence D_j.

If the quantum set is inside the polytope generated by the automata, see Fig. 3.1, the quantum correlation vector respects the half space inequalities (3.12). This implies that

$$\begin{aligned} \vec{a}_l.\vec{q} = \sum_j (\vec{a}_l)_j \cdot q_j &= \sum_j (\vec{a}_l)_j \cdot \mathrm{tr}(\rho Z_j) \leq \alpha_l \\ &\implies \mathrm{tr}\left[\rho(\alpha_l \mathbb{1} - \vec{a}_l \cdot \vec{Z})\right] \geq 0. \end{aligned} \tag{3.14}$$

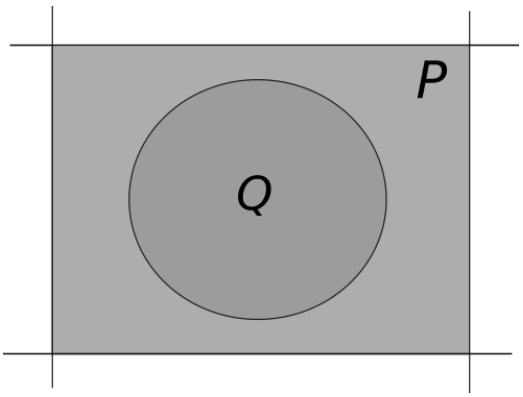

Figure 3.1: If Q is inside P, the quantum set respects the same inequalities as the classical set with memory.

for every ρ, in which we defined $\vec{Z} \equiv (Z_j)_{j=1}^{63}$. In accordance with the positivity criterion, this leads us to define the set of witnesses

$$W_l := \alpha_l \mathbb{1} - \vec{a_l} \cdot \vec{Z}. \tag{3.15}$$

Therefore, if all quantum vectors of correlations respect the automaton inequalities, for all l we have $W_l \geq 0$, i.e. they are positive semidefinite. Conversely, if the W_l's are all positive semidefinite, then any vector of correlations from the set Q respects all of the polytope inequalities (3.12) and thus it is inside the polytope. Therefore, we need to make the transformation from V-representation to H-representation, and then check if the obtained W_l's are positive semidefinite.

In principle, we constructed the vertices for the V-representation, and there are algorithms which make the transformation between V-representation and H-representation. Therefore, in principle we should be able to test for positive definiteness of the W_l's. The technical problem at this point is that the transformation between representations is a computationally hard problem. Indeed, we were not able to find a direct solution to the problem. We proceed by reducing the dimension of the polytope generated by the automata, intersecting it with the affine space of the quantum set. An affine space is a linear space displaced by some constant vector $\vec{b}$. More specifically, Q is contained in the affine space $\vec{b} + U \equiv \{\vec{b} + \vec{u} \mid \vec{u} \in U\}$, in which $b_j = \mathrm{tr}(\rho Z_j)$, for some fixed $\rho = \rho_0$ — for example $\rho_0 = \mathbb{1}/4$ — and $U = \{\vec{u} \mid u_j = \mathrm{tr}(GZ_j), \text{for some Hermitian operator } G, \mathrm{tr}(\rho_0 G) = 0\}$. This is true

since we can always write $\rho = \rho_0 + G$. The dimension of U is found to be 9, by the linear independence relations of the Z_l operators. In this way, $Q \subset P$ is the same as $Q \subset P \cap (\vec{b} + U)$.

Treating this intersection problem of convex sets as its proper extension to the corresponding cones, the equivalent problem might be formulated as follows. Let $\mathcal{P}$ be the extension to a cone associated to P, i.e., the set $\mathcal{P} = \{\vec{q} \mid \vec{q} = A\vec{r}, \vec{r} \succeq 0\}$; note that normalization condition was set aside. Notation $\vec{r} \succeq 0$ means $r_i \geq 0, \forall i$. The extension to the cone $\mathcal{P}$ is made by adding the vector $\vec{e} = (1, \ldots, 1)$ to the vectors $\vec{v}_j$ and dropping the normalization condition on $\vec{p}$. Consequently, A is the matrix with rows $[e, \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{63}]$. Therefore, $(1, \vec{q}) \in \mathcal{P} \iff \vec{q} \in P$. We also define the extension to U as

$$\mathcal{U} = \{(\lambda, \lambda\vec{b} + U) \mid \lambda \in \mathbb{R}, \vec{u} \in U\} \tag{3.16}$$

so that $(1, \vec{x}) \in \mathcal{U} \iff \vec{x} \in (\vec{b} + U)$. Now for some matrix K — whose kernel is a subspace associated to the linear space U — we construct some matrix F the range of which is the kernel of KA. It follows that

$$\begin{aligned}
\mathcal{P} \cap ker K &= \{A\vec{r} \mid \vec{r} \succeq 0, K(A\vec{r}) = \vec{0}\} \\
&= \{A\vec{r} \mid \vec{r} \succeq 0, \vec{r} = F\vec{s}, \text{ for some } \vec{s}\} \\
&= \{AF\vec{s} \mid F\vec{s} \succeq 0\} \\
&= AF\{\vec{s} \mid F\vec{s} \succeq 0\} \\
&= AF\{F'\vec{s} \mid \vec{s} \succeq 0\} \\
&= \{AFF'\vec{s} \mid \vec{s} \succeq 0\}.
\end{aligned} \tag{3.17}$$

The structure of the above reasoning is the following. The intersection with the plane of the quantum predictions is made through the additional condition that $KA\vec{r} = 0$. Thus, $\vec{r}$ is on the Kernel of KA, consequently it can be seen as a mapping through the matrix F. Until now, the set is written in the V-representation. The next step is to transform to the H-representation, to get the matrix F', and with these matrices we are able to find the inequalities that define the polytope $P \cap (\vec{b} + U)$.

Finding some K such that $\ker(K) = \mathcal{U}$, we choose some matrix F with range $\ker(KA)$. The matrix $B := AFF'$ is computed easily for this scenario. The next step is to revert the representations again to get the inequalities:

$$\{B\vec{s} \mid \vec{s} \succeq 0\} = \{\vec{s} \mid B'\vec{s} \succeq 0\}. \tag{3.18}$$

Now, this routine is already done with *cddlib* [77] to transform the representations. To generate the matrices K and F, we use *iml* [78]. Both packages work with unlimited integer precision, and thus the computation is exact. We verify independently the results using *porta* [79].

The next step is to use B' to obtain the inequalities to generate the W_l's. We have that $(1, \vec{q}) \in \mathcal{P} \cap \mathcal{U}$ if and only if

$$B'_{l,1} + \sum_j B'_{l,j+1} q_j \geq 0, \tag{3.19}$$

and we find our witnesses — c.f. Eq. (3.14) — to be

$$W_l = B'_{l,1}\,\mathbb{1} - \sum_j B'_{l,j+1} Z_j. \tag{3.20}$$

Proceeding exactly like described above, we found all the W_l's to be positive semidefinite. Therefore we can conclude that the Peres-Mermin scenario on two qubits can be simulated by a classical machine with only three internal states of memory. This is remarkable, since we expected somehow the memory cost to increase, with possibility to go beyond the Holevo bound.

Concluding, it remains an interesting question whether the memory cost for quantum contextuality can be greater than the Holevo bound for some scenario. The negative answer would mean that a classical system can efficiently simulate a contextuality scenario, while the positive answer would mean that there is a memory advantage of using quantum systems to perform sequential operations.

In the PM scenario, we found that a trit can be used to violate the corresponding NC inequality. Therefore, a qutrit could be used to play the same role as the trit, with manipulation on the computational basis, for example. It is possible that we still have a quantum advantage for this scenario? If a qubit could simulate the classical trit on this scenario, the answer is affirmative. In the next section we use this motivation to extend the analysis to the question whether or not there is contextuality on a qubit. When considering joint projective measurements on a two-level quantum system, there exists noncontextual models to explain the correlations. But we consider a broader scenario which uses instruments as the measurement model. Interestingly, the analysis made so far converged to a line of research explored also in Refs. [80–85]. In our analysis, presented in the next section, we are concerned with the possibility of the manifestation of contextuality using compatible and repeatable instruments.

3.2 A Qubit in a Quantum Automaton

The model of a quantum automaton proposed here works as a sequence of quantum instruments. The automaton operates on a subset of quantum states $\mathcal{S}(\mathcal{H})$, and it starts in an initial state ρ_0. It receives an input belonging to the discrete set $\mathcal{I}$, and returns an element of the discrete set of outputs $\mathcal{O}$. Also, there is an update rule $\Lambda_{x,X} : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})$ for every input–output pair. This update rule and the probability for this event to happen are modelled by instruments, and their corresponding Kraus decomposition, explained in chapter 1. For dichotomic measurements with outcomes $\{-1, +1\}$, this decomposition has the general form

$$\begin{aligned} \Lambda_{+,X}(\rho) &= \sum_i K_i \rho K_i^\dagger, \\ \Lambda_{-,X}(\rho) &= \sum_i Q_i \rho Q_i^\dagger, \end{aligned} \tag{3.21}$$

with the trace preserving condition

$$\sum_i K_i^\dagger K_i + \sum_j Q_j^\dagger Q_j = \mathbb{1}\,. \tag{3.22}$$

3.2.1 Conditions on the Instrument

If we do not impose compatibility and repeatability, we will find sequences of measurements on a qubit that violate a noncontextuality (NC) inequality. We can even violate the triangle inequality (2.23) which cannot be violated by projective measurements on quantum states. The violation can be achieved preparing a quantum state $|\psi\rangle\langle\psi|$ and assigning sharp observables to A and B such that $\{A(+1) = |\psi\rangle\langle\psi|, A(-1) = 1 - |\psi\rangle\langle\psi|\}$, $\{B(+1) = 1 - |\psi\rangle\langle\psi|, B(-1) = |\psi\rangle\langle\psi|\}$ and to C a deterministic output + together with a transformation that takes $|\psi\rangle\langle\psi| \mapsto |\psi^{\perp}\rangle\langle\psi^{\perp}|$. This will give us maximum algebraic violation, using the three sequences $\{AB, BC, CA\}$. Consequently, the measurements should be tested for compatibility and repeatability to give meaningful results.

In order to construct this model for a quantum machine, we only need to use a subset of qubit states. We assume that the machine uses at least two distinct states at some point during the sequence of measurements, otherwise it is only a convex mixture over classical probabilities. We assume the initial state ρ_0 to be a pure state $|\psi\rangle\langle\psi|$, because an initial mixed state can be decomposed into a mixture of pure states, and therefore a strategy which uses mixed states can be thought of as a mixture of strategies involving only pure states initially. We also need at least instrument which changes the state from ρ_0 to ρ_1, with $\rho_1 \neq \rho_0$, since otherwise the automaton would not use any memory and hence be noncontextual. The canonical convex decomposition of ρ_1 contains at least one projection $|\tilde{\psi}\rangle\langle\tilde{\psi}|$ which is different from $|\psi\rangle\langle\psi|$.

In terms of eq. (3.21), the condition in which the outcomes of a measurement must be repeatable in a sequence, i.e., each round gives one classical output for each measurement, reads

$$\mathrm{tr}\left[\Lambda_+\left(\Lambda_-(\rho)\right)\right] = \mathrm{tr}\left[\Lambda_-\left(\Lambda_+(\rho)\right)\right] = 0, \tag{3.23}$$

i.e. the probability to obtain contradicting outcomes is zero. Therefore

$$\mathrm{tr}\left(\sum_{i,j} Q_j K_i |\psi\rangle\langle\psi| K_i^{\dagger} Q_j^{\dagger}\right) = \mathrm{tr}\left(\sum_{i,j} K_i Q_j |\psi\rangle\langle\psi| Q_j^{\dagger} K_i^{\dagger}\right) = 0. \tag{3.24}$$

The trace of a sum is the sum of the trace, and as each term in the sum is nonnegative and we can see $Q_j K_i |\psi\rangle$ and $K_i Q_j |\psi\rangle$ as new vectors, we find, that for all i and j,

$$Q_j K_i |\psi\rangle = K_i Q_j |\psi\rangle = 0 \tag{3.25}$$

must hold. In this way,

1. if either $K_i = 0 \ \forall i$ or $Q_j = 0 \ \forall j$, we have a deterministic output and a general quantum channel to update the state;

2. if both maps are not zero, i.e. in the case of a *dichotomic measurement*, we will split in the cases in which it changes the input state, and in which it does not. Defining A as the state-changing measurement — the existence of which was assumed above — we express it through the maps $A_+(\rho) = \sum_i K_i \rho K_i^\dagger$ and $A_-(\rho) = \sum_i Q_i \rho Q_i^\dagger$. The repeatability argument on a sequence AA for the initial state $|\psi\rangle$ leads to (3.25) and on the last two measurements on AAA, which act on the part $|\tilde{\psi}\rangle \langle\tilde{\psi}|$, to

$$Q_j K_i |\tilde{\psi}\rangle = 0. \tag{3.26}$$

Together with equation (3.25), and noting that $|\psi\rangle$ and $|\tilde{\psi}\rangle$ span the two dimensional space, this leads to

$$K_i Q_j = Q_j K_i = 0, \quad \forall i, j \tag{3.27}$$

Because these relations do not depend on the indices, and the supports of the Kraus operators must be non-empty acting in the two dimensional Hilbert space, they might be defined as rank one operators

$$\begin{aligned} K_i = K_0 &= |a\rangle \langle b| , \\ Q_j = Q_0 &= |b^\perp\rangle \langle a^\perp| , \end{aligned} \tag{3.28}$$

in which $|a\rangle$ and $|b\rangle$ are some vectors in the qubit space. The normalization con-

dition (3.22) implies $|a\rangle = |b\rangle$, up to an arbitrary phase. Therefore, we can define $K_0 = |a\rangle\langle a|\,; Q_0 = 1 - |a\rangle\langle a|$. Consequently, A *is projective.*

Note, that a dichotomic measurement A' compatible with A also must be projective by the same reasoning, considering the sequences $A'A'$ and $AA'A'$. Furthermore, as we work with a two dimensional space, either they are the same projective measurement, or the complementary, i.e. a projective measurement in which the effects for each outcome are interchanged with respect to A.

Now let B be a measurement which does not change the input state. If there are two distinct input states, say $|\psi\rangle\langle\psi|$ and $|\psi'\rangle\langle\psi'|$, we can use the same argument above — the consequences of measuring the sequence BB for both states — to say that B is projective, or argument 1 if it is a channel.

The only case missing is a quantum automaton in which only one initial state $|\psi\rangle\langle\psi|$ occurs and a dichotomic non state-changing measurement B. Imposing repeatability means that the outcome for this measurement must be predefined for the input state. Now take a compatible measurement, say C. It also cannot change the input state, otherwise measuring BCB could lead to opposite outcomes on B, as it is dichotomic and the mappings are defined on a two dimensional space. Repeatability and not changing the state, therefore, leads to predefined outputs for this context, which can be modeled by a deterministic classical strategy. To violate a noncontextuality inequality, on the other hand, a context dependent result is necessary, i.e. an observable which gives "+" in one context and "−" in any other. Obviously this is not possible if all the measurements are non state changing. Without loss of generality, this deterministic part will be replaced by an identity map together with a fixed outcome, which is included in the class of general channels.

So far we found that in the scenario with repeatable and compatible instruments on a particular set of qubit states, we must work only with projections and general channels. The projections in the same context must be compatible, therefore the measurement of the observables of this context can be implemented in a joint measurement. We already know, as mentioned in Sec. 1.4.4, that projective measurements on a qubit have a classical — memoryless

— simulation. The deterministic assignments do not introduce any trouble if they are measured after the projections, since they must not disturb the outcomes of the projective measurements. The only possibility to disturb the probabilities would be to start by a measurement represented by a general map, follwed by projections. Thus, if the projections in a given context are composed by the effects $\{|\phi\rangle\langle\phi|, \mathbb{1} - |\phi\rangle\langle\phi|\}$, a compatible general map Λ must preserve its eigenspace, i.e.

$$\Lambda(|\phi\rangle\langle\phi|) = |\phi\rangle\langle\phi|, \tag{3.29a}$$

$$\Lambda(|\phi^\perp\rangle\langle\phi^\perp|) = |\phi^\perp\rangle\langle\phi^\perp|. \tag{3.29b}$$

Using the Kraus decomposition for $\Lambda(\rho) = \sum_i K_i \rho K_i^\dagger$, with

$$K_i = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix}, \tag{3.30}$$

in the basis $\{|\phi\rangle, |\phi^\perp\rangle\}$, Eqs. (3.29) are written as

$$\sum_i \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_i^* & c_i^* \\ b_i^* & d_i^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{3.31a}$$

$$\sum_i \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_i^* & c_i^* \\ b_i^* & d_i^* \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{3.31b}$$

Considering (3.31a) first,

$$\sum_i \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_i^* & c_i^* \\ b_i^* & d_i^* \end{bmatrix} = \sum_i \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \begin{bmatrix} a_i^* & c_i^* \\ 0 & 0 \end{bmatrix} = \sum_i \begin{bmatrix} |a_i|^2 & a_i c_i^* \\ a_i^* c_i & |c_i|^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.32}$$

Consequently, we conclude that

$$\sum_i |a_i|^2 = 1$$
$$\sum_i |c_i|^2 = 0 \implies c_i = 0. \tag{3.33}$$

Now considering (3.31b), an analogous reasoning leads to

$$\sum_i |d_i|^2 = 1$$
$$b_i = 0. \tag{3.34}$$

Applying this map to a general state ρ, in which

$$\rho = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}, \tag{3.35}$$

we get

$$\Lambda(\rho) = \sum_i \begin{bmatrix} a_i & 0 \\ 0 & d_i \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} a_i^* & 0 \\ 0 & d_i^* \end{bmatrix} = \sum_i \begin{bmatrix} a_i & 0 \\ 0 & d_i \end{bmatrix} \begin{bmatrix} \alpha a_i^* & \beta d_i^* \\ \gamma a_i^* & \delta d_i^* \end{bmatrix} = \sum_i \begin{bmatrix} \alpha |a_i|^2 & \beta a_i d_i^* \\ \gamma a_i^* d_i & \delta |d_i|^2 \end{bmatrix}$$
$$= \begin{bmatrix} \alpha \sum_i |a_i|^2 & \beta \sum_i a_i d_i^* \\ \gamma \sum_i a_i^* d_i & \delta \sum_i |d_i|^2 \end{bmatrix} = \begin{bmatrix} \alpha & \beta \sum_i a_i d_i^* \\ \gamma \sum_i a_i^* d_i & \delta \end{bmatrix}. \tag{3.36}$$

We can see that the probabilities of the two outcomes given the input state are the same given the state transformed by the map, i.e. $\mathrm{tr}\left[\rho \left|\phi\right\rangle \left\langle\phi\right|\right] = \mathrm{tr}\left[\Lambda(\rho) \left|\phi\right\rangle \left\langle\phi\right|\right] = \alpha$ and $\mathrm{tr}\left[\rho \left|\phi^\perp\right\rangle \left\langle\phi^\perp\right|\right] = \mathrm{tr}\left[\Lambda(\rho) \left|\phi^\perp\right\rangle \left\langle\phi^\perp\right|\right] = \delta$. We can see that this state after the channel cannot change the probability of the sequences obtained if instead we first measure the projections. Therefore, the argument ends due to the existence of classical distributions for projective measurements on a qubit, which cannot violate *any* NC inequality.

Conclusions and Perspectives

In this thesis, we saw intriguing differences between quantum and classical theories for physical phenomena. These differences lead to experimentally testable bounds on noncontextuality inequalities, which include Bell-type inequalities. When realizing the experiments in a sequential way, physicists measuring compatible observables from QT can model the experiment not in terms of density matrices and instruments, but by using classical systems with memory. This memory is related to the minimum number of classical states which must be accessible during the sequence of measurements. This has implications in better computational resources, which could be more efficient than the classical ones. Furthermore, there is an ontological query if quantum states associated to Hilbert spaces of dimension d might be in fact described by d distinct classical states, associated to the probabilistic automata. Up to now, there is no result showing a necessarily higher number of classical states when one is restricted to contextuality scenarios. Inside this topic, and the main result of this thesis, we showed that the Peres-Mermin scenario, in which $d = 4$, can be simulated by only three classical states, still less than Holevo's bound. We have thus raised the question if this is a general feature of contextuality or it depends on each situation. In the scenarios analysed so far, there is no example of memory advantage in the contextuality paradigm; on the other hand, there are other scenarios which need less than d states, like the CHSH.

If we can violate the PM inequality, i.e. Eq. (1.70), with three classical states, we can also do so with a qutrit and classical manipulation, since they transmit the same amount of information between two measurements. Thus, the next step would be to see if a sequential modeling of a *qubit* in sequential measurements might violate PM inequality, using the

instrument approach. We found a more general claim, that a qubit cannot violate *any* non-contextuality inequality. This is a stronger form of part of the analysis of the hidden-variable problem by Kochen, Specker and Bell. Thus, if one tests compatibility and repeatability in sequential measurements and still violates a noncontextuality inequality, then one can conclude that the system tested needs at least a three-dimensional Hilbert space to be described by quantum theory.

Besides the quest for memory advantage of quantum contextuality, it is interesting to investigate whether the qubit is equivalent to the bit or other generalized states in GPT's, regarding the probabilities which might be obtained. Partial results we obtained, which are not in the scope of this thesis, show that there are differences between the bit and the qubit, which can be tested experimentally, using noncompatible measurements in a sequence. On the other hand, calculations show that with a *g-bit*, i.e. a generalized bit, it is possible to violate NC inequalities, respecting both repeatability and compatibility. Consequently, an intriguing question that appears at this point is why it is not possible to violate NC inequalities for the qubit, when considering repeatable and compatible instruments.

Bibliography

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement," *Rev. Mod. Phys.*, vol. 81, pp. 865–942, 2009. http://link.aps.org/doi/10.1103/RevModPhys.81.865.

[2] B. Amaral, *The exclusivity principle and the set of quantum correlations*. PhD thesis, 2015. http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/EABA-9KJJ7H.

[3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," *Rev. Mod. Phys.*, vol. 86, pp. 419–478, 2014. https://link.aps.org/doi/10.1103/RevModPhys.86.419.

[4] J. I. de Vicente, "On nonlocality as a resource theory and nonlocality measures," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 42, p. 424017, 2014. https://doi.org/10.1088/1751-8113/47/42/424017.

[5] M. O. Scully, B.-G. Englert, and H. Walther, "Quantum optical tests of complementarity," *Nature*, vol. 351, no. 6322, pp. 111–116, 1991. http://dx.doi.org/10.1038/351111a0.

[6] M. Howard, J. Wallman, V. Veitch, and J. Emerson, "Contextuality supplies the 'magic' for quantum computation," *Nature (London)*, vol. 510, no. 7505, pp. 351–355, 2014. http://dx.doi.org/10.1038/nature13460.

[7] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, "Nonlocality and communication complexity," *Rev. Mod. Phys.*, vol. 82, pp. 665–698, 2010. https://link.aps.org/doi/10.1103/RevModPhys.82.665.

[8] O. Gühne, C. Budroni, A. Cabello, M. Kleinmann, and J.-Å. Larsson, "Bounding the quantum dimension with contextuality," *Phys. Rev. A*, vol. 89, p. 062107, 2014. https://link.aps.org/doi/10.1103/PhysRevA.89.062107.

[9] G. Brassard, R. Cleve, and A. Tapp, "Cost of exactly simulating quantum entanglement with classical communication," *Phys. Rev. Lett.*, vol. 83, pp. 1874–1877, 1999. https://link.aps.org/doi/10.1103/PhysRevLett.83.1874.

[10] M. Steiner, "Towards quantifying non-local information transfer: finite-bit non-locality," *Physics Letters A*, vol. 270, no. 5, pp. 239 – 244, 2000. http://www.sciencedirect.com/science/article/pii/S0375960100003157.

[11] E. F. Galvão and L. Hardy, "Substituting a qubit for an arbitrarily large number of classical bits," *Phys. Rev. Lett.*, vol. 90, no. 8, p. 087902, 2003. https://link.aps.org/doi/10.1103/PhysRevLett.90.087902.

[12] B. F. Toner and D. Bacon, "Communication cost of simulating Bell correlations," *Phys. Rev. Lett.*, vol. 91, no. 18, p. 187904, 2003. https://link.aps.org/doi/10.1103/PhysRevLett.91.187904.

[13] S. Pironio, "Violations of bell inequalities as lower bounds on the communication cost of nonlocal correlations," *Phys. Rev. A*, vol. 68, p. 062102, 2003. https://link.aps.org/doi/10.1103/PhysRevA.68.062102.

[14] C. Branciard and N. Gisin, "Quantifying the nonlocality of Greenberger-Horne-Zeilinger quantum correlations by a bounded communication simulation protocol," *Phys. Rev. Lett.*, vol. 107, p. 020401, 2011. https://link.aps.org/doi/10.1103/PhysRevLett.107.020401.

[15] M. Kleinmann, O. Gühne, J. R. Portillo, J.-Å. Larsson, and A. Cabello, "Memory cost of quantum contextuality," *New J. Phys.*, vol. 13, no. 11, p. 113011, 2011. http://stacks.iop.org/1367-2630/13/i=11/a=113011.

[16] S. Brierley, A. Kosowski, M. Markiewicz, T. Paterek, and A. Przysiezna, "Nonclassicality of temporal correlations," *Phys. Rev. Lett.*, vol. 115, no. 12, p. 120404, 2015. https://link.aps.org/doi/10.1103/PhysRevLett.115.120404.

[17] J. B. Brask and R. Chaves, "Bell scenarios with communication," *Journal of Physics A: Mathematical and Theoretical*, vol. 50, no. 9, p. 094001, 2017. http://stacks.iop.org/1751-8121/50/i=9/a=094001.

[18] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," *Rev. Mod. Phys.*, vol. 74, pp. 145–195, 2002. https://link.aps.org/doi/10.1103/RevModPhys.74.145.

[19] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?," *Phys. Rev.*, vol. 47, pp. 777–780, 1935. https://link.aps.org/doi/10.1103/PhysRev.47.777.

[20] D. Bohm, *Quantum theory*. Courier Corporation, 1951.

[21] D. Bohm, "A suggested interpretation of the quantum theory in terms of "hidden" variables. I," *Phys. Rev.*, vol. 85, pp. 166–179, 1952. https://link.aps.org/doi/10.1103/PhysRev.85.166.

[22] E. Specker, "Die Logik nicht gleichzeitig entscheidbarer Aussagen," *Dialectica*, vol. 14, no. 2–3, pp. 239–246, 1960. http://dx.doi.org/10.1111/j.1746-8361.1960.tb00422.x.

[23] E. Specker, "The logic of propositions which are not simultaneously decidable," *The Logico-Algebraic Approach to Quantum Mechanics. The University of Western Ontario Series in Philosophy of Science*, vol. 5, pp. 135–140, 1975.

[24] J. S. Bell, "On the einstein-podolsky-rosen paradox," *Physics*, vol. 1, no. 3, pp. 195–200, 1964.

[25] J. S. Bell, "On the problem of hidden variables in quantum mechanics," *Reviews of Modern Physics*, vol. 38, no. 3, p. 447, 1966. https://link.aps.org/doi/10.1103/RevModPhys.38.447.

[26] S. Kochen and E. P. Specker, "The problem of hidden variables in quantum mechanics," *J. Math. Mech.*, vol. 17, no. 1, pp. 59–87, 1967.

[27] A. Cabello, S. Severini, and A. Winter, "Graph-theoretic approach to quantum correlations," *Phys. Rev. Lett.*, vol. 112, p. 040401, Jan 2014. https://link.aps.org/doi/10.1103/PhysRevLett.112.040401.

[28] R. Rabelo, C. Duarte, A. J. Lpez-Tarrida, M. T. Cunha, and A. Cabello, "Multigraph approach to quantum non-locality," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 42, p. 424021, 2014. http://stacks.iop.org/1751-8121/47/i=42/a=424021.

[29] A. Cabello, "Experimentally testable state-independent quantum contextuality," *Phys. Rev. Lett.*, vol. 101, p. 210401, 2008. https://link.aps.org/doi/10.1103/PhysRevLett.101.210401.

[30] M. Kleinmann, C. Budroni, J.-Å. Larsson, O. Gühne, and A. Cabello, "Optimal inequalities for state-independent contextuality," *Phys. Rev. Lett.*, vol. 109, p. 250402, 2012. https://link.aps.org/doi/10.1103/PhysRevLett.109.250402.

[31] S. Yu and C. H. Oh, "State-independent proof of kochen-specker theorem with 13 rays," *Phys. Rev. Lett.*, vol. 108, p. 030402, 2012. https://link.aps.org/doi/10.1103/PhysRevLett.108.030402.

[32] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Phys. Rev. Lett.*, vol. 23, pp. 880–884, Oct 1969. https://link.aps.org/doi/10.1103/PhysRevLett.23.880.

[33] A. Peres, "Incompatible results of quantum measurements," *Phys. Lett. A*, vol. 151, no. 3–4, pp. 107–108, 1990. https://doi.org/10.1016/0375-9601(90)90172-K.

[34] N. D. Mermin, "Simple unified form for the major no-hidden-variables theorems," *Phys. Rev. Lett.*, vol. 65, no. 27, pp. 3373–3376, 1990. https://link.aps.org/doi/10.1103/PhysRevLett.65.3373.

[35] B. S. Cirel'son, "Quantum generalizations of Bell's inequality," *Letters in Mathematical Physics*, vol. 4, pp. 93–100, Mar 1980. https://doi.org/10.1007/BF00417500.

[36] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," *Probl. Inf. Transm.*, vol. 9, no. 3, pp. 177–183, 1973.

[37] R. Raussendorf, "Contextuality in measurement-based quantum computation," *Phys. Rev. A*, vol. 88, p. 022322, Aug 2013. https://link.aps.org/doi/10.1103/PhysRevA.88.022322.

[38] Č. Brukner, M. Żukowski, J.-W. Pan, and A. Zeilinger, "Bell's inequalities and quantum communication complexity," *Phys. Rev. Lett.*, vol. 92, p. 127901, Mar 2004. https://link.aps.org/doi/10.1103/PhysRevLett.92.127901.

[39] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués, "Operational framework for nonlocality," *Phys. Rev. Lett.*, vol. 109, p. 070401, Aug 2012. https://link.aps.org/doi/10.1103/PhysRevLett.109.070401.

[40] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, "Nonlocal correlations as an information-theoretic resource," *Phys. Rev. A*, vol. 71, p. 022101, Feb 2005. https://link.aps.org/doi/10.1103/PhysRevA.71.022101.

[41] A. Cabello, "The role of bounded memory in the foundations of quantum mechanics," *Foundations of Physics*, vol. 42, no. 1, pp. 68–79, 2012. https://doi.org/10.1007/s10701-010-9507-2.

[42] A. Cabello, M. Gu, O. Gühne, and Z.-P. Xu, "Optimal simulation of state-independent quantum contextuality," *arXiv preprint arXiv:1709.07372*, 2017. https://arxiv.org/abs/1709.07372.

[43] J. Thompson, A. J. Garner, V. Vedral, and M. Gu, "Using quantum theory to simplify input–output processes," *npj Quantum Information*, vol. 3, no. 1, p. 6, 2017. https://www.nature.com/articles/s41534-016-0001-3.

[44] M. S. Palsson, M. Gu, J. Ho, H. M. Wiseman, and G. J. Pryde, "Experimentally modeling stochastic processes with less memory by the use of a quantum processor," *Science Advances*, vol. 3, no. 2, 2017. http://advances.sciencemag.org/content/3/2/e1601302.

[45] J. Preskill *arXiv preprint arXiv:1801.00862*, 2018. https://arxiv.org/abs/1801.00862.

[46] G. Fagundes and M. Kleinmann, "Memory cost for simulating all quantum correlations from the peres–mermin scenario," *Journal of Physics A: Mathematical and Theoretical*, vol. 50, no. 32, p. 325302, 2017. http://stacks.iop.org/1751-8121/50/i=32/a=325302.

[47] B. Grünbaum, *Convex polytopes*. New York: Springer, 2nd ed., 2003. .

[48] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.

[49] Wikipedia, "Convex polytope," 2017. [Online; url="https://en.wikipedia.org/w/index.php?title=Convex_polytope&oldid=804402400", accessed 17-November-2017].

[50] T. Heinosaari and M. Ziman, *The mathematical language of quantum theory: from uncertainty to entanglement*. Cambridge University Press, 2012.

[51] P. Busch, P. J. Lahti, and P. Mittelstaedt, *The quantum theory of measurement*. Lecture Notes in Physics Monographs, Springer, 2nd rev. ed., 1996.

[52] C. Cohen-Tannoudji, B. Diu, and F. Laloe, *Quantum Mechanics: 1*. Wiley, 1995.

[53] J. J. Sakurai and J. J. Napolitano, *Modern quantum mechanics*. Pearson Higher Ed, 2014.

[54] A. Fine, "Hidden variables, joint probability, and the bell inequalities," *Phys. Rev. Lett.*, vol. 48, pp. 291–295, Feb 1982. https://link.aps.org/doi/10.1103/PhysRevLett.48.291.

[55] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. Vermeulen, R. N. Schouten, C. Abellán, *et al.*, "Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres," *Nature*, vol. 526, no. 7575, pp. 682–686, 2015. http://dx.doi.org/10.1038/nature15759.

[56] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-Å. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger, "Significant-loophole-free test of bell's theorem with entangled photons," *Phys. Rev. Lett.*, vol. 115, p. 250401, Dec 2015. https://link.aps.org/doi/10.1103/PhysRevLett.115.250401.

[57] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, "Strong loophole-free test of local realism," *Phys. Rev. Lett.*, vol. 115, p. 250402, Dec 2015. https://link.aps.org/doi/10.1103/PhysRevLett.115.250402.

[58] G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, and C. F. Roos, "State-independent experimental test of quantum contextuality," *Nature (London)*, vol. 460, no. 7254, pp. 494–497, 2009. http://dx.doi.org/10.1038/nature08172.

[59] E. Amselem, M. Rådmark, M. Bourennane, and A. Cabello, "State-independent quantum contextuality with single photons," *Phys. Rev. Lett.*, vol. 103, no. 16, p. 160405, 2009. https://link.aps.org/doi/10.1103/PhysRevLett.103.160405.

[60] R. Łapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger, "Experimental non-classicality of an indivisible quantum system," *Nature (London)*, vol. 474, no. 7352, pp. 490–493, 2011. http://dx.doi.org/10.1038/nature10119.

[61] X. Zhang, M. Um, J. Zhang, S. An, Y. Wang, D.-l. Deng, C. Shen, L.-M. Duan, and K. Kim, "State-independent experimental test of quantum contextuality with a single trapped ion," *Phys. Rev. Lett.*, vol. 110, no. 7, p. 070401, 2013. https://link.aps.org/doi/10.1103/PhysRevLett.110.070401.

[62] V. D'Ambrosio, I. Herbauts, E. Amselem, E. Nagali, M. Bourennane, F. Sciarrino, and A. Cabello, "Experimental implementation of a Kochen-Specker set of quantum tests," *Phys. Rev. X*, vol. 3, no. 1, p. 011012, 2013. https://link.aps.org/doi/10.1103/PhysRevX.3.011012.

[63] M. Jerger, Y. Reshitnyk, M. Oppliger, A. Potočnik, M. Mondal, A. Wallraff, K. Goodenough, S. Wehner, K. Juliusson, N. K. Langford, and A. Fedorov, "Contextuality without nonlocality in a superconducting quantum system," *Nat. Commun.*, vol. 7, p. 12930, 2016. http://dx.doi.org/10.1038/ncomms12930.

[64] G. Cañas, E. Acuña, J. Cariñe, J. F. Barra, E. S. Gómez, G. B. Xavier, G. Lima, and A. Cabello, "Experimental demonstration of the connection between quantum contextuality and graph theory," *Phys. Rev. A*, vol. 94, p. 012337, 2016. https://link.aps.org/doi/10.1103/PhysRevA.94.012337.

[65] G. Lüders, "Über die Zustandsänderung durch den Meßprozeß," *Ann. Phys. (Leipzig)*, vol. 443, no. 8, pp. 322–328, 1951. http://dx.doi.org/10.1002/andp.19504430510.

[66] J.-Å. Larsson, "Loopholes in Bell inequality tests of local realism," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 42, p. 424003, 2014. http://stacks.iop.org/1751-8121/47/i=42/a=424003.

[67] A. Montina and S. Wolf, "Information-based measure of nonlocality," *New Journal of Physics*, vol. 18, no. 1, p. 013035, 2016. http://stacks.iop.org/1367-2630/18/i=1/a=013035T1.

[68] T. Vértesi and E. Bene, "Lower bound on the communication cost of simulating bipartite quantum correlations," *Phys. Rev. A*, vol. 80, p. 062316, Dec 2009. https://link.aps.org/doi/10.1103/PhysRevA.80.062316.

[69] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 10th anniversary ed. ed., 2010.

[70] A. Cabello, M. Gu, O. Gühne, J.-Å. Larsson, and K. Wiesner, "Thermodynamical cost of some interpretations of quantum theory," *Phys. Rev. A*, vol. 94, p. 052127, Nov 2016. https://link.aps.org/doi/10.1103/PhysRevA.80.062316.

[71] M. Holcombe, *Algebraic automata theory*. Cambridge studies in advanced mathematics 1, Cambridge University Press, 1982.

[72] O. Gühne, M. Kleinmann, A. Cabello, J.-A. Larsson, G. Kirchmair, F. Zähringer, R. Gerritsma, and C. F. Roos, "Compatibility and noncontextuality for sequential measurements," *Phys. Rev. A*, vol. 81, p. 022121, Feb 2010.

[73] M. Araújo, M. T. Quintino, C. Budroni, M. T. Cunha, and A. Cabello, "All noncontextuality inequalities for the n-cycle scenario," *Phys. Rev. A*, vol. 88, p. 022118, 2013. https://link.aps.org/doi/10.1103/PhysRevA.88.022118.

[74] N. Barnett and J. P. Crutchfield, "Computational mechanics of input–output processes: structured transformations and the ϵ-transducer," *J. Stat. Phys.*, vol. 161, pp. 404–451, 2015. https://doi.org/10.1007/s10955-015-1327-5.

[75] A. Paz, *Introduction to Probabilistic Automata*. New York, London: Academic Press, 1971.

[76] M. Piccione and A. Rubinstein, "On the interpretation of decision problems with imperfect recall," *Games and Economic Behavior*, vol. 20, no. 1, pp. 3–24, 1997. https://doi.org/10.1006/game.1997.0536.

[77] K. Fukuda, "cddlib v094h." https://www.inf.ethz.ch/personal/fukudak/cdd_home/.

[78] Z. Chen, A. Storjohann, and C. Fletcher, "Integer Matrix Library v1.0.5." https://cs.uwaterloo.ca/~astorjoh/iml.html.

[79] T. Christof and A. Löbel, "POlyhedron Representation Transformation Algorithm v1.4.1." http://porta.zib.de/.

[80] P. Busch, "Quantum states and generalized observables: a simple proof of gleason's theorem," *Phys. Rev. Lett.*, vol. 91, p. 120403, 2003. https://link.aps.org/doi/10.1103/PhysRevLett.91.120403.

[81] C. M. Caves, C. A. Fuchs, K. K. Manne, and J. M. Renes, "Gleason-type derivations of the quantum probability rule for generalized measurements," *Foundations of Physics*, vol. 34, no. 2, pp. 193–209, 2004. https://doi.org/10.1023/B:FOOP.0000019581.00318.a5.

[82] A. Cabello, "Kochen-specker theorem for a single qubit using positive operator-valued measures," *Phys. Rev. Lett.*, vol. 90, p. 190401, 2003. https://link.aps.org/doi/10.1103/PhysRevLett.90.190401.

[83] F. De Zela, "Single-qubit tests of Bell-like inequalities," *Phys. Rev. A*, vol. 76, p. 042119, 2007. https://link.aps.org/doi/10.1103/PhysRevA.76.042119.

[84] A. Grudka and P. Kurzyński, "Is there contextuality for a single qubit?," *Phys. Rev. Lett.*, vol. 100, p. 160401, 2008. https://link.aps.org/doi/10.1103/PhysRevLett.100.160401.

[85] R. Kunjwal and S. Ghosh, "Minimal state-dependent proof of measurement contextuality for a qubit," *Phys. Rev. A*, vol. 89, p. 042118, 2014. https://link.aps.org/doi/10.1103/PhysRevA.89.042118.