Temporal Correlations in Quantum Theory

Masterarbeit
zur Erlangung des akademischen Grades
Master of Science
(M. Sc.)

der Universität Siegen

Department Physik

vorgelegt von
Jannik Hoffmann

Oktober 2016

UNIVERSITY OF SIEGEN
Department of physics
Hoffmann, Jannik: Temporal Correlations in Quantum Theory
M. Sc. Thesis *89 pages*.
Theoretical quantum optics
October 2016

Contents

Introduction

In the year 1935, Einstein, Podolsky and Rosen (EPR) presented their famous EPR-argument [1]. The argument shows that either the description, given by the quantum state $|\psi\rangle$ is incomplete or that there exist nonlocal influences. As EPR did not believe in the existence of such influences, they concluded that $|\psi\rangle$ is not a complete description of the system. They assumed that the incomplete description follows from the existence of unknown *hidden variables*. The EPR argument led to the question, whether we can replace quantum mechanics by a *local hidden variable* (LHV) model.
30 years later Bell [2] showed that quantum mechanics cannot be replaced by LHV theories. Bell wrote down the explicit assumptions of LHV theories and derived the bounds on the possible strength of correlations that follow for such theories. These days, we know these bounds as Bell-inequalities. Quantum mechanics violates these inequalities and thus cannot be replaced by LHV theories. Nevertheless, it took additional 50 years until the first loophole free experiments, i.e. removing all possible classical explanations such as inefficient measurements and influences travelling slower than light, were performed [3, 4, 5]. The distinction between QM and local hidden variable theories furthermore initiated new research topics, to decide which features from QM are genuinely quantum and which can be simulated by local hidden variable models. The feature of entanglement, until recently thought to be genuine feature of QM, can actually be simulated with hidden variable models, e.g. by Spekkens' Toy Model [6].
Furthermore, the existence of so called "post-quantum" theories, i.e. theories, that still fulfill the nonsignalling conditions resulting from the locality conditions, that allow stronger correlations than QM, lead to the question why this kind of correlations are neither realized by quantum theory or in nature.

A second no-go theorem, to distinguish between quantum theory and classical theories was given by Kochen and Specker. They showed that quantum theory is contextual for dimensions $d \geq 3$ [7]. Here, noncontextuality means that a measurement result of some

observable A does not depend on the context that it is measured jointly with, where a context is a set of measurements that do not disturb/influence each other. Measurements that do not disturb each other are called *compatible*. In QM this notion corresponds to the case of mutually commuting projective measurements. In the same way, as for the Bell scenario, we can construct inequalities that are valid for noncontextual hidden variable models (NCHV), but are violated by quantum mechanics. In a sense, the Bell scenario and the noncontextual scenario are similar, with the Bell-CHSH scenario being a special case of the noncontextual scenario, as noncontextuality and locality coincide for the special case of space-like separated observables.
A fundamental difference between these two concepts is that in the noncontextual scenario the phenomenon of *state independent contextuality* occurs, where QM violates a given noncontextuality inequality for every quantum state ϱ, while Bell inequalities are only violated for specific quantum states.

In noncontextual scenarios, we can consider two different setups for the measurements. In the first setup, measurements are performed jointly, i.e. sets of measurements are performed at the same time. The second method is to consider noncontextuality in sequential measurements, i.e. the measurements are performed one after another. The second method is the one preferred in experiments since it is easier to measure observables one after another instead of measuring them at the same time. In this thesis, we will focus on the case of sequential measurements. For measurements sequences, the probabilities are allowed to have correlations that are signalling in time in one direction (past to future), we refer to them as *arrow of time* (AoT) constraints [8]. The polytope, that describes the range of allowed probability distributions for the signalling in time constraints is called the *temporal correlation polytope*. Under this basic framework, we will consider two different problems.
The first problem is the question whether QM is able to reach the extremal points of the temporal correlation polytope if we constrain the dimension of the quantum system. In the case of sequential measurements, we have the AoT constraints. AoT basically means that the first measurement in the sequence can influence the measurement outcomes of the later measurements but not the other way around. The AoT constraints define a polytope and it is well known that quantum mechanics can reach all extremal points of this polytope if we do not bound the dimension of the system or the type of measure-

ments that are allowed to be performed [9, 10]. The same holds for the case of infinite dimensions and projective measurements, as shown by Budroni it al. in [11]. In this thesis, we show that for the case of arbitrary general measurement and bounded dimension, QM cannot reach all extremal points. In particular, we find that for a given sequence of length l some extremal points can only be reached by systems of minimal dimension d. This allows us to use extremal points to construct dimension witnesses. A dimension witness is used to give a lower bound on the dimension of a given quantum system [12]. Dimension witnesses are important in the framework of quantum information theory, where for many quantum information protocols systems of some specific dimension are needed, e.g. to ensure security in QKD protocols [13, 14, 15].
The second problem we consider is to investigate the effect of two proposed correction terms that aim to close the *compatibility loophole*. To do so, we need the notion of temporal correlation polytope from the first problem. The compatibility loophole allow for a classical explanation of a contextuality experiment, based on the imperfect compatibility of measurements. In experiments the assumption of perfectly compatible observables is never fulfilled if we consider measurements performed on one quantum system. Due to errors in the measurement processes, the best we can achieve are nearly compatible observables. These incompatible measurements, that are by definition contextual, as they disturb one another, may lead to a violation of some noncontextuality inequality that would not be violated by compatible observables. A doubter of quantum contextuality can therefore use the compatibility loophole to argue that quantum mechanics is noncontextual and that contextuality only results from incompatible measurements. There are attempts to close the compatibility loophole, detecting the errors from incompatible measurements by means of correction terms. In this thesis, we will have a look at two proposed correction terms, one by Gühne et al. [16] and one by Kujala et al. [17]. We find that the correction terms by Gühne et al. work for projective measurements and qubit systems and do not work (i.e. do not recover quantum bound for compatible measurements) for general measurements and finite dimensional systems of some minimal dimension d, while the correction terms by Kujala et al. do not work in either case.

Let us at this point give a short overview of the structure of the thesis. In chapter 1, we start with the mathematical preliminaries and basic mathematical concepts that are needed in the context of this thesis. In particular, we will have a look at the differ-

ences between projective and general measurements, and review the theory of quantum channels and instruments to describe post-measured states. In chapter 2, we discuss quantum and post-quantum theories. We will have investigate the EPR argument and its implications. We explain the concepts behind LHV theories and have a look at Bell's theorem and the Bell-inequality derived by Clauer-Horne-Shimony-Holt (CHSH) [18]. In the context of the CHSH inequality, we will also have a look at the quantum bound for this theory, the Tsirelson bound [19] and compare the results for QM with post-quantum theories, the Popescu and Rohrlich (PR) boxes [20]. Furthermore, we discuss the Kochen-Specker theorem and examine the different types of contextuality, state independent and state dependent contextuality. We also present examples for noncontextuality inequalities for both cases. Chaper 3 is again a technical chapter, where we define mathematical concepts that were used in the two main chapters of the thesis. For instance, we present the definition of a polytope/polyhedron. Additionally, we define two maximization techniques, namely *linear programming* and *quadratic programming* and explain the term dual program that exists for every maximization program and hence also for a linear and a quadratic program. In chapter 4, we have a look at the polytope defined by the AoT constraints. We determine the form and number of the extremal points of this polytope for arbitrary setups (i.e. arbitrary length of the measurements sequence, number of measurement settings per time step and number of outcomes per measurement). We further show that quantum systems with bounded dimension cannot in general reach all extremal points of the temporal correlation polytope. We use this result to construct a dimension witness. In chapter 5, using the notion of the temporal correlation polytope we will discuss the correction terms proposed by Gühne et al. and Kujala et al. and test their ability to close the compatibility loophole. We test the correction terms for simple scenarios, i.e. one qubit with projective measurements, then in the general framework of temporal correlations defined by the temporal correlation polytope. Finally in chapter 6, we will present a summary of this thesis, together with some concluding remarks and discuss some open problems that give rise to further research.

1 Mathematical Preliminaries

In this chapter, we will introduce the basic mathematical concepts that are used in the thesis. In the following, only finite dimensional systems, as well as measurements with finite numbers of outcomes are considered. For the definitions in the infinite dimensional case, the reference [21] is recommended.

1.1 The State Space

In quantum mechanics, pure states are generally described by vectors $|v\rangle$ a Hilbert space $\mathcal{H}$. Pure states, however are not sufficient to describe all possible states in quantum mechanics. Consider for instance a machine M that prepares the pure states $|v_i\rangle$ with probability p_i. Such preparation schemes can be described by so called *density matrices*, which are the most general description of quantum states. A precise mathematical definition of the set of density operators is given in definition 1.

Definition 1. *The convex set of states, also called density operators, is defined in the following way.*

$$\mathcal{S}(\mathcal{H}) := \{\varrho \in \mathcal{O}(\mathcal{H}) \,|\, \varrho \geq 0, \mathrm{tr}[\varrho] = 1\} \tag{1.1.1}$$

In the definition above, $\mathcal{O}(\mathcal{H})$ marks the set of operators (matrices) on the Hilbert space $\mathcal{H}$. The subset of the pure states, are the extremal points of the convex set of states (i.e. they cannot be written as a convex combination of other states) and fulfill additionally $\mathrm{tr}\left[\varrho^2\right] = \mathrm{tr}[\varrho] = 1$. Any element of $\mathcal{S}(\mathcal{H})$, which is not an extreme element, is called a *mixed state*[1].
For pure states we can further write down useful properties, which will be discussed in Lemma 2.

[1] Note that the decomposition of the mixed states into pure states is not unique.

Lemma 2. [21] *For any state $\varrho \in \mathcal{S}(\mathcal{H})$, the following conditions are equivalent:*

(*i*) *ϱ is a pure state,*

(*ii*) *$\varrho = |\psi\rangle\langle\psi|$ is a one-dimensional projection*[2]*,*

(*iii*) $\mathrm{tr}\left[\varrho^2\right] = 1$.

1.2 Bi- and Multipartite Systems

In quantum theory it is often not sufficient to consider single systems. If the system that we want to describe is a composed system (e.g. two qubits) it is often useful to describe the complete system as a multipartite system of several individual systems instead of one larger system. E.g. describe two separate spin-$\frac{1}{2}$-particles as two qubits and not as one four level system. Mathematically this process is described by the tensor product and we describe the joint Hilbert space as follows.
Consider a system of n subsystems $A, B, ...L$ with the Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B, ...\mathcal{H}_L$. Let $|a_i\rangle$ $(i = 1, ..., d_A)$ be a basis of $\mathcal{H}_A$ and analogously for the other systems. The Hilbert space of the complete system is then

$$\mathcal{H}_{Ges} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes ... \otimes \mathcal{H}_L \tag{1.2.1}$$

$$= \{\sum_{ij...k} \lambda_{ij...k} |a_i\rangle \otimes |b_j\rangle \otimes ... \otimes |l_k\rangle\} \quad \text{for} \quad \lambda_{ij...k} \in \mathbb{C}. \tag{1.2.2}$$

Let us have a look at this construction for the two party case. In this case, the basis vectors are given by $|e_{ij}\rangle = |a_i b_j\rangle = |a_i\rangle \otimes |b_j\rangle$ for $i \in \{1, ..., d_A\}$ and $j \in \{1, ...d_B\}$. According to equation 1.2.1, a general vector of the Hilbert space $\mathcal{H}_a \otimes \mathcal{H}_B$ is given by a linear combination of these basis vectors. The summation and of two vectors $|v_1\rangle = \sum_{ij} \lambda_{ij} |a_i\rangle \otimes |b_j\rangle$ and $|v_2\rangle = \sum_{ij} \mu_{ij} |a_i\rangle \otimes |b_j\rangle$ is then given by

$$|v\rangle = c_1 \cdot |v_1\rangle + c_2 \cdot |v_2\rangle = \sum_{ij} (c_1 \cdot \lambda_{ij} + c_2 \cdot \mu_{ij}) |a_i\rangle \otimes |b_j\rangle , \tag{1.2.3}$$

[2] A projection is an operator that fulfils $P^2 = P$

for all $c_1, c_2 \in \mathbb{C}$. We can further show that $|e_{ij}\rangle$ is indeed a basis of the joint Hilbert space. We have

$$\sum_{ij} |e_{ij}\rangle\langle e_{ij}| = \sum_{ij} \left(|a_i\rangle \otimes |b_j\rangle\right) \left(\langle a_i| \otimes \langle b_j|\right) \tag{1.2.4}$$

$$= \sum_{ij} |a_i\rangle\langle a_i| \otimes |b_j\rangle\langle b_j| = \mathbb{1}_{d_A} \otimes \mathbb{1}_{d_B} = \mathbb{1}. \tag{1.2.5}$$

To give a simpler example, let us have a look at the joint Hilbert space of two qubit systems.

Example 3. *Consider two qubit systems with the basis vectors* $|0\rangle$ *and* $|1\rangle$*. The basis of the joint system* $\mathcal{H}_4 = \mathcal{H}_2 \otimes \mathcal{H}_2$ *takes the form*
$|0\rangle \otimes |0\rangle = |00\rangle$, $|01\rangle$, $|10\rangle$ *and* $|11\rangle$*. Which we can also redefine to* $|0\rangle$, $|1\rangle$, $|2\rangle$ *and* $|3\rangle$*.*

Sometimes, we are only interested in the statistics on a partial system. E.g. a system is coupled to an environment and we are only interested in the statistics of the system without taking the environment into account. We can describe such systems by the so called *reduced density matrix*. The reduced density matrix can be calculated by taking the state of the complete system ϱ_{AB} and taking the partial trace over one of the systems.

Definition 4. *Let* $\varrho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ *be a state,* $|\psi_i\rangle$ *be a basis of* $\mathcal{H}_A$ *and* $|\varphi_j\rangle$ *a basis of* $\mathcal{H}_B$*. The reduced density matrix on the system* $\mathcal{H}_A$ *is defined as*

$$\varrho_A = \mathrm{tr}_B\left[\varrho_{AB}\right] = \sum_j \langle\varphi_j|\varrho_{AB}|\varphi_j\rangle\,, \tag{1.2.6}$$

where tr_B *denotes the partial trace over the system* B*.*

Example 5 illustrates the concept of reduced density matrices.

Example 5. *Consider the pure two-qubit state* $|\psi\rangle = |01\rangle$ *and the corresponding density matrix* $\varrho = |01\rangle\langle 01|$*. The reduced density matrix for the system* A *is given by*

$$\varrho_A = \mathrm{tr}_B\left[\varrho\right] = \sum_i |0\rangle\langle 0| \otimes \langle i||1\rangle\langle 1||i\rangle = |0\rangle\langle 0|\,, \tag{1.2.7}$$

the reduced density matrix for the system B *by*

$$\varrho_B = \mathrm{tr}_A[\varrho] = \sum_i \langle i||0\rangle\langle 0||i\rangle \otimes |1\rangle\langle 1| = |1\rangle\langle 1|. \tag{1.2.8}$$

For composite systems we can furthermore distinguish between separable and entangled states. For pure states we have the following simple definition.

Definition 6. *A pure quantum state $|\psi\rangle$ is separable if it can be written as*

$$|\psi\rangle = |\varphi_1\rangle \otimes |\varphi_2\rangle. \tag{1.2.9}$$

Otherwise the state is called entangled.

Example 7 shows some simple examples of separable and mixed states.

Example 7. *Consider the states $|\lambda\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$ and $|\Psi_-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. The state $|\lambda\rangle$ is a separable state, since we have*

$$\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) = \frac{1}{2}((|0\rangle + |1\rangle) \otimes |0\rangle + (|0\rangle + |1\rangle) \otimes |1\rangle) \tag{1.2.10}$$

$$= |x_+\rangle \otimes |x_+\rangle. \tag{1.2.11}$$

The state $|\Psi_-\rangle$ on the other hand is an entangled state. We can prove this by showing that there do not exist two states $|\alpha\rangle$ and $|\beta\rangle$ such that $|\Psi_-\rangle = |\alpha\beta\rangle$.
Let

$$|\alpha\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle, \tag{1.2.12}$$

and

$$|\beta\rangle = \beta_0|0\rangle + \beta_1|1\rangle, \tag{1.2.13}$$

be two-qubit states. The tensor product of the states $|\alpha\rangle$ and $|\beta\rangle$ is given by

$$|\alpha\beta\rangle = \alpha_0\beta_0|00\rangle + \alpha_0\beta_1|01\rangle + \alpha_1\beta_0|10\rangle + \alpha_1\beta_1|11\rangle. \tag{1.2.14}$$

When we now try to set the coefficients α_i and β_i such that $|\Psi_-\rangle = |\alpha\beta\rangle$, we see that $\alpha_1\beta_0 = 1$ and $\alpha_0\beta_1 = -1$. However, this implies that all coefficients are unequal to zero

and hence we cannot find coefficients that allow us to write $|\Psi_-\rangle$ as a product of two qubit states.

The state $|\Psi_-\rangle$ belongs to the set of Bell states (Maximally entangled states)

$$|\Phi_+\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right) \tag{1.2.15}$$

$$|\Phi_-\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle - |11\rangle\right) \tag{1.2.16}$$

$$|\Psi_+\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle + |10\rangle\right) \tag{1.2.17}$$

$$|\Psi_-\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right) \tag{1.2.18}$$

and is called the singlet state.

As we have seen in section 1.1, the set of pure states is not sufficient to describe all preparation procedures in quantum mechanics. To describe all possible states we also need the mixed quantum states. Hence we have to introduce the concept of entanglement for mixed states. For mixed states, we define entanglement in the following way:

Definition 8. *A two party quantum state ϱ is called separable if it can be written as*

$$\varrho = \sum_i p_i \varrho_A^{(i)} \otimes \varrho_B^{(i)}, \tag{1.2.19}$$

where $\sum_i p_i = 1$. Otherwise the state is called entangled.

The definition of separability for mixed states is chosen such that states, where one prepares product states $\varrho_A^{(i)} \otimes \varrho_B^{(i)}$ with probability p_i are considered separable . This also includes the *product states*, i.e. separable states where $p_i = 1$ and $p_j = 0$ if $i \neq j$.

1.3 Measurements in Quantum Mechanics

In the standard lectures on quantum mechanics, measurements are identified as self-adjoint (hermitian) operators. Operators of this form can be written in their spectral decomposition $A = \sum_i \lambda_i |v_i\rangle\langle v_i|$ with the eigenvalues λ_i and the eigenvectors $|v_i\rangle$. The probability of measuring the value λ_i is then simply given as $\mathrm{tr}\left[\varrho\,|v_i\rangle\langle v_i|\right] = \mathrm{tr}\left[\varrho\, P_i\right]$.

Example 9 illustrates the concept of standard quantum measurements for a sharp spin measurement.

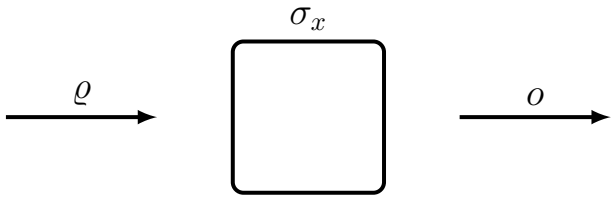

Figure 1.1: *PVM*: Spin measurement in x-direction with the outcome $o \in \{0, 1\}$, described the selfadjoint operator $\sigma_x = S_x(+) - S_x(-)$, with the respective projections $S_x(\pm)$.

Example 9. *Consider for instance the measurement process shown in figure* 1.1*, where a machine measures σ_x. The possible measurement outcomes are ± 1. We can decompose the self-adjoint operator σ_x into a combination of two operators $\sigma_x = S_x(+) - S_x(-)$ where $S_x(\pm)$ are the corresponding effects (see definition* 11*) of the eigenvalues ± 1 respectively. In general, when we describe a measurement, we associate an operator, called effect, to every outcome. The operators/effects can then be used to determine the probability to measure a certain outcome for every state ϱ. In this example, the operators $S_x(\pm)$ can be written in the form $S_x(\pm) = \frac{1}{2}(\mathbb{1} \pm \sigma_x)$ and the probability of obtaining the result ± 1 is given by*

$$p(\pm 1) = \mathrm{tr}\left[S_x(\pm)\,\varrho\right]. \tag{1.3.1}$$

As the operators $S_x(\pm)$ are projections, we call them Projection Valued Measure (PVM) elements of σ_x. We say that the measurement process itself is described by the PVMs.

However, similarly to the case for quantum states, where we find that pure states are not enough to describe all possible quantum states, the set of *PVM*s is not enough to describe all possible measurements. It is straightforward to find some measurements that cannot be described by the *PVM* formalism. One simple example is a spin measurement, where some machine M measures σ_x with a probability λ and σ_y with a probability $1-\lambda$.

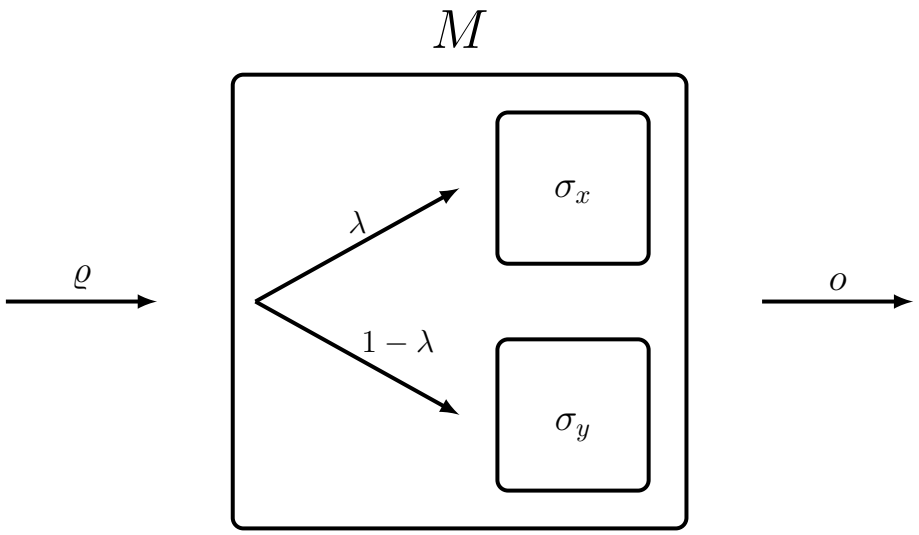

Figure 1.2: A machine M measures either $\sigma_x = S_x(+) - S_x(-)$ with a probability λ or $\sigma_y = S_y(+) - S_y(-)$ with a probability $1-\lambda$ and gives the result $0 \in \{0, 1\}$. The measurement is not a PVM since $M(\pm) = \lambda S_x(\pm) + (1-\lambda) S_y(\pm)$ is not a projection.

Example 10. *Consider the measurement process shown in figure* 1.2. *This time a machine M measures either σ_x with a probability λ or the operator σ_y with a probability $1-\lambda$. The probability of getting the outcome ± 1 is given as*

$$p(\pm 1) = \mathrm{tr}\left[M(\pm)\,\varrho\right], \tag{1.3.2}$$

with

$$M(\pm) = \lambda S_x(\pm) + (1-\lambda) S_y(\pm). \tag{1.3.3}$$

We then have $M^2(\pm) \neq M(\pm)$, meaning that the effects $M(\pm)$ are not projections. Hence the measurement performed by the machine M cannot be described by a PVM.

Example 10 shows us that we need a definition of generalized measurements that reduces to the usual PVM definition if the measurement can be described as a standard quantum measurement. The generalization of measurements is called *Positive Operator Valued Measure* ($POVM$). Instead of all effects being projections, the $POVM$ formalism only

demands the effects to be bounded positive operators ($E \leq \mathbb{1}$).
Since we only consider measurements with a finite, discrete number of outcomes, we can give a simplified definition of a $POVM$, where we only need a finite number of positive operators, called effects, to describe a measurement.

Definition 11. *A $POVM$ is described by sets of operators $\{E_1, ..., E_n\}$, called effects, satisfying the relations*

$$\forall j = 1, ..., n : \ 0 \leq E_j \leq \mathbb{1}, \tag{1.3.4}$$

$$\sum_{j=1}^{n} E_j = \mathbb{1}, \tag{1.3.5}$$

i.e. a $POVM$ is a set of positive matrices summing up to one.

We can now specify the conditions under which a $POVM$ can be described by the PVM formalism of standard quantum mechanics.

Definition 12. *A PVM is a $POVM$, where all effects E_i are projections. The corresponding observable is called a sharp observable.*

Even though PVMs are not the most general description of quantum measurement for a given quantum system, i.e. we can theoretically describe measurements that cannot be described by PVMs, like in example 10, we can always describe a $POVM$ as a PVM in a higher dimensional Hilbert-Space. This is known as the Neumark extension.

Theorem 13. [22, 23] ***Neumark's Theorem:*** *Consider a $POVM$ with the effects $E_i = |e_i\rangle\langle e_i|$ $(i = 1, ..., N)$ that act on $\mathcal{H}$. Then there exists an orthonormal basis $|\psi_i\rangle$ of $\mathcal{H}^A \oplus \mathcal{H}$ such that each $|e_i\rangle$ is a restriction of $|\psi_i\rangle$ to $\mathcal{H}$.*

Proof. [23] Let $E_i = |e_i\rangle\langle e_i|$ be the effects of the $POVM$, $\{|k\rangle, k = 1, ...d\}$ a basis of $\mathcal{H}$ and $\{|l\rangle, l = d+1, ...N\}$ be a basis of $\mathcal{H}^A$. One ansatz for the vectors $|\psi_i\rangle$ is given by

$$|\psi_i\rangle := |e_i\rangle + |f_i\rangle = \sum_k e_i^k |k\rangle + \sum_l f_i^l |l\rangle . \tag{1.3.6}$$

Since the vectors $|\psi_i\rangle$ are chosen to be orthonormal, we have

$$\sum_k \left(e_j^k\right)^* e_i^k + \sum_l \left(f_j^l\right)^* f_i^l = \delta_{ij}. \tag{1.3.7}$$

Due to the normalization of the $POVM$ elements ($\sum_i E_i = \mathbb{1}$), we further know

$$\sum_i \sum_{k,k'} e_i^k \left(e_i^{k'}\right)^* \left|k\right\rangle\!\left\langle k'\right| = \sum_{k,k'} \delta_{k,k'} \left|k\right\rangle\!\left\langle k'\right|, \tag{1.3.8}$$

and hence $\sum_i e_i^k \left(e_i^{k'}\right)^* = \delta_{k,k'}$. Consider now the $N \times N$ -matrix

$$X = \begin{pmatrix} e_1^1 & \dots & e_1^d & f_1^{d+1} & \dots & f_1^N \\ e_2^1 & \dots & e_2^d & f_2^{d+1} & \dots & f_1^N \\ \vdots & & \vdots & \vdots & & \vdots \\ e_N^1 & \dots & e_N^d & f_N^{d+1} & \dots & f_N^N \end{pmatrix}. \tag{1.3.9}$$

Due to the orthonormality condition (equation 1.3.7) we know that the matrix X has to be unitary (i.e. its columns and rows have to be orthogonal). Furthermore, equation 1.3.8 ensures that the first d columns of X are orthogonal. We now have to choose the remaining $N-d$ columns such that the orthogonality relation is fulfilled for every column.

□

1.4 Channels, Stinespring and Choi-Jamiołkowski

Up to now, we have only defined quantum states and measurements. In reality, we have to consider the fact that quantum states my change with time. We therefore have to introduce a notion to describe the time evolution of a quantum state in a way, that both unitary evolutions, as well as post-measured states are covered by this notion. Mathematically, state transformations are described by completely positive maps and channels. The complete positivity is important to ensure that the transformed state is still positive after the transformation.

Definition 14. *A linear mapping $\mathcal{E} : \mathcal{O}(\mathcal{H}) \to \mathcal{O}(\mathcal{H})$ is said to be completely positive if $(\mathcal{E}_A \otimes \mathcal{I}_B)(O) \geq 0$ for all $O \in \mathcal{O}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and for all finite dimensional extensions $\mathcal{H}_B$.*

Now, we can give a definition for the physical reasonable state transformations.

Definition 15. [21] *A mapping $\mathcal{E} : \mathcal{O}(\mathcal{H}) \to \mathcal{O}(\mathcal{H})$ is an operation (a channel) if the following conditions hold:*

(i) $\mathcal{E}$ is linear,

(ii) $\mathcal{E}$ is completely positive,

(iii) $\mathcal{E}$ is trace nonincreasing (trace preserving).

The conditions in definition 15 can be physically justified. The completely positivity of the mapping ensures that a positive operator on the higher dimensional system $\mathcal{H}_A \otimes \mathcal{H}_B$ is still positive after the linear mapping on the system $\mathcal{H}_A$. This is especially important if we look at mappings that transform states. Without completely positivity, it is possible that the state after the mapping is no longer positive, which makes the transformation non-physical. The third condition, accounts for the fact, that an operation can destroy fractions of the system. For this *subnormalized states* (i.e. positive operators with $\mathrm{tr}[\varrho] \leq 1$) the predicted probabilities sum up to a number less than one and describe the loss of systems in an experiment. In other words, the normalization factor $\mathrm{tr}[\varrho] \leq 1$ describes the probability of obtaining the state ϱ after the application of the operation.
To illustrate definition 15, have a look at example 16 that shows that the standard unitary time evolution is a channel in accordance with definition 15.

Example 16. [21] *Consider the unitary time evolution from text-book quantum mechanics, i.e. $\sigma_U(O) := UOU^\dagger$. Let O be a positive operator acting on $\mathcal{H}_A \otimes \mathcal{H}_B$. For every vector $|\varphi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, we have*

$$\langle\varphi|(\sigma_U \otimes \mathcal{I}_B)(O)|\varphi\rangle = \langle\varphi|(U \otimes I)\, O \left(U^\dagger \otimes I\right)|\varphi\rangle = \langle\varphi_U|O|\varphi_U\rangle\,, \tag{1.4.1}$$

with $|\varphi_U\rangle = \left(U^\dagger \otimes I\right)|\varphi\rangle$. Since the operator O is positive, we know that $(\sigma_U \otimes \mathcal{I}_B)(O)$ is a positive operator as well. Additionally, we have $\mathrm{tr}\left[UOU^\dagger\right] = \mathrm{tr}[O]$, and hence we know that $\sigma_U(O)$ is a channel.

The unitary channels, introduced in example 16 play an important role in quantum theory. We know that the time evolution of closed quantum systems is of this form. In general, this statement is not true for single systems (open systems), where we have to consider channels and operations. It is however always possible to describe a quantum channel acting on the Hilbert space $\mathcal{H}$, as an unitary channel acting on some extended

Hilbert space $\mathcal{H}\otimes\mathcal{H}_E$. This specific feature of quantum channels is known as Stinespring's dilation theorem.

Theorem 17. [21] ***Stinespring's Theorem**: Let $\mathcal{E} : \mathcal{O}(\mathcal{H}) \to \mathcal{O}(\mathcal{H})$ be a quantum channel. We then can always find an Hilbert space extension $\mathcal{H}_E$, some unitary operator U and an initial state of the extension $\zeta \in \mathcal{S}(\mathcal{H}_E)$ such that*

$$\mathcal{E}(\varrho) = \mathrm{tr}_E \left[U (\varrho \otimes \zeta) U^\dagger \right] \tag{1.4.2}$$

for all $\varrho \in \mathcal{S}(\mathcal{H})$. The set $\langle \mathcal{H}_E, U, \zeta \rangle$ is called Stinespring dilation of the channel $\mathcal{E}$.

Although definition 15 and theorem 17 can describe the state transformation in quantum mechanics, it is useful to have a description of channels by using operators on the Hilbert space of the system, without using theorem 17 to describe the channel as a unitary channel in a higher dimensional system. Indeed, we can always describe a linear mapping by a set of operators. as we can see in theorem 18.

Theorem 18. [21, 22] ***Kraus Representation**: A linear mapping $\mathcal{E} : \mathcal{O}(\mathcal{H}) \to \mathcal{O}(\mathcal{H})$ is an operation iff there exists a (finite or infinite) set of operators K_i, $i = 1, 2, ...$ such that:*

$$\mathcal{E}(O) = \sum_i K_i O K_i^\dagger \tag{1.4.3}$$

This decomposition has the following properties:

(i) $\mathcal{E}$ is a channel iff $\sum_i K_i^\dagger K_i = \mathbb{1}$,

(ii) $\mathcal{E}$ is unital iff $\sum_i K_i K_i^\dagger = \mathbb{1}$,

(iii) The minimal number of Kraus operators (Kraus rank) is $r \leq d^2$,

(iv) There exists always a representation with r Hilbert-Schmidt orthogonal Kraus operators (I.e. $\mathrm{tr}\left[K_i^\dagger K_j\right] \propto \delta_{ij}$),

(v) Two sets of Kraus operators $\{K_i\}$ and $\{L_j\}$ represent the same mapping iff there exists a unitary matrix U such that $K_i = \sum_j u_{ij} L_j$.

We now know that we can describe (completely positive) mappings either as unitary mappings in higher dimensional Hilbert spaces or by using the Kraus representation.

However, if we want to decide whether a given positive mapping is completely positive, it is reasonable to describe the mapping by the Choi-Jamiołkowski isomorphism, i.e. a complex matrix in higher dimension (see theorems 19 and 20). The first notable result from the Choi-Jamiołkowski isomorphism relates the completely positivity of a mapping to the Choi matrix. This is called Choi's theorem.

Theorem 19. [21] ***Choi's Theorem**: Let $\mathcal{E} : \mathcal{O}(\mathcal{H}_d) \to \mathcal{O}(\mathcal{H}_{d'})$ be a positive linear mapping and $\{|\varphi_i\rangle\}$ be an orthonormal basis of $\mathcal{H}_d$. The statements (i) to (iii) are then equivalent:*

(i) The mapping $\mathcal{E}$ is completely positive,

(ii) $\mathcal{E} \otimes \mathcal{I}_d$ is a positive map for all finite d,

(iii) The matrix

$$C_{\mathcal{E}} = \begin{pmatrix} \mathcal{E}(|\varphi_1\rangle\langle\varphi_1|) & \dots & \mathcal{E}(|\varphi_1\rangle\langle\varphi_d|) \\ \vdots & \ddots & \vdots \\ \mathcal{E}(|\varphi_d\rangle\langle\varphi_1|) & \dots & \mathcal{E}(|\varphi_d\rangle\langle\varphi_d|) \end{pmatrix} \tag{1.4.4}$$

is positive. We call $C_{\mathcal{E}}$ Choi matrix of the mapping $\mathcal{E}$.

Choi's theorem 19 depicts a simple way to test, whether a linear mapping is completely positive. Since Choi's theorem holds for all orthonormal bases of $\mathcal{H}_d$, it is sufficient to show that the Choi matrix is positive for one specific entangled quantum state. We can choose the maximally entangled state

$$P_+ = \frac{1}{d} \sum_{j,k=1}^{d} |\varphi_j\rangle\langle\varphi_k| \otimes |\varphi_j\rangle\langle\varphi_k|, \tag{1.4.5}$$

and calculate

$$C_{\mathcal{E}} = d \cdot (\mathcal{E} \otimes \mathcal{I}_d)(P_+) = \sum_{j,k=1}^{d} \mathcal{E}\left(|\varphi_j\rangle\langle\varphi_k|\right) \otimes |\varphi_j\rangle\langle\varphi_k| \tag{1.4.6}$$

to test whether the resulting matrix C is positive or not.

A second important result is that the state P_+ allows us to define an isomorphism that relates a linear mapping $\mathcal{E} : \mathcal{O}(\mathcal{H}_d) \to \mathcal{O}(\mathcal{H}_{d'})$ to states on a $d \cdot d'$ dimensional (a system composed of a d dimensional system and a d' dimensional system) system.

Theorem 20. [21] ***Choi-Jamiołkowski isomorphism****: Consider a linear map $\mathcal{E} : \mathcal{O}(\mathcal{H}_d) \to \mathcal{O}(\mathcal{H}_{d'})$. Let $|\varphi_i\rangle$ be an orthonormal basis of $\mathcal{H}_d$. The mapping*

$$\mathcal{C} : \mathcal{E} \mapsto \Omega_{\mathcal{E}} := (\mathcal{E} \otimes \mathcal{I})(P_+) \tag{1.4.7}$$

is the Choi-Jamiołkowski isomorphism between the linear map $\mathcal{E}$ on the d dimensional Hilbert space and the operators on the $d' \cdot d$ dimensional Hilbert space. The inverse mapping is defined as

$$\mathcal{C}^{-1} : \Omega \mapsto \mathcal{E}_\Omega[X] = d \operatorname{tr}_2\left[\left(I \otimes X^T\right)\Omega\right], \tag{1.4.8}$$

with tr_2 being the partial trace over the first system. This definition further leads to the following properties:

(i) $\Omega_{\mathcal{E}} = \Omega_{\mathcal{E}}^\dagger$ iff $\mathcal{E}\left(O^\dagger\right) = \mathcal{E}(O)^\dagger$ for all $O \in \mathcal{O}(\mathcal{H}_d)$,

(ii) $\mathcal{E}$ is trace-preserving iff $\operatorname{tr}_1[\Omega_{\mathcal{E}}] = \frac{1}{d}\mathbb{1}_d$.

Up to this point, we have generally defined how linear mappings can be described in quantum mechanics. In the next section, we want to consider the special case of quantum measurements and how to describe the post-measurement state. This will lead us to the notions of *measurement models* and *instruments*.

1.5 Measurement Models and Instruments

If we perform a measurement on a quantum system, the system is coupled to the measurement apparatus for the duration of the measurement. This means that, in principle, a mathematical description of a quantum measurement should include this coupling in its description. To derive the notion of measurement model, we assume that we couple the system and the apparatus, and after some joint evolution, the system is decoupled and the measurement is performed solely on the system of the apparatus. The outcomes of the measurement on the apparatus' system and the quantum system are correlated

due to their coupling before the measurement and hence the measurement result gives us information about the initial and final state of the quantum system. We can give a precise definition of this concept, which is called *measurement model*.

Definition 21. [21] *Let A be an observable on the Hilbert space $\mathcal{H}$. The quadruple $\mathcal{M} = \langle \mathcal{K}, \zeta, \mathcal{V}, F \rangle$ is a measurement model of the observable A, if the probability reproducibility condition*

$$\operatorname{tr}\left[\varrho\, A\left(x\right)\right] = \operatorname{tr}\left[\mathcal{V}\left(\varrho \otimes \zeta\right)\left(I \otimes F\left(x\right)\right)\right] \tag{1.5.1}$$

holds for all outcomes x and $\varrho \in \mathcal{S}\left(\mathcal{H}\right)$ and we have

(i) $\mathcal{K}$ is the Hilbert space associated with the measurement apparatus,

(ii) ζ is the initial state of the apparatus,

(iii) $\mathcal{V}$ is the channel describing the interaction between the system and the apparatus,

(iv) F is the observable measured on the apparatus' system, which is assumed to have the same outcome space as the observable A.

As we see in definition 21, there is no difference in the outcome probabilities if we describe a measurement as a measurement directly performed on the system, or in the sense of a measurement model. The notion of measurement models however gives us a simple way to describe multiple measurements performed on the same quantum system. Let us have a look at a measurement sequence of length two, i.e. we first measure observable A and later measure observable B on the same system. We can then describe the first measurement as a measurement model by the quadruple $\mathcal{M} = \langle \mathcal{K}, \zeta, \mathcal{V}, F \rangle$. The second measurement on the other hand can be described as a measurement performed directly on the system. The joint probabilities for the measurement sequence are then given by

$$p\left(x, y\right) = \operatorname{tr}\left[\mathcal{V}\left(\varrho \otimes \zeta\right)\left(B\left(y\right) \otimes F\left(x\right)\right)\right]. \tag{1.5.2}$$

Formula 1.5.2 takes the easy form

$$p\left(x, y\right) = \operatorname{tr}\left[\mathcal{I}_x^{\mathcal{M}}\left(\varrho\right) B\left(y\right)\right], \tag{1.5.3}$$

with the operator $\mathcal{I}_x^{\mathcal{M}}(\varrho) := \mathrm{tr}_{\mathcal{K}}[\mathcal{V}(\varrho \otimes \zeta)(I \otimes F(x))]$[3]. The mapping $\varrho \mapsto \mathcal{I}_x^{\mathcal{M}}(\varrho)$ is called *instrument* and defined as follows:

Definition 22. [21] *A mapping* $\mathcal{I} : \mathcal{O}(\mathcal{H}) \to \mathcal{O}(\mathcal{H})$ *is called an instrument if it satisfies*

(i) the mapping $\mathcal{I}_x^{\mathcal{M}}$ *is an operation for all* x*,*

(ii) $\mathrm{tr}\left[\mathcal{I}_\Omega^{\mathcal{M}}(\varrho)\right] = 1$ *and* $\mathcal{I}_\emptyset^{\mathcal{M}}(\varrho) = 0 \; \forall \, \varrho \in \mathcal{S}(\mathcal{H})$*,*

(iii) $\mathrm{tr}\left[\mathcal{I}_{\cup_j x_j}^{\mathcal{M}}(\varrho)\right] = \sum_j \mathrm{tr}\left[\mathcal{I}_{x_j}^{\mathcal{M}}(\varrho)\right]$*,*

where Ω *is the set of outcomes for the measurement model* $\mathcal{M}$ *and* x *is an element of the set* $\mathcal{P}(\Omega)$*, where* $\mathcal{P}(\Omega)$ *is the power set of* Ω*.*

As indicated by the indice $\mathcal{M}$ an instrument $\mathcal{I}_x^{\mathcal{M}}$ is uniquely determined by a given measurement model. In the same way, a given instrument $\mathcal{I}$ uniquely determines an observable $A^{\mathcal{I}}$ trough

$$\mathrm{tr}\left[\varrho\, A^{\mathcal{I}}(x)\right] = \mathrm{tr}\left[\mathcal{I}_x(\varrho)\right]. \tag{1.5.4}$$

The reverse statement however is not true. For a fixed observable A, we call an instrument A-compatible if $A = A^{\mathcal{I}}$ but each observable admits infinitely A-compatible instruments as is shown in the following example.

Example 23. [21] *Consider the observable* A *and let* ζ *be some state. A possible* A*-compatible instrument is then given by*

$$\mathcal{I}_x(\varrho) = \mathrm{tr}\left[\varrho\, A(x)\right]\zeta. \tag{1.5.5}$$

It is easy to check that equation 1.5.5 *is indeed an* A*-compatible instrument, since*

$$\mathrm{tr}\left[\mathcal{I}_x(\varrho)\right] = \mathrm{tr}\left[\varrho\, A(x)\right]. \tag{1.5.6}$$

Furthermore, the state ζ *is arbitrary and hence we can conclude that for each observable* A *we can find infinitely many* A*-compatible instruments.*

[3] The instrument is uniquely determined by the given measurement model

We see that a given measurement model uniquely defines an instrument and a given instrument uniquely defines an observable. But a given observable defines a whole equivalence class of instruments and a given instrument defines an equivalence class of measurement models. It is therefore not sufficient to know the $POVM$ elements to describe the post-measurement state which, for a given instrument $\mathcal{I}$, takes the form

$$\tilde{\varrho}_x = \frac{1}{\mathrm{tr}\left[\mathcal{I}_x\left(\varrho\right)\right]}\mathcal{I}_x\left(\varrho\right). \tag{1.5.7}$$

2 Quantum and Post-Quantum Theories

In this chapter, we will have a look at the differences between quantum theory and hidden variable theories. In particular, we will have a look at the Bell scenario, which gives rise to Bell inequalities that can differentiate between quantum theory and local hidden variable theories (LHV). We will also have a look at the concept of noncontextuality, which can be used do differ between QM and noncontextual hidden variable theories (NCHV).

2.1 Hidden Variable Theories

The discussion about the basic structures of quantum mechanics is almost as old as quantum mechanics itself [24] and resulted in the differentiation between the Copenhagen interpretation of quantum mechanics and the De Broiglie-Bohm-Theory.[24, 25] For a more detailed view over the different interpretations of quantum mechanics the "Map of Madness" by A. Cabello can be recommended [26].
In the thirties, Einstein, Podolsky and Rosen (EPR) dissatisfied with the seemingly random measurement results in quantum theory (e.g. measurement of the spin in x-direction of a particle originally oriented with its spin along the z-axis leads to perfectly random measurement results.) decided to investigate the question whether quantum mechanics is a complete theory or not. The work of EPR resulted in their famous argument for the noncompleteness of the wave function and hence quantum theory from the year 1935 [1].

To complete the theory in the sense of EPR, one assumes that there exists some variable (λ), which cannot be determined by any measurement, called *hidden variable* which had it been known, would avoid the randomness from quantum mechanics, leaving any measurement result to be deterministic, permitting definite values for all physical quantities for a given state $|\psi\rangle$.
According to such a theory, states $|\psi\rangle$ are described as an average over the hidden variables as we have no perfect control over the hidden variable λ. For each given λ occurring

in the average, we can assign predetermined values for the physical quantities and the average describes the relative frequencies for the appearances of a specific state hidden variable λ under a preparation procedure.
Afterwards it took almost 30 years until the Bell (1964) [2] and the Kochen-Specker-theorem (1967) [7] ruled out the possibility of completing quantum mechanics in a local hidden variable theory and noncontextual hidden variable theory respectively.
However, even another 40 years after the Bell- and the Kochen-Specker theorem, the question, in what sense quantum physics is different from classical probability theories remains an important research topic in modern physics. Modern research topics include e.g. finding new nonlocality/noncontextuality inequalities [27], generalizing the concepts of nonlocality and noncontextuality to $POVM$s [28, 29, 30] and investigation of these concepts under from the point of view of sequential measurements [31, 32].

For the comprehensibleness of the thesis, let us first have a closer look at the EPR argument, which was the starting point of hidden variable theories.
Whereas the original argument was derived using the operators $\hat{x}$ and $\hat{p}$, (i.e. the position and the momentum operator), in this thesis a simplified version derived by D. Bohm will be presented [33], which instead uses the Pauli spin operators σ_x and σ_z.

2.2 EPR Argument

According to Einstein, Podoslky and Rosen, in a complete theory any element of physical reality needs to have a counterpart in the physical theory. EPR define the term element of physical reality as follows:

Definition 24. *If the value of a physical observable can be determined with certainty (i.e. $p \equiv 1$) without disturbing the system, then this value is an element of physical reality.*

Using their definition for elements of physical reality Einstein, Podolsky and Rosen derive a definition for a complete theory.

Definition 25. *A theory is said to be complete if every element of reality corresponds to an element of the theory.*

With these two definitions, there are only two possibilities for the quantum mechanical wave function $|\psi\rangle$

(i) The wave function $|\psi\rangle$ is not a complete description of the system,

(ii) The wave function $|\psi\rangle$ assigns a fixed value for all elements of physical reality.

2.2.1 EPR Argument for Spin-1/2 Particles

Consider two spin-1/2 particles int the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right) = \frac{1}{\sqrt{2}}\left(\left|x^+x^-\right\rangle - \left|x^-x^+\right\rangle\right), \tag{2.2.1}$$

and the observables

$$A_1 = \sigma_z \otimes \sigma_z \quad A_2 = \quad \sigma_x \otimes \sigma_x.$$

It is obvious that a measurement of the form $\sigma_z \otimes \mathbb{1}$ or $\sigma_x \otimes \mathbb{1}$ leaves the second qubit in an eigenstate of σ_z or σ_x respectively. Since the same argument also works for the opposite measurement order (i.e first measure on the second qubit), all four measurements σ_z^1,σ_x^1,σ_z^2 and σ_x^2 are elements of the physical reality in the sense of EPR. First, their values can be predicted with certainty depending on the basis of the first performed measurement. Second, the first measurement cannot disturb the system as the measurements are performed on space-like separated particles. According to EPR this implies that the measurement results for all four measurements should be known beforehand.
On the other hand, the operators σ_z and σ_x do not commute, meaning that in quantum theory it is impossible to construct a wave function that is an eigenstate of all four operators simultaneously. Quantum mechanics thus fails to assign a value for all elements of reality. The consequence of this is that the wave function $|\psi\rangle$ cannot assign a value for all elements of reality. It follows that assumption (ii) is false and hence assumption (i) is true (i.e. the wave function is not a complete description of the system). EPR assume the idea that quantum theory can be completed by introducing additional hidden variables, leading to assignments for all elements of physical reality.

2.2.2 Consequences of the EPR Argument

As a consequence of the argument given above (see section 2.2.1), there were attempts to introduce hidden variables λ to create a complete theory in the sense of EPR. A completion of quantum theory in this way can be described from the mathematical point of

view via means of an classical probability theory.
A classical probability theory is defined by Kolmogorov's axioms [34] and described by a so called probability space (Λ, Σ, μ), where Λ is the set of hidden variables, Σ is a sigma algebra on the set Λ and μ a normalized measure on Σ.
There have been various attempts to embed quantum theory in such a hidden variable (classic probability) theory [35, 25], as well as many impossibility proofs [36, 7, 37, 38, 39]. As the attempts for hidden variable models try to explain different scenarios, wee distinguish between two different models, namely

(i) Local hidden variable models (LHV),

(ii) Noncontextual hidden variable models (NCHV).

In the following the concepts of local and noncontextual hidden variable theories shall be explained, starting with local hidden variable models.

2.3 Local Hidden Variable Models and Bells Theorem

The aim of local hidden variable theories is to describe the statistics of measurements, performed on separate systems. A special case occurs, if the two systems are space-like separated. In such a scenario, a measurement being performed one one system cannot have any causal influence on the measurement result of the other system, i.e., the probabilities for outcomes of measurements on separated systems have to be statistically independent, once the hidden variable is fixed.

2.3.1 Construction of a Local Hidden Variable Model

To introduce the concept of a local hidden variable theory, contemplate the following simple setting. Two experimenters, Alice and Bob, are spaced far apart and perform space-like separated measurements on their respective system. Each of them can choose between two different measurements (A_i, B_j $i, j \in \{0, 1\}$) with the outcomes $a, b \in \{-1, 1\}$.
A LHV theory describes the conditional probabilities $p(ab|ij)$, i.e. the probability of getting the outcomes a and b, if Alice performs measurement i and Bob measurement j, as a probability distribution of the form

$$p\left(ab|ij\right) = \int_{\Lambda} p\left(\lambda\right) p\left(a|i,\lambda\right) p\left(b|j,\lambda\right) d\lambda. \tag{2.3.1}$$

In this kind of theory, the probabilities $p\left(a|i,\lambda\right)$ and $p\left(b|j,\lambda\right)$ can be chosen as deterministic functions of the hidden variables λ, since whenever $p\left(a|i,\lambda\right)$ is nondeterministic, (i.e. $p\left(a|i,\lambda\right) \in [0,1]$) it is possible to introduce additional hidden variables λ' such that

$$p\left(a|i,\lambda\right) = \int d\lambda' p'\left(a|i,\lambda,\lambda'\right) \tag{2.3.2}$$

with probabilities $p'\left(a|i,\lambda,\lambda'\right) \in \{0,1\}$ (i.e. deterministic). Moreover, the conditional probability $p\left(ab|ij,\lambda\right)$ factorizes for fixed values of λ, implying the independency of Alices and Bobs measurements.

An additional feature of a LHV theory is the independence of $p\left(\lambda\right)$ from the chosen measurements (i,j). This feature is usually called *free will* and means that the experimenters Alice and Bob can choose their measurements independent of the hidden variable λ.

The assumptions made by such a local hidden variable theory can be summarized as follows:
(i) Realism: Observables have predetermined values, whether they are measured or not. The values are simply revealed by the measurement. In equation 2.3.1 the values are determined once the hidden variable λ is fixed.
(ii) Locality: Since the speed of light is the maximum velocity with which information can propagate, two space-like separated events cannot have a causal influence on each other. Therefore equation 2.3.1 ensures the statistical independence of the probabilities for distinct system, once the variable λ is specified.
(iii) Free will: The experimenter can choose his measurement setting "freely", i.e. independent of the source of randomness of the system preparation. This assumption is realized in equation 2.3.1 in the form that the probability distribution over the hidden variables $p\left(\lambda\right)$ is independent of the measurement basis chosen by A and B.

2.3.2 CHSH Inequality

The notation for a hidden variable theory presented in equation 2.3.1 endows us to calculate different bounds for correlations of such an hidden variable theory. The most well-known bound is given by the CHSH inequality (Clauser-Horne-Shimony-Holt) [18]. In this scenario, we consider four measurement A_0, A_1, B_0, B_1 with outcomes $\{-1, 1\}$, where A_i denote the measurement settings of Alice and B_j the measurement settings of Bob. The CHSH inequality then reads as

$$\langle \mathcal{B} \rangle = \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \overset{\text{LHV}}{\leq} 2. \tag{2.3.3}$$

The bound of an hidden variable theory can be proven as follows:

Proof. First, define $f_{A_i} = p\,(+1|i,\lambda) - p\,(-1|i,\lambda)$ as the expectation value for the measurements A_i. The expectation values for the measurements B_j are defined analogously. The functions f_{A_i} and f_{B_j} are deterministic functions of the hidden variable λ, i.e., once λ is fixed, the functions have either the value $+1$ or the value -1. With this functions it is possible to rewrite equation 2.3.3 as

$$\begin{aligned}
\langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle &= \int_\Lambda p\,(\lambda)\, f_{A_0}\,(\lambda)\, f_{B_0}\,(\lambda)\, d\lambda + \int_\Lambda p\,(\lambda)\, f_{A_0}\,(\lambda)\, f_{B_1}\,(\lambda)\, d\lambda \\
&+ \int_\Lambda p\,(\lambda)\, f_{A_1}\,(\lambda)\, f_{B_0}\,(\lambda)\, d\lambda - \int_\Lambda p\,(\lambda)\, f_{A_1}\,(\lambda)\, f_{B_1}\,(\lambda)\, d\lambda \\
&= \int_\Lambda p\,(\lambda)\, [f_{A_0}\,(\lambda)\, f_{B_0}\,(\lambda) + f_{A_0}\,(\lambda)\, f_{B_1}\,(\lambda) \\
&+ f_{A_1}\,(\lambda)\, f_{B_0}\,(\lambda) - f_{A_1}\,(\lambda)\, f_{B_1}\,(\lambda)] d\lambda \\
&\leq \int_\Lambda p\,(\lambda) \max_\lambda [f_{A_0}\,(f_{B_0}\,(\lambda) + f_{B_1}\,(\lambda)) \\
&+ f_{A_1}\,(\lambda)\,(f_{B_0}\,(\lambda) - f_{B_1}\,(\lambda))] d\lambda = 2.
\end{aligned} \tag{2.3.4}$$

□

The above proven bound for the LHV theory, allows us to state Bell's theorem.

Theorem 26. *(Bell 1964) No local hidden variable theory can reproduce all the predictions of quantum mechanics.* [2]

Proof. As equation 2.3.3 provides an upper bound for any LHV-theory, it is sufficient to show that there exists some quantum state $|\psi\rangle$ that violates this bound.
In the following consider the total anticorrelated Bell-state

$$\left|\Psi^-\right\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right), \tag{2.3.5}$$

where $|0\rangle$ and $|1\rangle$ are the eigenstates of the Pauli-σ_z-operator. If the measurements bases of Alice and Bob are chosen as $A_0 = \sigma_z$, $A_1 = \sigma_x$ and $B_0 = \frac{1}{\sqrt{2}}(\sigma_z + \sigma_x)$, $B_1 = \frac{1}{\sqrt{2}}(\sigma_z - \sigma_x)$ respectively the expectation value for the Bell operator $\mathcal{B}$ is given as

$$\langle A_0 B_0\rangle + \langle A_0 B_1\rangle + \langle A_1 B_0\rangle - \langle A_1 B_1\rangle = 2\sqrt{2} > 2. \tag{2.3.6}$$

□

The value $2\sqrt{2}$ is actually the maximal violation quantum mechanics can achieve for the measurement operator $\mathcal{B}$. The bound is named Tsirelson bound after B.S. Tsirelson. In section 2.3.3 we will have a closer look at its derivation and implications.

2.3.3 The Tsirelson Bound

We see that we can violate the CHSH inequality up to the value of $2\sqrt{2}$ using two qubit systems. However, this does not necessarily mean that the bound is independent of the dimension of the quantum system. In the year 1980, this question was finally answered by Tsirelson [19], who proved that the value $2\sqrt{2}$ was actually the maximal value achievable by quantum mechanics independently of the dimension or the type of measurements. Here we will present a simple version of the argument, first presented by Landau [40].

Proof. We can assume the observable A_i and B_j to be projective measurements with the outcomes ± 1. (According to theorem 13, we can always describe a $POVM$ as a PVM on higher dimensional systems.) We then have $A_i^2 = B_j^2 = \mathbb{1}$. If we now take the square of the Bell operator $\mathcal{B}$, we end up with

$$\mathcal{B}^2 = 4\mathbb{1} - [A_0, A_1] \otimes [B_0, B_1]. \tag{2.3.7}$$

Since A_i and B_j are ± 1 valued observables, we know that $\|A_i\| = \|B_j\| = 1$ and hence

$\|[A_0, A_1]\| = 2$. We can now use the fact that $\|A \otimes B\| = \|A\| \cdot \|B\|$ do derive the bound for the Bell operator

$$\langle \mathcal{B} \rangle^2 \leq \langle \mathcal{B}^2 \rangle \leq \left\| \mathcal{B}^2 \right\| = 8 \tag{2.3.8}$$

$$\Rightarrow \langle \mathcal{B} \rangle \overset{\text{QM}}{\leq} 2\sqrt{2}. \tag{2.3.9}$$

□

2.4 Post-Quantum Theories

We have seen that quantum mechanics violates the CHSH inequality up to the value of $2\sqrt{2}$. We can still ask, though, whether the Tsirelson bound is the maximal value achievable by physical reasonable theories. Popescu and Rohrlich showed that it is indeed possible to reach the algebraic maximum for nonsignalling theories [20].

2.4.1 Nonsignalling Theories

In the CHSH case, we consider space-like separated quantum systems. A reasonable physical restriction for the outcome probabilities is given by the nonsignalling constraints,

$$p\left(a|x\right) = \sum_b p\left(ab|xy\right) = \sum_b p\left(ab|xy'\right) \quad \forall y, y', \tag{2.4.1}$$

$$p\left(b|y\right) = \sum_a p\left(ab|xy\right) = \sum_a p\left(ab|x'y\right) \quad \forall x, x', \tag{2.4.2}$$

$$p\left(ab|xy\right) \geq 0. \tag{2.4.3}$$

Here, equations 3.1.6 and 3.1.7 account for the fact that for space like separated systems the measurement result of A should not depend on the observable B and vice versa. As we will see in section 3.1, these constraints actually describe a polytope, the nonsignalling polytope.

Popescu and Rohrlich found an assignment to the probabilities $p\left(ab|xy\right)$, which fulfills the equations 3.1.6-3.1.8 and reaches the algebraic maximum for the Bell operator. These nonsignaling assignments are called PR boxes [20].

2.4.2 PR Boxes

One simple example of a PR box (nonsignalling box) is given by the assignment $p_{PR}(ab|xy) = \frac{1}{2}\delta_{a\oplus b,xy}$[1]. The probability assignments are displayed in table 2.1.

		A_0		A_1	
		$+1$	-1	$+1$	-1
B_0	$+1$	$\frac{1}{2}$	0	$\frac{1}{2}$	0
	-1	0	$\frac{1}{2}$	0	$\frac{1}{2}$
B_1	$+1$	$\frac{1}{2}$	0	$\frac{1}{2}$	0
	-1	$\frac{1}{2}$	0	$\frac{1}{2}$	0

Table 2.1: Probability assignments in a PR-box

It is easy to check that the probability assignments in table 2.1 do not violate the nonsignalling constraints and indeed, we have $\langle A_i \otimes B_j \rangle = 1$ for all terms in $\mathcal{B}$ up to the last one, where we have $\langle A_1 \otimes B_1 \rangle = -1$, hence we have

$$\langle \mathcal{B} \rangle \overset{\text{LHV}}{\leq} 2 \overset{\text{QM}}{\leq} 2\sqrt{2} \overset{\text{NS}}{\leq} 4. \tag{2.4.4}$$

2.5 Noncontextual Hidden Variable Models

Another branch of hidden variable models is given by the so called noncontextual models. In this kind of models one does not consider two systems, on which space-like separated measurements are performed, as in local hidden variable models. Instead one considers one system and measurements $\{A_i\}$. To describe noncontextuality, we first need to give a proper definition of what we mean, when we call two observables *compatible*.

Definition 27. *Two observables A and B are said to be compatible if the measurement of observable A does not disturb the measurement of observable B and vice versa. ($p(b|B) = p(b|AB)$ for all preparations.)*

In quantum mechanics, this definition amounts to commutativity for projective measurements. For general measurements, it is difficult to give a proper definition of compatible observables. There exist some approaches, (e.g. *joint measurability* or *coexistence* [21])

[1] Here, $a \oplus b$ means the sum of the outcomes a and b mod2

but all of them are not defined in the sense of definition 27, which is the motivation for the discussion of the correction terms in chapter 5.
Further a definition of a *measurement context* is needed to describe noncontextuality.

Definition 28. *A context is a set of compatible observables.*

A noncontextual theory is a theory in which the outcome/description of every measurement $\{A_i\}$ is independent of its context. (I.e. If the observable A_i exists in two different contexts, e.g. $\mathcal{C}_1 = \{A_i, A_j\}$ and $\mathcal{C}_2 = \{A_i, A_k\}$ with $A_i \neq A_j \neq A_k$, the measurement result of observable A_i does not depend whether it is measured in the context $\mathcal{C}_1$ or $\mathcal{C}_2$.) Noncontextual theories can hence be defined as follows.

Definition 29. *A theory is said to be noncontextual if the following conditions hold.*

(*i*) *Realism: Observables have well defined values, whether they are measured or not and only revealed during a measurement process,*

(*ii*) *Noncontextuality: The value of the observable is independent of the chosen measurement context* $\mathcal{C}$*,*

(*iii*) *Free will: The experimenter chooses his experiment freely (i.e. the probability of the choice of the distribution of the hidden variable is independent of the choice of the measurement basis).*

Kochen and Specker (KS) showed that such models are impossible for quantum systems of dimension $d \geq 3$ [7]. The Kochen-Specker theorem is as follows:

Theorem 30. *Consider a quantum system with dimension* $d \geq 3$*. Then it is impossible to assign the values* 0 *or* 1 *to all projectors such that*

(*i*) *In every complete set of projectors, exactly one projector is assigned the value* 1*,*

(*ii*) *The assignment is independent of the complete set of projectors.*

KS assumptions correspond to the restriction of the notion of noncontextuality, given in definition 29 to the case of projective measurements (PVMs) and measurement contexts are described as complete sets of orthonormal vectors $|\varphi_i\rangle$. (I.e. a basis of orthonormal vectors.) The proof of Kochen-Specker for dimension $d = 3$ is quite complicated and needs 117 vectors to show that such an assignment is impossible [7]. For dimension $d = 4$, Cabello et al. derived a proof that only uses 18 different vectors and constructs 9 different bases (contexts), where each vector occurs in two different bases (contexts) [41].

Proof. Choose the following 18 vectors, arranged in 9 different bases in $d = 4$.

$$\begin{array}{ccccccccc} \mathcal{B}_1 & \mathcal{B}_2 & \mathcal{B}_3 & \mathcal{B}_4 & \mathcal{B}_5 & \mathcal{B}_6 & \mathcal{B}_7 & \mathcal{B}_8 & \mathcal{B}_9 \\ 1000 & 001\bar{1} & 1111 & 010\bar{1} & \bar{1}111 & 01\bar{1}0 & 0001 & 0011 & 11\bar{1}\bar{1} \\ 0100 & 1\bar{1}00 & 1\bar{1}1\bar{1} & 1010 & 111\bar{1} & 0110 & 0100 & 1\bar{1}00 & 1\bar{1}1\bar{1} \\ 0011 & 11\bar{1}\bar{1} & 10\bar{1}0 & 11\bar{1}1 & 1001 & 0001 & 10\bar{1}0 & 11\bar{1}1 & 1001 \\ 001\bar{1} & 1111 & 010\bar{1} & \bar{1}111 & 01\bar{1}0 & 1000 & 1010 & 111\bar{1} & 0110 \end{array},$$

with the short hand notation $\bar{1} := -1$. When we try to assign a noncontextual, predetermined outcomes to these vectors/ bases, we first note that, since every $\mathcal{B}_i$ is an orthonormal basis, we have to assign the value $+1$ to exactly one vector of $\mathcal{B}_i$. On the other hand, the bases are constructed such that each vector occurs exactly in two of them. This means that we have to assign an even number of ones to the nine bases, which is a contradiction, that in each of the nine bases exactly one vector is assigned the value one, i.e. uneven number of ones. □

2.5.1 The Peres-Mermin Square

The Peres-Mermin square is an easy example for a KS theorem [42, 43]. It consists of nine different projective measurements $A, a, \alpha, B, b, \beta, C, c, \gamma$ that are arranged in a square.

$$\begin{pmatrix} A & B & C \\ a & b & c \\ \alpha & \beta & \gamma \end{pmatrix} := \begin{pmatrix} \sigma_z \otimes \mathbb{1} & \mathbb{1} \otimes \sigma_z & \sigma_z \otimes \sigma_z \\ \mathbb{1} \otimes \sigma_x & \sigma_x \otimes \mathbb{1} & \sigma_x \otimes \sigma_x \\ \sigma_z \otimes \sigma_x & \sigma_x \otimes \sigma_z & \sigma_y \otimes \sigma_y \end{pmatrix} \tag{2.5.1}$$

The arrangement of the measurements inside the matrix is such that each observable commutes with the other observables in the same row and the same column, implying that observables in the same rows and columns can be measured jointly. In addition, the observables are chosen such that the product of the three observables inside a row, i.e. $ABC, abc, \alpha\beta\gamma$ yields the identity $\mathbb{1}$. The same statement also holds for the columns up to the last column, $Cc\gamma$, which gives $-\mathbb{1}$. This gives us some restrictions for the measurement outcomes. If one for instance denotes the outcomes as $\nu(A), \nu(a) ... \nu(\gamma)$, in a NCHV model these outcomes should satisfy $\nu(A)\nu(B)\nu(C) = 1$, $\nu(C)\nu(c)\nu(\gamma) = -1$ etc. If one wants to describe the measurements in the sense of a noncontextual theory, according to condition (ii) in definition 29, each observable gets assigned a value ± 1

independently of their measurement context.
To show the contradiction, consider the product

$$\mathcal{P} = R_1 R_2 R_3 C_1 C_2 C_3, \tag{2.5.2}$$

i.e. multiplying all rows and columns. If this measurement could be described by a noncontextual theory, $\mathcal{P} = 1$ should hold as each observable occurs exactly twice, once in a row and once in a column and gets assigned a fixed value $\nu(O_i)$ independently whether O_i is measured in a row or a column.

$$\mathcal{P}_{contextual} = \langle R_1\rangle\langle R_2\rangle\langle R_3\rangle\langle C_1\rangle\langle C_2\rangle\langle C_3\rangle = \langle \nu(A)^2 \cdot ... \cdot \nu(\gamma)^2\rangle = 1 \tag{2.5.3}$$

In quantum mechanics however, we know that the product of each row and column is 1, except for the last column, where the product is -1.

$$\mathcal{P}_{quantum} = \langle R_1\rangle\langle R_2\rangle\langle R_3\rangle\langle C_1\rangle\langle C_2\rangle\langle C_3\rangle = \langle -\mathbb{1}\rangle = -1 \tag{2.5.4}$$

Thus quantum mechanics assigns $\mathcal{P} = -1$, leading to a contradiction.

2.5.2 State Dependent Contextuality

While the Kochen-Specker theorem and the Peres-Mermin square show contradictions between quantum theory and noncontextual theories, these kind of contradictions are not experimentally testable, in contrast to the Bell inequalities (like the CHSH inequality) for local hidden variable theories. It is however possible to derive Bell like inequalities for noncontextual theories that can be probed in experiments. For this reason we need to introduce the notion of noncontextuality inequality. Noncontextuality inequalities can show two different types of contextuality. *State dependent contextuality*, where the violation of the NCHV bound depends on the state ϱ and *state independent contextuality*, where quantum mechanics violates the noncontextual bound for every possible state. In this section, we will have a look at one inequality for state dependent contextuality. This inequality was derived by Klyachko, Can, Binicioğlu and Shumovsky (KCBS) [44] and takes the form

$$\langle A_0 A_1\rangle + \langle A_1 A_2\rangle + \langle A_2 A_3\rangle + \langle A_3 A_4\rangle + \langle A_4 A_0\rangle \overset{\text{NCHV}}{\geq} -3, \tag{2.5.5}$$

with two-outcome measurements A_i. Two measurements that occur in an expectation value (e.g. $\langle A_0 A_1 \rangle$) are compatible, i.e. their measurement operators commute. The classical bound -3 can be proven by testing all possible noncontextual ± 1-assignments for the observables A_i.

The violation of equation 2.5.5 is state dependent and reaches a maximal violation of $5 - 4\sqrt{5} \approx -3.94$ for the state $|\psi\rangle = (1, 0, 0,)^T$ and measurement settings $A_i = 2\,|v_i\rangle\langle v_i| - \mathbb{1}$, with $|v_i\rangle = (\cos\vartheta, \sin\vartheta \cos\left(j\pi\frac{4}{5}\right), \sin\vartheta \sin\left(j\pi\frac{4}{5}\right))^T$, with

$$\cos^2\vartheta = \frac{\cos\left(\frac{\pi}{5}\right)}{\left(1 + \cos\left(\frac{\pi}{5}\right)\right)}. \tag{2.5.6}$$

2.5.3 State Independent Contextuality

Let us now have a look noncontextuality inequalities, that show state independent contextuality. We see that each set of projectors not admitting a noncontextual truth-assignment and hence leading to a proof of the Kochen-Specker Theorem, also provides a violation of a specific noncontextual inequality for any quantum state [45], [46], [47]. This phenomenon was named *state independent contextuality* (SIC).
One simple example of SIC can be constructed out of the Peres-Mermin square [48]. Consider for instance the expression

$$\langle \chi_{PM} \rangle = \langle ABC \rangle + \langle abc \rangle + \langle \alpha\beta\gamma \rangle + \langle Aa\alpha \rangle + \langle Bb\beta \rangle - \langle Cc\gamma \rangle. \tag{2.5.7}$$

Here the measurements in each of the six sequences are compatible. For NCHV theories, the expectation value is upper bounded by $\langle \chi_{PM} \rangle \leq 4$, as can be easily seen by trying all 2^9 noncontextual assignments to the observables. For a four dimensional quantum system it is possible to use the actual observables from the PM square, given in section 2.5.1. These observables lead to a state independent violation of $\langle \chi_{PM} \rangle = 6$.
Another example for SIC was given by Yu and Oh [49]. To write down this SIC inequality, one considers the following 13 vectors in $\mathbb{C}^3$, given in table 2.2.

$v_1 = (1,0,0)$	$v_5 = (1,0,-1)$	$v_A = (-1,1,1)$
$v_2 = (0,1,0)$	$v_6 = (1,-1,0)$	$v_B = (1,-1,1)$
$v_3 = (0,0,1)$	$v_7 = (0,1,1)$	$v_C = (1,1,-1)$
$v_4 = (0,1,-1)$	$v_8 = (1,0,1)$	$v_D = (1,1,1)$
	$v_9 = (1,1,0)$	

Table 2.2: The 13 vectors of the Yu-Oh graph

For each of the vectors v_i one writes a ± 1 valued observable $A_i = 2\,|v_i\rangle\langle v_i| - \mathbb{1}$. The compatibility relations (orthogonal vectors v_i correspond to compatible observables A_i) between the variables A_i are summarized in figure 2.1.

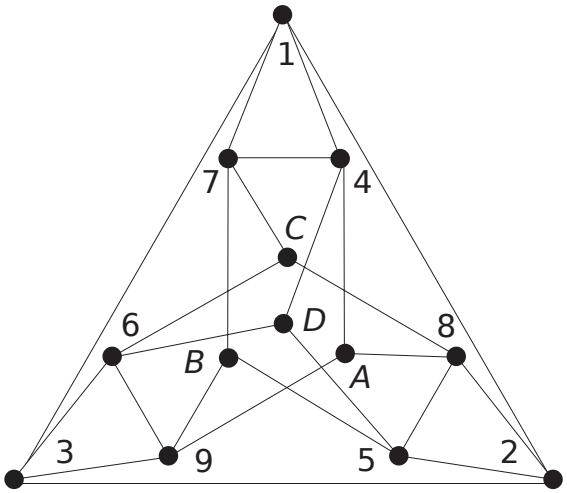

Figure 2.1: [50] Orthogonality relation in the Yu-Oh graph

In figure 2.1, two measurements A_i and A_j are connected by an edge ij if they are compatible (i.e. commute). This permits us to write down the following NC inequality

$$L_{NCHV} = \langle a_i \rangle - \frac{1}{2} \sum_{edges} \langle a_i a_j \rangle \leq 8. \tag{2.5.8}$$

In the equality given above, the a_i are classical noncontextual random variables. In a NCHV theory, the a_i are determined by the hidden variables beforehand and do not depend on the fact whether they are measured alone or together with other compatible measurements. The NCHV bound 8 can then simply be determined by trying all 2^{13} possible noncontextual assignments for a_i.

In the quantum mechanical case, it is possible to calculate an operator L for inequality

2.5.8. Using the explicit vectors given in table 2.2 one gets

$$L_{QM} = \sum_i A_i - \frac{1}{2} \sum_{edges} A_i A_j = \frac{25}{3} \mathbb{1}. \tag{2.5.9}$$

It follows that quantum mechanics violates inequality 2.5.8 state independently, as

$$\langle L \rangle_\varrho = \frac{25}{3} > 8 \tag{2.5.10}$$

holds for every state ϱ.

2.5.4 Experimental Tests of Noncontextuality

The Kochen-Spekker Theorem tells us that quantum theory is contextual for quantum systems of dimension $d \geq 3$. Since contextuality is an intrinsic element of quantum theory, we want to perform experiments to test noncontextuality inequalities, like the Peres-Mermin inequality. Recently, a number of experiments were performed to test noncontextuality inequalities [51, 48, 52]. In [53], an experimental violation of the noncontextuality inequalities for the n-cycle scenario[2] have been observed for even n up to the value of 14. Another experiment tested the cycle with $n = 5$ (KCBS inequality) using superconducting qutrits [54]. However, even though the experimental results seem to rule out noncontextual theories, there exists some loopholes, similarly as for Bell inequality tests. The loophole with the highest impact is the compatibility loophole. The compatibility loophole comes from the fact that all noncontextuality inequalities assume the measurements that are measured together to be perfectly compatible. In experiments this can never be achieved. The best we can reach with experiments is to have nearly perfectly compatible measurements, which means that every noncontextuality test suffers from this loophole. In experiments the attempt to close the loophole is often to slightly alter the inequality to consider the fact that the measurements are not perfectly compatible [16, 17, 54]. In chapter 5, we will have a closer look at two suggested correction terms and investigate whether they are able to recover the bound of the original bound if the measurements are not perfectly compatible.

[2] The n-cycle scenario describes a family of noncontextuality inequalities. E.g. $n = 4$ equals the CHSH inequality for the noncontextuality scenario

2.6 Dimension Witnesses

For many quantum information tasks, knowledge about the dimensions of quantum systems is essential. One example is given by quantum key distribution (QKD). In the simplest protocols, the correlations between the parties A and B are supposed to come from measurements of quantum states of a certain dimension d [13, 14]. Actually, the assumption of a well known d dimensional system is the basis for most security proofs of QKD protocols [15]. For practical implementation of QKD protocols, we have to be able to give a lower bound for the dimension of the quantum system used for the protocol. If this bound of the quantum system is lower than the required dimension for the protocol, we know that we cannot implement the protocol securely. As for the Bell-CHSH scenario, we typically write down an operator, whose expectation value can be measured in experiments. We call an operator W, whose expectation value gives rise to a lower bound for the dimension of the quantum system, a *dimension witness*. Therefore, a dimension witness is unable to determine the real dimension of a quantum system, instead it ensures us that the quantum system has at least dimension d_b.

Due to its importance for quantum information science, the field of dimension witnesses, in particular to find new witnesses and to optimize bounds, have been an active research area in the last few years. In [12], Brunner et al. gave two examples of dimension witnesses that can distinguish between systems of dimension $d = 2$ and $d \geq 3$ for Bell type scenarios. In this case, the dimension witness may be able to test whether an experimenter is able to entangle a certain amount of quantum levels. Another attempt is to give a lower bound for a single quantum system (e.g. the number of energy levels of an ion the experimenter is able to manipulate). To get a lower bound on the dimension of a single quantum system, the system is normally measured sequentially, which is the type of measurement used for noncontextuality proofs. In [55] Gühne et al. showed that noncontextuality inequalities[3] can be used as state independent dimension witnesses. Furthermore in [56, 57, 58, 59] dimension witnesses for arbitrary dimensional systems, which are in addition able to distinguish between classical and quantum systems of the same dimension were proposed.

[3] In particular, they considered the KCBS inequality and the Peres-Mermin inequality

3 Polytopes, Linear and Quadratic Programming

In this chapter, we will introduce polytopes to describe range of allowed probability distributions for different kinds of correlations in theories. E.g. the nonsignalling correlations define a polytope, that describes the set of probability distributions that are non-signalling (do not violate special relativity for space-like separated systems). The second part of the chapter deals with numerical optimization techniques. In our case, we are especially interested in the linear and quadratic optimization techniques, which were used to maximize some functions in section 5.4.

3.1 Polytopes and Polyhedrons

Polytopes are a powerful tool to describe extreme correlations in theories. E.g. the nonsignalling assignment that reaches the maximal violation of the CHSH inequality is given by an extremal point of the nonsignalling polytope. A polytope allows us to calculate the extremal correlations, which we then can compare with the result from quantum mechanics. A polytope can be defined in two different ways. In the first way, we already know the extremal points of the polytope and the polytope itself is described as a convex combination of the extremal points. Precisely we have:

Definition 31. *Consider the vectors $v_1, ...v_n \in \mathbb{R}^d$. The polytope spanned by this set of vectors is defined as their convex combination*

$$P_V = \text{conv}\,(v_1, ..., v_n) = \{x \in \mathbb{R}^d | x = \sum_i \lambda_i v_i, \lambda_i \geq 0, \sum_i \lambda_i = 1\}. \tag{3.1.1}$$

Additionally, we need to define a polyhedron as a set of points fulfilling a set of linear inequalities.

Definition 32. *We define a polyhedron in the following way:*

$$P_V = \{x \in \mathbb{R}^d | Ax \leq b\}, \tag{3.1.2}$$

where the inequalities are represented by the matrix A and the vector b. A further constriction for the vectors x is $\max|x| < \infty$ for $x \in P_V$. If the vectors x are not bounded, we call the set a polyhedral cone.

Since both definitions are equivalent [60], we can convert one definition into the other. If we are given the extremal points of the polytope, we can calculate the inequalities and if we are given the inequalities, we can determine the extremal points of the polytope. In most cases, we consider the second calculation. We usually know the linear inequalities (e.g. the nonsignalling inequalities 3.1.6-3.1.8) and want to determine the extremal points (e.g. the assignment that allows the maximal violation of the CHSH inequality).
To get a better grasp of the concept of polytopes, let us now have a look at a simple example.

Example 33. *Let us consider the $\mathbb{R}^2$ with the vectors $x = (p_1, p_2)^T$ and the inequalities*

$$p_1 \geq 0, \tag{3.1.3}$$

$$p_2 \geq 0, \tag{3.1.4}$$

$$p_1 + p_2 = 1. \tag{3.1.5}$$

The inequalities define the polytope shown in figure 3.1. The polytope defined by these constraints is actually a one dimensional line, with the extremal points $(1,0)^T = |v_1\rangle$ and $(0,1)^T = |v_2\rangle$. It is easy to see that alternatively, we can describe the same polytope by defining the extremal points $|v_1\rangle$ and $|v_2\rangle$ instead in the sense of definition 31.

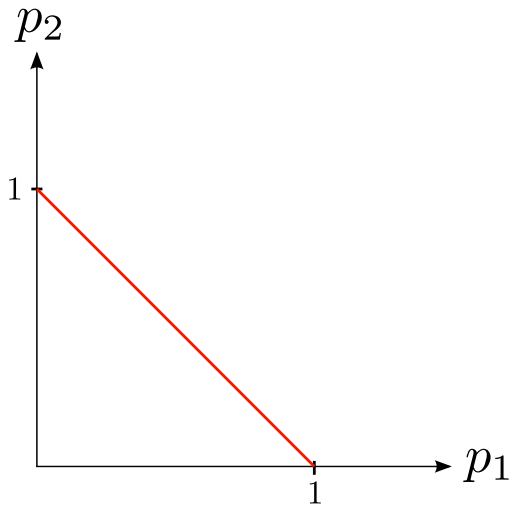

Figure 3.1: The red line marks the one dimensional polytope described by the variables p_1 and p_2.

As a second example, let us have a look at the polytope defined by the nonsignalling constraints.

Example 34. *Consider the nonsignalling case. Here, the vector x consists of the probabilities $p(ab|xy)$, explicitly, we have $x = (p(00|00), ..., p(11|11))^T$. The inequalities are given by the nonsignalling constraints.*

$$p(a|x) = \sum_b p(ab|xy) = \sum_b p\left(ab|xy'\right) \quad \forall y, y', \tag{3.1.6}$$

$$p(b|y) = \sum_a p(ab|xy) = \sum_a p\left(ab|x'y\right) \quad \forall x, x', \tag{3.1.7}$$

$$p(ab|xy) \geq 0. \tag{3.1.8}$$

It is clear that the constraints take the form of a polyhedron/polytope as in definition 32. The extremal points of this polytope can be calculated for example with the `PANDA` [61]. `PANDA` *is a software that was developed by members of the university of Heidelberg (Main developer: Stefan Lörwald), for transforming the descriptions of polyhedra and polytopes. (In our case, we know the matrix A used to describe the polyhedron and want to determine the extremal points of the polytope).*
One extremal point is given by the probability assignment described in section 2.4.2 in the context of PR-boxes. As we have seen there, this assignment is able to reach the algebraic

maximum of the Bell inequality.

3.2 Linear and Quadratic Programming

In this section, we will introduce a technique to maximize/minimize some linear function on a polytope. (I.e. the variables in the linear function are subjected to constraints that describe a polytope). Although there are several different techniques (e.g. linear programming, semi-definite programming etc.), we will only consider linear and quadratic programming as these are the only techniques which are actually used in the later part of the thesis. For more information about other maximization methods, we can advise the reference [60]. To start with, let us give a precise definition of what we mean with *linear programing.*

Definition 35. *A linear program is defined as a minimization of a linear function on a polyhedral set, more precisely,*

$$\text{minimize} \qquad c^T x \tag{3.2.1}$$

$$\text{subject to} \qquad Ax \leq b \tag{3.2.2}$$

Where A is a $m \times m$ real matrix, b is a m-dimensional vector and the inequality sign $\geq$ is intended componentwise.

In the definition above, we call the expression that is to be minimized $c^T x$, *objective function.* The inequalities $Ax \leq b$ specify a convex polytope (see definition 32) over which the objective function shall be optimized.
An important question is whether we can trust the result of a liner optimization. It is possible to show that any feasible solution of a liner program is a valid maximization over the polytope [60]. This is because, for every linear program we can define a dual program, that is feasible iff the original linear program was feasible and the solution of the dual program gives a valid lower bound for the solution of the linear program. If we use a computer algebra program to optimize a given objective function, the program automatically solves the dual program as well. Let us have a look, how to determine the dual problem for a linear program. For more information see [60].

3.2.1 The Lagrange Dual Function

Let us consider an optimization problem of the standard form[1]

$$\begin{aligned} &\text{minimize} && f_0(x) && (3.2.3)\\ &\text{subject to} && f_i(x) \leq 0, \quad i = 1, ..., m && (3.2.4)\\ &\text{subject to} && h_i(x) = 0, \quad i = 1, ...p. && (3.2.5)\\ & && && (3.2.6) \end{aligned}$$

with $x \in \mathbb{R}^n$. The domain $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom} f_i \cap \bigcap_{i=0}^{p} \mathbf{dom} h_i$ is assumed to be non-empty and the optimal solution is denoted by p^*. The Lagrangian of this problem is defined as

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x), \tag{3.2.7}$$

where λ_i is the Lagrange multiplier associated with the ith inequality $f_i(x) \leq 0$ and ν_i as the Lagrange multiplier associated with the ith equality $h_i(x) = 0$.
The *Lagrange dual function*, or simply *dual function* $g(\lambda, \nu)$ is defined as the infimum of the Lagrangian over the vector x. Concretely, we have

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right). \tag{3.2.8}$$

The dual function gives us valid lower bounds for the optimal solution p^*. For any $\lambda \geq 0$ and for all ν we then have

$$g(\lambda, \nu) \leq p^*. \tag{3.2.9}$$

Proof. The proof is straight forward. Suppose that y is a feasible point for the original problem, i.e. $f_i(y) \leq 0$, $h_i(y) = 0$ and consider $\lambda \geq 0$. It follows that

$$\sum_{i=1}^{m} \lambda_i f_i(y) + \sum_{i=1}^{p} \nu_i h_i(y) \leq 0, \tag{3.2.10}$$

and therefore

[1] Note that this is not necessarily a linear program, but we can always rewrite a linear program in this way.

$$L\left(y,\lambda,\nu\right)=f_0\left(y\right)+\sum_{i=1}^{m}\lambda_i f_i(y)+\sum_{i=1}^{p}\nu_i h_i(y)\leq f_0(y). \tag{3.2.11}$$

Since we defined $g(\lambda,\nu)=\min_x L(x,\lambda,\nu)$, we hence have $g(\lambda,\nu)\leq f_0(y)$ for all feasible points y of the original problems, which, applied to the optimal solution, gives us equation 3.2.9. □

Let us now explicate this concept of the Lagrange dual function for the example of a linear program.

Example 36. *Consider the linear program*

$$\begin{aligned} &\text{minimize} && c^T x && (3.2.12)\\ &\text{subject to} && Ax=b && (3.2.13)\\ & && x\geq 0, && (3.2.14) \end{aligned}$$

with the inequality constraints $f_i(x)=-x_i$ for $i=1,...,n$. The Lagrangian of this linear program takes the form

$$L(x,\lambda,\nu)=c^T x-\sum_{i=1}^{n}\lambda_i x_i+\nu^T\left(Ax-b\right)=-b^T\nu+\left(c+A^T\nu-\lambda\right)^T x. \tag{3.2.15}$$

We get the dual function by minimizing over all feasible x,

$$g(\lambda,\nu)=\inf_x L(x,\lambda,\nu)=-b^T\nu+\inf_x\left(c+A^T\nu-\lambda\right)^T x. \tag{3.2.16}$$

The dual function, given in equation [3.2.16]*, can be easily determined analytically. A linear function is bounded below only when it is identically zero. Hence, we have*

$$g(\lambda,\nu)=\begin{cases}-\infty, & \left(c+A^T\nu-\lambda\right)\neq 0\\ -b^T\nu. & \left(c+A^T\nu-\lambda\right)=0\end{cases} \tag{3.2.17}$$

We see, that in this linear program we get a valid lower bound for the optimal solution exactly for the case when we have $\lambda\geq 0$ and $c+A^T\nu-\lambda=0$. All other choices of λ and ν do not result in a reasonable lower bound.

In section 5.4, we will use `PYTHON` to minimize two different functions. One of the functions is linear, while the other is not. Instead, we can rewrite the second function to take

a quadratic form. The first equation can obviously be maximized by a linear program, the second one by a quadratic program. Since `PYTHON` uses a different, but equivalent, definition of a linear program, which is the following:

Definition 37. *A linear program is defined as a minimization of a linear function, more precisely,*

$$\begin{aligned} &\text{minimize} && c^T x && (3.2.18)\\ &\text{subject to} && Gx + s = h && (3.2.19)\\ & && Ax = b && (3.2.20)\\ & && s \succeq 0 && (3.2.21) \end{aligned}$$

Where G and A are $m \times m$ and $n \times m$ real matrices, c, h and b are m, m and n-dimensional vectors and the inequality sign $\geq$ is intended componentwise.

The definition of a quadratic program is somewhat similar to that of a linear program, but the minimization part contains an additional term of the form $x^T P x$. The precise definition for a quadratic program is as follows:

Definition 38. *A quadratic program is defined as a minimization of a quadratic function, under the constraints*

$$\begin{aligned} &\text{minimize} && \frac{1}{2} x^T P x + q^T x\\ &\text{subject to} && Gx \preceq h\\ & && Ax = b \end{aligned}$$

Where P, G and A are $m \times m$, $n \times m$ and $o \times m$ real matrices, while q,h and b are m,n and o-dimensional vectors.

4 Temporal Correlations

In chapter 2, we looked at the Bell scenario, i.e. spatially separated systems on which different measurements are performed. The non-signaling constraints describe a polytope, whose extremal points are able to reach the algebraic maximum of the Bell-CHSH inequality, whereas quantum mechanics cannot. Another important scenario is given by a temporal scenario, where we consider a single quantum system and perform one measurement after the other. The constraints that describe such set-ups are known as the *Arrow of Time* (AoT) constraints [8] and define the temporal correlation polytope. It is known that quantum mechanics is able to reach the extremal points of the polytope if neither the dimension, nor the type of measurements are restricted [62].
In this chapter, we want to characterize the temporal correlations that occur in such systems, under the special case of finite dimension, arbitrary $POVM$s and instruments. To do so, we consider the polytope that is described by the temporal correlation constraints and test whether we can reach all its extremal points for measurement procedures of this form.
The structure of this chapter is as follows: In section 4.1, we describe the AoT constraints. In section 4.2, we show that the extremal points of the temporal correlation polytope for sequences of arbitrary length are given by deterministic assignments and give a formula to calculate the number of extremal points of the polytope for sequences of length n. In section 4.3, we describe the quantum mechanical setup. In section [4.5], we show that the quantum mechanical setup, for two measurement sequences with two settings and two outcomes, cannot reach all extremal points for two-dimensional systems and in section 4.6, we give an explicit example that is able to reach the extremal point for qutrits. In section 4.7, we determine the equivalence classes of the extremal points for the temporal correlation polytope and determine which equivalence classes can be reached with qubit systems and which with qutrit systems. In section 4.8, we define a temporal Bell operator that can be used to distinguish between two- and three-dimensional quantum systems. Finally, in section 4.9 we consider an application of the temporal Bell operator

as a dimension witness.

4.1 Temporal Correlation Constraints

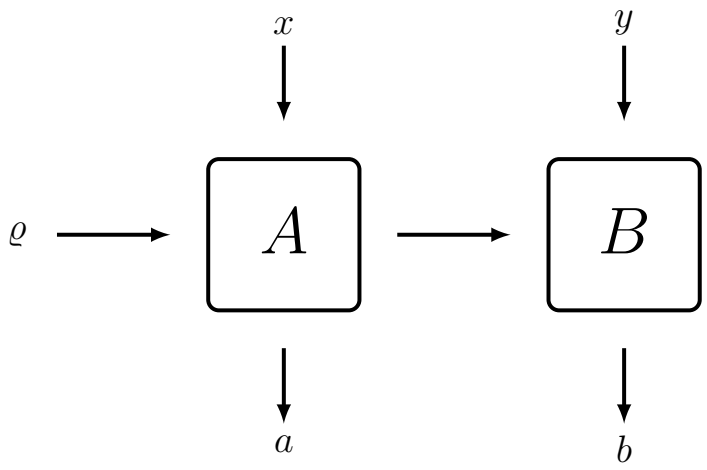

Figure 4.1: Temporal measurement sequence

We consider measurement sequences of the form, shown in figure 4.1. The initial state of the system is ϱ. At the time t_1, the experimenter chooses between one of two possible measurements A_0, A_1, depending on the input x and gets the outcome a. At a later time t_2, the experimenter again chooses between two possible measurements B_0, B_1, depending on the input y and gets the output b. The probabilities describing such measurement sequences take the form $p\,(ab|xy)$, depending on the number of outcomes for the individual measurements. The probabilities further fulfill the AoT constraints, i.e. the first measurement can influence the second measurement, but not the other way around. In the AoT conditions, this is included in the fact that we can define the marginal probability $p(a|x)$, (i.e. the probability to get outcome a if measurement x is performed) for the first measurement, which is independent of the setting y and the outcome b of the second measurement. In mathematical terms, we can write down the AoT conditions for two measurement sequences like,

$$\sum_b p(ab|xy) = \sum_b p(ab|xy') \quad \forall a, x, \tag{4.1.1}$$

$$\sum_{a,b} p\left(ab|xy\right) = 1 \quad \forall x, y, \tag{4.1.2}$$

$$p\left(ab|xy\right) \geq 0 \quad \forall a, b, x, y. \tag{4.1.3}$$

We can use condition 4.1.1 to define the probabilities $p(a|x)$

$$p\left(a|x\right) := \sum_b p\left(ab|xy\right) \quad \forall a, x, y. \tag{4.1.4}$$

It stands out that the arrow of time conditions take a similar form, as the nonsignalling conditions (3.1.6 - 3.1.8. We can also rewrite the conditions such that they take the form given in definition 32. The constraints can be expressed by the matrix A and the vector b such that $Av \leq b$, where the vector v is a vector consisting of the probabilities $p(ab|xy)$. Therefore, the constraints define a polytope via the matrix A and the vector b. We call the polytope defined by the AoT constraints *temporal correlation polytope*. In the following, we will label the polytope as $P_l^{O,S}$, where O is the number of outcomes per measurement setting, S is the number of measurement settings and l is the length of the measurement sequence.

4.2 Extremal Points of the Temporal Correlation Polytope

4.2.1 Sequences of Arbitrary Length

For sequences of length two, the extremal points can be directly determined by using `PANDA` [61]. It turns out that the 64 extremal points are all deterministic assignments (i.e. assignments of probability 0 or 1) that fulfill the arrow of time conditions 4.1.
For sequences of length $l \geq 3$, we could no longer determine the extremal points with `PANDA`. This is because the number of variables and constraints increases drastically for longer sequences. For instance, the number of variables is $16 = 2^4$ for sequences of length two, but $2^6 = 64$ for sequences of length three[1]. Similarly, the number of constraints increases from 24, to 110. However, we show analytically in proposition 41 that even for longer sequences the extremal points of the polytope are given by deterministic assign-

[1]The numbers given here, are for the case of two measurement settings with two outcomes each.

ments, which are compatible with the generalized AoT conditions. To start, let us have a look at the generalized AoT conditions for measurement sequences of length n,

$$p\left(a|x\right) = \sum_{b,c,\ldots} p\left(abc\ldots|xyz\ldots\right) \quad \forall a, x, y, z, \ldots, \tag{4.2.1}$$

$$p\left(ab|xy\right) = \sum_{c,d,\ldots} p\left(abcd\ldots|xyz\ldots\right) \quad \forall a, b, x, y, z, \ldots, \tag{4.2.2}$$

and all the way down to the last measurement in the sequence.
For the probabilities $p\left(abc\ldots|xyz\ldots\right)$, we can write down the following lemma.

Lemma 39. *If $p\left(abc\ldots|xyz\ldots\right)$ fulfills the arrow of time conditions, we can write it as*

$$p\left(abc\ldots|xyz\ldots\right) = p\left(a|x\right) p\left(b|axy\right) p\left(c|abxyz\right) \ldots, \tag{4.2.3}$$

with $p\left(a|x\right)$, $p(b|axy)$, $p(c|abxyz)$ etc. being probability distributions with respect to the variables a, b, c etc.

Proof. We will show the proof of Lemma 39 for the case $n = 3$, it is however straightforward to generalize the proof to sequences of arbitrary length. The marginal probabilities $p(a|x)$ are well defined by the AoT conditions. We just have to introduce two conditional probabilities $p(b|axy)$ and $p(c|abxyz)$. For $p(b|axy)$, we have

$$p\left(b|axy\right) := \frac{p\left(ab|xy\right)}{\sum_b p\left(ab|xy\right)} = \frac{p\left(ab|xy\right)}{p\left(a|x\right)}, \tag{4.2.4}$$

for $p\left(a|x\right) \neq 0$, and

$$p\left(b|axy\right) := 0, \tag{4.2.5}$$

for $p\left(a|x\right) = 0$. It is easy to show that $p(b|axy)$ is a valid probability distribution, if $p(a|x) \neq 0$. $p(b|axy)$ is positive since $p(ab|xy)$ and $p(a|x)$ are positive and we have

$$\sum_b p(b|axy) = \frac{\sum_b p(ab|xy)}{p(a|x)} = 1. \tag{4.2.6}$$

The conditional probability $p(c|abxyz)$ is defined as

$$p(c|abxyz) = \frac{p(abc|xyz)}{p(ab|xy)}, \tag{4.2.7}$$

if $p(ab|xy) \neq 0$ and zero otherwise. In an analogous way as for $p(b|axy)$, we can show that $p(c|abxyz)$ is a valid probability distribution if $p(ab|xy) \neq 0$. We then simply have

$$p\left(abc|xyz\right) = p\left(a|x\right) p\left(b|axy\right) p\left(c|abxyz\right) \tag{4.2.8}$$

$$= p\left(a|x\right) \frac{p\left(ab|xy\right)}{p\left(a|x\right)} \frac{p\left(abc|xyz\right)}{p\left(ab|xy\right)}. \tag{4.2.9}$$

□

The reverse statement is also true, as is shown in lemma 40.

Lemma 40. *Let $p\left(a|x\right), p\left(b|axy\right)$, $p(c|abxyz)$ etc. be consistent conditional probabilities (i.e. if $p(a|x) = 0$ for some a and x, the probabilities $p(b|axy)$ and $p(c|abxyz)$ etc. have to be zero as well). The product of the probabilities*

$$p\left(abc...|xyz...\right) = p\left(a|x\right) p\left(b|axy\right) p\left(c|abxyz\right) ... \tag{4.2.10}$$

fulfils the AoT conditions.

Proof. As before, we will show the proof for sequences of length three and mention that the proof can be generalized for arbitrary sequences. We need to show that the construction in equation 4.2.10 fulfills the AoT conditions. We have

$$\sum_{b,c} p\left(abc|xyz\right) = \sum_{b,c} p\left(a|x\right) p\left(b|axy\right) p\left(c|abxyz\right) \tag{4.2.11}$$

$$= p\left(a|x\right) \sum_{b} p\left(b|axy\right) \sum_{c} p\left(c|abxyz\right) \tag{4.2.12}$$

$$= p\left(a|x\right) \quad \forall a, x, y, z, \tag{4.2.13}$$

and

$$\sum_{c} p\left(abc|xyz\right) = \sum_{c} p\left(a|x\right) p\left(b|axy\right) p\left(c|abxyz\right) \tag{4.2.14}$$

$$= p\left(a|x\right) p\left(b|axy\right) \sum_{c} p(c|abxyz) \tag{4.2.15}$$

$$= p\left(a|x\right) \frac{p\left(ab|xy\right)}{p\left(a|x\right)} = p\left(ab|xy\right) \quad \forall a, b, x, y, z, \tag{4.2.16}$$

hence the probabilities $p\left(abc|xyz\right) = p\left(a|x\right)p\left(b|axy\right)p\left(c|abxyz\right)$, defined by the conditional probabilities fulfill the AoT conditions.

□

We now have everything we need to prove proposition 41.

Proposition 41. *The extremal points of the temporal correlation polytope are given by the deterministic assignments that fulfil the AoT conditions.*

Proof. To prove proposition 41, we need to show that (i) all deterministic assignments are extremal and (ii) every v (i.e. vector of the probabilities $p(abc...|xyz...)$) can be written as a convex combination of deterministic assignments.
The proof of (i) is trivial in the sense that a deterministic assignment for the vector v can never be written as a convex combination of other vectors. For (ii), we show the proof for the two measurement sequences, with two measurement outcomes, however, we can always generalize the method to longer sequences, and larger sets of outcome.
First note that probability distributions of the form $p\left(a|x\right)$, take the form shown in example 33, for a fixed value of x. Example 33 shows us, that the extremal points of the polytope described by this probability distribution are given by the deterministic assignments, hence for each distribution $p\left(a|x\right)$ the deterministic assignments are the extremal points. Every other probability assignment for $p(a|x)$ can be written as a convex combination. Let us define the vector

$$v_x := \begin{pmatrix} p(0|x) \\ p(1|x) \end{pmatrix} = c\begin{pmatrix} 1 \\ 0 \end{pmatrix} + (1-c)\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} c \\ 1-c \end{pmatrix}, \tag{4.2.17}$$

which is a convex combination of the two extremal points $(1,0)^T$ and $(0,1)^T$, respectively describing probability 1 for outcome 0 and probability 1 for outcome 1.
Let us for the moment assume that $c \neq 0, 1$. Then the conditional probabilities $p(b|axy)$ also take the form, given in example 33 if we fix a, x and y. In this case however, we have to take into account that the convex combination for the probabilities $p(b|axy)$ depends on the outcome a. Therefore, we define the vector

$$v_{xy} := \begin{pmatrix} p(0|0xy) \\ p(1|0xy) \\ p(0|1xy) \\ p(1|1xy) \end{pmatrix} = d \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + (1-d) \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + e \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + (1-e) \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} d \\ 1-d \\ e \\ 1-e \end{pmatrix}. \tag{4.2.18}$$

The first two entries describe the probability distribution for $a = 0$, with the convex coefficient d and the last two, the probability distribution for $a = 1$ with the convex coefficient e. Both convex combinations are independent of each other.

We now want to show that for fixed x and y the probability distribution $p(ab|xy)$ is always a convex combination if the conditional probabilities are non-deterministic. For this, let us define the vector

$$v := \begin{pmatrix} p(00|xy) \\ p(01|xy) \\ p(10|xy \\ p(11|xy) \end{pmatrix} = \begin{pmatrix} p(0|x)p(0|0xy) \\ p(0|x)p(1|0xy) \\ p(1|x)p(0|1xy \\ p(1|x)p(1|1xy) \end{pmatrix}, \tag{4.2.19}$$

where we used the fact that we can factorize the probabilities $p(ab|xy)$ into the conditional probabilities $p(a|x)$ and $p(b|axy)$. If we replace the probabilities $p(a|x)$ and $p(b|axy)$ with the respective coefficients in the vectors in equations 4.2.17 and 4.2.18, we get

$$v = \begin{pmatrix} cd \\ c(1-d) \\ (1-c)e \\ (1-c)(1-e) \end{pmatrix} = cd \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + c(1-d) \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + (1-c)e \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + (1-c)(1-e) \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \tag{4.2.20}$$

which is a convex combination of four different vectors if $0 < c < 1$, $0 \leq d \leq 1$ and $0 \leq e \leq 1$, since $cd, c(1-d), (1-c)e, (1-c)(1-e) \geq 0$ and $cd+c(1-d)+(1-c)e+(1-c)(1-e) = 1$. In this proof, we restricted ourselves to the case, where $p(a|x) \neq 0$ for all a, given x, otherwise, there would exist an a for which all probabilities $p(b|axy) = 0$ for all b.

Now consider the case of a deterministic assignment for $p(a|x)$ and fixed x, i.e. $c = 0$ or $c = 1$. Let w.l.o.g. be $c = 1$. From the AoT constraints it follows that $p(b|axy)$ is zero

for $a = 0$. The vector v_{xy} then takes the form

$$v_{xy} := \begin{pmatrix} p(0|0xy) \\ p(1|0xy) \\ p(0|1xy) \\ p(1|1xy) \end{pmatrix} = d \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + (1-d) \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}. \tag{4.2.21}$$

For the vector v, we get

$$v = \begin{pmatrix} d \\ (1-d) \\ 0 \\ 0 \end{pmatrix} = d \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + (1-d) \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \tag{4.2.22}$$

which is still a convex combination of deterministic assignments.
Since we can construct vectors like this for every choice of x and y, we find that all non-deterministic assignments for the vector v are convex combinations of deterministic assignments.

□

4.2.2 Number of Extremal Points for n Measurement Sequences.

As we have seen in section 4.2.1, the 64 extremal points for the two measurement sequence can be derived by choosing the marginal probabilities $p(a|x)$ to be deterministic. For each probability that is set to one, we then can assign a deterministic outcome for the second measurement (e.g. $p(0|0) = 1$ and we assign $p(01|01) = p(00|00) = 1$). One example of such a deterministic assignment that fulfills the AoT conditions is given by $p(01|00) = p(00|01) = p(10|10) = p(10|11) = 1$ and 0 otherwise.
In section 4.2.1, we further showed that the extremal points for sequences of length $n > 2$ can also be determined in the same way. However, we have not yet discussed the number of extremal points for sequences of length n. It turns out that there exists a simple formula to determine the number of extremal points for sequences of arbitrary length. To derive this formula, let us first have a look at the case of two sequential measurements. Here we first have to assign deterministic values for the marginals $p(a|x)$. For this probability distribution, there exist four different deterministic assignments, i.e. we have four different possibilities for the assignment $p(a|x) = 0, 1$. If we now include the

second measurement, we have to assign deterministic assignments for the probabilities $p\,(ab|xy)$ such that this assignment fulfills the marginal condition. For each distribution $p(a|x) = 0, 1$, we can assign $4^2 = 16$ assignments for $p(b|abxy)$. This is easily showed by some example.

Example 42. *Consider the assignment* $p(0|0) = 1$ *and* $p(1|1) = 1$ *for the marginal probabilities. For the second measurement, we have to assign deterministic outcomes for the conditional probabilities* $p(b|axy)$*. From section* 4.2.1*, we know that* $p(b|10y)$ *and* $p(b|01y)$ *are equal to zero for all* b *and* y*. For* $p(b|00y)$ *we have four choices for a deterministic assignment, the same holds true for* $p(b|11y)$*. Altogether, this results in* $4 \cdot 4 = 16$ *different choices for the probability distributions* $p(b|axy)$ *for a fixed assignment for* $p(a|x)$*. The deterministic assignment of* $p(ab|xy)$ *follows from* $p(ab|xy) = p(a|x)p(b|axy)$*.*

It is straightforward to generalize this way of counting the number of extremal points for sequences of arbitrary length. The generalized formula is given by

$$N = \prod_{i=0}^{n-1} 4^{\left(2^i\right)} = 4^{\sum_{i=0}^{n-1} 2^i} = 4^{2^n - 1}, \tag{4.2.23}$$

where the last part of equation 4.2.23 follows from the geometric series.
Equation 4.2.23 is restricted to the case of two measurements settings and two outcomes. We may also be interested in other setups, with a larger number of settings/outcomes. In this case, we can generalize equation 4.2.23 to arbitrary number of settings and outcomes. For simplicity reasons, we restrict ourselves to the reasonable case where every observable has the same number of outcomes, however a generalization to the case where different observables have different numbers of outcomes is possible. Let O be the number of outcomes, and S be the number of settings. We then have O^S ways to choose the marginal probability distribution $p(a|x)$. Using the same argument, as before, we find that the number of choices for $p(b|axy)$ is given by $\left(O^S\right)^S$. For arbitrary numbers of outcomes and settings, the general formula takes the form

$$N_{O,S} = \prod_{i=0}^{n-1} \left(O^S\right)^{\left(S^i\right)} = \left(O^S\right)^{\sum_{i=0}^{n-1} S^i} = \left(O^S\right)^{\left(\frac{S^n-1}{S-1}\right)}. \tag{4.2.24}$$

4.3 Quantum Mechanical Setup

In this section, we want to describe the quantum mechanical setup, used to prove the main theorem of the thesis. Again, we consider a measurement process as shown in figure 4.2.

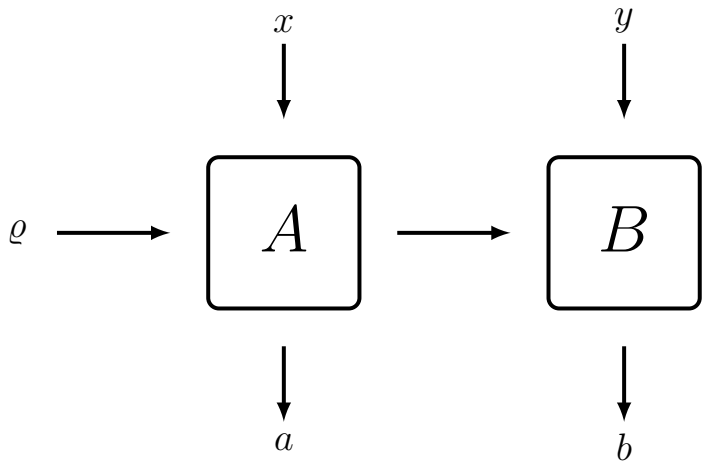

Figure 4.2: Temporal measurement sequence

At the times t_1 and t_2, the experimenter chooses between two dichotomic observables with outcomes $0, 1$, which are described by the effects E_{ax} and E_{by}, for $a, b, x, y \in \{0, 1\}$. Due to the normalization condition of the effects, we have $E_{0x} + E_{1x} = E_{0y} + E_{1y} = \mathbb{1}$. We describe the state after the first measurement by an instrument in the Kraus representation

$$\varrho \to \frac{\mathcal{I}(\varrho)}{\operatorname{tr}[\mathcal{I}(\varrho)]} = \frac{\sum_i K_i \, \varrho \, K_i^\dagger}{\operatorname{tr}\left[\sum_i K_i \, \varrho \, K_i^\dagger\right]} = \varrho', \tag{4.3.1}$$

where the Kraus-operators K_i have to fulfill the condition

$$\sum_i K_{axi}^\dagger K_{axi} = E_{ax}, \tag{4.3.2}$$

since $\mathcal{I}_{ax}(\varrho)$ has to be an A_x-compatible instrument (see equation 1.5.4). The measurement probabilities $p(ab|xy)$ will be described according to equation 1.5.3 and take the form

$$p\left(ab|xy\right) = \mathrm{tr}\left[\mathcal{I}_{ax}\left(\varrho\right)E_{by}\right] = \mathrm{tr}\left[\sum_i K_{axi}\,\varrho\,K_{axi}^{\dagger}E_{by}\right]. \tag{4.3.3}$$

4.4 Extremal Points in Quantum Mechanics

Quantum mechanics is a theory that fulfills the AoT constraints. When we describe measurement sequences in quantum theory, we find that indeed, the measurement result of the first measurement $p(a|x)$ is independent of the second measurement, while the second measurement can be influenced by the first one. This means that we can factorize the probabilities according to the AoT constraints.

$$p\left(abc|xyz\right) = p\left(a|x\right)p\left(b|axy\right) = \mathrm{tr}\left[E_{ax}\,\varrho\right]\mathrm{tr}\left[E_{by}\,\varrho_{ax}\right], \tag{4.4.1}$$

with the state evolution $\varrho \to \varrho_{ax}$.
For the extremal points of the polytope $P_2^{2,2}$, we have $p(a|x) \in \{0,1\}$ for all a, x and $p(b|axy) \in \{0,1\}$ for all $a,,b,x,y$. If we do not bound the dimension of the quantum system, we can write down an explicit model to reach all extremal points. For this, assume a set of orthogonal states $|\psi\rangle, |\psi_{ax}\rangle$ for all a, x, where $|\psi\rangle$ is the initial state of the system and $|\psi_{ax}\rangle$ are the states after the measurement x is performed and the outcome a is obtained. We can define the effects for the first measurement to be

$$E_{ax} = |\psi\rangle\langle\psi|, \tag{4.4.2}$$

if $p(a|x) = 1$ and set

$$E_{a'x} = \mathbb{1} - E_{ax}, \tag{4.4.3}$$

which automatically implies that $p(a'|x) = 0$. For the second measurement, we can in the same way define the effects

$$E_{by} = \sum_{a,x} |\psi_{ax}\rangle\langle\psi_{ax}| \tag{4.4.4}$$

for all a, x for which $p(b|axy) = 1$ and

$$E_{b'y} = \mathbb{1} - E_{by} \tag{4.4.5}$$

to ensure that $p(b'|axy) = 0$ is fulfilled. This model is able to reach all extremal points of the polytope $P_2^{2,2}$. However, this does not necessarily mean that quantum mechanics is able to reach all extremal points of the temporal correlation polytope for finite dimensions, since the number of orthogonal states $|\psi\rangle, |\psi_{ax}\rangle$ is limited. We prove that the ability to reach all extremal points of the polytope $P_l^{2,2}$ depends on the dimension d. In the following, we use the reasonable assumption that ($E_{ax} = E_{by}$ if $a = b$ and $x = y$).[2] This assumption leads us to the statement of theorem 43.

Theorem 43. *For quantum systems with two measurement bases with arbitrary instruments, and bounded memory (dimension constraint), it is not generally possible to reach all extremal points of the temporal correlation polytope $P_l^{O,S}$.*

In this thesis, we are particularly interested in the case of two-measurement sequences (i.e. $l = 2$) and qubit systems. In the next section, we will prove that we need at least a three level system to reach all extremal points of the polytope $P_2^{2,2}$.

4.5 Sequences of Qubit Measurements of Length 2

In this section, we want to show explicitly that it is impossible to reach all extremal points of the temporal correlation polytope $P_2^{2,2}$ with qubit systems under the constraint ($E_{ax} = E_{by}$ if $a = b$ and $x = y$).

Proof of Theorem 43 : The extremal point used for the counterexample is given by the following assignment. We have the marginals $p(0|0) = p(1|1) = 1$ and we assign the value 1 to $p(01|00)$, $p(00|01)$, $p(10|10)$ and $p(10|11)$, while the other remaining 12 terms get assigned the value 0.

The first step to get a counterexample is to show that an extremal point can only be achieved if the initial state is pure. (I.e. we have $\varrho = |\psi\rangle\langle\psi|$). To show this, let us assume a mixed state ($\varrho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$) and show that we can only reach the given (or any) deterministic assignment if we set $p_i = 1$ and $p_j = 0 \quad \forall j \neq i$ (i.e. the state ϱ is pure).

From the assignment $p(0|0) = 1$, we get

[2] The experimenter can choose between two measurement settings and can either repeat the same measurement again or measure in the other setting

$$1 = \sum_i \mathrm{tr}\left[K_{00i}\,\varrho\,K_{00i}^\dagger\right] \tag{4.5.1}$$

$$= \sum_{ij} p_j\,\mathrm{tr}\left[K_{00i}\,|\psi_j\rangle\langle\psi_j|\,K_{00i}^\dagger\right] \tag{4.5.2}$$

$$= \sum_{ij} p_j\,\mathrm{tr}\left[K_{00i}^\dagger K_{00i}\,|\psi_j\rangle\langle\psi_j|\right] \tag{4.5.3}$$

$$= \sum_{j} p_j\,\mathrm{tr}\left[E_{00}\,|\psi_j\rangle\langle\psi_j|\right], \tag{4.5.4}$$

which can only be fulfilled if we have $E_{00}\,|\psi_j\rangle = \sum_i K_{00i}^\dagger K_{00i}\,|\psi_j\rangle = |\psi_j\rangle$ for all j. In other words, either E_{00} is the identity $\mathbb{1}$ or the vectors $|\psi_j\rangle$ are all eigenvectors of the operator E_{00} with the eigenvalue 1. In the qubit case, in which we are interested here, every operator has two eigenvalues. It is straightforward to show that the only self-adjoint 2×2-matrix with twice the eigenvalue $+1$ is the identity matrix $\mathbb{1}$.

Hence, we either have $E_{00} = \mathbb{1}$, or that the initial state ϱ was a pure state $\varrho = |\psi\rangle\langle\psi|$, where the vector $|\psi\rangle$ is an eigenvector of the operator E_{00} with the eigenvalue $+1$. (I.e. we have $E_{00}\,|\psi\rangle = \sum_i K_{00i}^\dagger K_{00i}\,|\psi\rangle = |\psi\rangle$)

It is easy to show that we cannot set $E_{00} = \mathbb{1}$. For this, let us define the instrument induced by observable E_{00} as $\mathcal{I}_{00}(\varrho) = \sum_i K_{00i}\,\varrho\,K_{00i}^\dagger$. The contradiction follows from $p(0|0) = \mathrm{tr}\left[\mathcal{I}_{00}(\varrho)\right] = 1$ and $p(00|00) = \mathrm{tr}\left[E_{00}\mathcal{I}_{00}(\varrho)\right] = 0$. It is obvious that this assignment is impossible if we set $E_{00} = \mathbb{1}$.

We therefore know that we have $E_{00} \neq \mathbb{1}$ and $\varrho = |\psi\rangle\langle\psi|$. The argument to show that we still cannot reach the desired extremal point is a bit more subtle. We can use the probability $p(1|1) = 1$ in the exact same way as before to deduce that the vector $|\psi\rangle$ is additionally an eigenvector to the observable E_{11}. We hence have $E_{11}\,|\psi\rangle = \sum_i K_{11i}^\dagger K_{11i}\,|\psi\rangle = |\psi\rangle$.

These two equations however are not enough to reach a contradiction. To get more constraints, let us have a look at the post-measurement state. We know that each set of Kraus-Operators K_{ax} describes an instrument $\mathcal{I}_{ax}(\varrho) = \widetilde{\varrho}$ for given a, x. From $p(0|0) = \mathrm{tr}\left[\mathcal{I}_{00}(\varrho)\right] = 1$ and $p(1|1) = 1$ we further know that the initial state $\varrho = |\psi\rangle\langle\psi|$ is transformed trace-preserving[3]. Since we have to assign the value $+1$ to the probability

[3] Note that the instrument $\mathcal{I}_{ax}$ is not necessarily a channel, we only know that the initial state $\varrho = |\psi\rangle\langle\psi|$ is transformed trace-preserving.

$p(01|00)$, we have

$$p(01|00) = \mathrm{tr}\left[E_{10}\mathcal{I}_{00}(\varrho)\right] \tag{4.5.5}$$

$$= \mathrm{tr}\left[E_{10}\widetilde{\varrho}\right] = 1. \tag{4.5.6}$$

We can conclude that the post measurement state $\widetilde{\varrho}$ must also be a pure state. We can similarly argue for the instrument $\mathcal{I}_{11}(\varrho)$. Hence we can define

$$\mathcal{I}_{00}(|\psi\rangle\langle\psi|) = \sum_i K_{00i}|\psi\rangle\langle\psi|K_{00i}^\dagger = |\varphi\rangle\langle\varphi|, \tag{4.5.7}$$

and

$$\mathcal{I}_{11}(|\psi\rangle\langle\psi|) = \sum_i K_{11i}|\psi\rangle\langle\psi|K_{11i}^\dagger = |\lambda\rangle\langle\lambda|. \tag{4.5.8}$$

Inserting equation 4.5.7 in $p(01|00)$ and $p(00|01)$ and equation 4.5.8 in $p(10|10)$ and $p(10|11)$ gives us the additional constraints $E_{10}|\varphi\rangle = \sum_i K_{10i}^\dagger K_{10i}|\varphi\rangle = E_{01}|\varphi\rangle = \sum_i K_{01i}^\dagger K_{01i}|\varphi\rangle = |\varphi\rangle$ and $E_{00}|\lambda\rangle = \sum_i K_{00i}^\dagger K_{00i}|\lambda\rangle = E_{01}|\lambda\rangle = \sum_i K_{01i}^\dagger K_{01i}|\lambda\rangle = |\lambda\rangle$. Summing up, we have the following constraints

$$E_{00}|\psi\rangle = E_{11}|\psi\rangle = |\psi\rangle, \tag{4.5.9}$$

$$E_{00}|\lambda\rangle = E_{01}|\lambda\rangle = |\lambda\rangle, \tag{4.5.10}$$

$$E_{10}|\varphi\rangle = E_{01}|\varphi\rangle = |\varphi\rangle, \tag{4.5.11}$$

$$E_{00}|\varphi\rangle = E_{11}|\varphi\rangle = 0, \tag{4.5.12}$$

$$E_{10}|\lambda\rangle = E_{11}|\lambda\rangle = 0. \tag{4.5.13}$$

Finally, we construct the contradiction from these constraints. From the equations 4.5.9 and 4.5.12 we know that the operator E_{00} has the eigenvector $|\psi\rangle$ for the eigenvalue $+1$ and the eigenvector $|\varphi\rangle$ for the eigenvalue 0. Since a 2×2-matrix only has two eigenvalues it follows that the effect E_{00} has to be a projection for qubit systems. Hence, we can write $E_{00} = |\psi\rangle\langle\psi|$. From the same equations, we find out that the operator E_{11} shares both eigenvectors and eigenvalues with the operator E_{00}. Thus, for two dimensional systems,

we have $E_{00} = E_{11}$. Let us now have a look at the equations 4.5.12 and 4.5.13. This gives us $E_{00}\,|\varphi\rangle = E_{11}\,|\lambda\rangle = E_{00}\,|\lambda\rangle = 0$. Since we know that the eigenvector of the operator E_{00} for the eigenvalue 0 is $|\varphi\rangle$, this implies that $|\varphi\rangle = |\lambda\rangle$. Finally, the contradiction follows from equation 4.5.10 from which we can infer the chain

$$0 = E_{00}\,|\lambda\rangle \neq |\lambda\rangle = E_{00}\,|\lambda\rangle\,, \tag{4.5.14}$$

which is completes the argument.

□

4.6 Sequences of Qutrit Measurements of Length 2

The contradiction given above shows that there exists at least one extremal point, which cannot be reached with qubits systems, however it does not certify us that we can reach them with qutrits. In this section, we will give an explicit example that is able to reach the extremal point used in section 4.5 for qutrit systems. To ensure that we really can reach the assignment, we use not only the constraints from equations 4.5.9 to 4.5.13, but also additional constrains which follow from the remaining probabilities, which were not used in the counterargument before. This gives us the following extra assumptions.

$$E_{10}\,|\psi\rangle = 0, \tag{4.6.1}$$

$$E_{01}\,|\psi\rangle = 0. \tag{4.6.2}$$

These assumptions come from the fact that we have to set $p\,(1b|0y)$ and $p\,(0b|1y)$ to zero for all b and y. To reach the extremal point, we can now define the operators E_i as

$$E_{00} = |\psi\rangle\langle\psi| + |\lambda\rangle\langle\lambda|\,, \tag{4.6.3}$$

$$E_{10} = |\varphi\rangle\langle\varphi|\,, \tag{4.6.4}$$

$$E_{01} = |\varphi\rangle\langle\varphi| + |\lambda\rangle\langle\lambda|\,, \tag{4.6.5}$$

$$E_{11} = |\psi\rangle\langle\psi|\,. \tag{4.6.6}$$

It is easy to check, that the operators E_i defined in this way fulfill all constraints given

above.

4.7 Equivalence Classes

Finally, we want to check, whether we are able to reach all extremal points with qutrit systems and whether we are able to reach some extremal points with qubit systems as well. To do so, we considered the three symmetry transformations T_1 (change of the outcome if the measurement setting 1 was used), T_2 (change of the outcome for setting 0), T_3 ($(x, y) \rightarrow (x', y')$, i.e. both settings x and y are changed) and their combinations to assign the 64 extremal points into sets of equivalence classes. The ten different equivalence classes and one representative per class are shown in table 4.7. In

Equivalence class	Extremal point	Qutrits	Qubits
1	$p(10\|00), p(10\|01), p(01\|10), p(00\|11)$	✓	✓
2	$p(10\|00), p(10\|01), p(01\|10), p(01\|11)$	✓	✗
3	$p(10\|00), p(10\|01), p(10\|10), p(11\|11)$	✓	✗
4	$p(10\|00), p(10\|01), p(10\|10), p(10\|11)$	✓	✓
5	$p(10\|00), p(10\|01), p(11\|10), p(11\|11)$	✓	✓
6	$p(10\|00), p(10\|01), p(11\|10), p(10\|11)$	✓	✗
7	$p(10\|00), p(10\|01), p(00\|10), p(00\|11)$	✓	✓
8	$p(11\|00), p(10\|01), p(01\|10), p(00\|11)$	✓	✓
9	$p(11\|00), p(10\|01), p(10\|10), p(11\|11)$	✓	✗
10	$p(11\|00), p(10\|01), p(11\|10), p(11\|11)$	✓	✓

Table 4.1: The ten equivalence classes of extremal points. The second column gives an exemplary extremal point for each equivalence class. The probabilities mentioned there are the ones that are set to one. The last two columns depict whether we can reach the extremal points with qutrit- or qubit-systems.

the sections 4.5 and 4.6, we have already seen an example of an extremal point that can only be reached by a three-level-system. We now want to show an example for an extremal point that can be reached with a two-level-system.

Example 44. *Let us have a look at equivalence class number one. If we write down an explicit model to reach the exemplary extremal point in the same way as in section* 4.6*, we get*

$$E_{00} = |\varphi\rangle\langle\varphi|\,, \tag{4.7.1}$$

$$E_{10} = |\psi\rangle\langle\psi| + |\lambda\rangle\langle\lambda|\,, \tag{4.7.2}$$

$$E_{01} = |\psi\rangle\langle\psi| + |\varphi\rangle\langle\varphi| + |\lambda\rangle\langle\lambda| = \mathbb{1}, \tag{4.7.3}$$

$$E_{11} = 0, \tag{4.7.4}$$

with and orthonormal basis consisting of the initial state $|\psi\rangle$, the state $|\varphi\rangle$ after getting the outcome $(1|0)$ for the first measurement and $|\lambda\rangle$ as the state after getting $(0|1)$ for the first measurement. This however is obviously not the only possible choice for a model. We can for example set $|\psi\rangle = |\lambda\rangle$, without creating any disagreements from the effects, since $|\psi\rangle$ and $|\lambda\rangle$ always occur in the same effect. This implies that we can write a model that only uses two orthogonal states and hence can be reached by qubit systems. For the qubit case, we have

$$E_{00} = |\varphi\rangle\langle\varphi|\,, \tag{4.7.5}$$

$$E_{10} = |\psi\rangle\langle\psi|\,, \tag{4.7.6}$$

$$E_{01} = |\psi\rangle\langle\psi| + |\varphi\rangle\langle\varphi| = \mathbb{1}, \tag{4.7.7}$$

$$E_{11} = 0. \tag{4.7.8}$$

4.8 Temporal Bell Inequality

In section 4.5, we showed that with the quantum mechanical set-up, described in section 4.3, we cannot reach all extremal points of the temporal correlation polytope (i.e. there exists at least one extremal point which cannot be reached) for qubit systems, while this is possible for qutrit systems. However, this counterexample cannot be tested in experiments. If we want to test the counterexample experimentally, we need to have a finite, non-vanishing difference between the qubit case and the qutrit case, like in the Bell-CHSH scenario, where the LHV bound is 2 and the quantum bound is $2\sqrt{2}$. In this section we therefore want to derive some sort of a "Temporal Bell Inequality", which can actually be tested in experiments to distinguish between two- and three-level systems.
Let us start with the four probabilities used in the original counterexample for the qubit

system and arrange them as in equation 4.8.1 to get a first version of the inequality[4].

$$\begin{aligned}\langle\mathcal{B}\rangle = &\, p\,(01|00) + p\,(00|01) + p\,(10|10) + p\,(10|11) \\ &- p\,(00|00) - p\,(01|01) - p\,(11|10) - p\,(11|11)\,.\end{aligned} \tag{4.8.1}$$

In the qubit case, the counterargument tells us that we cannot set the first four terms to one, and set the last four terms to zero at the same time. For three-level systems this is however possible, hence we get the inequality

$$\langle\mathcal{B}\rangle \overset{\text{Qubit}}{\leq} b_{Qubit} \overset{\text{Qutrit}}{<} 4, \tag{4.8.2}$$

with the qubit bound $0 \leq b_{Qubit} < 4$.
However, the form of the inequality, given in equation 4.8.1 cannot be used to calculate the precise qubit bound b_{Qubit}. To do so, we reconstruct the Bell operator $\mathcal{B}$ by rewriting the probabilities $p\,(ab|xy)$ in terms of the initial state ϱ and the corresponding Kraus operators. Since we are interested in the maximal achievable value for qubit systems, we use the extremal representation of measurements (i.e. each instrument is described by a single Kraus-operator). Let us at this point give a short argumentation for this. First note that we want to maximize the term in 4.8.1, hence we can assume extremal completely positive maps. An extremal cp-map takes the form

$$\Phi(\varrho) = K\,\varrho\,K^\dagger. \tag{4.8.3}$$

For each map Φ, we can create an instrument $\mathcal{I}$ by introducing a variable α such that

$$\frac{K^\dagger K}{\alpha} \leq \mathbb{1}, \tag{4.8.4}$$

and defining $K_0^\dagger K_0 = E_0$. The instrument can then be completed to a channel by defining

$$E_1 = \mathbb{1} - E_0 = \mathbb{1} - K_0^\dagger K_0, \tag{4.8.5}$$

with $E_1 = \sum_i K_{1i}^\dagger K_{1i}$.
Since we later use the normalization condition for effects to eliminate all Kraus operators

[4] We sum up the probabilities which we have to set to one and subtract four probabilities which we have to set to zero

except K_{00i} and K_{11j}, we can choose extremal instruments such that all instruments are described by one Kraus operator.
This gives us

$$\begin{aligned}\langle\mathcal{B}\rangle = &\operatorname{tr}\left[K_{00}\,\varrho\,K_{00}^{\dagger}E_{10} + K_{00}\,\varrho\,K_{00}^{\dagger}E_{01} + K_{11}\,\varrho\,K_{11}^{\dagger}E_{00} + K_{11}\,\varrho\,K_{11}^{\dagger}E_{01}\right] \\ &- \operatorname{tr}\left[K_{00}\,\varrho\,K_{00}^{\dagger}E_{00} + K_{00}\,\varrho\,K_{00}^{\dagger}E_{11} + K_{11}\,\varrho\,K_{11}^{\dagger}E_{10} + K_{11}\,\varrho\,K_{11}^{\dagger}E_{11}\right].\end{aligned} \tag{4.8.6}$$

We can further simplify this equation by using the normalization condition of the effects

$$K_{0x}^{\dagger}K_{0x} + K_{1x}^{\dagger}K_{1x} = E_{0x} + E_{1x} = \mathbb{1}, \tag{4.8.7}$$

to eradicate the effects E_{10} and E_{01}. This gives us the final result

$$\begin{aligned}\mathcal{B} = &2\left(K_{00}^{\dagger}K_{00} - K_{00}^{\dagger}\left(K_{00}^{\dagger}K_{00} + K_{11}^{\dagger}K_{11}\right)K_{00}\right) \\ &+ 2\left(K_{11}^{\dagger}\left(K_{00}^{\dagger}K_{00} - K_{11}^{\dagger}K_{11}\right)K_{11}\right).\end{aligned} \tag{4.8.8}$$

4.8.1 Numerical Determination of the Qubit Bound

Up to this point, we were not able to determine an analytical upper bound for b_{Qubit}. Instead, we can maximize the norm of the operator $\mathcal{B}$ with `Mathematica`. We performed numerical maximization for several different numerical methods. Two of the different methods give the value $\langle\frac{\mathcal{B}}{2}\rangle_{max} \leq 1.28171$, while a third method gives us the value $\langle\frac{\mathcal{B}}{2}\rangle_{max} \leq 1.25$.
This gives us a rough estimate of the qubit bound b_{Qubit} of,

$$b_{Qubit} \simeq 2.6. \tag{4.8.9}$$

4.9 Application as a Dimension Witness

In the previous section, we have seen that equation 4.8.2 is able to distinguish between two- and higher dimensional systems. This implies that we can use the "temporal Bell-operator", defined in equality 4.8.8 as a dimension witness. We can compare the operator from equation 4.8.2 with the previous results described in section 2.6. We see that our

operator needs only one quantum system and considers sequential measurements, similarly as in [55]. Our case is state dependent, however the measurements are not fixed. For several different kinds of measurements, we can find an optimal state.
Our operator has advantages and disadvantages compared to other proposed dimension witness tests. On the one hand, the operator involves a single quantum system, which means that we do not need to have two space-like separated laboratories in the experiment, as opposed to the dimension witnesses derived by Brunner et al. in [12]. Furthermore, our operator is simple, since the it consists only of two independent Kraus-operators and we only need to implement two different measurement settings. For the CHSH scenario, we e.g. need four different measurement settings. Another advantage is that, according to the simulation, it seems as if a qutrit system reaches the maximal value of four, only if all the effects are projections, which means that we can perform the measurements in experiments without having to take $POVM$s into account.
One disadvantage of our operator is that the violation for systems of dimension $d \geq 2$ is not state independent.
All in all, we find that we can use the "temporal Bell operator" as a simple, resource-saving, state-independent dimension witness to determine whether an experimenter is limited to manipulate two levels of a quantum system or whether he is able to manipulate more than two levels.

4.10 Summary

In this section, we will give a short summary of the most important results of chapter 4. We found out that the AoT constraints define the temporal correlation polytope $P_l^{O,S}$. In contrast to the non-signalling polytope, the extremal points of the temporal correlation polytopes are given by deterministic assignments (i.e. zero-one assignments) that fulfill the AoT constraints. We have seen that for a quantum model with two measurement settings, two outcomes and length two, qubit systems are not able to reach all extremal points of the polytope $P_2^{2,2}$ under the reasonable assumption $E_{ax} = E_{by}$ if $a = b$ and $x = y$, i.e. only two measurement settings, $x = y$ implies the repetition of measurement x. For the extremal point used in the counterexample, we were able to give a direct example for the effects E_{ax} such that the extremal point can be reached with qutrits. This feature was used to derive a "temporal Bell inequality" that can be used as a dimension witness.

5 Correction Terms for Noncontextuality Tests

In the sections 2.5.2 and 2.5.3, we introduced state dependent and state independent contextuality, the corresponding inequalities for these cases are both derived for the case of perfectly compatible observables.
If we however want to test noncontextuality inequalities experimentally, we have to account for the fact that perfectly compatible measurements are an idealization and that actual measurements are not perfect. In measurements it may happen that, with a certain probability p, one measures a different observable than intended. (E.g. The wanted observable σ_z is measured with probability $1-p$, whereas with probability p some other observable A is measured.) As a result, we can never achieve prefect compatibility for observables acting on the same system, while we can reach compatibility for observables acting on different systems. The best we can achieve for two observables acting on one system is to be nearly compatible. If we test noncontextuality inequalities with non-compatible observables, it may happen that the quantum bound for this observables is different than the bound for compatible observables and thus violate the inequality even though the quantum bound for compatible measurements would confirm the NCHV bound. Hence, a measurement of the original inequality is subjected to the compatibility loophole and cannot be properly used to test noncontextuality. Therefore, if we want to test noncontextuality inequalities, we need to estimate the error, which results from using imperfect measurements and correct this errors by the means of additional correction terms in the original inequality.
In the following section, we will introduce two different approaches to estimate these errors. The structure of this chapter is as follows: In section 5.1, we will have a look at some correction terms which were proposed by Gühne et al. in 2010 [16]. In section 5.2, we will investigate a different form of correction terms proposed by Kujala et al. in 2015 [17]. In section 5.3, we will perform an analytical calculation of the correction

terms for qubit systems and projective measurements for both kinds of correction terms. Finally, in section 5.4, we will simulate the correction terms for quantum systems of infinite dimensions and general dichotomic measurements.

5.1 Correction Terms by Gühne et al.

In their paper from the year 2010, Gühne et al. considered the problem that the measurements performed for noncontextuality inequality tests are imperfect and hence incompatible even though the undisturbed measurements may be compatible. Thus a violation of the noncontextuality inequalities can be interpreted as a failure of the assumption that the observables measured are perfectly compatible. In other words, the noisy observables disturb each other and a measurement of one observable influences the outcome of the measurement of the second observable. They named this problem the "compatibility loophole" and derived additional correction terms to account for the imperfectness of experimental measurements [16].
The correction terms introduced in their paper shall give an upper bound on the influence that two non-compatible observables have on each other. For compatible observables A and B, we find some state $|\psi\rangle$ such that

$$p(ab) = \langle\psi|A_a B_b|\psi\rangle = 1, \tag{5.1.1}$$

while all other combinations of a and b have probability zero. For projective measurements, this is equivalent to A and B having a common eigenbasis. If the observables however are not compatible, we cannot assign a fixed outcome for both observables at the same time, which means that the measurements disturb each other.
In their paper, Gühne et al. consider sequences of measurements and bound the error, resulting from incompatible measurements, by flipping probabilities (e.g. the value of observable A is flipped if observable B is measured in between). The reason to use the flipping probabilities as an upper bound for the correction terms follows from the assumption of cumulative noise (the disturbance created by measuring two incompatible observables becomes larger if the sequence of alternating measurements is extended).

The formula for the correction terms can be derived in several steps. First note that

$$\begin{aligned} & p\left[\left(A_1^+|A_1\right) \text{ and } \left(B_1^+|B_1\right)\right] \\ & \leq p\left[A_1^+, B_2^+|A_1B_2\right] + p\left[\left(B_1^+|B_1\right) \text{ and } \left(B_2^-|A_1B_2\right)\right]. \end{aligned} \tag{5.1.2}$$

Here $p\left[\left(A_1^+|A_1\right) \text{ and } \left(B_1^+|B_1\right)\right]$ is the probability that one gets the measurement result A^+ if A is measured first and B^+ if B is measured first. These kind of probabilities are inaccessible by experiments as we have to decide on a measurement sequence, but are well defined in NCHV models. The term $p\left[A_1^+, B_2^+|A_1B_2\right]$ describes the probability that we get the outcomes A_1^+ and B_2^+ if we measure the sequence A_1B_2, where the subscripts define the place in the measurement sequence. Finally $p\left[\left(B_1^+|B_1\right) \text{ and } \left(B_2^-|A_1B_2\right)\right]$ is the probability that a measurement of observable B at time t_1 gives us the outcome B_1^+, while a measurement of B_2 in the sequence A_1B_2 gives the outcome B_2^-. Inequality 5.1.2 is true because if the hidden variable λ is such that if it contributes to $p\left[\left(A_1^+|A_1\right) \text{ and } \left(B_1^+|B_1\right)\right]$, then either the value of B does not change when measuring the sequence A_1B_2 (i.e. λ contributes to $p\left[A_1^+, B_2^+|A_1B_2\right]$) or the value of B is flipped (λ contributes to $p\left[\left(B_1^+|B_1\right) \text{ and } \left(B_2^-|A_1B_2\right)\right]$).
The other estimation they use is that the expectation value $\langle AB\rangle$ can be rewritten as

$$\begin{aligned} \langle AB\rangle &= p\left[\left(A_1^+|A_1\right) \text{ and } \left(B_1^+|B_1\right)\right] + p\left[\left(A_1^-|A_1\right) \text{ and } \left(B_1^-|B_1\right)\right] \\ &\quad - p\left[\left(A_1^+|A_1\right) \text{ and } \left(B_1^-|B_1\right)\right] - p\left[\left(A_1^-|A_1\right) \text{ and } \left(B_1^+|B_1\right)\right] \\ &= 1 - 2p\left[\left(A_1^+|A_1\right) \text{ and } \left(B_1^-|B_1\right)\right] - 2p\left[\left(A_1^-|A_1\right) \text{ and } \left(B_1^+|B_1\right)\right]. \end{aligned} \tag{5.1.3}$$

Following the steps from [16], we can insert inequality 5.1.2 in equality 5.1.3 to derive the new inequality

$$\langle \mathcal{B}_{CHSH}\rangle \leq 2\left(1 + p^{flip}\left[AB\right] + p^{flip}\left[CB\right] + p^{flip}\left[CD\right] + p^{flip}\left[AD\right]\right), \tag{5.1.4}$$

where $p^{flip}\left[AB\right]$ is defined as

$$p^{flip}\left[AB\right] = p\left[\left(B_1^+|B_1\right) \text{and} \left(B_2^-|A_1B_2\right)\right] + p\left[\left(B_1^-|B_1\right) \text{and} \left(B_2^+|A_1B_2\right)\right]. \tag{5.1.5}$$

The flip probability $p^{flip}\left[AB\right]$ is the sum of the probabilities $p\left[\left(B_1^+|B_1\right) \text{and} \left(B_2^-|A_1B_2\right)\right]$

(i.e. The value of the observable B is B^+ if B is measured first, but is flipped to B^- if the sequence A_1B_2 is measured) and $p\left[\left(B_1^-|B_1\right) \text{and} \left(B_2^+|A_1B_2\right)\right]$ (i.e. The value is B^- if B is measured first, but B^+ in the sequence A_1B_2). In other words $p^{flip}\left[AB\right]$ describes the probability that the outcome of observable B is flipped by the observable A.
The correction terms $p^{flip}\left[AB\right]$ in equation 5.1.4 however are not measurable in experiments. Probabilities like $p\left[\left(B_1^+|B_1\right) \text{and} \left(B_2^-|A_1B_2\right)\right]$ cannot be determined experimentally as it is impossible to measure the sequences B_1 and A_1B_2 at the same time. In experiments one has to decide whether to measure first observable B or observable A. As we want correction terms that can be used to describe imperfect measurements in real experiments, we need to estimate them in such a way that they can also be tested experimentally. If one assumes that the terms in the probability $p\left[\left(B_1^+|B_1\right) \text{and} \left(B_2^-|A_1B_2\right)\right]$ describe some sort of disturbance of the measurement B by the measurement A, (i.e. if B would be measured first, the value B_1^+ is returned, however if the sequence A_1B_2 is measured, the value for B gets disturbed by the measurement A_1 and one gets the result B_2^- instead) we can us the cumulative noise assumption and estimate [16]

$$\begin{aligned} p\left[\left(B_1^+|B_1\right) \text{and} \left(B_2^-|A_1B_2\right)\right] &\leq p\left[\left(B_1^+|B_1\right) \text{ and } \left(B_1^+B_3^-|B_1A_2B_3\right)\right] \\ &\equiv p\left[B_1^+, B_3^-|B_1A_2B_3\right]. \end{aligned} \tag{5.1.6}$$

In inequality 5.1.6, the left-hand side describes the amount of disturbance of observable B created by observable A. The right-hand side quantifies the disturbance of observable B when the sequence $B_1A_2B_3$ is measured. It can be expected that in real experiments the disturbance in the sequence $B_1A_2B_3$ is larger than in the sequence A_1B_2, due to additional experimental procedures being involved.
Error terms of the form, given in equation 5.1.6 are experimentally feasible and are also of the right form to be used in the simulation carried out in section 5.5. Replacing the p^{flip} terms in equality 5.1.4 with the estimation shown in equality 5.1.6, gives us finally

$$\langle \mathcal{B}_{CHSH} \rangle \leq 2\left(1 + p^{err}\left[B_1A_2B_3\right] + p^{err}\left[B_1C_2B_3\right] + p^{err}\left[D_1C_2D_3\right] + p^{err}\left[D_1A_2D_3\right]\right), \tag{5.1.7}$$

whereby the notation

$$p^{err}\left[B_1A_2B_3\right] = p\left[B_1^+, B_3^-|B_1A_2B_3\right] + p\left[B_1^-, B_3^+|B_1A_2B_3\right] \tag{5.1.8}$$

was used.

5.2 Correction Terms by Kujala et al.

In their paper form the year 2015, Kujala et al. derived error terms for noncontextuality inequalities under the assumption of inconsistent connectedness [17]. The main idea of the paper is to give a definition of noncontextuality such that contextuality cannot be explained through systematic measurement errors or signalling. They call noncontextuality of this form *maximally noncontextual*. Since their definition of maximally noncontextuality is such that systematic measurement errors (e.g. non-compatible measurements) do not lead to a violation of an inequality, they basically derive correction terms for noncontextuality inequalities, in the same sense as Gühne et al., (i.e. noncontextual behavior should not be explainable by non-compatible measurements).
Let us start with a short look at the derivation of their results. Their basic concept is different from the one used by Gühne et al. in [16]. Kujala et al. start with a finite set of physical properties/measurements $\mathcal{Q} = \{q_1, ...q_n\}$. The physical properties are measured in subsets of $\mathcal{Q}$, called context $\{c_1, ...c_m\}$. The result of a measurement of property q in the context c is described by a random variable R_q^c. They call the set of random variables, that represent the same propertiy q in different contexts a *connection* (for q). If all random variables, for each connection are identically distributed, the system is called *consistently connected*, otherwise it is called *inconsistently connected*. A system is then said to have a noncontextual description if there exists a joint distribution of these random variables in which any two variables representing the same property q in different contexts c_i (e.g. $R_q^{c_1}$ and $R_q^{c_2}$) are equal with probability one. Otherwise the system is contextual. This definition however only includes consistent connected systems. Kujala et al. generalize this definition to include inconsistent connected systems. As these kind of systems are always contextual according to the usual definition, they use their definition of *maximally noncontextuality* to distinguish between systems that are genuine contextual and systems, where the contextuality follows form measurement errors or signalling, where the later are still called *maximally noncontextual*.
For cyclic systems (i.e. A system of n observables A_0, A_{n-1}, where always two observ-

ables with the indices i and $i+1$ are in a measurement context, e.g. A_0 and A_1), they find the following theorem, which connects the definition of maximally contextuality and cyclic systems.

Theorem 45. *A cyclic system of rank $n > 1$ with dichotomic random variables has a maximally noncontextual description if and only if*

$$\mathbf{s}_1\left(\langle R_i^i R_{i\oplus 1}^i\rangle, 1 - \left|\langle R_i^i\rangle - \langle R_i^{i\ominus 1}\rangle\right| : i = 1, ..., n\right) \leq 2n - 2, \tag{5.2.1}$$

where $\mathbf{s}_1$ is defined as

$$\mathbf{s}_1\left(x_1, ..., x_n\right) = \max_{l_1,...,l_n, \prod_k l_k = -1} \sum_k l_k x_k. \tag{5.2.2}$$

Equation 5.2.1 is constructed such that it is reduced to the simple form

$$\mathbf{s}_1\left(\langle R_i^i R_{i\oplus 1}^i\rangle : i = 1, ..., n\right) \leq n - 2, \tag{5.2.3}$$

under the assumption of noncontextuality. Then one has (i.e.$\langle R_i^i\rangle = \langle R_i^{i\ominus 1}\rangle$), hence one comes back to the original inequality. For $n = 3$ one recovers the LG inequality, for $n = 4$ the CHSH inequality and for $n = 5$ the KCBS inequality.

In [17] it was also shown that equality 5.2.1 can be further simplified, which results in the following equation.

$$\mathbf{s}_1\left(\langle R_i^i R_{i\oplus 1}^i\rangle : i = 1, ..., n\right) - \sum_{i=1}^{n} \left|\langle R_i^i\rangle - \langle R_i^{i\ominus 1}\rangle\right| \leq n - 2. \tag{5.2.4}$$

The correction terms introduced in section 5.1 and section 5.2 are supposed to detect non-perfectly compatible measurements. A reasonable constraint for the error terms is that they should "restore" the original quantum bound for compatible measurements. For the simple operator

$$\mathcal{T} = \langle AB\rangle + \langle BC\rangle - \langle AC\rangle \tag{5.2.5}$$

the quantum bound for compatible measurements is equal to the classical bound $\Omega_{\mathrm{NCHV}} = \Omega_{\mathrm{QM}}^{\mathrm{comp}} = 1$. We call the corresponding inequality *triangle inequality*.

Now that we have introduced the correction terms, we want to investigate, which effect they have on equality 5.2.5 if we do not have compatible measurements and consider the question: Under which assumptions do the correction terms behave as expected? This means that for a given inequality we recover the quantum bound for compatible observables, if we add them to the original inequality. If the correction terms manage to recover the bound, we know that a violation of the inequality actually is equivalent to the contextuality of quantum mechanics. In other words, we close the compatibility loophole.

5.3 Analytical Calculation of the Correction Terms

In this section, we want to calculate the correction terms presented in section 5 analytically for the special case of PVMs for qubit systems.
For the density matrix ϱ the general Bloch sphere representation [21]

$$\varrho = \frac{1}{2}\left(\mathbb{1} + \vec{\varrho}\vec{\sigma}\right) \tag{5.3.1}$$

was used, while the observables were written in the form

$$P_B^s = \frac{1}{2}\left(\mathbb{1} + sB\right), \tag{5.3.2}$$

where the projectors for the outcome $+1$ is defined as $P_B^+ = P_B^{s=1}$ and the projector for the outcome -1 as $P_B^- = P_B^{s=-1}$. The measurement direction B is noted in the Bloch sphere notation as $B = \vec{b}\vec{\sigma}$ for some unit vector $\vec{b}$.

For both correction term variants, the modified variant of the triangle inequality is considered.

$$\langle AB\rangle + \langle BC\rangle - \langle AC\rangle + \Gamma_i \leq 1, \tag{5.3.3}$$

where Γ_0 stands for the correction terms derived by Gühne et al. and Γ_1 for the correction terms from Kujala et al. The calculation for the correction terms presented by Kujala et al. is performed in section 5.3.1. The calculation for the error terms by Gühne et al. can be found in section 5.3.2.

5.3.1 Calculation for the Correction Terms by Kujala et al.

For the correction terms introduced in section 5.2, the modified triangle inequality 5.3.3 takes the form

$$\langle AB\rangle + \langle BC\rangle - \langle AC\rangle - |\langle A\rangle_B - \langle A\rangle_C| - |\langle B\rangle_A - \langle B\rangle_A| - |\langle C\rangle_A - \langle C\rangle_B| \leq 1. \quad (5.3.4)$$

When we calculate the correction terms, we have to take into account that the measurement sequences are fixed in the order, shown in equation 5.3.4, i.e. the measurement A is always performed first, and the measurement C always last. This fixed measurement sequence has consequences for the correction terms, in the sense that the first correction terms is simple to calculate. Hence let us start with the terms of the form $\langle A\rangle_i$. Because A is always the first measurement that is carried out, it follows that

$$\begin{aligned}
\langle A\rangle_B &= \sum_{a,b} a \cdot p\,(A = a, B = b) = \sum_a a \cdot p_B\,(A = a) \\
&= \sum_{a,b} a \,\mathrm{tr}\,[P_a \varrho P_a P_b] \\
&= \sum_a a \,\mathrm{tr}\left[P_a \varrho P_a \sum_b P_b\right] \\
&= \sum_a a \,\mathrm{tr}\,[P_a \varrho P_a] \\
&= \frac{1}{2}\sum_a \mathrm{tr}\,[\varrho\,(a\mathbb{1} + A)] \\
&= \mathrm{tr}\,[\varrho A] = (\vec{\varrho}\vec{a})\,. \qquad (5.3.5)
\end{aligned}$$

The term $\langle A\rangle_C$ is calculated analogously and gives the same result as $\langle A\rangle_B$, which means that for the first correction term $|\langle A\rangle_B - \langle A\rangle_C| = 0$ is true independently of the chosen measurements A, B, C or initial state ϱ.

To get the other terms needed to calculate the correction terms, is it sufficient to determine the term $\langle C\rangle_B$, as $\langle C\rangle_A$ and $\langle B\rangle_A$ have to be calculated equivalently to $\langle C\rangle_B$, while $\langle B\rangle_C$ can be calculated in the same way as the terms $\langle A\rangle_i$. For the term $\langle C\rangle_B$, we have

$$
\begin{aligned}
\langle C\rangle_B &= \sum_{c,b} c\cdot p\left(C=c, B=b\right) = \sum_{b,c} c\,\mathrm{tr}\left[P_b\varrho P_b P_c\right] \\
&= \sum_{b,c} c\,\mathrm{tr}\left[\varrho P_b P_c P_b\right] \\
&= \frac{1}{8}\sum_{b,c} c\,\mathrm{tr}\left[\varrho\left(\mathbb{1}+bB\right)\left(\mathbb{1}+cC\right)\left(\mathbb{1}+bB\right)\right] \\
&= \frac{1}{8}\sum_{b,c} c\,\mathrm{tr}\left[\varrho\left(2\mathbb{1}+2bB+cC+cb\{C,B\}+cBCB\right)\right] \\
&= \frac{1}{2}\,\mathrm{tr}\left[\varrho\left(C+BCB\right)\right] \\
&= \frac{1}{4}\,\mathrm{tr}\left[\left(\mathbb{1}+\varrho_i\sigma_i\right)\left(c_j\sigma_j+b_l\sigma_l c_m\sigma_m b_n\sigma_n\right)\right] \\
&= \frac{1}{4}\,\mathrm{tr}\left[\varrho_i c_j\sigma_i\sigma_j+\varrho_i b_l c_m b_n\sigma_i\sigma_l\sigma_m\sigma_n\right] \\
&= \frac{1}{4}\,\mathrm{tr}\left[\varrho_i c_i+\varrho_i b_l c_m b_n\sigma_i\sigma_l\left(\delta_{mn}+i\epsilon_{mnk}\sigma_k\right)\right] \\
&= \frac{1}{4}\,\mathrm{tr}\left[\varrho_i c_i+\varrho_i b_i c_m b_m+\left(\delta_{ml}\delta_{ni}-\delta_{mi}\delta_{nl}\right)\varrho_i b_l c_m b_n\right] \\
&= \left(\vec{\varrho}\vec{b}\right)\left(\vec{c}\vec{b}\right). \qquad (5.3.6)
\end{aligned}
$$

For the calculations in equation 5.3.6, it was used that $c^2 = b^2 = 1$, $B^2 = C^2 = \mathbb{1}$, $\mathrm{tr}\left[\sigma_i\right] = 0$ and $\left\|\vec{b}\right\| = \|\vec{c}\| = 1$.
As mentioned, the terms $\langle B\rangle_A$ and $\langle C\rangle_A$ are calculated equally, due to symmetry reasons. The results are

$$\langle C\rangle_A = \left(\vec{\varrho}\vec{a}\right)\left(\vec{c}\vec{a}\right), \qquad (5.3.7)$$

and

$$\langle B\rangle_A = \left(\vec{\varrho}\vec{a}\right)\left(\vec{b}\vec{a}\right). \qquad (5.3.8)$$

For the last term $\langle B\rangle_C$, we get analogous to 5.3.5

$$\langle B\rangle_A = \left(\vec{\varrho}\vec{b}\right). \qquad (5.3.9)$$

Finally, we get the modified triangle inequality by inserting the correction terms calculated above and setting $\langle AB\rangle = \vec{a}\vec{b}$, which is true in general for the qubit case with

projective measurements

$$\begin{aligned}
&\langle AB\rangle + \langle BC\rangle - \langle AC\rangle - |\langle A\rangle_B - \langle A\rangle_C| - |\langle B\rangle_A - \langle B\rangle_A| - |\langle C\rangle_A - \langle C\rangle_B| \\
&= \vec{a}\vec{b} + \vec{b}\vec{c} - \vec{a}\vec{c} - \left|\vec{\varrho}\vec{b} - (\vec{\varrho}\vec{a})\left(\vec{a}\vec{b}\right)\right| - \left|(\vec{\varrho}\vec{a})\,(\vec{a}\vec{c}) - \left(\vec{\varrho}\vec{b}\right)\left(\vec{c}\vec{b}\right)\right| \\
&\leq 1. \qquad (5.3.10)
\end{aligned}$$

It is obvious that the Bloch vectors of the measurements cannot be chosen such that the correction terms vanish independently of the qubit state ϱ. However, it is possible to avoid the additional terms in equation 5.3.10 by choosing appropriate Bloch vectors for the state and the measurement directions. If one e.g. chooses $\vec{\varrho} \perp \vec{a}$ and $\vec{\varrho} \perp \vec{b}$ (i.e. the state is orthogonal to the plane spanned by the measurements A and B), all of the correction terms are set to zero.

One possibility is to choose $\vec{\varrho} = \vec{e}_y$, and the measurement vectors $\vec{i}$ to be in the x-z-plane. We can then describe the vectors $\vec{i}$ by angles φ_i. We have $\vec{a} = \sin\varphi_1\vec{e}_x + \cos\varphi_1\vec{e}_z$, $\vec{b} = \sin\varphi_2\vec{e}_x + \cos\varphi_2\vec{e}_z$ and $\vec{c} = \sin\varphi_3\vec{e}_x + \cos\varphi_3\vec{e}_z$. As the correction terms vanish in this configuration, inequality 5.3.10 takes the easy form

$$\cos(\varphi_1 - \varphi_2) + \cos(\varphi_2 - \varphi_3) - \cos(\varphi_1 - \varphi_3) \leq 1. \qquad (5.3.11)$$

We can now investigate, whether the correction terms work as expected, i.e. whether we recover the quantum bound for compatible measurements. For this, we try to maximize equation 5.3.11. Since equation 5.3.11 only depends on the differences of the angles (only depends on $\varphi_i - \varphi_j$), we can set w.l.o.g. the angle $\varphi_1 = 0$. Therefore, we have to maximize

$$\cos\varphi_2 + \cos(\varphi_2 - \varphi_3) - \cos\varphi_3\,. \qquad (5.3.12)$$

The maximal value under this configuration is 1.5, which is equal to the maximal violation of the triangle inequality without correction terms in the qubit case with projective measurements. The angles for which we reach the maximal violation are $\varphi_1 = 0$, $\varphi_2 = \frac{\pi}{3}$ and $\varphi_3 = \frac{2\pi}{3}$. We find that the correction terms do not work, even for the simplest case.

5.3.2 Calculation for the Correction Terms by Gühne et al.

Let us now have a look at the correction introduced in section 5.1. If we apply these terms to equality 5.3.3, we get

$$\langle AB\rangle + \langle BC\rangle - \langle AC\rangle \leq 1 + 2\left(p^{err}\left(BAB\right) + p^{err}\left(CBC\right) + p^{err}\left(CAC\right)\right). \tag{5.3.13}$$

Due to symmetry reasons, it is sufficient to calculate the first correction term $p^{err}\left(BAB\right)$. The other terms are calculated analogously. We have

$$\begin{aligned}
p^{err}\left[B_1A_2B_3\right] &= p\left[B_1^+, B_3^- | B_1A_2B_3\right] + p\left[B_1^-, B_3^+ | B_1A_2B_3\right] \\
&= \sum_a \mathrm{tr}\left[\varrho\left(P_+^B P_a P_-^B P_a P_+^B + P_-^B P_a P_+^B P_a P_-^B\right)\right] \\
&= \frac{1}{32}\sum_a \mathrm{tr}\left[\varrho\left(\left(\mathbb{1}+B\right)\left(\mathbb{1}+aA\right)\left(\mathbb{1}-B\right)\left(\mathbb{1}+aA\right)\left(\mathbb{1}+B\right)\right)\right] \\
&+ \mathrm{tr}\left[\varrho\left(\left(\mathbb{1}-B\right)\left(\mathbb{1}+aA\right)\left(\mathbb{1}+B\right)\left(\mathbb{1}+aA\right)\left(\mathbb{1}-B\right)\right)\right] \\
&= \frac{1}{16}\left(4 - 2\left(\mathrm{tr}\left[\varrho BABA\right] + \mathrm{tr}\left[\varrho ABAB\right]\right)\right).
\end{aligned} \tag{5.3.14}$$

For simplicity, we consider the terms $\mathrm{tr}\left[\varrho BABA\right]$ and $\mathrm{tr}\left[\varrho ABAB\right]$ individually. For the first term, we have

$$\begin{aligned}
2\,\mathrm{tr}\left[\varrho BABA\right] &= \mathrm{tr}\left[\left(\mathbb{1} + \varrho_i\sigma_i\right)\left(b_j\sigma_j a_k\sigma_k b_l\sigma_l a_m\sigma_m\right)\right] \\
&= \mathrm{tr}\left[\left(\mathbb{1} + \varrho_i\sigma_i\right)\left(b_j a_k b_l a_m\left(\delta_{jk} + i\epsilon_{jkn}\sigma_n\right)\left(\delta_{lm} + i\epsilon_{lmo}\sigma_o\right)\right)\right] \\
&= \mathrm{tr}\left[\left(\mathbb{1} + \varrho_i\sigma_i\right)\left(2\left(\vec{b}\vec{a}\right)^2 - \mathbb{1} + 2i\epsilon_{lmo}b_l a_m\sigma_o\left(\vec{b}\vec{a}\right)\right)\right] \\
&= 4\left(\vec{b}\vec{a}\right)^2 - 2 + 4i\left(\left(\vec{b}\times\vec{a}\right)\vec{\varrho}\right)\left(\vec{b}\vec{a}\right).
\end{aligned} \tag{5.3.15}$$

For the second term, we get after a similar calculation

$$2\,\mathrm{tr}\left[\varrho ABAB\right] = 4\left(\vec{a}\vec{b}\right)^2 - 2 + 4i\left(\left(\vec{a}\times\vec{b}\right)\vec{\varrho}\right)\left(\vec{a}\vec{b}\right). \tag{5.3.16}$$

Inserting 5.3.15 and 5.3.16 into equation 5.3.14 gives us the final result for the correction

term, which is

$$p^{err}\left[B_1A_2B_3\right] = \frac{1}{2}\left(1 - \left(\vec{b}\vec{a}\right)^2\right). \tag{5.3.17}$$

It stands out that the error terms are independent of the state ϱ and vanish for compatible measurements. (In the case of projective measurements for qubits, two measurements are compatible (i.e. they commute) if the Bloch vectors describing the measurements direction are parallel or antiparallel.)

As explained above, the other correction terms can be calculated analogously. We are now able to rewrite inequality 5.3.13 in the following way

$$\vec{a}\vec{b}\left(1 + \vec{a}\vec{b}\right) + \vec{b}\vec{c}\left(1 + \vec{b}\vec{c}\right) - \vec{a}\vec{c}(1 - \vec{a}\vec{c}) \leq 4. \tag{5.3.18}$$

Again, we are interested in the maximum we can reach with projective, non-commuting observables. To find the maximal value of equation 5.3.18, let us consider the following function.

$$x = \cos\vartheta\,(1 + \cos\vartheta) + \cos\eta\,(1 + \cos\eta) - \cos(\xi)\,(1 - \cos(\xi)), \tag{5.3.19}$$

with $\vec{a}\vec{b} = \cos\vartheta$, $\vec{b}\vec{c} = \cos\eta$ and $\vec{a}\vec{c} = \cos(\xi)$ with the additional constraint $\xi \leq \vartheta + \eta$. We can maximize equation 5.3.19 numerically, using `Mathematica`. The maximal value that can be achieved is 4, which is the result of several different numerical methods, and is equal to the bound for compatible measurements. The maximal value is achieved for $\vartheta = \eta = \xi = 0$. This however means that basically all three measurements are the same, hence it is trivial that the noncontextual bound (the quantum bound for compatible measurements) is reached for this scenario. We find that the correction terms by Gühne et al. work for this case.

5.4 Simulation of the Correction Terms

In section 5.3, we calculated the correction terms introduced in sections 5.1 and 5.2 for the qubit case and thereby restricted ourselves to PVMs. However, generally neither are quantum systems restricted to the qubit case, nor are the measurements performed on the system necessarily described by Projective Measurements. Therefore, we want to

analyze the correction terms for a more general setup, where the dimension of the system is not bounded and the measurements can be arbitrary $POVM$s. (The measurements considered in the following simulation are dichotomic observables. This is due to the fact that the error terms by Kujala et al. [17] are derived specifically for the case of dichotomic measurements. Further, we can always perform a coarse graining for a given $POVM$, to describe a measurement as a dichotomic measurement.)
In the following subsections, configurations of this form are simulated using the $CVXOPT$ module for `PYTHON`.

5.4.1 Commonalities of the Simulations

In all simulations x is a vector, where each element x_i corresponds to one probability of the form $p\left(O_aO_bO_c|S_aS_bS_c\right)$. (I.e. the set of probabilities, completely describing the system.) Each element of the vector x thus describes a probability for a given measurement scenario. $p\left(O_aO_bO_c|S_aS_bS_c\right)$ is the probability, that measuring the sequence of observables S_a, S_b, S_c results in the outcomes O_a, O_b, O_c. There is however a difference between the two simulation. In the simulation of the error terms by Gühne et al., $S_i \in \{0,1,2,3\}$ $\forall i \in a,b,c$, where $S_i = 1$ means that observable A is measured, while $S_i = 2$ means measurement of observable B and $S_i = 3$ measurement of observable C. $S_i = 0$ means that no measurement is performed (E.g. $p\left(O_aO_bO_c|132\right) = p\left(O_aO_bO_c|ACB\right)$ etc.). This is necessary to be able to describe the correction terms of the form $p\left(B_1^+B_3^-|B_1A_2B_3\right)$. The simulation of the correction terms by Kujala et al. was performed in two different ways. In the first simulation, $S_i \in 0,1$, where $S_i = 1$ means that observable i is measured. $S_i = 0$ again means that no measurement is performed (E.g. $p\left(O_aO_bO_c|101\right) = p\left(O_aO_bO_c|A\mathbb{1}C\right)$). In the second simulation of the error terms by Kujala et al., the configuration is the same as in the simulation of the error terms by Gühne et al., where S_c is omitted since we only consider measurement sequences of length two.

In all simulations, the matrix G, together with the vector h describe the positivity condition of the probabilities x_i. G is a diagonal matrix, with diagonal elements -1, while the vector $h = 0$ is the zero vector. This gives us the condition $Gx = -x \leq 0$. Comparing this with the definition of a linear program (see definition 37), we see that G and h describe the inequality part $Gx + s = h$ with $s \succeq 0$.

Another consistency in all simulations is the matrix A, which together with the vector b describes the constraints on the elements of x, imposed by signalling in time in one direction. These constraint are depicted by the following three equations

$$1 = p\left(000|000\right) = \sum_a p\left(a00|S_a00\right) \qquad \forall S_a, \tag{5.4.1}$$

$$p\left(a00|S_a00\right) = \sum_b p\left(ab0|S_aS_b0\right) \qquad \forall a, S_a, S_b, \tag{5.4.2}$$

$$p\left(ab0|S_aS_b0\right) = \sum_c p\left(abc|S_aS_bS_c\right) \qquad \forall a, b, S_a, S_b. \tag{5.4.3}$$

The matrix A is constructed such that $Ax = b$ recovers the equations 5.4.1-5.4.3. (E.g. we set $A_{1j}x_j = \sum_a p\left(a00|100\right) = b_1 = 1$ and so forth). Together, A and b describe the equality conditions in definition 37.
After explaining the commonalities between the different simulations, the results of the individual simulations will be shown in the next three sections.

5.5 Simulation of the Correction Terms by Gühne et al.

As explained in the previous section, the vector x consists of all possible probabilities $p\left(O_aO_bO_b|S_aS_bS_c\right)$ with $S_i \in \{0,1,2,3\}$ and $i \in \{0,1\}$.
Since the linear program, as defined in definition 35 minimizes a function, the considered function in this simulation is no longer equation 5.3.13. Instead, the following function, which is simply equation 5.3.13 multiplied by a factor -1 is minimized.

$$-\langle AB\rangle - \langle BC\rangle + \langle AC\rangle + 2\left(p^{err}\left(BAB\right) + p^{err}\left(CBC\right) + p^{err}\left(CAC\right)\right) \geq -1 \tag{5.5.1}$$

In this simulation, an expectation value is e.g. described by

$$\langle AB\rangle = p\left(000|120\right) + p\left(110|120\right) - p\left(100|120\right) - p\left(010|120\right), \tag{5.5.2}$$

while the error terms are written like

$$p^{err}(BAB) = p(001|212) + p(100|212) + p(011|212) + p(110|212). \tag{5.5.3}$$

The vector c consists of elements x_i taking the values $\{0, \pm 1\}$ such that $c^T x$ assigns the pre-factors $\{0, \pm 1\}$ to the elements x_i such that inequality 5.5.1 is recreated.
The result of the simulation is

$$\max\left[\langle AB\rangle + \langle BC\rangle - \langle AC\rangle - 2\left(p^{err}(BAB) - p^{err}(CBC) - p^{err}(CAC)\right)\right] = 3 > 1, \tag{5.5.4}$$

which is the algebraic maximum of the original triangle inequality and of equation 5.3.13. It follows that all error terms can be cancelled (set to zero), while the rest of the terms can be chosen such that the maximum is reached. Obviously, the correction terms fail for the case of general measurements and infinite dimension.

5.6 Simulation of the Correction Terms by Kujala et al. Version 1

In this simulation, the vector x consists of all probabilities $p(O_a O_b O_b | S_a S_b S_c)$ with $S_i \in \{0, 1\}$ and $i \in \{0, 1\}$. As before, we do not maximize equation 5.3.4, instead we minimize we multiply equation 5.3.4 by a factor -1 and minimize the resulting function. For convenience however, we write down the original inequality instead.
When we want to simulate the modified triangle inequality 5.3.4, we have the problem that neither a linear program, nor a quadratic program is able to simulate the terms with the absolute value, i.e. the correction terms. We therefore have to estimate the correction terms in a form that we can simulate by the means of a quadratic program. The estimation used in the following two simulations is simple and is given as

$$|a| \leq \frac{1}{2}a^2 \quad \forall a \in [-2, 2], \tag{5.6.1}$$

with $a = \langle A\rangle_B - \langle A\rangle_C$ etc. The estimation terms of the form $\frac{1}{2}a^2$ can be simulated by a quadratic program by choosing an appropriate matrix P such that $x^T P x$ assigns the correct pre-factors $0, \pm 1$ for the probabilities (i.e. elements of x)[1]. The matrix

[1] The term $x^T P x$ describes the quadratic part of the function to be optimized in definition 38

P was constructed in the following way. First, in the same way as in the previous section, we constructed two vectors k and l, consisting of $0, \pm 1$-values elements such that a multiplication with the vector x leaves only the probabilities needed to describe the correction terms

$$\begin{aligned} \langle B\rangle_A - \langle B\rangle_C &= k^T x && (5.6.2)\\ &= p\,(000|110) - p\,(010|110) + p\,(100|110) - p\,(110|110) && (5.6.3)\\ &\quad - \left(p\,(000|010) + p\,(010|010)\right). && (5.6.4) \end{aligned}$$

The correction term $\langle C\rangle_A - \langle C\rangle_B$ is described analogously by $l^T x$. The matrix $P = K + L$ then consists of the two matrices K and L where

$$K\,[i,j] = k\,[i]\,k\,[j]\,, \tag{5.6.5}$$

and

$$L\,[i,j] = l\,[i]\,l\,[j]\,. \tag{5.6.6}$$

Here P is splitted in two parts, because the two vectors describing the error terms should not mix with each other.
The rest of inequality 5.3.4 is created by the vector q in the same way as the inequality is created by the vector c in section 5.5.
The maximal value determined by the simulation is shown in equation 5.6.7

$$\begin{aligned} &\max\left[AB\rangle + \langle BC\rangle - \langle AC\rangle - \frac{1}{2}\left(\langle A\rangle_B - \langle A\rangle_C\right)^2 - \frac{1}{2}\left(\langle B\rangle_A - \langle B\rangle_A\right)^2 - \frac{1}{2}\left(\langle C\rangle_A - \langle C\rangle_B\right)^2\right]\\ &= 3 > 1. \end{aligned} \tag{5.6.7}$$

Again, the maximally achievable value is equal to the algebraic bound of the unmodified triangle inequality. As in section 5.5, it is possible to cancel the error terms in this most general set-up, while the rest of the terms in the inequality can be chosen such that the algebraic bound of the unmodified LG inequality is reached. Again the correction terms fail for the case of general measurements and infinite dimension.

5.7 Simulation of the Correction Terms by Kujala et al. Version 2

As explained in section 5.4.1, we use the same form for the vector x as in section 5.5. The probabilities p are described as in section 5.5 with $S_a, S_b \in \{0, 1, 2, 3\}$. The measurement direction S_c is omitted as we only want to describe measurement sequences of length two. The matrix P is defined exactly the same way as in section 5.6, adjusted to the different description of the probabilities. The final result is, as expected, equal to the result of the first version of the simulation in section 5.6

$$\max \left[AB\rangle + \langle BC\rangle - \langle AC\rangle - \frac{1}{2}\left(\langle A\rangle_B - \langle A\rangle_C\right)^2 - \frac{1}{2}\left(\langle B\rangle_A - \langle B\rangle_A\right)^2 - \frac{1}{2}\left(\langle C\rangle_A - \langle C\rangle_B\right)^2 \right] = 3 > 1. \tag{5.7.1}$$

We find that the correction terms by Kujala et al. do not change the algebraic bound, independently of the description of the system (the chosen simulation). This was to be expected, since we showed in section [5.3.1] that the correction terms already do not recover the quantum bound for qubit systems and projective measurements.

5.8 Summary

In section 5.3.1, we have seen that the correction terms proposed by Kujala et al. are unable to recover the quantum bound for compatible measurements even for the simple case of qubit systems and projective measurements. Instead, they even allow the maximal violation. On the other hand, we have seen in section 5.1 that the correction terms proposed by Gühne et al. do not allow quantum mechanics (Qubits and PVMs) to violate inequality 5.3.3 and therefore recover the bound for compatible measurements.
In section 5.4, we see that the simulations give us the algebraic maximum for equation 5.3.3, independently of the form of the correction terms. Essentially this means that both correction terms are unable to recover the quantum bound for compatible measurements if we allow general $POVM$s and consider quantum systems without dimension constraints. Noisy/incompatible measurements can contribute to the violation of an noncontextuality inequality. This possibility can not be ruled out (or corrected) by this correction terms.

Since the correction terms by Kujala et al. are already unable to recover the bound for PVMs and two-dimensional systems it seems unlikely that these correction terms can recover the bound if we somehow give additional constraints on the system. For the terms proposed by Gühne et al. however, it is still possible that the correction terms may recover the quantum bound under certain additional assumptions, which we were not able not simulate, e.g. for a fixed dimension and arbitrary $POVM$ or for an unbounded dimension and PVMs.

6 Conclusion and Outlook

In this chapter, we want to summarize the main results of this thesis and give an outlook to several open problems.

Firstly, we showed in proposition 41 that the extremal points of the temporal correlation polytope $P_l^{O,S}$ are given by deterministic assignments for arbitrary numbers of outcomes O, settings S and length l. In section 4.2.2, we further presented a way to compute all extremal points, i.e. write them explicitly and presented a formula to calculate the number of extremal points of the polytope for generalized settings.

The main result of chapter 4 is that although unrestricted quantum mechanics (i.e. no bound on the dimension) fills out the temporal correlation polytope $P_l^{O,S}$, a qubit is unable to reach all extremal points of $P_2^{2,2}$ for qubit systems. To show this we used the reasonable assumption $E_{ax} = E_{by}$ if $a = b$ and $x = y$, i.e. two measurement settings and we can either perform the same measurement again or perform the other one. In table 4.7, we presented the 10 equivalence classes of extremal points of the polytope $P_2^{2,2}$, of which 6 can be reached with qubit systems, while the remaining 4 equivalence classes under symmetry transformations can only be reached by qutrit systems or higher.

Since there exists some extremal points that cannot be reached by qubit systems, we were further able to use these points to construct dimension witnesses to distinguish between systems of dimension two or higher. For this, we presented an explicit example of a dimension witness, which we called temporal Bell inequality in section 4.8. Although we were not able to find an analytical upper bound for the temporal Bell-operator, we performed a numerical maximization with `Mathematica` in section 4.8.1.

Let us now recap the results of chapter 5. In this chapter, we analyzed the effects of two proposed correction terms for not perfectly compatible measurements in noncontextuality experiments. In section 5.3, we found that the correction terms proposed by Gühne et al. [16], are able to recover the quantum bound for compatible measurements for qubit-systems and projective measurements. On the other hand, the correction terms

by Kujala et al. [17], cannot recover the quantum bound even for this most simple scenario. For these correction terms it was easy to find a violation, without looking for the global maximum. The simulations from section 5.4 reached the algebraic maximum for both kinds of correction terms. However, this result was to be expected. On the one hand, the correction terms by Kujala et al. already failed for the most simple setup and on the other hand, we know from that quantum mechanics without bounds on the dimension fills out the temporal correlation polytope so it was to be assumed that the maximum can be reached without any restrictions of the quantum system.

The thesis also leaves some open questions, which may be of further interest. Again, we can divide the outlooks in two parts, one dealing with open questions from chapter 4 and one with the ones from chapter 5.
Let us first have a look at chapter 4. For sure, one the most interesting open questions is the question, whether we can use the temporal correlation polytope to create further dimension witnesses that can distinguish between systems of dimension three and four. E.g. In this thesis, we found that some extremal points of $P_2^{2,2}$ can only be reached by qutrit-systems, so it is reasonable to assume that if we extend the length to 3 that some points can only be reached with ququarts. Another open problem is to find an analytical upper bound for the temporal Bell-operator introduced in section 4.8 such that experiments testing the inequality can be performed.
Chapter 5 also leaves us with some open questions. We know that the correction terms proposed by Gühne et al. work for projective measurements on qubit systems and no longer work for quantum systems of infinite dimension and general $POVM$s but we do not know yet anything about the cases in between these two extremes. It would be interesting to test the terms e.g. for higher, but finite dimensions and general $POVM$s or for infinite dimensions and PVMs. Another open problem is the question, whether we can construct "better" correction terms (i.e. correction terms that work for more cases than the ones investigated in this thesis).

Bibliography

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

[2] J. S. Bell: On the Einstein-Podolsky-Rosen paradox. *Physics*, 195 (2004)

[3] B. Hensen et al.: Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometers. *Nature*, **526**, 682 (2015)

[4] M. Giustina et al: Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons. *Phys. Rev. Lett.*, **115**, 250401 (2015)

[5] L. K. Shalm, et al.: Strong Loophole-Free Test of Local Realism. *Phys. Rev. Lett.*, **115**, 250402 (2015)

[6] R. W. Spekkens: Evidence for the epistemic view of quantum states: A toy theory. *Phys. Rev. A*, **75**, 032110 (2007)

[7] S. Kochen, E. Specker: The Problem of Hidden Variables in Quantum Mechanics. *Indiana Univ. Math. J.*, **17**, 59 (1968)

[8] J. Kofler, Č. Brukner: Condition for macroscopic realism beyond the Leggett-Garg inequalities. *Phys. Rev. A*, **87**, 052115 (2013)

[9] T. Fritz: Quantum correlations in the temporal Clauser–Horne–Shimony–Holt (CHSH) scenario. *New Journal of Physics*, **12**, 083055 (2010)

[10] L. Clemente, J. Kofler: No Fine Theorem for Macrorealism: Limitations of the Leggett-Garg Inequality. *Phys. Rev. Lett.*, **116**, 150401 (2016)

[11] C. Budroni, T. Moroder, M. Kleinmann, O. Gühne: Bounding Temporal Quantum Correlations. *Phys. Rev. Lett.*, **111**, 020403 (2013)

[12] N. Brunner, S. Pironio, A. Acín, N. Gisin, A. A. Méthot, V. Scarani: Testing the Dimension of Hilbert Spaces. *Phys. Rev. Lett.*, **100**, 210503 (2008)

[13] P. W. Shor, J. Preskill: Simple Proof of Security of the BB84 Quantum Key Distribution Protocol. *Phys. Rev. Lett.*, **85**, 441 (2000)

[14] B. Kraus, N. Gisin, R. Renner: Lower and Upper Bounds on the Secret-Key Rate for Quantum Key Distribution Protocols Using One-Way Classical Communication. *Phys. Rev. Lett.*, **95**, 080501 (2005)

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

[16] O. Gühne, M. Kleinmann, A. Cabello, J.-Å. Larsson, G. Kirchmair, F. Zähringer, R. Gerritsma, C. F. Roos: Compatibility and noncontextuality for sequential measurements. *Phys. Rev. A*, **81**, 022121 (2010)

[17] J. V. Kujala, E. N. Dzhafarov, J.-Å. Larsson: Necessary and Sufficient Conditions for an Extended Noncontextuality in a Broad Class of Quantum Mechanical Systems. *Phys. Rev. Lett.*, **115**, 150401 (2015)

[18] J. F. Clauser, M. A. Horne, A. Shimony, R. A. Holt: Proposed Experiment to Test Local Hidden-Variable Theories. *Phys. Rev. Lett.*, **23**, 880 (1969)

[19] B. S. Cirel'son: Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, **4**, 93 (1980)

[20] S. Popescu, D. Rohrlich: Quantum nonlocality as an axiom. *Foundations of Physics*, **24**, 379 (1994)

[21] T. Heinosaari, M. Ziman: The Mathematical Language of Quantum Theory. *Cambridge University Press* (2012)

[22] M. M. Wolf: Quantum Channels and Operations. *Script* (2012)

[23] O. Gühne: Grundlagen der Quantenmechanik. *Script* (2013)

[24] W. Heisenberg: The Physical Principles of the Quantum Theory. *Courier Corporation* (1930)

[25] D. Bohm: A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. I. *Phys. Rev.*, **85**, 166 (1952)

[26] A. Cabello: Interpretations of quantum theory: A map of madness. *arXiv:1509.04711v1 [quant-ph]* (2015).

[27] M. Kleinmann, C. Budroni, J.-A. Larsson, O. Gühne, A. Cabello: Optimal Inequalities for State-Independent Contextuality. *Phys. Rev. Lett.*, **109**, 250402 (2012)

[28] Y.-C. Liang, R. W. Spekkens, H. M. Wiseman: Specker's parable of the overprotective seer: A road to contextuality, nonlocality and complementarity. *Physics Reports*, **506**, 1 (2011)

[29] S. Yu, C. H. Oh: Quantum contextuality and joint measurement of three observables of a qubit. *arXiv:1312.6470v1 [quant-ph]* (2013).

[30] R. W. Spekkens: Contextuality for preparations, transformations, and unsharp measurements. *Phys. Rev. A*, **71**, 052108 (2005)

[31] S. Brierley, A. Kosowski, M. Markiewicz, T. Paterek, A. Przysiężna: Nonclassicality of Temporal Correlations. *Phys. Rev. Lett.*, **115**, 120404 (2015)

[32] M. Markiewicz, P. Kurzyński, J. Thompson, S.-Y. Lee, A. Soeda, T. Paterek, D. Kaszlikowski: Unified approach to contextuality, nonlocality, and temporal correlations. *Phys. Rev. A*, **89**, 042109 (2014)

[33] D. Bohm, Y. Aharonov: Discussion of Experimental Proof for the Paradox of Einstein, Rosen, and Podolsky. *Phys. Rev.*, **108**, 1070 (1957)

[34] A. N. Kolmogorov: Foundations of the Theory of Probability. *Chelsea Publishing Company* (1950)

[35] L. de Broglie: An introduction to the Study of Wave Mechanics. *E. P. Dutton and Company Inc., New York* (1930)

[36] J. S. Bell: On the Problem of Hidden Variables in Quantum Mechanics. *Rev. Mod. Phys.*, **38**, 447 (1966)

[37] J. von Neumann: Mathematische Grundlagen der Quantenmechanik. *Springer* (1932)

[38] J. M. Jauch, C. Piron: Can Hidden Variables be Excluded in Quantum Mechanics. *Helv. Phys. Acta*, **36**, 827 (1963)

[39] A. Gleason: Measures on the Closed Subspaces of a Hilbert Space. *Indiana Univ. Math. J.*, **6**, 885 (1957)

[40] L. J. Landau: On the violation of Bell's inequality in quantum theory. *Physics Letters A*, **120**, 54 (1987)

[41] A. Cabello, J. Estebaranz, G. García-Alcaine: Bell-Kochen-Specker theorem: A proof with 18 vectors. *Physics Letters A*, **212**, 183 (1996)

[42] N. D. Mermin: Simple unified form for the major no-hidden-variables theorems. *Phys. Rev. Lett.*, **65**, 3373 (1990)

[43] A. Peres: Incompatible results of quantum measurements. *Physics Letters A*, **151**, 107 (1990)

[44] A. A. Klyachko, M. A. Can, S. Binicioğlu, A. S. Shumovsky: Simple Test for Hidden Variables in Spin-1 Systems. *Phys. Rev. Lett.*, **101**, 020403 (2008)

[45] A. Cabello: Experimentally Testable State-Independent Quantum Contextuality. *Phys. Rev. Lett.*, **101**, 210401 (2008)

[46] P. Badziag, I. Bengtsson, A. Cabello, I. Pitowsky: Universality of State-Independent Violation of Correlation Inequalities for Noncontextual Theories. *Phys. Rev. Lett.*, **103**, 050401 (2009)

[47] X.-D. Yu, Y.-Q. Guo, D. M. Tong: A proof of the Kochen–Specker theorem can always be converted to a state-independent noncontextuality inequality. *New Journal of Physics*, **17**, 093001 (2015)

[48] G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, C. F. Roos: State-independent experimental test of quantum contextuality. *Nature*, **460**, 494 (2009)

[49] S. Yu, C. H. Oh: State-Independent Proof of Kochen-Specker Theorem with 13 Rays. *Phys. Rev. Lett.*, **108**, 030402 (2012)

[50] C. Budroni: Temporal Quantum Correlations and Hidden Variable Models. *Springer Theses* (2016)

[51] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, A. Zeilinger: Experimental non-classicality of an indivisible quantum system. *Nature*, **474**, 490 (2011)

[52] J. Ahrens, E. Amselem, A. Cabello, M. Bourennane: Two Fundamental Experimental Tests of Nonclassicality with Qutrits. *Scientific Reports*, **3**, 2170 (2013).

[53] G. Borges, M. Carvalho, P.-L. de Assis, J. Ferraz, M. Araújo, A. Cabello, M. T. Cunha, S. a. Pádua: Quantum contextuality in a Young-type interference experiment. *Phys. Rev. A*, **89**, 052106 (2014)

[54] M. Jerger, Y. Reshitnyk, M. Oppliger, A. Potočnik, M. Mondal, A. Wallraff, K. Goodenough, S. Wehner, K. Juliusson, N. K. Langford, A. Fedorov: Contextuality without nonlocality in a superconducting quantum system. *arXiv:1602.00440v2 [quant-ph]* (2016).

[55] O. Gühne, C. Budroni, A. Cabello, M. Kleinmann, J.-A. Larsson: Bounding the quantum dimension with contextuality. *Phys. Rev. A*, **89**, 062107 (2014)

[56] N. Brunner, M. Navascués, T. Vértesi: Dimension Witnesses and Quantum State Discrimination. *Phys. Rev. Lett.*, **110**, 150501 (2013)

[57] R. Gallego, N. Brunner, C. Hadley, A. Acín: Device-Independent Tests of Classical and Quantum Dimensions. *Phys. Rev. Lett.*, **105**, 230501 (2010)

[58] J. Ahrens, P. Badziąg, A. Cabello, M. Bourennane: Experimental estimation of the dimensions of classical and quantum systems. *Nature Phys.*, **8**, 588 (2012)

[59] M. Hendrych, R. Gallego, M. Mičuda, N. Brunner, A. Acín, J. P. Torres: Experimental estimation of the dimensions of classical and quantum systems. *Nature Phys.*, **8**, 588 (2012)

[60] S. Boyd, L. Vandenberghe: Convex Optimization. *Cambridge University Press* (2004)

[61] S. Lörwald, G. Reinelt: PANDA: a software for polyhedral transformations. *EURO Journal on Computational Optimization*, **3**, 297 (2015)

[62] T. Fritz: Beyond Bell's theorem: correlation scenarios. *New Journal of Physics*, **14**, 103001 (2012)

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie Zitate und Ergebnisse anderer kenntlich gemacht habe.

Ort Datum Unterschrift