Generalized Bell Inequalities and Quantum Entanglement

Dissertation

zur Erlangung des Grades eines Doktors
der Naturwissenschaften

vorgelegt von

Fabian Bernards

eingereicht bei der Naturwissenschaftlich-Technischen Fakultät
der Universität Siegen,
Siegen 2022

Betreuer und erster Gutachter

Prof. Otfried Gühne
Universität Siegen

Zweiter Gutachter

Prof. Adán Cabello
Universität Sevilla

Angehörige der Promotionskommission

Prof. Otfried Gühne
Prof. Adán Cabello
Prof. Mohamed Barakat
Dr. Michael Johanning

Tag der mündlichen Prüfung
Montag, 30. Mai 2022

Abstract

The focus of this thesis lies on Bell inequalities. We introduce the concept of generalizations of a Bell inequality, which are Bell inequalities that by construction perform at least as well at any given task as the Bell inequality they generalize. Further, we present the cone-projection technique that we use to find such generalizations of certain Bell inequalities. Specifically, we find all 3050 symmetric generalizations of the I3322 inequality to three parties and study their quantum mechanical properties. Some of them detect nonlocality of quantum states, for which all two-setting Bell inequalities fail to do so. Moreover, we find generalizations of the Svetlichny inequality, the I4422 inequality, the Guess-Your-Neighbors-Input inequality as well as Bell inequalities that simultaneously generalize the I3322 inequality and the Clauser-Horne-Shimony-Holt inequality. We study the quantum mechanical properties of all of those inequalities.

Furthermore, we investigate different hybrid models and present Bell inequalities to test them. We numerically estimate the noise robustness for all of these Bell inequalities. We also construct a family of Bell inequalities for a particular class of hybrid models. To simplify research on Bell inequalities, we present Bellpy, which is a Python library to construct and investigate facet-defining Bell inequalities.

Besides our work on Bell inequalities, we also investigate absolute maximally entangled Werner states and show that such states only exist for systems of two qubits and three qutrits. Finally, we analyze a variation of the Bose-Marletto-Vedral experiment, where two quantum-mechanically described beads interact gravitationally.

Zusammenfassung

Diese Dissertation ist hauptsächlich Bell'schen Ungleichungen gewidmet. Wir führen das Konzept der Verallgemeinerung einer Bell'schen Ungleichung ein. Eine Verallgemeinerung einer Bell'schen Ungleichung ist selbst eine Bell'sche Ungleichung. Das Konzept ist so definiert, dass jede Verallgemeinerung einer Bell'schen Ungleichung automatisch für jede Aufgabe mindestens ebenso gut geeignet ist, wie die Bell'sche Ungleichung, die sie verallgemeinert. Wir stellen außerdem eine Methode vor, mit der sich solche Verallgemeinerungen von Bell'schen Ungleichungen finden lassen.

Mithilfe dieser Methode finden wir alle 3050 symmetrischen Verallgemeinerungen der I3322 Ungleichung für drei Parteien und analysieren ihre quantenmechanischen Eigenschaften. Einige dieser Ungleichungen detektieren Nichtlokalität in Quantenzuständen, deren Nichtlokalität von keiner Bell'schen Ungleichung erkannt wird, bei der jede Partei nur die Wahl zwischen zwei verschiedenen Messeinstellungen hat. Darüber hinaus finden wir Verallgemeinerungen der I4422 Ungleichung, der Guess-Your-Neighbors-Input Ungleichung, als auch Bell'sche Ungleichungen für drei Parteien, die gleichzeitig sowohl die Clauser-Horne-Shimony-Holt Ungleichung als auch die I3322 Ungleichung verallgemeinern. Wir untersuchen die quantenmechanischen Eigenschaften all dieser Ungleichungen.

Wir betrachten außerdem verschiedene Hybridmodelle und stellen Bell'sche Ungleichungen vor, mit deren Hilfe sich diese Hybridmodelle testen lassen. Darüber hinaus konstruieren wir eine Familie von Bell'schen Ungleichungen für eine spezielle Klasse von Hybridmodellen. Um die Forschung an Nichtlokalität zu erleichtern, stellen wir Bellpy vor, ein Pythonmodul zur Konstruktion und Analyse von Bell'schen Ungleichungen.

Zusätzlich zu unserer Forschung an Nichtlokalität untersuchen wir absolut-maximal verschränkte Wernerzustände und zeigen, dass solche Zustände nur in Systemen mit zwei Qubits oder drei Qutrits existieren können. Zum Schluss analysieren wir eine Variante des Bose-Marletto-Vedral Experiments, bei dem zwei kleine, quantenmechanisch beschriebene Kugeln durch Schwerkraft miteinander wechselwirken.

Contents

Introduction

Bell nonlocality is a phenomenon that is exhibited in so-called Bell experiments. A Bell experiment involves at least two spatially separated laboratories, called parties. Besides the number of parties, Bell experiments are further distinguished by the number of measurement settings per party and the number of possible outcomes for each measurement. The basic description of a Bell experiment that arises from this information is called a scenario. A Bell experiment is conducted over several rounds. In each round, each party chooses a measurement setting and records this choice together with the obtained outcome. One may be tempted to assume that the measurement outcomes of each party are predetermined and one party's choice of measurement does not affect the outcome of another party's measurement. This assumption is called local realism, while a violation of local realism is called nonlocality [1]. Nonlocality is indicated by the violation of a so-called Bell-inequality [2], which is predicted in quantum mechanics. Quantum mechanics predicts nonlocality even if the parties are far apart such that no signal can be transmitted from one party to another during each round of the experiment according to the laws of special relativity. Experimentally, this prediction has been confirmed [3, 4, 5, 6].

Besides ruling out local realism, Bell inequalities have also been used to study the communication complexity for computational tasks that involve several parties [6, 7, 8]. Moreover, Bell inequalities can be used to demonstrate the difference between quantum mechanics and no-signaling theories in multi-player games [9]. In the field of encryption, Bell inequalities are useful for multiparty conference key agreement [10] and in the device independent testing of quantum states and measurements [11].

The focus of this thesis lies on Bell inequalities and more specifically on so-called facet-defining Bell inequalities. For any given scenario, nonlocality can be completely characterized using facet-defining Bell inequalities. However, their number drastically increases with the complexity of the scenario and finding all facet-defining Bell inequalities quickly becomes prohibitively challenging. Moreover, if the goal is not the complete characterization of nonlocality for a certain scenario, it may be unnecessary as well as impracticable to test all of the facet-defining Bell inequalities. This is for example the case if an experiment is set up to prove that a system exhibits nonlocality and a quantum description of this system is known at least approximately. In this case one would be interested in the Bell inequality that is expected to show the largest violation for the given system or the strongest robustness to noise that may

affect the system. In short, the goal is to find Bell inequalities that perform well for some predefined tasks.

We make two main contributions to achieving this goal. Firstly, we introduce the concept of generalizations of a Bell inequality. These are Bell inequalities that are at least as strong as the Bell inequality they generalize. Secondly, we devise a method to find all facet-defining Bell inequalities that satisfy a list of affine constraints. We call this method the cone-projection technique. In particular, the cone-projection technique can be used to find generalizations of a facet-defining Bell inequality.

The thesis is organized in the following way. Before we begin presenting the results of our research, we introduce the mathematical and physical notions that are most important for our research. In Chapter 1, we discuss convex polyhedra. The latter are important for the characterization of nonlocality. While investigating the properties of the Bell inequalities we found, we employ techniques from the field of convex optimization, which are discussed in Chapter 2. Nonlocality is predicted and explained in quantum mechanics through the concept of entanglement. We therefore briefly present the formalism of quantum mechanics in Chapter 3 with an emphasis on entanglement theory. With this we are prepared to give a more detailed explanation of Bell nonlocality in Chapter 4.

Starting from Chapter 5, we present the results of our own research. In Chapter 5, we introduce the most important concept of our work: generalizations of Bell inequalities. Further, we introduce the cone-projection technique that allows us to find generalizations of Bell inequalities.

We then prove the power of our method in numerous applications in Chapters 6 and 7. In Chapter 6, we find all 3050 symmetric generalizations of the I3322 inequality to three parties. For the simplest of these inequalities, we study their quantum mechanical properties and demonstrate that they detect nonlocality of quantum states, for which all two-setting Bell inequalities fail to do so.

In Chapter 7, we find generalizations of the I4422 inequality, the GYNI inequality, as well as Bell inequalities that simultaneously generalize the I3322 inequality and the CHSH inequality. We study the quantum mechanical properties of all of those inequalities including the generalizations of the I3322 inequality presented in Chapter 6.

In Chapter 8, we investigate different hybrid models and present Bell inequalities to test them. We numerically estimate the noise robustness for all of the Bell inequalities we found. Further, we construct a family of Bell inequalities for a particular class of hybrid models. Lastly, we present two generalizations of Svetlichny's Bell inequality.

In Chapter 9, we present Bellpy, which is a Python module to construct and investigate facet-defining Bell inequalities.

In Chapters 10 and 11 we shift our attention away from Bell inequalities. In Chapter 10 we investigate absolutely maximally entangled Werner states and show that such states only exist for systems of two qubits and three qutrits. In Chapter 11 we investi-

gate a variation of the proposed BMV experiment, where two quantum-mechanically described beads interact gravitationally. We analyse the entanglement dynamics of the system and confirm that the beads entangle if the gravitational coupling of the beads is strong enough. We further compute the point in time for which the entanglement between the beads is maximal.

1 Convex polyhedra

This chapter is dedicated to convex polyhedra and particularly to two of their subclasses, namely polyhedral cones and convex polytopes. From now on, we refer to these objects simply as polyhedra, cones and polytopes throughout this chapter. The results and definitions presented in this chapter are also covered in the book by Ziegler [12] or the lecture notes by Paffenholz [13] or by Fukuda [14] unless other references are provided. All of the former are excellent recources for this topic. The theory of convex polyhedra has vast applications. Polyhedra emerge for example as the domains of linear optimization problems, which include famous problems such as the traveling salesman problem [12, 15]. Crucially for our purposes, however, they are a central concept in the field of Bell nonlocality, as we will see later [16, 17].

1.1 Generators of convex polyhedra

The perhaps two most fundamental concepts in the study on convex polyhedra are *conic combinations* and *convex combinations*. Conic combinations are linear combinations with non-negative coefficients. Convex combinations are linear combinations with non-negative coefficients that sum up to one. Given a matrix $V \in \mathbb{R}^{d\times n}$, interpreted as a finite set $V = \{v_1, v_2, \ldots, v_n\}$ of column vectors, the set of all conic (convex) combinations of the elements in V is called the convex (conic) hull of V. The convex and conic hulls are denoted

$$\mathrm{conv}(V) := \left\{ \sum_{i=1}^{n} \lambda_i v_i \,|\, \sum_i \lambda_i = 1, \lambda_i \geq 0 \;\forall i \in \{1,\ldots,n\} \right\} \tag{1.1}$$

$$\mathrm{cone}(V) := \left\{ \sum_{i=1}^{n} \lambda_i v_i \,|\, \lambda_i \geq 0 \;\forall i \in \{1,\ldots,n\} \right\}. \tag{1.2}$$

Also, the convex (conic) hull of V is said to be generated by V under convex (conic) combinations and the elements of V are referred to as *generators*. If V is finite, its convex (conic) hull is called *finitely generated*.

These notions give rise to *cones* and *polytopes*. The convex hull of a finite set of vectors is called a polytope; the conic hull of a finite set of vectors, called *rays*, is called a cone. Given a cone or a polytope, the set of its generators V is in general not unique, as one can add convex combinations of the generators to the generating set without changing

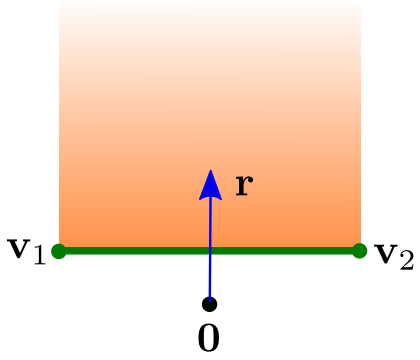


Figure 1.1: The vertices $\mathbf{v}_1$, $\mathbf{v}_2$ define a one-dimensional polytope (bold line). The Minkowski sum of this polytope and the ray $\mathbf{r}$ (blue) is a polyhedron (orange), which extends infinitely to the top. The polyhedron has three facets, which are the lines that confine it to the bottom, the left, and the right. This figure appeared in our paper [B].

the polytope or cone. Generators are called *redundant*, if they can be expressed as convex (conic) combinations of other generators. Otherwise, they are called *extremal*. The extremal generators together form the minimal generating set of the polytope or cone. This set is unique, whereby vectors v_1, v_2 that appear as generators of a cone are considered equivalent if $v_1 = lv_2$ for some $l > 0$. The extremal points of a polytope are called *vertices*.

If a cone is generated by the column vectors of a matrix $R \in \mathbb{R}^{d\times k}$, then R is called the *generating matrix* of that cone [18]. The expression $\mathrm{conv}(V)$ is called the *V-representation* of the polytope defined by it. Also, the expression $\mathrm{cone}(R)$ is referred to as the V-representation of the cone defined by it.

The notion of polytopes and cones can be unified using the concept of *polyhedra*. First, given two sets A and B, one can define their so-called Minkowski sum as

$$A + B = \{a + b \,|\, a \in A, b \in B\}. \tag{1.3}$$

The sum of a cone and a polytope is called a polyhedron. Fig. (1.1) illustrates the Minkowski sum of a line segment and a ray.

Although polyhedra are more general than polytopes and cones, it suffices to consider cones. The reason is that for every polyhedron $P \subset \mathbb{R}^d$ one can construct its *associated cone* $C \subset \mathbb{R}^{d+1}$. This works as follows. Consider a polyhedron

$$P = \mathrm{conv}(V) + \mathrm{cone}(R), \tag{1.4}$$

which is defined in terms of its V-representation, where $V \in \mathbb{R}^{d\times n}$ and $R \in \mathbb{R}^{d\times m}$.

Then, its *associated cone* is defined as

$$C = \operatorname{cone} \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ V & R \end{pmatrix}, \tag{1.5}$$

where $\mathbf{1} \in \mathbb{R}^{1\times n}$ is a matrix the entries of which are all one and $\mathbf{0} \in \mathbb{R}^{1\times m}$ is a matrix for which all entries are zero [12]. The associated cone owes its name to the fact that P can be retrieved from C by intersecting it with the hyperplane that has a fixed first coordinate equal to one. In this way, properties of a polyhedron can be inferred from properties of its associated cone.

1.2 Fourier-Motzkin elimination

Before we continue, we need an important tool. The Fourier-Motzkin elimination is an algorithm to eliminate variables from a set of affine inequalities S in these variables, very much like the Gaussian algorithm is a method to eliminate variables from a set of affine equations. In fact, S may also contain affine equations. However, without loss of generality, we assume this not to be the case, since every equation can be replaced by two inequalities that together are equivalent to the equation.

The method works as follows. In every step of the algorithm one variable is eliminated. The algorithm is terminated, when all variables that are selected for elimination are eliminated. One elimination step works like this: Given a variable, say x, that shall be eliminated. Then, x partitions the set of affine inequalities into three subsets: 1) Inequalities that define an upper bound on x, 2) inequalities that define a lower bound on x and 3) inequalities that are independent of x. From this partition, a new set of affine inequalities S' is constructed that does not contain x such that S' admits a solution if and only if the original set of inequalities S admits a solution.

The central observation for this construction is the following: Every lower bound on x is also a valid lower bound on every upper bound on x if and only if there exists a solution for x. For this reason, S' contains all the inequalities that arise from combinations of lower and upper bounds on x additionally to those inequalities in S that are independent of x. As a consequence, the number of inequalities to consider can increase rapidly from one iteration to the next, which is the reason why Fourier-Motzkin elimination is typically not used for practical purposes. Rather, it is noteworthy mostly for its simplicity. Also it has a clear geometrical interpretation: The Fourier-Motzkin elimination is a projection into the orthogonal complement of the subspace spanned by the eliminated variables.

For computational purposes however, other, more refined algorithms are preferable, such as the double description method or pivoting algorithms such as the reverse search vertex enumeration method, which are covered Section 1.9.

1.3 Halfspace representation of a polyhedron

In the previous section, we defined cones in terms of their rays. Using the Fourier-Motzkin elimination, one can show that any cone C also admits a second representation

$$C = \text{rep}(A) := \{x \in \mathbb{R}^d \mid Ax \leq 0\}, \tag{1.6}$$

where $A \in \mathbb{R}^{m \times d}$ and the inequality is to be understood element-wise, so that each of the m linear inequalities defines a halfspace. The expression $\text{rep}(A)$ is called a *halfspace representation* of C or *H-representation* for short. A is called a *representation matrix* of C [18]. The result that every finitely generated cone admits an H-representation is known as *Weyl's theorem* and it can be proved as follows. Any finitely generated cone

$$C = \text{cone}(R) = \{x \in \mathbb{R}^d \mid \exists\, \lambda \geq 0 : x = R\lambda\} \tag{1.7}$$

with $R \in \mathbb{R}^{d \times k}$ and $\lambda \in \mathbb{R}^k$, is the projection of the set

$$D = \{(x, \lambda) \mid \exists\, \lambda \geq 0 : x = R\lambda\} \tag{1.8}$$

onto the first d dimensions. Since D clearly is the set of solutions to a system of affine inequalities and equations, this projection can be carried out using Fourier-Motzkin elimination. The resulting system of affine inequalities will then be an H-representation of the cone C. Note that for any matrix A, $\text{rep}(A)$ defines a convex cone, that is a set that is closed under conic combinations. However, it is a non-trivial result that any such cone is a polyhedral cone (in this chapter simply called "cone"), which means that it is finitely generated. This will be explained in a later section.

One can extend Weyl's theorem to polyhedra. As discussed earlier, any polyhedron

$$P = \text{conv}(V) + \text{cone}(R) \tag{1.9}$$

has an associated cone

$$C = \text{cone}\left(\begin{bmatrix} \mathbf{1} & \mathbf{0} \\ V & R \end{bmatrix}\right). \tag{1.10}$$

Due to Weyl's theorem for cones, C has an H-representation that allows us to write

$$C = \text{rep}((-b, A)) = \{(x_0, x) \in \mathbb{R} \times \mathbb{R}^d \mid Ax \leq bx_0\}, \tag{1.11}$$

where we split up the representation matrix of C, such that $-b$ is its first column and $A \in \mathbb{R}^{m \times d}$ is the submatrix that contains the remaining columns. One now retains the polyhedron P by intersecting C with the plane $x_0 = 1$, which yields

$$P = \{x \in \mathbb{R}^d \mid Ax \leq b\}. \tag{1.12}$$

1.4 Farkas' lemma

The Farkas Lemma exists in many variants [12], but we only need one of these variants for the following discussion. Given a finitely generated cone cone(R), it states that any point x either (1) is contained in the cone or (2) there is a linear half-space a that contains the cone but not x. This follows from the fact that cone(R) has a half-space representation and x is contained in the cone if and only if it is contained in all half-spaces that contain the cone. In other words, x is in the cone if and only if there is no half-space that contains the cone but not x.

Farkas Lemma can be generalized to arbitrary convex bodies. This result is known as the *separating hyperplane theorem* and plays an important role in the field of convex optimization.

1.5 The polar (dual) of a cone

Given a cone $C \in \mathbb{R}^d$, the normal vectors of the halfspaces that contain C form a cone themselves: the polar cone

$$C^* = \{a \in \mathbb{R}^d \mid a^T x \leq 0, \forall\, x \in C\}. \tag{1.13}$$

It owes its name to the natural association of dual vectors $a^T \in (\mathbb{R}^d)^*$ with inequalities $a^T x \leq 0$. The dual cone usually defined as the reflection at the origin of C^* as defined above. For convenience, we do not make this distinction in this chapter and instead use both terms synonymously.

Let C be a cone that is finitely generated by the matrix $R \in \mathbb{R}^{d\times k}$. Then we can write

$$C = \text{cone}(R) = \{R\lambda \mid \lambda \geq 0\}. \tag{1.14}$$

If we express the dual cone of C in terms of R, that is

$$\begin{aligned} C^* &= \{a \mid a^T R\lambda \leq 0\, \forall \lambda \geq 0\} \\ &= \{a \mid a^T R \leq 0\} \\ &= \text{rep}(R^T), \end{aligned} \tag{1.15}$$

we find that it is represented by R^T.

Vice versa, given a convex cone

$$C = \text{rep}(A) = \{x \mid Ax \leq 0\} \tag{1.16}$$

in its H-representation, its dual cone

$$C^* = \{a \mid a^T x \leq 0,\ \ \forall x \in C\} \tag{1.17}$$

is the finitely generated cone

$$D = \{A^T\lambda \mid \lambda \geq 0\} = \text{cone}(A^T). \tag{1.18}$$

Showing that $D \subseteq C^*$ is trivial. To show $C^* \subseteq D$, consider $a \notin D$. Then with Farkas Lemma there exists $b \in \mathbb{R}^d$ such that the halfspace $h = \{x \mid b^T x \leq 0\}$ contains D but not a. This means that $Ab \leq 0$ and $a^T b > 0$. $Ab \leq 0$ implies $b \in C$, which in turn implies $a \notin C^*$. Therefore, we can write

$$C = \text{rep}(A) \Rightarrow C^* = \text{cone}(A^T). \tag{1.19}$$

With this, one can prove the *Minkowski theorem*, which states that every cone $C = \text{rep}(A)$ is finitely generated. For this, we write C together with its dual and double dual cone

$$C = \text{rep}(A) \tag{1.20}$$

$$C^* = \text{cone}(A^T) = \text{rep}(R^T) \tag{1.21}$$

$$C^{**} = \text{rep}(A) = \text{cone}(R). \tag{1.22}$$

In the second line we first use Eq. (1.19). Then the existence of R^T is guaranteed by Weyl's theorem. In the third line, we then first use Eq. (1.15), which shows that $C = C^{**}$ and then we use Eq. (1.19) once again which proves Minkowski's theorem.

Given a cone $C = \text{cone}(R) = \text{rep}(A)$, the pair (A, R) is called a *double description pair*. Since the dual of C is $C^* = \text{cone}(A^T) = \text{rep}(R^T)$, the pair (R^T, A^T) is a double description pair, too. Hence, any algorithm that takes a generator matrix of a cone as input and outputs a representation can also perform the opposite task of converting a representation matrix to a generator matrix.

1.6 Faces of polyhedra

Let us finally introduce the concept of a face of a polyhedron. With every halfspace we associate the inequality that defines it. If a halfspace contains the polyhedron, the associated inequality is called *valid*. Consider a valid inequality and the hyperplane h, which bounds its associated halfspace. If the intersection between the polyhedron and h is non-empty, this intersection is called a *face* of the polyhedron. Faces are classified by their affine dimension, which is the dimension of the smallest affine subspace that contains the face. A face that has an affine dimension of d_a is called an d_a-face. Any face is the Minkowski sum of the polytope that is generated by the vertices that are contained in h and the cone that is generated by the rays that are parallel to h. This has mainly three consequences: (1) Faces of polyhedra are themselves polyhedra, (2) any face with dimension $d_a < d - 1$ is the intersection of higher dimensional faces, and (3) the 0-faces of a polyhedron are exactly its vertices and the 1-faces of a cone are

exactly its rays. The 1-faces of a polytope are called *edges*. Further, the $(d-1)$-faces of a polyhedron are called *facets*.

Facet-defining inequalities take a special role among the valid inequalities of a polyhedron: They are the only non-redundant inequalities in the H-representation of a polyhedron P. For simplicity we prove this for a cone $C = \text{cone}(R)$ with $\dim(C) = d$. Consider the dual cone of C

$$C^* = \{a \mid a^T x \leq 0,\ \forall x \in C\}. \tag{1.23}$$

Since every halfspace defined by an inequality $a^T x \leq 0$ contains the origin, which is the zero-face of C, any such halfspace contains a face of C. Denoting the set of faces of C as $\text{Faces}(C)$, we can therefore write

$$C^* = \bigcup_{f \in \text{Faces}(C)} \{a \in C^* \mid a^T x = 0,\ \forall x \in f\}. \tag{1.24}$$

Any set

$$f^* = \{a \in C^* \mid a^T x = 0,\ \forall x \in f\} \tag{1.25}$$

is a face of the dual cone C^* and as such is a cone itself. Due to Minkowski's theorem it is therefore generated by its one-dimensional faces. However, if $\dim(f^*) = 1$ then $\dim(f) = d - 1$. Consequently, C^* is generated by the facet-normal vectors of C. From duality it then follows that C is represented by its facet-defining inequalities.

Apart from their geometrical meaning, faces are important as they establish the combinatorial structure of the polyhedron: Two d_a-faces are called *adjacent* in P if they are contained in the same $(d_a + 1)$-face of P.

The concept of duality can not only be applied to cones but to polyhedra of any kind. In the case of polytopes, every facet-normal vector of a polytope that contains the origin corresponds to a vertex of its dual.

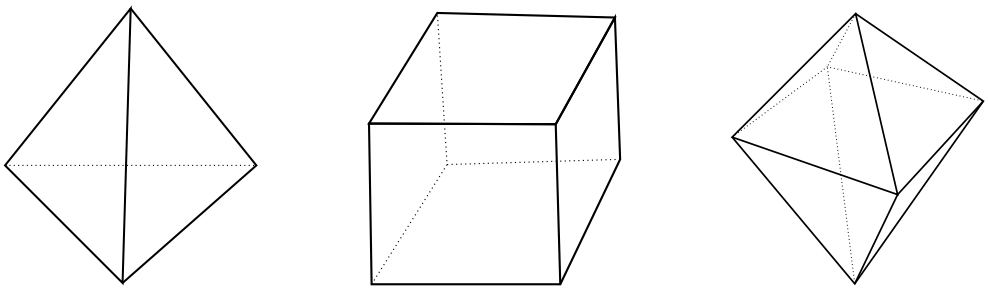
1.7 Simple and simplicial polytopes

Consider a set of points in $\mathbb{R}^d$, such that there is no $d-1$ dimensional hyperplane that contains all of these points. Then, these points are called *affinely independent*. Consider a set of $d+1$ affinely independent points. Then the polytope defined as the convex hull of these points is called a *d-simplex* or *simplex* for short[12]. This is a generalization of a tetrahedron in $\mathbb{R}^3$.

Similarly, cubes can be generalized to hypercubes, the standard version of which is defined as [12]

$$C_d = \text{conv}(\{-1, 1\}^{\times d}) \tag{1.26}$$

$$= \{x \in \mathbb{R}^d \mid -1 \leq x_i \leq 1,\ i = 1, \ldots, d\}. \tag{1.27}$$



(a) (b) (c)

Figure 1.2: The figure shows three polytopes in $\mathbb{R}^3$. These are the (a) tetrahedron, (b) cube, and (c) octahedron.

The polar of a hypercube is a crosspolytope

$$C_d^* = \text{conv}(\{e_1, -e_1, \ldots, e_d, -e_d\}) \tag{1.28}$$

$$= \{x \in \mathbb{R}^d \mid \sum_i |x_i| \leq 1\}, \tag{1.29}$$

where e_i denotes the i-th cartesian basis vector. The crosspolytope is a generalization of the octahedron. The crosspolytope and the hypercube demonstrate by example that the number of facets of a polytope can be exponential in the number of vertices and vice versa.

A polytope of dimension d is called *simple*, if every vertex is contained in no more than d facets. The hypercube is an example for a simple polytope. A d-dimensional polytope is called *simplicial*, if every facet contains no more than d vertices. The crosspolytope is an example for a simplicial polytope. Further, the polar of a simple polytope is simplicial polytope and vice versa. If the dimension $d \geq 3$, then the only polytopes that are both simple and simplicial are simplices [12].

1.8 Important examples

In this section, we briefly discuss classes of polytopes that are relevant in the field of Bell nonlocality, namely correlation polytopes, cut polytopes, and more generally, 0/1 polytopes.

Given a graph (V, E), the correlation polytope of this graph is a polytope in the real vector space $\mathbb{R}^{|V|+|E|}$ [19]. It is thus natural to label the dimensions with the vertices and edges of the graph. The correlation polytope is defined in V-representation and from each subset $I \subseteq V$, we obtain one vertex v^I of the polytope, where the coefficients

of v^I are defined as

$$v^I_u = \begin{cases} 1, \text{ if } u \in I, \\ 0 \text{ else} \end{cases} \tag{1.30}$$

for $u \in V$ and

$$v^I_{wu} = \begin{cases} 1, \text{ if } \{w, u\} \subseteq I, \\ 0 \text{ else} \end{cases} \tag{1.31}$$

for $wu \in E$. For the same graph, the cut polytope is defined as a polytope in the vector space $\mathbb{R}^{|E|}$. For each subset $I \subseteq V$ one obtains one vertex v^I of the polytope, the coefficients of which are defined as

$$v^I_{wu} = \begin{cases} 1, \text{ if exactly one of } w, u \text{ is in } I, \\ 0 \text{ else} \end{cases} \tag{1.32}$$

for $\{w, u\} \in E$.

Both correlation polytopes and cut polytopes have vertices with coefficients in $\{0, 1\}$. Polytopes of this kind are called 0/1 polytopes [12] and there exist algorithms that are especially dedicated to this class of polytopes [20].

1.9 Algorithms

In this section, we discuss algorithms that solve the following task: Given either a representation matrix or a generating matrix of a polyhedron, find a double description pair for that polyhedron. As a consequence of duality, any algorithm that can solve one of the two problems can also solve the other. It is therefore sufficient to consider the problem of finding a generating matrix for a cone in H-representation. For polytopes, this task is called *vertex enumeration*. In general, this problem is hard, as one can easily see from the example of a hypercube. For hypercubes, the number of vertices grows exponentially with the number of facet-defining halfspaces that define it. Since the latter are the input of the vertex enumeration algorithm, even the task of merely printing the output has exponential time complexity. For a meaningful complexity analysis, one hence states the complexity in terms of input and output size [18, 21]. The runtime of an algorithm as a function of both input and output size is then referred to as the *total time* of the algorithm [21].

For simple polyhedra, the *reverse search vertex enumeration method* solves the vertex enumeration problem in time $O(mvd)$ for m halfspaces, dimension d, and v vertices. However it is known that time complexity can scale super-polynomially in the number of vertices for non-simple polyhedra [22]. For simplicial polyhedra, the primal-dual method – a method similar to the reverse search enumeration problem but formulated for the facet enumeration problem – enumerates facets efficiently [23, 22].

Further, there exists a polynomial total time algorithm for the vertex enumeration of 0/1-polytopes [20].

In the following, we describe two well known algorithms for polyhedral representation conversion: the double description method [24, 18] and the above mentioned reverse search vertex enumeration method [25]. All results regarding these methods can be found in the above cited references, if no other reference is provided. The descriptions of the methods that are offered in this section are intended to aid understanding of the principles the methods rely on. For this reason, we try to avoid technical difficulties as much as possible. The disadvantage of this approach is that the descriptions are not detailed enough to write a competitive implementation of these algorithms based on them. However, excellent detailed descriptions of these algorithms exist in the works referenced throughout this section.

1.9.1 Double description method

The double description method is a method to solve the following task: Given an H-representation $C = \text{rep}(A)$ of a cone, find a minimal V-representation $C = \text{cone}(R)$ of the same cone. The method was originally invented by Motzkin, Raiffa, Thompson and Thrall [24] and later refined and implemented by Fukuda and Prodon [18].

The algorithm works in the following way: First, one ensures that the cone is pointed. A cone C is pointed iff for any vector $v \in C$ it holds that $-v \notin C$. If C is not pointed, then $Av \geq 0$ and $A(-v) \geq 0$, so v lies in the kernel of A. It is therefore easy to check the pointedness of the cone by checking the rank of the representation matrix. The kernel of the representation matrix is called the *lineality space* of the cone. It is easy to see that any cone is the Minkowski sum of its projection on the orthogonal complement of its lineality space C_P and its lineality space. In order to characterize a cone in terms of its extremal rays, it is therefore sufficient to find a V-representation for its projection C_P, which is pointed.

Therefore one can always assume without loss of generality that the cone C is pointed. Then, given that C has dimension d, one can choose d halfspaces associated with d linearly independent rows of A to define a cone C_s. This cone is the associated cone of a simplex, so one can readily find its minimal V-representation by inverting its representation matrix. This serves as the starting point for the double description method.

A minimal V-representation for C is obtained iteratively, by defining a finite sequence of cones $C_0, \ldots, C_N$ with $C_0 = C_s$ and $C_N = C$ for each of which a minimal V-representation is found. The cones C_i are defined recursively: C_{i+1} is the cone that is the intersection of C_i with a halfspace present in the H-representation of C that is not present in the H-representation of C_i. The extremal rays of C_{i+1} are computed from the extremal rays of C_i. In the following we explain how this is accomplished. If

$C_{i+1} = C_i \cap h$, where $h = \{x \mid a^T x \geq 0\}$ is some halfspace, then one can distinguish three types of extremal rays of C_i:

1. strictly feasible rays, that is rays that strictly lie in h,
2. strictly infeasible rays, that is rays that do not lie in h and
3. barely feasible rays, that is rays that lie on the hyperplane $h_p = \{x \mid a^T x = 0\}$ that separates h from its complement.

Extremal rays of type (1) and (3) remain extremal rays of C_{i+1} while rays of type (2) are by definition not rays of C_{i+1}. Besides the feasible rays of C_i with respect to C_{i+1} (types (1) and (3)), C_{i+1} may have new extremal rays. Concerning these rays, some observations are helpful: (a) these rays are contained in the subspace h_p, (b) the rays are contained in two-faces of C_i, (c) these two-faces are generated by pairs of adjacent extremal rays, one of which is strictly feasible and one of which is strictly infeasible.

Observation (a) follows from the fact that extremal rays are one-faces. They are therefore uniquely defined by the intersection of $d-1$ linearly independent non-redundant halfspaces of C_{i+1}. If an extremal ray is the intersection of $d-1$ linearly independent non-redundant halfspaces of C_i, then it is an extremal ray of C_i. Thus, the only rays that are extremal in C_{i+1} but not in C_i lie on h_p.

To prove observation (b) consider a ray $r \in h_p$ of C_i, such that the lowest dimensional face of C_i that contains r is a k-face. Then, the lowest dimensional face of C_{i+1} that contains r is either a k-face or a $(k-1)$-face. Thus, if r is an extremal ray in C_{i+1} then it is either extremal in C_i or it must be contained in a two-face of C_i.

Observation (c) directly follows from (a) and (b): Two-faces are generated by two extremal rays, which are thereby adjacent. Since the two-face must contain an extremal ray in h_p that is different from the rays that generate it, it must be generated by a strictly feasible and a strictly infeasible extremal ray of C_i. With this, we established that any extremal ray of C_{i+1} that is not an extremal ray of C_i satisfies observations (a), (b), and (c). In fact, the converse is also true. Choosing a two-face of C_i that is generated by one strictly feasible extremal ray r_{feas} and one strictly infeasible extremal ray r_{inf} with respect to h ensures that the intersection of h_p with the two-face is a one-face and therefore an extremal ray of C_{i+1}. Given r_{feas} and r_{inf}, this new extremal ray r' of C_{i+1} is readily computed as

$$r' = (a^T r_{\text{feas}}) r_{\text{inf}} - (a^T r_{\text{inf}}) r_{\text{feas}}. \tag{1.33}$$

In one iteration one therefore computes a minimal V-representation for C_{i+1} from a minimal V-representation of C_i.

The method can be and has been optimized with respect to the order in which halfspaces are added when constructing the sequence C_i [18]. Experimentally, static orderings, that is orderings that are fixed from the beginning, have been found to be

superior to dynamic orderings [18]. For technical details, including data structures how to keep track of adjacencies, the reader is also referred to [18]. Recently, a method called pyramidal decomposition has been developed, with which the double description method can be parallelized [26]. The worst case time complexity of the algorithm is exponential in both input and output, yet it has been found useful especially for low dimensional problems $d \leq 12$ or highly degenerate polyhedra [18]. Implementations are available for example in the cddlib library [27], the Parma Polyhedra Library [28], PORTA [29] and with parallel computing in normaliz [30].

1.9.2 Reverse search vertex enumeration

There is a second class of algorithms for representation conversion of convex polyhedra that is of practical importance: Variations of the reverse search vertex enumeration algorithm, which is implemented for example in the program lrs [25, 31].

As it is the case in the double description method, the objective of the reverse search vertex enumeration method is to find the vertices of a polyhedron given in H-representation. To simplify the following discussion, consider the special case of a convex polytope P. The *adjacency graph* $G = (V, E)$ of P is the graph consisting of the vertices and edges of P. One way to find the vertices of P is thus to construct and traverse the adjacency graph P. This can be done using a graph traversion algorithm such as depth first search or breadth first search in conjunction with a method to find the adjacent vertices of some starting vertex.

The latter is accomplished with *pivoting*. In its simplest form, the procedure can be described like this: Any vertex v of the polytope is the unique solution of d linearly independent inequalities drawn from the H-representation of the polytope. If one inequality is removed from the system, the solution is a line l through v. If the H-representation of P is not minimal, that is it incorporates redundant halfspaces, then this line may intersect P only in this one vertex v. Note that the line l does never intersect the interior of P since it is defined by inequalities that are satisfied for all points in P. However, if the H-representation is minimal, then every halfspace corresponds to a facet of P and therefore the line contains an edge e of P. This edge is the convex hull of two vertices, one of which, v is already known. The other vertex can be found by intersecting the line with the bounding hyperplane of another halfspace of the H-representation of P. To ensure that the intersection is a point, the corresponding inequality is chosen linearly independent from the set of $d - 1$ inequalities that define the line l. Further, one chooses the inequality such that it is not saturated by v, in order to ensure that a point p different from v is found. Now two cases remain to be distinguished: If p is feasible, then it is the vertex adjacent to v along the edge e. Otherwise, one chooses a different bounding hyperplane of P to intersect l with.

Note that any total order of the inequalities in the H-representation of P induce a

total order of the neighbors of any vertex v, since any neighbor of v is characterized by the edge e it shares with v and this edge can be identified with the inequality that is saturated for v but not e. This observation will become important for the algorithm later.

In principle one could find all vertices of a polytope using pivoting in conjunction with a graph traversal algorithm such as depth first search. Depth-first search is a recursive algorithm. Starting with a vertex, it successively discovers undiscovered adjacent vertices. Once a new vertex is discovered, it is stored in memory and a new instance of a depth first search is initialized at the new vertex. Discovered vertices must be stored in memory to avoid discovering the same vertex multiple times along different paths and especially to avoid loops in the graph. This, as it turns out, is an inefficiency that can be avoided.

The key part of the reverse search vertex enumeration algorithm is the insight how to arrange the – yet to be discovered – vertices of a polytope in a graph that has more structure that helps to traverse the graph more easily without the need to keep the discovered vertices in memory. This graph is constructed as follows: Its vertex set is the set of vertices of P. The edges of the graph are defined with the help of a local search function f. Two vertices (v_1, v_2) form a directed edge, iff $v_2 = f(v_1)$. The vertex v_2 is then called the parent of v_1. The local search function f further has the property that for any vertex v there exists a finite number n, such that $f^n(v) = r$, where r is the root of the graph, which is the only vertex of G that does not have a parent.

One possible choice for the local search function is the function that corresponds to one step in the simplex method. The *simplex method* is an algorithm to solve a linear program. Consider a linear program with the polytope P as feasible region and a linear objective that singles out one vertex of the polytope as optimal solution. Then, starting from any vertex, the simplex method finds a new vertex that yields an equal or better objective value than the previous vertex using pivoting. Since linear programs are convex optimization problems and the linear program under consideration is bounded, the simplex algorithm terminates once the optimal vertex is found. This vertex is the root of the graph G. How the simplex algorithm works in detail is described for example in Ref. [12].

The reverse search vertex enumeration algorithm enumerates the vertices of the polytope by traversing the graph G while constructing it. Note that this is simpler than traversing the adjacency graph of P. For one, the time complexity of depth first search scales linearly in the number of vertices and edges of a graph [32] and G has strictly fewer edges than the adjacency graph of P. Additionally, it has a simpler structure: It is a rooted, spanning, oriented tree of the adjacency graph of P: G spans the adjacency graph of P, since their vertex sets coincide, the local search function defines the orientation of every edge, and G is an oriented tree, since its underlying undirected graph is a tree, which means that any two vertices are connected by exactly one

path [33]. Moreover, G can be equipped with a total order: Parents precede their children and siblings are ordered according to the pivoting rule as discussed above. This has two benefits. First, since the vertices are in a predefined order, the possibility of visiting a vertex twice is automatically excluded. Second, the vertex enumeration can be interrupted and the only information necessary to pick up the process later on is the last visited vertex. In particular, this makes the algorithm economical with respect to consumed memory in comparison with other algorithms such as the double description method. Moreover, it was shown that the reverse search vertex enumeration algorithm is efficient for simple polytopes in the sense that its runtime scales linearly in the number of halfspaces in the H-representation and in the number of vertices [25]. Also, the algorithm can be parallelized as implemented in the library ZRAM [34] or in the program mplrs as described in [31]. The code mplrs has been found to be more scalable than other polyhedral representation conversion codes. A comparison can be found in Ref. [31].

2 Optimization techniques

In this chapter, we discuss optimization techniques that are relevant to this work. These include convex optimization techniques such as linear programming and semidefinite programming, which we describe in the first two sections. We then briefly discuss two useful techniques for non-convex optimization problems: the seesaw technique and the Navascués-Pironio-Acín hierarchy of semidefinite relaxations.

2.1 Convex optimization

Convex optimization problems are problems that involve the minimization of a convex function over a convex domain. Equivalently, we also refer to the maximization of a concave function over a convex domain as convex optimization problem. This chapter gives an introduction to this class of problems with a focus on linear optimization problems and semidefinite optimization problems. If no other source is given, facts stated in the first two sections of this chapter can be found in the book *Convex Optimization* by Boyd and Vandenberghe [15].

Convex optimization problems are an important class of problems for mainly two reasons: First, these problems naturally occur in a wide variety of different fields. Especially the subclass of semidefinite programms (SDPs) is important in the field of quantum information for tasks such as deciding the separability of a multipartite quantum state [35], discriminating between different quantum states using measurements while minimizing the error rate [36], or bounding the violation of a Bell inequality [37]. The second reason why convex optimization problems are important is that they can be solved efficiently and the solution is usually guaranteed to be optimal. Since for a convex optimization problem every local minimum is automatically also a global minimum these problems can be solved using local optimization methods such as gradient decent methods, Newton's method, or interior point methods. Furthermore, any convex optimization problem has a so-called dual problem. Under certain conditions, which we will discuss later, one can certify the optimality of the solution of the original problem by solving its dual.

In the simplest case, a convex optimization problem for a convex *objective function*

$f_0 : \mathbb{R}^n \supset \text{dom}(f_0) \to \mathbb{R}$ in a variable $x \in \mathbb{R}^n$ has the form

$$\begin{aligned} P: \quad & \min_x f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned} \tag{2.1}$$

where $A \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$, and $f_i : \mathbb{R}^n \supset \text{dom}(f_i) \to \mathbb{R}$ are convex functions. The intersection of the domains of the objective function f_0 and all the constraint functions f_i is called the *domain of the convex optimization problem* P [15]. We denote it as $\text{dom}(P)$.

A *convex set* is a set that contains all line segments between any two points in the set, so if $x, y \in S$, then $\lambda x + (1-\lambda)y \in S, \forall\ \lambda \in [0,1]$. A function f is *convex* iff its epigraph is a convex set, that is $f(\lambda x + (1 - \lambda y)) \leq \lambda f(x) + (1-\lambda) f(y)$. If all constraint functions f_i in problem Eq. (2.1) are convex, then the set of points x that satisfy the constraints is convex. The set of points x that satisfy the affine equality constraint $Ax = b$ is convex, too. The set of points that simultaneously satisfy all the constraints is called the feasible set $\text{feas}(P)$. The feasible set is convex as the intersection of convex sets. This shows that the above problem $\min_{x \in \text{feas}(P)} f_0(x)$ is indeed a convex optimization problem. Any point $y \in \text{val}(P) := \{f_0(x) \mid x \in \text{feas}(P)\}$ is called a *value* of the optimization problem P. In the following we assume that the infimum over $\text{val}(P)$ is attained. The value $y^* = \min \text{val}(P)$ is called the *optimal value* of P. Any point $x^* \in \text{feas}(P)$ such that $f_0(x^*) = y^*$ is called a *solution* of P.

One can consider a more general case of a convex optimization problem where the constraining functions f_i with $i > 0$ do not map to the real numbers, but instead $f_i : \mathbb{R}^n \supseteq \text{dom}(f_i) \to V_i, i = 1, \ldots, m$, where V_i is a euclidean, finite dimensional vector space. In this case, an order $\succeq$ must be defined on V_i for each inequality constraint of the problem Eq. (2.1). Such an order is however not arbitrary but should satisfy some conditions, which we state in the following. For this, we rely on some definitions regarding closed convex cones. A *closed convex cone* is a closed subset $C \subseteq \mathbb{R}^d$ which is closed under positive scaling and addition. If C has a non-empty interior is called *solid*. If $0 \in C$ and C does not contain a line, then it is called *pointed* [15]. A solid pointed closed convex cone is called *proper cone* [15]. A proper cone C induces an order $\succeq_C$ through the relationship

$$b \succeq_C a \ \Leftrightarrow b - a \in C. \tag{2.2}$$

Analogously one defines

$$a \preceq_C b \ \Leftrightarrow b - a \in C. \tag{2.3}$$

One can show that this order is a partial order. A *partial order* $\preceq$ on a set S has three properties: reflexivity, antisymmetry, and transitivity. Reflexivity means that $x \preceq x$ for all $x \in S$. Antisymmetry means that $x \preceq y$ and $y \preceq x$ together imply $x = y$ for $x, y \in S$.

Finally, transitivity means that $x \preceq y$ and $y \preceq z$ imply $x \preceq z$. Since a proper cone contains the origin, its induced order is reflexive. Because a proper cone is pointed, it does not contain a line and hence its induced order is antisymmetric. Finally, since a proper cone is convex, its induced order is transitive. Moreover, $\preceq_C$ also has other properties that are not required by the definition of a partial order. First, it is preserved under non-negative scaling. This means that $x \preceq_C y$ and $\alpha \geq 0$ imply $\alpha x \preceq_C \alpha y$. This is a consequence of C being closed under non-negative scaling. Second, $\preceq_C$ is preserved under addition. If $x \preceq_C y$ and $u \preceq_C v$ then also $x + u \preceq_C y + v$ holds. This follows from C being closed under addition. Finally, since C is a closed set, its induced order is preserved under limits. Inequalities with respect to an order which is induced by a proper cone are called *generalized inequalities* [15].

Using generalized inequalities, one can write a more general form of a convex optimization problem.

$$\begin{aligned} & \min_x f_0(x) \\ \text{s.t.} \quad & f_i(x) \preceq_{C_i} 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{aligned} \tag{2.4}$$

where $A \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$ and $C_i \subset V_i$ are proper cones.

In practice, we usually do not need to consider the most general case of convex optimization problems. Also, most solvers cannot handle problems of the most general type. It is therefore useful to consider the most important special cases. *Cone programs* are convex optimization problems that are constrained by only one affine generalized inequality and have a linear objective function. *Semidefinite programs* are cone programs where the cone that defines the order is the cone of positive semidefinite matrices. A semidefinite program thus has the form

$$\begin{aligned} & \min_x c^T x \\ \text{s.t.} \quad & \sum_i x_i F_i + G \preceq 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{aligned} \tag{2.5}$$

where G and the F_i are symmetric matrices and $x, c \in \mathbb{R}^n$. *Linear programs* are cone programs where the order inducing cone is the nonnegative orthant. Hence, linear programs have the form

$$\begin{aligned} & \min_x c^T x \\ \text{s.t.} \quad & Gx - h \preceq 0, \\ & Ax = b, \end{aligned} \tag{2.6}$$

where $G \in \mathbb{R}^{m \times n}$, and $h \in \mathbb{R}^m$.

2.2 Duality

Duality is an important concept in convex optimization theory. For every convex optimization problem, one can construct a related problem, the so called *dual problem*. The original problem is then also called *primal problem*.

The dual problem is a maximization problem and its optimal value can never exceed the minimal value of the primal problem. This property is called *weak duality* [15]. The difference between the maximal value of the dual problem and the minimal value of the primal problem is called the *duality gap*. For some convex optimization problems the duality gap is zero. This is called *strong duality* [15]. Strong duality is useful to certify that a solution of a convex optimization problem has been found. For this reason it is desirable to know in advance, whether strong duality holds for a convex optimization problem. To find out, one can check *Slater's condition*. If it is met, strong duality holds.

In this section we explain the aforementioned facts and state all necessary definitions.

First, we need to define the dual cone. Given a cone $K \subset V$ in a vector space V, the dual cone

$$K^* = \{y \in V^* \,|\, y(x) \geq 0 \text{ for all } x \in K\} \tag{2.7}$$

contains all vectors in the dual space V^* that map all vectors in the cone to the non-negative reals.

This notion allows us to define the *Lagrangian* of the optimization problem 2.4. The *Lagrangian* is defined as

$$L(x, \lambda, \nu) = f_0(x) + \sum_i \lambda_i (f_i(x)) + \nu^T (Ax - b). \tag{2.8}$$

Here, $\nu \in \mathbb{R}^p$ and λ is a tuple with elements $\lambda_i \in V_i^*$ such that $\lambda_i \succeq_{C_i} 0$. This ensures that

$$L(x, \lambda, \nu) \leq f_0(x) \tag{2.9}$$

for all feasible x and all (λ, ν). The *Lagrange dual function* is defined as

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu). \tag{2.10}$$

The dual problem is then defined as

$$\begin{aligned} &\max_{\lambda,\nu} g(\lambda, \nu) \\ &\text{s. t.} \quad \lambda_i \succeq_{C_i} 0 \quad i = 1, \ldots, m. \end{aligned} \tag{2.11}$$

We now discuss Slater's constraint qualification which gives two conditions that guarantee strong duality if they are jointly fulfilled. To state the conditions, we first need

the definition of a relative interior. The *relative interior* of a set is the interior with respect to its affine hull – the lowest dimensional hyperplane that contains the set. In the following, we denote the domain of the convex optimization problem under consideration as D.

The first condition of Slater's criterion is that there is a point $\tilde{x}$ in the relative interior of D. The second condition is that the point $\tilde{x}$ is strictly feasible, so it satisfies [1]

$$f_i(\tilde{x}) < 0, \quad i = 1, \ldots, m \tag{2.12}$$

$$A\tilde{x} - b = 0. \tag{2.13}$$

In the following we show a proof, that Slater's constraint qualification implies strong duality. This proof is an adapted version of the one found in the book *Convex optimization*[15]. The proof relies on the separating hyperplane theorem. It states that for any two disjoint nonempty convex sets A, B, there exists a hyperplane defined by $a \in \mathbb{R}^n, \alpha \in \mathbb{R}$ with $a \neq 0$ that separates them, so

$$\langle a, x \rangle \geq \alpha, \quad \forall x \in A \tag{2.14}$$

$$\langle a, x \rangle \leq \alpha, \quad \forall x \in B. \tag{2.15}$$

Here, $\langle \, , \, \rangle$ denotes the dot-product. To prove that Slater's constraint qualification implies strong duality, the separating hyperplane theorem is applied to two specific sets A and B, with

$$A = \{(u, v, t) \,|\, \exists x \in D : u_i \succeq_{C_i} f_i(x), Ax - b = v, t \geq f_0(x)\}, \tag{2.16}$$

$$B = \{(0, 0, s) \,|\, s < p^*\}, \tag{2.17}$$

where p^* is the optimal value of the primal problem. We can assume that the optimal value is finite, since otherwise weak duality already implies strong duality. Note that A, B do not intersect. The separating hyperplane theorem states that there are $\lambda_i \in V_i^*$, $\nu \in \mathbb{R}^p$, $\mu \in \mathbb{R}$, $\alpha \in \mathbb{R}$, such that

$$\sum_i \lambda_i(u_i) + \nu^T v + \mu t \geq \alpha \tag{2.18}$$

and

$$\mu s \leq \alpha \tag{2.19}$$

hold for all $(u, v, t) \in A$ and $(0, 0, s) \in B$. This implies that $\lambda_i \succeq_{C_i *} 0$, $\mu \geq 0$ and that in particular also

$$\sum_i \lambda_i(f_i(x)) + \nu^T(Ax - b) + \mu f_0(x) \geq \alpha \tag{2.20}$$

[1] There is a refined version of Slater's condition where the first $1 \leq m_a \leq m$ inequality constraints do not have to be strictly satisfied if they are not only convex but affine. For simplicity, we do not show a proof of this refined version here.

and

$$\mu p^* \leq \alpha \tag{2.21}$$

hold for all x that lie in the domain of the problem. Combining the two inequalities yields

$$\sum_i \lambda_i(f_i(x)) + \nu^T(Ax - b) + \mu(f_0(x) - p^*) \geq 0. \tag{2.22}$$

If $\mu > 0$, then we can divide the inequality by μ which implies strong duality. Indeed, $\mu > 0$ holds if Slater's condition is fulfilled. For the Slater point $\tilde{x}$, we have

$$\sum_i \lambda_i(f_i(\tilde{x})) < 0 \tag{2.23}$$

$$\nu^T(A\tilde{x} - b) = 0 \tag{2.24}$$

and

$$\mu(f_0(\tilde{x}) - p^*) \geq 0. \tag{2.25}$$

We assume that there are no redundant or contradicting equality constraints. This means that A has full row-rank, so $\nu^T A \neq 0$ for $\nu \neq 0$. Then, since $\tilde{x}$ is a strictly feasible point in the relative interior of the problem's domain D, there exists a point $\tilde{x}' \in D$, such that

$$\sum_i \lambda_i(f_i(\tilde{x}')) < 0 \tag{2.26}$$

$$\nu^T(A\tilde{x}' - b) < 0 \tag{2.27}$$

and

$$\mu(f_0(\tilde{x}') - p^*) \geq 0. \tag{2.28}$$

If $\lambda_i \neq 0$ or $\nu \neq 0$ this necessitates $\mu > 0$ for Eq. (2.22) to hold. Conversely, if all $\lambda_i = 0$ and $\nu = 0$, then also $\mu > 0$, since the hyperplane $(\lambda, \nu, \mu) \neq (0, 0, 0)$ exists.

For cone programs, the relation between primal and dual problem is symmetric. In this case, one can equivalently check Slater's condition for the dual problem to establish strong duality.

2.3 Beyond convex optimization problems

The manageability of convex optimization problems and the maturity of software packages for solving linear and semidefinite programs makes the attempt desirable to reduce solving optimization problem, possibly even a non-convex one, to solving a series of linear or semidefinite programs instead. In quantum information theory, most prominently two approaches are employed for this purpose: seesaw algorithms [38, 39, 40] and the Navascués-Pironio-Acín hierarchy of semidefinite programms [41, 35, 37].

2.3.1 Seesaw algorithms

Seesaw algorithms work in the following way [38]: Let the domain of the optimization problem be a subset of a vector space V. Then, the first step is to find subspaces V_1, V_2 of V, such that $V = V_1 \oplus V_2$. Instead of solving the optimization problem

$$\begin{aligned} & \min_x f(x) \\ \text{s.t.} \quad & x \in \mathrm{dom}(f) \end{aligned} \tag{2.29}$$

one then alternates between solving the problem

$$\begin{aligned} & x_1^{i+1} = \underset{x_1}{\mathrm{argmin}} f(x_1 + x_2^i) \\ \text{s.t.} \quad & x_1 \in V_1 \\ & x_1 + x_2^i \in \mathrm{dom}(f) \end{aligned} \tag{2.30}$$

and

$$\begin{aligned} & x_2^{i+1} = \underset{x_2}{\mathrm{argmin}} f(x_1^{i+1} + x_2) \\ \text{s.t.} \quad & x_2 \in V_2 \\ & x_1^{i+1} + x_2 \in \mathrm{dom}(f) \end{aligned} \tag{2.31}$$

until the values of the sequence $(f(x_1^i + x_2^i))_i$ do not grow by more than some predefined value ϵ from step to step. Analogously, the method can easily be adapted to include any number of steps instead of two. For the above problem the goal is to find V_1, V_2 such that the optimization problems Eq. (2.30) and Eq. (2.31) are either linear or semidefinite programs. Clearly this is not always possible. A typical use case are optimization problems with an objective function that is bilinear in its arguments $x_1 \in V_1$, $x_2 \in V_2$ that is optimized over both arguments simultaneously, which means that $\mathrm{dom}(f) \subseteq V_1 \oplus V_2$.

The the value $f(x_1^{i_{max}} + x_2^{i_{max}})$ obtained from a seesaw algorithm may depend strongly on the initial values x_1^0, x_2^0 corresponding to different local minima of the objective function. In practise one therefore runs a seesaw algorithm several times with different initial values in order to achieve a good upper bound for the global minimum.

2.3.2 Navascués-Pironio-Acín hierarchy

For some non-convex optimization problems there exists a sequence – a so-called hierarchy – of semidefinite programms, such that the optimal values of these semidefinite optimization problems converge monotonically to the optimal value of the original problem. One such hierarchy is the Navascués-Pironio-Acín hierarchy, or NPA-

hierarchy for short [37, 42, 43]. It is applicable to problems of the form

$$\begin{aligned} p^* = \inf_{X,\phi} & \langle \phi,\, p(X)\phi \rangle \\ \text{s.t.} \quad & ||\phi|| = 1 \\ & q_i(X) \succeq 0, \quad i = 1, \ldots, i_q \\ & r_i(X)\phi = 0, \quad i = 1, \ldots, i_r \\ & \langle \phi, s_i(X)\phi \rangle \geq 0, \quad i = 1, \ldots, i_s, \end{aligned} \tag{2.32}$$

where $p, q_i, r_j, s_k \in \mathbb{C}[x, x^*]$ are polynomials, $\phi \in H$ is a vector in a Hilbert space H and $X = (X_1, \ldots, X_n)$ is an indexed family of operators on H [43]. Further, $\succeq$ denotes positivity with respect to the cone of positive semidefinite operators and p as well as all s_i and all q_i are hermitian, that is $p(X)^\dagger = p(X)$, which ensures that the objective and the inequality constraints

$$\langle \phi, s_i(X)\phi \rangle \geq 0, \quad i = 1, \ldots, i_s \tag{2.33}$$

are real valued and $q_i(X)$ are hermitian operators. Lastly, one also assumes that the polynomials q_i satisfy the so-called archimedean assumption

$$C - \sum_{k=1}^{2n} X_k X_k^\dagger = \sum_i f_i(X) f_i(X)^\dagger + \sum_{i,j} g_{ij}^\dagger(X) q_i(X) g_{ij}(X) \tag{2.34}$$

for some positive constant C. This ensures that any feasible operators X are bounded. What makes problem Eq. (2.32) hard is the polynomial dependence of the objective and the constraints on the variables X. The NPA-hierarchy relies mainly on two ideas to linearize and relax the problem. First, we find the maximal polynomial degree d among the polynomials in the problem. The set W_d of monomials with degree at most d then forms a basis of the vector space of polynomials of degree d, allowing us to express the objective as

$$\langle \phi, p(X)\phi \rangle = \sum_{w \in W_d} p_w \langle \phi, w(X)\phi \rangle. \tag{2.35}$$

Similarly, we can write the constraints in terms of the monomial basis. Now, the objective depends linearly on the so-called moments $m(w) = \langle \phi, w(X)\phi \rangle$, which makes it desirable to treat the latter as variables. However, they are not independent, as considering monomials $u, v, w \in W_d$ with $u(X)v(X) = w(X)$ shows. This is where the second idea of the method, the relaxation, comes into play, which centers on the properties of the moment matrix. The moment matrix $M \in \mathbb{C}^{|W_k| \times |W_k|}$ of order k is indexed in the monomials with elements defined as

$$M_{vw} := m(v^* w) = \langle \phi, v^*(X) w(X) \phi \rangle, \tag{2.36}$$

where we dropped the dependences of ϕ and X in the notation for convenience. Moment matrices are readily shown to be positive semidefinite. The relaxation of the problem now consists of substituting complex variables $\gamma(w)$ for the moments $m(w)$. The interdependence of the variables $\gamma(w)$ is then captured by constructing a matrix Γ analogously to the moment matrix with

$$\Gamma_{vw} = \gamma(v^* w) \tag{2.37}$$

and demanding $\Gamma \succeq 0$. Additionally, the constraints are reformulated and relaxed in very much the same fashion. The constraints $q_i(X) \succeq 0$ imply that the localizing matrices L^{q_i} with elements

$$\begin{aligned} L^{q_i}_{vw} &= \langle \phi, v(X)^* q_i(X) w(X) \phi \rangle & (2.38)\\ &= \sum_{u \in W_{d_i}} q_{i,u} \langle \phi, v(X)^* u(X) w(X) \phi \rangle & (2.39)\\ &= \sum_{u \in W_{d_i}} q_{i,u} m(v^* u w) & (2.40) \end{aligned}$$

are positive semidefinite, where d_i is the degree of q_i and $q_i = \sum_u q_{i,u} u$. Again, we substitute the variables $\gamma(w)$ for the moments and replace the constraint $q_i(X) \succeq 0$ with $\Lambda^{q_i} \succeq 0$, where

$$\Lambda^{q_i}_{vw} = \sum_{u \in W_{d_i}} q_{i,u} \gamma(v^* u w). \tag{2.41}$$

Analogously, we replace the constraint

$$\langle \phi, s_i(X), \phi \rangle \geq 0 \tag{2.42}$$

with

$$\sigma_i = \sum_w s_{i,w} \gamma(w) \geq 0, \tag{2.43}$$

where $s_i = \sum_w s_{i,w} w$. The constraint $r_i(X)\phi = 0$, with $r_i = \sum_w r_{i,w} w$ implies

$$\begin{aligned} 0 &= \langle \phi v(X), r_i(X) \phi \rangle & (2.44)\\ &= \sum_w r_{i,w} m(vw), & (2.45) \end{aligned}$$

which we formulate in terms of the variables $\gamma(w)$ as

$$\sum_w r_{i,w} \gamma(vw) = 0. \tag{2.46}$$

With this, the relaxed problem reads

$$
\begin{aligned}
& p^k = \min_{\gamma_w, w \in W_{2k}} \sum_w p_w \gamma_w && (2.47)\\
\text{s.t.} \quad & \gamma(1) = 1 \\
& \Gamma \succeq 0 \\
& \Lambda^{q_i} \succeq 0,\ i = 1, \ldots, i_q \\
& \sum_w r_{i,w} \gamma(vw) = 0\ \forall v \in W_k \\
& \sum_w s_{i,w} \gamma(w) \geq 0,\ i = 1, \ldots, i_s,
\end{aligned}
$$

where the maximal monomial degree $2k$ has to be greater or equal to the maximal polynomial degree d of the polynomials involved in the problem. The number k is also referred to as the level of the hierarchy. If the infimum in the original problem Eq. (2.32) is achieved, then any relaxation in the NPA-hierarchy must yield an optimal value p^k that is lower or equal to the optimal value p^* of the original problem. Further, $(p^k)_k$ is a monotonically increasing sequence. Let $k' > k$ and let $p^{k'}$ be attained for $(\gamma_w)_{w \in W_{2k'}}$. Then, for the same γ, $(\gamma_w)_{w \in W_{2k}}$ is a feasible point for the k-th level of the hierarchy and will yield $p^{k'}$ as well. Furthermore, it has been proved that

$$
\lim_{k \to \infty} p^k = p^*. \tag{2.48}
$$

Conveniently, the NPA-hierarchy in practise oftentimes converges already for small k [37] and a stopping criterion, the so-called rank-loop has been derived to detect that the NPA-hierarchy has indeed converged.

3 Quantum Mechanics

This chapter is not a comprehensive exposition of quantum mechanics. Instead, we focus on the basics and select topics that are relevant to this thesis. The contents covered in this chapter can also be found in the book by Nielsen and Chuang [44] or in the book by Heinosaari and Ziman [45], if no other reference is given.

Quantum mechanics offers a description of nature that has three components: The first component, a *state*, is the description of a physical system, which is a part of the physicist surroundings declared by the physicist to be of special interest. The second component, a *channel*, describes the dynamics the system is subject to. The third component is the description of a measurement, which links states and observations. In the following we take a closer look at each part individually.

3.1 States

In quantum mechanics, systems are associated with Hilbert spaces. A Hilbert space $\mathcal{H}$ is a complete inner product space. The inner product is positive definite, linear in the second argument and conjugate symmetric under exchange of the arguments. Positive definite means, that for any vector $|\psi\rangle \in \mathcal{H}$ the inner product of this vector with itself $\langle\psi|\psi\rangle > 0$ is strictly positive, unless $|\psi\rangle = 0$, in which case the inner product vanishes. This ensures that the inner product induces a norm, $\| |\psi\rangle \| = \sqrt{\langle\psi|\psi\rangle}$. While Hilbert spaces can have any dimension – in fact they can be infinite dimensional – in this work we are only concerned with finite dimensional Hilbert spaces, in particular $\mathbb{C}^d$. In the following, we write $|i\rangle$ with $i \in \{0, \ldots, d-1\}$ to denote the basis vectors of a chosen orthonormal basis of $\mathbb{C}^d$.

Just as several subsystems can be composed to form a larger system, so can Hilbert spaces. Given a system Σ that is composed of subsystems $s \in S$ with Hilbert spaces $\mathcal{H}_s$, the Hilbert space $\mathcal{H}_\Sigma$ is the tensor product of the Hilbert spaces associated with the subsystems, that is

$$\mathcal{H}_\Sigma = \bigotimes_{s \in S} \mathcal{H}_s. \tag{3.1}$$

We are now prepared to introduce quantum states. The term quantum state can refer to two mathematically different objects. It can either refer to a so-called density matrix, or to a state vector. A state vector is a unit vector in a Hilbert space. State vectors

can be added and the normalized sum of state vectors yields a state vector called a *superposition* of the state vectors is made up of. Given the set of subsystems S, each described by a state vector $|\psi_s\rangle$, the state $|\psi\rangle$ of Σ is given by the state vector

$$|\psi\rangle = \bigotimes_{s\in S} |\psi_s\rangle. \tag{3.2}$$

In systems that are composed of subsystems, superposition gives rise to *entanglement*. A state vector $|\psi\rangle \in \mathcal{H}_\Sigma$ is said to be entangled, if it is not a *product state*, that is if there does not exist a sequence of states $|\psi_s\rangle \in \mathcal{H}_s$, such that Eq. (3.2) holds. In this case, the subsystems cannot be described by state vectors independently, which is why the subsystems are called entangled. An important example for entangled states are the *Bell states*

$$\begin{aligned}
|\phi^+\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle),\\
|\phi^-\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle),\\
|\psi^+\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle),\\
|\psi^-\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle),
\end{aligned} \tag{3.3}$$

which form a basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$, where $|01\rangle = |0\rangle \otimes |1\rangle$.

As mentioned earlier, the term quantum state can also refer to a density matrix. A density matrix ϱ is a hermitian positive semidefinite operator on a Hilbert space $\mathcal{H}$ with unit trace, that is

$$\varrho^\dagger = \varrho, \varrho \geq 0, \operatorname{Tr}(\varrho) = 1. \tag{3.4}$$

We denote the set of density matrices acting on $\mathcal{H}$ as $S(\mathcal{H})$. Further, every state vector $|\psi\rangle$ corresponds to a density matrix, which is defined as the rank-one projector on that state vector, $|\psi\rangle\langle\psi|$. The converse is not true however. The density matrix formalism is more expressive. While one cannot assign a subsystem of an entangled system a state vector, one can assign it a density matrix. Provided a state $\varrho \in S(\mathcal{H}_\Sigma)$ and a subset $T \subseteq S$ of systems, the state ϱ_T of the system that consists of the subsystems in T is computed using the partial trace

$$\varrho_T = \operatorname{Tr}_{S\setminus T}(\varrho). \tag{3.5}$$

The partial trace is defined uniquely as the linear operator that maps

$$\operatorname{Tr}_{S\setminus T} : A \otimes B \mapsto \operatorname{Tr}(A)\, B, \tag{3.6}$$

where A is an operator on $\mathcal{H}_{S\setminus T}$ and B is an operator on $\mathcal{H}_T$. Besides describing the marginals of entangled states, density matrices can also be used to describe mixtures

of pure states. In this way, one can account for the case, in which a system is known to be in some state $|a\rangle$ with a probability p or else in a different state $|b\rangle$. In this case, the system is described by the state $\varrho = p|a\rangle\langle a| + (1-p)|b\rangle\langle b|$. In this case, the rank of the state is greater than one and it is called *mixed*. In contrast, states that are rank one are called *pure*. Whether a state is pure or mixed can be conveniently decided by computing the *purity* $\mathrm{Pur}(\varrho) = \mathrm{Tr}\left(\varrho^2\right)$ of that state. A state is pure, iff its purity is equal to one. Mixed states are related to pure states in (at least) two ways. Any mixed state can be diagonalized; it can be expressed as a convex combination of pure states. Alternatively, any state $\varrho_A = \sum_i p_i |i\rangle\langle i|$ can be obtained as a partial trace $\varrho_A = \mathrm{Tr}_B\left(|\psi_{AB}\rangle\langle\psi_{AB}|\right)$ of its *purification* $|\psi_{AB}\rangle = \sum_i \sqrt{p_i}\,|ii\rangle$.

The impurity of a state ϱ can alternatively also be captured by quantifying the lack of information about a quantum system in that state. This is done using the Von Neumann entropy, which is defined as

$$S(\varrho) = -\mathrm{Tr}\left(\varrho \log \varrho\right). \tag{3.7}$$

The von Neumann entropy of a state is non-negative quantity that vanishes for pure states and only for pure states. Further, the von Neumann entropy is concave.

3.1.1 Bloch vectors and operator bases

We introduced density matrices as positive semidefinite operators with trace one acting on an underlying finite dimensional Hilbert space $\mathcal{H}$. However, there is a second perspective. Let $\mathcal{B}(\mathcal{H})$ be the real vector space of operators acting on $\mathcal{H}$. Then the set of density matrices arises as the intersection of the convex cone of positive semidefinite operators with the hyperplane of trace-one operators in $\mathcal{B}(\mathcal{H})$. We denote the subspace of $\mathcal{B}(\mathcal{H})$ that contains the hermitian matrices as $\mathcal{B}^H(\mathcal{H})$. For practical purposes, it is convenient to choose a basis for $\mathcal{B}^H(\mathcal{H})$. In the case of qubits, one such basis consists of the Pauli matrices

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{3.8}$$

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{3.9}$$

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{3.10}$$

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.11}$$

Every qubit state can be expressed in the Pauli basis as

$$\varrho = \frac{1}{2}\left(\sigma_0 + \sum_{i=1,2,3} r_i \sigma_i\right). \tag{3.12}$$

Hence, every qubit state can be associated with a point in the euclidian vector space $\mathbb{R}^3$. This point has the coordinates r_i in the cartesian basis. The position vector of this point is called the *Bloch vector* of the state ϱ. For pure states it holds that $\|\vec{r}\| = 1$. Thus, the set of Bloch vectors of all qubit states $S(\mathbb{C}^2)$ describes a three-dimensional ball. This ball is called the *Bloch ball* and it is full. Points on the boundary represent pure states, while points inside the ball represent mixed states.

The concept of the Bloch vector can be generalized for states in higher dimensional Hilbert spaces. However, the geometric structure of the state set $S(\mathcal{H})$ becomes more complicated in this case [46].

3.2 State evolution

The perhaps simplest time evolution a system can undergo is from a pure state $|\psi\rangle \in \mathcal{H}$ to another pure state $|\psi'\rangle \in \mathcal{H}$. This time evolution is described by a unitary matrix U acting on $\mathcal{H}$. If the dynamics of the system are described by a constant Hamiltonian H, the unitary U can be computed from the Hamiltonian H as

$$U = \exp\left(\frac{-iHt}{\hbar}\right), \tag{3.13}$$

where t is the parameter accounting for time and $\hbar$ is the reduced Planck constant. This is the solution to the Schrödinger equation

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\psi(t)\rangle = H |\psi(t)\rangle. \tag{3.14}$$

While this may be a sufficient description of an isolated system, it is not, if one aims to describe the time evolution of a subsystem of a larger system Σ. Such a system is also called an *open* quantum system. The complement of the open quantum system S with respect to Σ is called the *environment* of S. Consider for example a two-qubit system that is initially in a product state $|00\rangle$ and undergoes a transition into a Bell state. In this case, the time evolution of a single qubit cannot be described with a unitary, as its reduced state after the transition is mixed and any unitary maps pure states to pure states.

The most general time evolution of a system is described by a *quantum channel*. A quantum channel, or *channel* for short, is a map

$$C : S(\mathcal{H}) \to S(\mathcal{H}), \tag{3.15}$$

which is completely positive, linear, and trace preserving. Such a map is called a CPTP map. If a map is completely positive, this means that for any finite dimensional Hilbert space $\mathcal{H}'$ with identity operator id acting on $\mathcal{B}(\mathcal{H}')$, the map $\mathrm{id} \otimes C$ maps a positive semidefinite operator in $\mathcal{B}(\mathcal{H}' \otimes \mathcal{H})$ to another positive semidefinite operator in

$\mathcal{B}(\mathcal{H}' \otimes \mathcal{H})$ [45]. An important example for a map that is positive but not completely positive is the partial transpose, which we will discuss in more detail in the section dedicated to entanglement. Complete positivity as well as trace preservation are properties that directly follow from the fact that channels are supposed to map states to states. Linearity is motivated by supposing that acting with a channel on a mixture of states yields the same result as mixing the states that were obtained by acting with the same channel on the components of the mixed state [44], that is

$$C\left(\sum_i p_i |i\rangle\langle i|\right) = \sum_i p_i C(|i\rangle\langle i|). \tag{3.16}$$

Earlier, we discussed unitary time evolution, the corresponding channel of which is called a unitary channel. Given that the time evolution of a closed system is described by a unitary channel [44], this elicits the question whether for any open quantum system S, one can construct a larger, closed quantum system that contains S, such that the time evolution of S is compatible with a unitary time evolution of the closed system it is embedded in. *Stinespring's dilation theorem* establishes, that this is indeed the case [45]: For any channel $C : S(\mathcal{H}_S) \to S(\mathcal{H}_S)$ acting on a system S in a state $\varrho \in S(\mathcal{H}_S)$ there exists a system E with Hilbert space $\mathcal{H}_E$, a pure state $\xi \in S(\mathcal{H}_E)$ and a unitary $U : \mathcal{H}_S \otimes \mathcal{H}_E \to \mathcal{H}_S \otimes \mathcal{H}_E$, such that

$$C(\varrho) = \mathrm{Tr}_E\left(U(\varrho \otimes \xi)U^\dagger\right). \tag{3.17}$$

Every channel further admits a second representation, the *Kraus form* [45]. This means that every channel $C : S(\mathcal{H}) \to S(\mathcal{H})$ can be expressed with the help of a sequence of so-called Kraus operators $K_i : \mathcal{H} \to \mathcal{H}$, such that

$$C(\varrho) = \sum_i K_i \varrho K_i^\dagger, \tag{3.18}$$

where

$$\sum_i K_i^\dagger K_i = \mathbb{1}, \tag{3.19}$$

and $\mathbb{1}$ denotes the identity operator acting on $\mathcal{H}$. If $\dim(\mathcal{H}) = d$ then one can show that no more than d^2 Kraus operators are necessary to express the action of any channel [45].

We now discuss some examples for qubit channels: The bit flip channel, the phase flip channel, the bit-phase flip channel and finally the depolarizing channel. In short, the first three are examples for channels that describe an evolution in which the state is subject to a unitary evolution defined by one of the Pauli matrices for some probability and else remains unchanged. The depolarizing channel describes a process in which the state is swapped out for a maximally mixed state with some probability. The names

reflect the actions of the Pauli matrices on a state

$$|\psi\rangle = \sqrt{p}\,|0\rangle + \sqrt{1-p}\,e^{i\phi}\,|1\rangle \tag{3.20}$$

expressed in the eigenbasis of the Pauli Z matrix. Acting on this state with X exchanges the basis states $|0\rangle$ and $|1\rangle$ which corresponds to a bit flip. The Kraus operators are then simply

$$E_0 = \sqrt{q}\mathbb{1} \tag{3.21}$$

$$E_1 = \sqrt{1-q}X \tag{3.22}$$

for some probability $0 \leq q \leq 1$. In the same way that X leads to a bit flip, Z leads to a phase flip, that is a sign flip in the complex phase $\exp(i\phi)$. As $Y = iXZ$, the Pauli Y unitary essentially leads to a combination of bit flip and a phase flip. The depolarizing channel has the interpretation of mixing the state with white noise as the maximally mixed state contains no information about the system it describes.

Other than bit flips, phase flips, and mixture with white noise, one can also consider amplitude damping and phase damping. The amplitude damping channel is described with the Kraus operators

$$E_0 = |0\rangle\langle 0| + \sqrt{1-\gamma}\,|1\rangle\langle 1| \tag{3.23}$$

$$E_1 = \sqrt{\gamma}|0\rangle\langle 1| \tag{3.24}$$

and $0 \leq \gamma \leq 1$. This channel leaves the state $|0\rangle$ invariant while damping the amplitude of the $|1\rangle$ state. This process is physically relevant for example in order to describe spontaneous emission, where the states $|0\rangle$ and $|1\rangle$ are fock states. In this case γ can be interpreted as the probability for photon loss [44].

As a last example, we discuss a simple scattering model [47], where system and environment are initially in a product state

$$\varrho_{SE} = \varrho_S \otimes \varrho_E \tag{3.25}$$

and the dynamics between system and environment is modeled using a simple coupling Hamiltonian

$$H_{SE} = Z \otimes R. \tag{3.26}$$

With this one can calculate the time evolution of the system using the Schrödinger equation. One finds that the density matrix of the system evolves such that the diagonal elements remain unchanged, while the off-diagonal elements – called coherences – decay exponentially. Such a process is thus called *decoherence*.

3.3 Measurements

In this section we discuss the description of measurements in the simplest form, where only the measurement statistics matter.

Consider a measurement of a quantum system with Hilbert space $\mathcal{H}$ in the state $\varrho \in S(\mathcal{H})$. Further let the measurement yield outcomes $o \in \Omega$. We are interested in probabilities at which different events occur. This is formalized using a probability space, where Ω is the sample space. The event space F is a σ-algebra on Ω. Together with a probability measure $\mu_\varrho : F \to [0, 1]$, the sample space and event space form a probability space.

One essential part of a measurement is how different states give rise to different probability measures. This is described using a *positive operator-valued measure*, or POVM, for short. A POVM is a map

$$P : F \to E(\mathcal{H}) \tag{3.27}$$

where $E(\mathcal{H})$ is the set of effects acting on $\mathcal{H}$. An *effect* is a positive-semidefinite operator with eigenvalues smaller or equal to one [45]. Further, a POVM satisfies the conditions

$$\begin{aligned} P(\emptyset) &= 0 \\ P(\Omega) &= \mathbb{1} \\ P(\cup_i X_i) &= \sum_i P(X_i) \end{aligned} \tag{3.28}$$

for a sequence $(X_j)_j$ of disjoint sets $X_j \in F$ [45]. A POVM gives rise to the family of probability measures μ_ϱ through the Born rule, that is

$$\mu_\varrho(X) := \mathrm{Tr}\left(P(X)\varrho\right). \tag{3.29}$$

In the case where Ω is finite, one can simply choose the power set of Ω as the σ-algebra F and any POVM is sufficiently defined by the collection of effects $(E_o)_{o\in\Omega}$ associated with the measurement outcomes o.

A subclass of POVMs are projection-valued measures (PVMs). These are POVMs that map to effects that are projectors. Measurements that are described by PVMs are called *projective measurements*. If Ω is finite, projective measurements can also be described using *observables*. Observables are hermitian operators and are particularly useful, if only the expectation value of some quantity is of interest. An observable can be obtained from a PVM that is specified by a sequence of projectors $(\Pi_i)_{i\in\Omega}$ as

$$O = \sum_{i\in\Omega} o_i \Pi_i \tag{3.30}$$

after assigning a real value o_i to every outcome. Vice versa, every observable defines a PVM given by the projectors on the eigenspaces of that observable. These are well defined since any hermitian matrix possesses a spectral decomposition.

3.4 Entanglement

We now return to the quantum mechanical phenomenon of entanglement. The findings discussed in the sections about entanglement are also discussed in the excellent reviews [48, 49], if no other resource is cited. We already briefly discussed entanglement for pure states. Given a system Σ with Hilbert space $\mathcal{H}_\Sigma$ that consists of subsystems $s \in S$ with Hilbert spaces $\mathcal{H}_s$, a pure state $|\psi_\Sigma\rangle$ is called *fully separable* if it can be expressed as the product

$$|\psi_\Sigma\rangle = \bigotimes_{s\in S} |\psi_s\rangle \tag{3.31}$$

of states $|\psi_s\rangle \in \mathcal{H}_s$. A mixed state $\varrho_\Sigma \in S(\mathcal{H}_\Sigma)$ is called fully separable, if it is a convex combination

$$\varrho_\Sigma = \sum_k \lambda_k |\psi_{\Sigma,k}\rangle\langle\psi_{\Sigma,k}| \tag{3.32}$$

of fully separable pure states $|\psi_{\Sigma,k}\rangle$ for convex coefficients $0 \leq \lambda_k \leq 1$, $\sum_k \lambda_k = 1$ [48, 49].

The notion of full separability can be generalized to allow for a more nuanced description of quantum states: Consider a partition $P = \{c_1, c_2, \ldots, c_n\}$ of the set of subsystems S. This means that the cells $c_k \subset S$ are disjoint and nonempty and $\cup_{k=1}^n c_k = S$. A pure state $|\psi_\Sigma\rangle \in \mathcal{H}_\Sigma$ on a system Σ with subsystems in S is called *n-separable* [49], if there exists a partition P with n cells, such that for every cell $c \in P$ there exists a state $|\psi_c\rangle \in \bigotimes_{i\in c} \mathcal{H}_i$ so that

$$|\psi_\Sigma\rangle = \bigotimes_{c\in P} |\psi_c\rangle. \tag{3.33}$$

As before, this definition is extended to mixed states. A mixed state is called n-separable if it is the convex combination of n-separable pure states. It is important to note that the pure states in the convex decomposition do not need to be n-separable with respect to the same partition. Obviously, n-separablity implies $(n-1)$-separability. Therefore, 2-separablity – usually called biseparability – is the weakest form of separability. If a multipartite state is not biseparable, it is called *genuinely multipartite entangled*.

There are more ways to classify quantum states and quantify their entanglement. In the following we discuss common equivalence classes for quantum states and motivate the properties a function $E : S(\mathcal{H}_\Sigma) \to \mathbb{R}^+$ must have to qualify as an *entanglement measure*. First of all, it should hold that $\mu(\varrho) = 0$ if ϱ is a fully-separable state.

Second, two quantum states $\varrho, \gamma \in S(\mathcal{H}_\Sigma)$ are considered equivalent with regard to all of their entanglement properties, if there exist local unitary operators $U_s : \mathcal{H}_s \to \mathcal{H}_s$, such that

$$\varrho = \bigotimes_{s\in S} U_s \gamma \bigotimes_{s\in S} U_s^\dagger. \tag{3.34}$$

Such a local unitary operation only changes the local degrees of freedom of the state and is locally reversible. These equivalence classes can be coarse grained by defining a bigger class of transformations of states that preserve equivalence. A notable class of such transformation are *local operations and classical communication* (LOCC). LOCC operations encompass all state transformations that arise from the application of local channels $\Gamma_s : S(\mathcal{H}_s) \to S(\mathcal{H}_s)$ and the exchange of classical information between the parties, which allows the parties to coordinate their application of local channels. Since no entanglement can be created in this way, entanglement measures are defined to be non-increasing under LOCC operations [48]. Oftentimes, other or even stronger properties are demanded for entanglement measures, such as additivity $E(\varrho_1 \otimes \varrho_2) = E(\varrho_1) + E(\varrho_2)$, convexity, or to be non-increasing on average under LOCC operations [48]. In general, these axioms do not lead to a unique way to quantify entanglement. In fact, many different entanglement quantifiers, that do not necessarily exhibit all of the desirable properties for entanglement measures above, have been proposed. Two well-known entanglement measures are negativity and entanglement of formation. Negativity quantifies the violation of a separablity criterion called PPT criterion, which will be discussed later. Entanglement of formation is defined for bipartite pure states and exploits the fact that the marginal states of the subsystems are mixed, if the state is entangled. The entanglement between the two parties is then measured as the Von Neumann entropy of the marginals. Entanglement of formation can also be extended to mixed states using a so-called convex roof construction. This means that the entanglement of formation of a mixed state is $E(\varrho) = \inf_{|\psi_k\rangle, p_k} E(|\psi_k\rangle)$ such that $\varrho = \sum_k p_k |\psi_k\rangle\langle\psi_k|$.

3.4.1 Notable families of quantum states

From the viewpoint of entanglement of formation, the question which bipartite states $\varrho_{AB} \in S(\mathbb{C}^d \otimes \mathbb{C}^d)$ are maximally entangled can be answered: They are pure states with maximally mixed marginals $\varrho_A = \mathrm{Tr}_B(\varrho_{AB}) = \frac{1}{d}\mathbb{1}$ and $\varrho_B = \frac{1}{d}\mathbb{1}$. For two-qubit systems, these states are equivalent up to local unitaries to the Bell states Eq. (3.3). For bipartite systems with local dimension d, the states

$$|\phi_d^+\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle\,|i\rangle \tag{3.35}$$

satisfy this property and are generally referred to as the *maximally entangled states* of their respective systems.

In n-qubit systems, the GHZ state

$$|GHZ\rangle = \frac{1}{\sqrt{2}}(\,|0\ldots0\rangle + |1\ldots1\rangle) \tag{3.36}$$

is an example of a highly entangled state that leads to a maximal violation of many Bell inequalities [48, 50]. For three qubits, almost all genuinely tripartite entangled

states belong to the SLOCC class of the GHZ state [48]. SLOCC classes are a coarse graining of LOCC classes, where two states $|\psi\rangle, |\phi\rangle \in \mathcal{H}_\Sigma$ are regarded equivalent if they are related by an SLOCC operation, that is, if there exist invertible operators $A_s : \mathcal{H}_s \to \mathcal{H}_s$, such that

$$|\psi\rangle = \bigotimes_{s\in S} A_s \, |\phi\rangle. \tag{3.37}$$

The name SLOCC stands for *stochastic local operations and classical communication* and stems from the fact that SLOCC operations describe all operations that can be accomplished by means of LOCC operations with some probability. For three qubits, there is only one other SLOCC class of genuinely entangled states apart from the GHZ-class, which is represented by the W-state

$$|W\rangle = \frac{1}{\sqrt{3}}(\, |001\rangle + |010\rangle + |100\rangle \,). \tag{3.38}$$

The concept of maximally entangled states in bipartite systems is generalized by the concept of *absolutely maximally entangled* (AME) states, which are multipartite states that are maximally entangled with respect to every bipartition [51, 52]. Specifically, an AME-state $\mathrm{AME}(n,d) \in \mathcal{H}_\Sigma = \bigotimes_{i=1}^n \mathbb{C}^d$ is an n partite qudit state, such that for any subset $M \subset S$, $|M| \leq \lfloor \frac{n}{2} \rfloor$ of parties, the corresponding marginal state

$$\varrho_M = \mathrm{Tr}_{S\setminus M} \left(|\mathrm{AME}(n,d)\rangle\langle \mathrm{AME}(n,d)| \right) = \frac{1}{d^{|M|}} \mathbb{1} \tag{3.39}$$

is maximally mixed. Apart from the Bell states, the three party GHZ state is another example for an AME state. A table that indicates whether for given n, d an AME state exists, does not exist or is unknown can be found on the AME website [53, 54].

The last family of states we want to discuss in this section are *Werner states* [55, 56]. They are quantum states $\varrho \in S((\mathbb{C}^d)^{\otimes n})$, that remain invariant up to a complex phase under the action of an arbitrary local unitary $U : \mathbb{C}^d \to \mathbb{C}^d$ acting on every subsystem, so

$$\varrho = U^{\otimes n} \varrho \, (U^\dagger)^{\otimes n}. \tag{3.40}$$

In Chapter 10 we will investigate the question, for which systems there exist states that are Werner states and AME states at the same time.

3.4.2 Entangled pure states and the Schmidt decomposition

For pure states, deciding whether they are entangled or not is a simple task. If a pure state is fully separable, this means that it is a product state and hence all of its marginal states, that is the states of its subsystems, are pure, too. For bipartite states, one can also check, how close some given state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is to a product state. To do

this, one calculates the *Schmidt decomposition* of that state. Calculating the Schmidt decomposition amounts to finding local orthonormal basis vectors $|i\rangle_A \in \mathcal{H}_A$ and $|i\rangle_B \in \mathcal{H}_B$, such that the state can be decomposed as

$$|\psi\rangle = \frac{1}{\sqrt{d}} \sum_{k=0}^{d-1} \lambda_k |i_A i_B\rangle, \tag{3.41}$$

with positive *Schmidt coefficients* λ_k and *Schmidt rank* $d \leq \min(\dim(\mathcal{H}_A), \dim(\mathcal{H}_B))$.

For pure states, the purity of the marginal states is a simple, necessary and sufficient criterion for separability. However, it is in general difficult to decide whether a mixed state is entangled or not. In the following, we discuss practical entanglement criteria that give necessary conditions for separability.

3.4.3 PPT criterion

A common and useful method to detect entanglement is the *positive partial transpose* (PPT) criterion [57, 58]: The partial transpose – that is the transposition map acting on only some of the subsystems – of a separable state yields a positive semidefinite operator which is, in fact, another separable state. This is easily understood, since the transposition map is a positive trace preserving map which means that ϱ^T is a state if ϱ is a state. Applied to a separable state $\varrho \in S(\mathcal{H}_A \otimes \mathcal{H}_B)$ with

$$\varrho = \sum_k p_k \varrho_A^k \otimes \varrho_B^k, \tag{3.42}$$

one finds that

$$\varrho^{T_A} = \sum_k p_k (\varrho_A^k)^T \otimes \varrho_B^k \tag{3.43}$$

is a separable state as claimed above. Hence, entanglement is detected as soon as one of the partial transpositions of a state has a negative eigenvalue. For systems of two qubits or one qubit and one qutrit, the PPT criterion is even sufficient to certify separability [58]. In general however, it is not sufficient and there exist examples of entangled PPT states.

3.4.4 Entanglement witnesses

An *entanglement witness* is an observable W, such that

$$\mathrm{Tr}\,(\varrho W) \geq 0 \tag{3.44}$$

for all separable states ϱ and

$$\mathrm{Tr}\,(\varrho W) < 0 \tag{3.45}$$

for at least one entangled state ϱ.

For example, the observable

$$W = (|\psi\rangle\langle\psi|)^{T_A} \tag{3.46}$$

yields a positive expectation value for all separable states since

$$\mathrm{Tr}\,(\varrho W) = \mathrm{Tr}\left(|\psi\rangle\langle\psi|\varrho^{T_A}\right) \geq 0 \tag{3.47}$$

and every separable state is PPT. If there exists an entangled state that violates the inequality, then W is an entanglement witness. This can be ensured by choosing the eigenvector to a negative eigenvalue of an entangled state with negative partial transpose for $|\psi\rangle$.

3.4.5 Computable Crossnorm / Realignment criterion

We have discussed earlier how one can decide the separability of a bipartite pure state by calculating the Schmidt decomposition. For mixed states, there is a similar criterion called *computable crossnorm / realignment (CCNR) criterion*. The computable cross norm is defined as

$$\mathrm{CCN}(\varrho) = \sum_k \lambda_k, \tag{3.48}$$

where λ_k are the Schmidt coefficients of the state ϱ, if the latter is interpreted as a vector in the space $\mathcal{B}(\mathcal{H}_A) \otimes \mathcal{B}(\mathcal{H}_B)$. This means that

$$\varrho = \sum_k \lambda_k G_k^A \otimes G_k^B \tag{3.49}$$

where $(G_k^A)_k$ and $(G_k^B)_k$ are ordered, hermitian, orthonormal bases of $\mathcal{B}(\mathcal{H}_A)$ and $\mathcal{B}(\mathcal{H}_B)$ respectively.

The CCNR criterion states that for any bipartite, separable state $\varrho \in S(\mathcal{H}_A \otimes \mathcal{H}_B)$ it holds that

$$\mathrm{CCN}(\varrho) \leq 1. \tag{3.50}$$

The CCNR criterion is established by verfying that the computable crossnorm is a norm, which implies convexity. Then the claim follows from the fact that $\mathrm{CCN}(|\psi\rangle\langle\psi|) = 1$ for pure states $|\psi\rangle$.

3.4.6 Symmetric extension criterion

The symmetric extension criterion due to Doherty et al [35] is interesting for us to discuss for mainly two reasons. First of all, the term actually refers to a sequence of

ever more strict entanglement criteria, such that for any entangled state, there exists an index n, such that the criterion indexed by n detects this entanglement. Secondly, the idea on which these criteria rely, symmetric extensions, is versatile and can for example be helpful to decide whether a state can violate a Bell inequality. We will come back to this in Chapter 4, where we discuss nonlocality.

The symmetric extension criterion for a biseparable state $\varrho \in S(\mathcal{H}_A \otimes \mathcal{H}_B)$ indexed with an index n is the following: If ϱ is separable, then there exists a state $\varrho_n \in S(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B)$, called a *symmetric extension* of ϱ, which satisfies three conditions. The first two conditions are

$$\mathrm{Tr}_{A^{\otimes(n-1)}}(\varrho_n) = \varrho, \tag{3.51}$$

$$P(\varrho_n) = \varrho, \tag{3.52}$$

where $P : S(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B) \to S(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B)$ is an operator that performs an arbitrary permutation of the subsystems with Hilbert space $\mathcal{H}_A$. The third condition is that any partial transposition of ϱ_n is positive semidefinite.

For a separable state

$$\varrho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \otimes |\phi_i\rangle\langle\phi_i| \tag{3.53}$$

the state

$$\varrho_n = \sum_i p_i |\psi_i\rangle\langle\psi_i|^{\otimes n} \otimes |\phi_i\rangle\langle\phi_i| \tag{3.54}$$

is a symmetric extension. Trivially, for $n = 1$ a symmetric extension exists if and only if the state is PPT, so the symmetric extension criterion reduces to the PPT criterion in this case.

The symmetric extension criterion can be checked on a computer and finding a symmetric extension of a specific level amounts to solving a semidefinite program. Yet, in practise, deciding whether a state is separable or not is not easy, as it may be necessary to compute many levels. In fact, the separability problem is known to be NP-hard [35].

3.4.7 Bell inequalities as device independent tests of entanglement

Another tool to detect entanglement are Bell inequalities. Bell inequalities are designed to test nonlocality, which is a notion that is different from entanglement [59]. However, in quantum mechanics, nonlocality implies entanglement, which makes Bell inequalities a tool to detect entanglement in quantum systems. Bell inequalities will be discussed in greater detail in the Chapter 4, which is dedicated to Bell nonlocality. In this section, we briefly discuss Bell inequalities as a a tool for entanglement detection. One

example of a Bell inequality, which we state without proof or derivation, is the CHSH inequality [60]

$$\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \leq 2. \tag{3.55}$$

The CHSH inequality holds, if two parties, A and B perform measurements A_1, A_2 on subsystem A and B_1, B_2 on subsystem B and every measurement yields outcomes ± 1. Moreover, the expectation values $\langle A_i B_j \rangle$ have to be compatible with a so-called *local hidden-variable model* [2]. This means that they can be expressed in the form

$$\langle A_i B_j \rangle = \sum_\lambda p_\lambda \langle A_i \rangle_\lambda \langle B_j \rangle_\lambda. \tag{3.56}$$

.

If a Bell inequality is analyzed in the framework of quantum mechanics, the measurements should in general be described by POVM measurements. For example, if A_1 denotes a measurement that yields outcomes one out of finitely many outcomes $(o_i)_{i\in\Omega}$, then

$$\begin{aligned} \langle A_1 \rangle &= \sum_{i\in\Omega} o_i \mathrm{Tr}\left(E_i \varrho\right) \\ &= \left\langle \sum_{i\in\Omega} o_i E_i \right\rangle \\ &=: \langle \hat{A}_1 \rangle \end{aligned} \tag{3.57}$$

for a POVM defined by the effects $(E_i)_{i\in\Omega}$. Sometimes, if a Bell inequality is formulated in terms of expectation values of measurements like in Eq. (3.55), this notation is referred to as 'observable notation', although the measurement A_1 is in general not described by an observable. Nevertheless, one can associate the hermitian operator $\hat{A}_1$ with the measurement A_1, which is useful to calculate the expecatation value of A_1 in quantum mechanics. By associating a hermitian operator with every measurement of the CHSH inequality, we can write a quantum mechanical version of the inequality as

$$\langle \hat{A}_1 \hat{B}_1 + \hat{A}_1 \hat{B}_2 + \hat{A}_2 \hat{B}_1 - \hat{A}_2 \hat{B}_2 \rangle \leq 2. \tag{3.58}$$

The operator

$$\hat{S} = \langle \hat{A}_1 \hat{B}_1 + \hat{A}_1 \hat{B}_2 + \hat{A}_2 \hat{B}_1 - \hat{A}_2 \hat{B}_2 \rangle \tag{3.59}$$

is called the *Bell operator* the CHSH inequality and it depends on the particular choice of measurements. If a Bell inequality is violated for a particular choice of measurement settings, then the Bell operator of this inequality for this choice of measurement settings is an entanglement witness up to an affine offset. To meet the definition, the Bell operator must satisfy two conditions. First, the inequality must be violated. Second, the inequality holds for all separable states. Let $\varrho = \sum_\lambda p_\lambda \varrho_\lambda^A \otimes \varrho_\lambda^B$ be a separable state.

Then we can express all of the expectation values that appear in the CHSH inequality as

$$\langle \hat{A}_i \hat{B}_j \rangle = \sum_\lambda p_\lambda \mathrm{Tr}\left(\hat{A}_i \varrho_\lambda^A\right) \mathrm{Tr}\left(\hat{B}_j \varrho_\lambda^B\right). \tag{3.60}$$

Thus, the expectation values are described by a local hidden-variable model. Therefore, the CHSH inequality holds. This result extends to other Bell inequalities and the proof works in the same way [48]. Crucially, for separable states the Bell inequality holds for any measurements that yield appropriate outcomes. The implementation of the measurements does not matter. This feature is called *device independence* [48].

3.5 EPR argument

The Einstein-Podolski-Rosen argument [61], called EPR argument for short, is an argument in favor of the hypothesis that quantum mechanics is incomplete. EPR define that a theory is complete, if every element of reality has a counterpart in the theory. A physical quantity qualifies to be associated with an element of reality, if it can be predicted without disturbing the system. So, if a physical quantity can be predicted without disturbance, a theory is complete if and only if it predicts this quantity with certainty.

To reach their conclusion that quantum mechanics is incomplete, EPR investigate the relationship between the following two hypotheses:

H1 Quantum mechanics is incomplete.

H2 Non-commuting observables cannot correspond to elements of reality.

EPR formulate two assertions regarding these two hypotheses. Assertion one is that one of the hypotheses is true. The argument for this assertion is this: If H1 was true, there is nothing to show. Thus, assume that H1 is false and therefore quantum mechanics is complete. In quantum mechanics, the outcomes of two non-commuting observables are not predicted with certainty. Therefore, if quantum mechanics is complete, they cannot be predicted with certainty and H2 follows.

The second assertion is that if H1 is false, then also H2 must be false. In their argument, EPR make use of the position and momentum operator, but an equivalent argument can be formulated using the observables $X = \sigma_1$ and $Z = \sigma_3$ [62]. Consider the Bell state

$$\begin{aligned} |\phi^+\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\ &= \frac{1}{\sqrt{2}}(|++\rangle + |--\rangle), \end{aligned} \tag{3.61}$$

where $|+\rangle$, $|-\rangle$ are the eigenstates of X. If the first party, say Alice, measures observable X, then the second subsystem is described by an eigenstate of X and hence the measurement outcome of measuring X on the second system can be predicted with certainty. However, if Alice measures the observable Z, then the second subsystem is in an eigenstate of Z and therefore Z, too, corresponds to an element of reality. If one assumes that a measurement on the first subsystem does not cause a change in the second subsystem, then both X and Z simultaneously correspond to elements of reality. This is a contradiction to H2.

Since according to assertion one at least one of the hypotheses must be true, hypothesis H1 must be true and hence quantum mechanics is incomplete.

4 Bell Nonlocality

The concept of Bell nonlocality was introduced by John Bell as a reaction to the Einstein-Podolsky-Rosen (EPR) argument [2, 61] (see Chapter 3, Section 3.5). Specifically, it was conceived to make the assumption experimentally testable, whether quantum mechanics can be completed in the sense of EPR [2].

The situation that is considered is similar to the one considered in the EPR argument. We depict a simple experimental setup in Figure 4.1. At least two parties, typically called Alice and Bob, perform measurements on disjoint systems. Ideally, the parties are envisioned to be placed in distant laboratories. In this situation, Bell locality arises as the conjunction of three assumptions. The first assumption is *no-signaling*. No-signaling implies that the choice of measurement of one party does not affect the outcome of a measurement of another party. One justification for this assumption may be granted in the case in which the events of Alice choosing her measurement setting and Bob obtaining the outcome for his measurement are space-like separated. In this case, a signal sent out to disturb Bob's system when Alice chooses her measurement cannot reach Bob's system according to the laws of special relativity.

The second assumption is called *freedom of choice*. Freedom of choice implies that each party can choose her measurement settings independently of the state of the measured system or the measurement apparatuses.

The third assumption is that measurement outcomes are fundamentally predetermined. Einstein expressed his belief that this is indeed the case when he famously stated that god does not does play dice with the universe [63].

If Bell locality is violated, this phenomenon is called Bell nonlocality. In this chapter, we discuss the concept of Bell nonlocality in detail. Our focus lies on a rigorous presentation of the subject. To achieve this, we introduce some terminology that helps us to keep notation reasonably concise. It also lays the foundation for Bellpy, a Python library for nonlocality, which we introduce in Chapter 9.

4.1 Black-box experiments

The goal of this section is to formalize the description of an experiment that involves one or multiple systems, without imposing any physical assumptions. Such an experiment is known as a black-box experiment. There are only three pieces of information

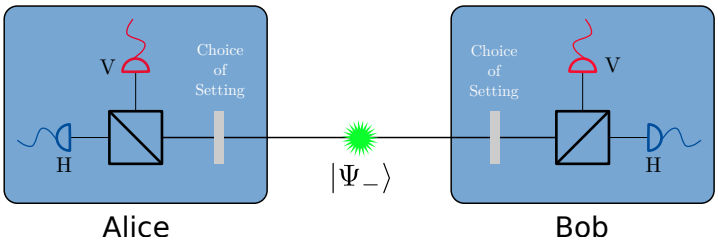


Figure 4.1: Simplified experimental setup for a Bell experiment – A pair of entangled photons is distributed to Alice and Bob, who measure the polarization, horizontal or vertical, of their respective photons after manipulating them using waveplates (gray). In this way, different settings are realized.

that are assumed to be known about any measurement in a black-box experiment. The first piece of information is the label k of the system that is affected by the measurement. Typically, these systems are envisioned to be distant laboratories, which are called *parties*. Parties are usually labeled using the letters A, B, C. If there are more than three parties, we choose $k \in \mathbb{N}$. The second piece of information about a measurement is a label $i \in \mathbb{N}$, called *input*. The third piece of information is a finite set O of *outputs* of the measurement. We hence define a *measurement* as a triple $m = (k, i, O)$. We also write $k(m), i(m), O(m)$ to refer to the system, the input, or the outcomes of m, respectively. An alternative term for measurement is *setting*.

The set S of settings that can be performed in an experiment is called the *scenario* of this experiment [64, 65]. A commonly considered scenario is the so-called *CHSH scenario*, which contains the settings

$$A_1 = (A, 1, \{-1, 1\}), \tag{4.1}$$
$$A_2 = (A, 2, \{-1, 1\}), \tag{4.2}$$
$$B_1 = (B, 1, \{-1, 1\}), \tag{4.3}$$
$$B_2 = (B, 2, \{-1, 1\}). \tag{4.4}$$

Let $J \subset S$ be a subset of measurements, such that every measurement in J affects a different party. Then these measurements can be performed jointly, which is why we call J a *joint measurement*. We define the set of joint measurements for a scenario S as

$$\mathcal{J}(S) = \{J \subset S \mid k(m) \neq k(m') \text{ for all } m, m' \in J\}. \tag{4.5}$$

In the case of the CHSH scenario S_{CHSH}, the set of joint measurements is

$$\mathcal{J}(S_{CHSH}) = \{\{A_1\}, \{A_2\}, \{B_1\}, \{B_2\}, \{A_1, B_1\}, \{A_1, B_2\}, \{A_2, B_1\}, \{A_2, B_2\}\}. \tag{4.6}$$

Given a joint measurement $J \in \mathcal{J}(S)$, we call any sequence $r = (o_m)_{m \in J}$ such that $o_m \in O(m)$ a *result* of J. We denote the set of possible results of a joint measurement J

as $R(J)$. An experiment is described by stating the probability $p(r \mid J)$ for each result r to occur after each joint measurement $J \in \mathcal{J}(S)$. We call every such pair $c = (r \mid J)$ of a joint measurement $J \in \mathcal{J}(S)$ and a result r of J a *record*, since this amounts to the information recorded in every round of the experiment. To stress the fact that $p(c)$ is the probability of r occurring under the condition that J was performed, we denote c as $(r \mid J)$ instead of (r, J). We denote the set of possible records in an experiment as

$$C(S) = \{(r \mid J) \mid r \in R(J), J \in \mathcal{J}(S)\}. \tag{4.7}$$

As an example, consider the records for the joint measurement $\{A_1, B_2\}$, which are

$$(-1,-1 \mid A_1, B_2),\ (-1,1 \mid A_1, B_2),\ (1,-1 \mid A_1, B_2),\ (1,1 \mid A_1, B_2), \tag{4.8}$$

where we have omitted parenthesis and curly brackets for better readability. In total, the CHSH scenario has 24 different records.

We now show how to encode the information of the probabilities $p(c)$ in a vector. We define V as the free vector space over $\mathbb{R}$ generated by the set of records $C(S)$. Further, we equip V with an inner product $\cdot : V \times V \to \mathbb{R}$, such that the records form an orthonormal basis. We now define the *behavior space* $\mathcal{N}$ as the affine subspace of V that contains all vectors $b \in V$ which satisfy the normalization constraints

$$\sum_{r \in R(J)} b \cdot (r \mid J) = 1 \tag{4.9}$$

for all $J \in \mathcal{J}(S)$. We say that $\mathcal{N}$ is the behavior space for the scenario S. If more than one scenario is defined, we also write $\mathcal{N}(S)$ for clarity.

Finally, we define the *set of behaviors* $\mathcal{B} \subset \mathcal{N}$ as the set of vectors $b \in \mathcal{N}$ that satisfy the constraints

$$b \cdot c \geq 0 \tag{4.10}$$

for all records $c \in C(S)$. $\mathcal{B}$ is the intersection of finitely many half-spaces in $\mathcal{N}$ and therefore a convex polytope (see Chapter 1 for reference). We therefore also refer to $\mathcal{B}$ as the *polytope of behaviors*. Any vector $b \in \mathcal{B}$ is called a *behavior*. The probability for a record $c \in C(S)$ can be obtained from a behavior as

$$p(c) = b \cdot c. \tag{4.11}$$

We already established that $\mathcal{B}$ is a convex polytope. More specifically, $\mathcal{B}$ is a product of standard simplices, as defined in [12]. To see this, consider the subspace $V_J \subset V$ for a joint measurement $J \in \mathcal{J}(S)$ which is the span of the records $(r \mid J)$, $r \in R(J)$. The projection of the polytope of behaviors into V_J is a standard simplex as defined in Ref. [12]. Further note that every behavior $b \in \mathcal{B}$ is the direct sum of vectors $b_J \in V_J$. This

proves the statement. The vertices of this polytope are known and correspond to the *deterministic behaviors*, which are the behaviors that satisfy

$$p(c) \in \{0, 1\} \tag{4.12}$$

for all $c \in C(S)$.

In the following we discuss behaviors that satisfy further conditions. Given such conditions, we refer to the set $\mathcal{M} \subseteq \mathcal{B}$ of behaviors that satisfy the conditions as a *model*. In this chapter we discuss four types of models: The model of no-signaling behaviors, local hidden-variable models, the model of quantum behaviors, and hybrid models. To each of these models we devote one section, starting with no-signaling behaviors.

4.2 No-signaling behaviors

If an experiment is conducted on a system that is composed of several subsystems and measurements are performed locally in a space-like separated manner, then the behaviors are subject to the no-signaling constraints [66]: If the event e_1 of choosing a measurement on one subsystem A is space-like separated from the event e_2 of obtaining an outcome on another subsystem B, then the choice of measurement on subsystem A does not affect the statistics of measurement outcomes obtained on B. Otherwise a message encoded in the change of measurement statistics could be sent from one point in space-time to another space-like separated one, which would imply superluminal signaling, or *signaling* for short. This argument relies on the assumption that the measurement on subsystem A is indeed chosen at e_1 rather than being predetermined. This assumption is known as the *freedom of choice* assumption.

For a bipartite system, where x (a) is the input (output) on subsystem A and y (b) is the input (output) on system B, the absence of signaling from party B to party A can be stated as

$$p(a \mid x) = \sum_b p(a, b \mid x, y), \tag{4.13}$$

for all a, x and y. The condition that excludes signaling from A to B can be stated similarly. These are the no-signaling conditions for the bipartite case. For multipartite systems, the no-signaling conditions are the conditions that exclude signaling from any subset of parties to any other subset of parties. The only information necessary to state them is the scenario.

Geometrically, the no-signaling constraints define a hyperplane $\mathcal{NS}$ in the affine vector space $\mathcal{N}$. The *set of no-signaling behaviors* M_{NS} is the intersection of this hyperplane with the polytope of behaviors B. Hence, M_{NS} is a convex polytope.

4.2.1 Dimension of no-signaling behaviors

Following the arguments of Tsirelson [65], one can calculate the dimension of the bipartite no-signaling polytope as the dimension of the affine subspace $\mathcal{NS}$ it is contained in. We call this subspace the *no-signaling subspace*.

We begin with a bipartite scenario $S = S^A \cup S^B$ and denote the measurements on party A (B) as S^A (S^B). Consider two measurements $x \in S^A$, $y \in S^B$ and assume $1 \in O(x)$, $1 \in O(y)$. The crucial point in the calculation of the dimension of $\mathcal{NS}$ for this case is the observation that if all the marginals $p(a \mid x)$ and $p(b \mid y)$ are known, then all those probabilities $p(a, b \mid x, y)$ that affect either the outcome $a = 1$ or $b = 1$ can be omitted because they are redundant. In fact, they are already defined by the no-signaling conditions as

$$p(a, 1 \mid x, y) = p(a \mid x) - \sum_{b \in O(y) \setminus \{1\}} p(a, b \mid x, y) \tag{4.14}$$

and

$$p(1, b \mid x, y) = p(b \mid y) - \sum_{a \in O(x) \setminus \{1\}} p(a, b \mid x, y). \tag{4.15}$$

Also, the marginal probabilities $p(1 \mid x)$ and $p(1 \mid y)$ are redundant since the marginal distributions are normalized and hence

$$p(1 \mid x) = 1 - \sum_{a \in O(x) \setminus \{1\}} p(a \mid x), \tag{4.16}$$

and accordingly for subsystem B. The probabilities $p(a, b \mid x, y)$ and marginal probabilities $p(a \mid x), p(b \mid y)$, except for the ones containing $a = 1$ or $b = 1$, form an independent set of parameters and together fully specify a behavior. One can therefore obtain the dimension of $\mathcal{NS}$ by counting these parameters as

$$\begin{aligned} \dim(\mathcal{NS}) &= \sum_{x \in S^A} (|O(x)| - 1) + \sum_{y \in S^B} (O(y) - 1) + \sum_{x \in S^A} (|O(x)| - 1) \sum_{y \in S^B} (O(y) - 1) \\ &= (o_A - m_A + 1)(o_B - m_B + 1) - 1, \end{aligned} \tag{4.17}$$

where $o_A = \sum_{x \in S^A} |O(x)|$ ($o_B = \sum_{y \in S^B} |O(y)|$) is the sum of the number of outcomes for all measurements on subsystem A (B). This can be generalized to an arbitrary number of systems $k \in \{1, \ldots, n\}$ with measurements in S^k as [67]

$$\dim(\mathcal{NS}) = \prod_{k=1}^{n} \left(\sum_{m \in S^k} (|O(m)| - 1) + 1 \right) - 1, \tag{4.18}$$

where n denotes the number of parties.

4.2.2 Expectation value behavior space

Consider a scenario $S = \bigcup_{k=1}^{n} S^k$ where S^k is the set of measurements of party k such that $O(m) = \{-1, +1\}$ for all measurements $m \in S$. In this case Eq. (4.18) simplifies to

$$\dim(\mathcal{NS}) = \prod_{k=1}^{n} (|S^k| + 1) - 1. \tag{4.19}$$

As independent parameters we choose all components of $\mathcal{NS}$ in the dimensions given by the records that only include outcomes $+1$.

Alternatively, any point $v \in \mathcal{NS}$ for this scenario can be characterized by stating for each joint measurement $J \in \mathcal{J}(S)$ the quantity

$$\langle J \rangle := \sum_{r \in R(J)} v \cdot (r \mid J) \prod_{o_m \in r} o_m \tag{4.20}$$

where we write $o_m \in r$ for elements o_m of the sequence r. If v is a behavior, $v \cdot (r \mid J)$ has the interpretation of the probability $p(r \mid J)$ and $\langle J \rangle$ can hence be interpreted as the expectation value of the product of the outcomes of $m \in J$.

We continue with a proof of the statement that no-signaling behaviors for such scenarios can indeed be inferred from the expectation values of all joint measurements. We define the *expectation value behavior space* as the free vector space V' over $\mathbb{R}$ generated by $\mathcal{J}(S)$. Further, we equip V' with a scalar product such that the joint measurements $J \in \mathcal{J}(S)$ form an orthonormal basis.

We now show that there exists an affine mapping $\alpha : \mathcal{NS} \to V'$, such that

$$\alpha(v) \cdot J = \langle J \rangle \tag{4.21}$$

for all $v \in \mathcal{NS}$ and that α is bijective and unique. As a sanity check, note that $\dim(\mathcal{NS}) = |\mathcal{J}(S)| \equiv \dim(V')$. It is trivial that an affine mapping between these two spaces exists. It remains to show that α is unique and bijective. We show this by induction over $|J|$. We parametrize the points $v \in \mathcal{NS}$ in terms of their coefficients $v \cdot (1 \ldots 1 \mid J)$ for all joint measurements $J \in \mathcal{J}(S)$. We begin with the one-body marginals, that is $|J| = 1$,

$$\begin{aligned} \langle J \rangle &= v \cdot (1 \mid J) - v \cdot (-1 \mid J) \\ &= 2v \cdot (1 \mid J) - 1, \end{aligned} \tag{4.22}$$

for which the statement holds trivially. Assume that the t-body expectation values can be stated as invertible affine functions of the s-body probabilities for $s \leq t$. We now show that if the assumption holds for $s < t$, then it also holds for $s + 1$. This follows from the fact that $\langle J \rangle$ always depends non-trivially on the probability $p(1, \ldots, 1 \mid J)$ and does not depend on any probability $p(1, \ldots, 1 \mid J')$ with $J' \neq J$. This ends the proof.

4.2.3 Extremal no-signaling behaviors

As a convex polytope, the no-signaling polytope can be characterized by its vertices. The vertices can be found by solving the vertex enumeration problem, which is discussed in detail in Chapter 1. If the number of parties, measurement settings and outcomes per measurement is not too large, it is practicable to solve the vertex enumeration problem on a computer. In the scenario consisting of two subsystems with two dichotomic measurements each, any extremal no-signaling behavior is either deterministic or it represents a so-called *Popescu-Rohrlich box*, short PR box [68, 69]. A PR box is a black-box experiment with two parties A, B for the scenario $M = S^A \cup S^B$ with

$$\begin{aligned} S^A &= \{(A, 1, \{0,1\}), (A, 2, \{0,1\})\}, \\ S^B &= \{(B, 1, \{0,1\}), (B, 2, \{0,1\})\}. \end{aligned} \tag{4.23}$$

that yields the behavior with components [70]

$$p(a,b \mid x,y) = \begin{cases} \frac{1}{2}, & a+b \mod 2 = (i(x)-1)(i(y)-1) \\ 0, & \text{otherwise} \end{cases}, \tag{4.24}$$

for all measurements $x \in S^A$ and $y \in S^B$. As defined in the beginning of the chapter, $i(x)$ denotes the input of x. In words, this means the following: If Alice or Bob (or both) choose to measure in setting 1, then they always obtain the same outcome, which locally is uniformly distributed. However, if both of the parties choose to measure in setting 2, then they will always obtain opposite outcomes that also locally are uniformly distributed. Needless to say, the labeling of the measurement settings and outcomes is entirely arbitrary and behaviors that arise from Eq. (4.24) by any relabeling are also PR behaviors.

The extremal no-signaling behaviors have also been identified in the bipartite scenarios with an arbitrary number of local measurement settings, each of which leads to one of two possible outcomes [70]. Moreover, the extremal no-signaling behaviors are known for bipartite scenarios with two measurement settings per party that yield an arbitrary number of different outcomes [71].

4.3 Local realism

Local realism is a principle that restricts the set of behaviors [72]. In the following, we adopt the common practise of referring to local realism simply as locality, although this notion implies more than only no-signaling. To understand locality, we must first discuss possible interpretations of behaviors.

If a theory or model correctly predicts a deterministic behavior without disturbing it, then the measured physical quantity corresponds to an element of reality in

the sense of EPR [61]. The observations are then compatible with the idea that the measurement revealed the value of the physical quantity and the value was predetermined. In the case of most general behaviors as we considered in Section 4.1, all the extremal behaviors are deterministic and therefore compatible with this interpretation. The non-deterministic behaviors can then be explained with the incompleteness of the theoretical description [61]. To see this, consider different elements of reality that have the same counterpart in the theory. In this case, the measurements that correspond to these different elements of reality are not distinguished in the theory. Instead, the theory considers only one physical quantity. If this physical quantity is measured, one may actually measure a different element of reality in each round of the experiment. In the evaluation of the experiment, one would then mix the statistics of experimental runs that actually correspond to measurements of different physical quantities. Mathematically, this corresponds to forming a convex combination of deterministic behaviors, which itself is in general not deterministic. In this way, the incompleteness of the theoretical description can lead to a behavior that is probabilistic, although the underlying processes are deterministic.

A behavior is a *local deterministic behavior*, if it obeys the no-signaling constraints and every setting corresponds to an element of reality [17]. If a behavior is in the convex hull of the local deterministic behaviors, it is called a *local behavior* [1, 72]. The local behaviors form a polytope M_{LHV} called *local polytope* [16].

In the following we write the local polytope for an arbitrary scenario S in vertex representation. For any measurement $x \in S$, any function $\chi_x : O(x) \to \{0, 1\}$ is called a *local deterministic assignment* [72]. Any sequence $\lambda = (\chi_x)_{x \in S}$ of local deterministic assignments gives rise to to a local deterministic behavior $b_\lambda \in M_{\mathrm{LHV}}$ through the relations

$$b_\lambda \cdot ((o_x)_{x \in J} \mid J) = \prod_{x \in J} \chi_x(o_x). \tag{4.25}$$

The sequence λ that determines the outcomes of every measurement in S is called *local hidden variable* [72]. Denoting the finite set of local hidden variables as Λ, one can state the local polytope as

$$M_{\mathrm{LHV}} = \mathrm{conv}(\{b_\lambda \mid \lambda \in \Lambda\}). \tag{4.26}$$

This model is called a *local hidden-variable model* [72, 1]. For any bipartite scenario S with exclusively dichotomic measurements, the local polytope coincides with the correlation polytope (see Section 1.8) for the graph $(S, \mathcal{J}(S))$ [73, 74].

Note that the assumption of locality is stronger than the assumption of no-signaling [68]. As we discussed in the previous section, there exist no-signaling behaviors that cannot be decomposed into a convex combination of no-signaling deterministic behaviors. The most prominent examples for such behaviors are the PR behaviors, which are

non-deterministic. The local polytope is therefore a proper subset of the no-signaling polytope. It is known that the local polytope has the same affine dimension as the no-signaling polytope [67].

4.4 Bell inequalities

Since the set of local behaviors is a convex polytope, it can be characterized by a finite set of affine inequalities. Any inequality that is satisfied for all local behaviors is called a *Bell inequality* [72]. Such inequalities were first developed by Bell [2] [1]. If a Bell inequality supports a facet of the local polytope, it is also called *tight* [74]. Together, the facet-defining Bell inequalities of a local polytope characterize it. The problem of finding the facet-defining inequalities of a polytope is known as the *facet-enumeration problem*, and there are algorithms with which such problems can in principle be solved [18, 25]. We discuss this topic in more detail in Chapter 1. From a practical point of view, it is worth mentioning that local polytopes are special instances of 0/1-polytopes, for which specially designed algorithms exist that may solve problems such as facet-enumeration more efficiently than algorithms that are designed for more general problems [20]. However, if one considers scenarios with increasingly many parties, measurement settings or measurement outcomes per setting, solving the facet enumeration problem quickly becomes impracticable for the local polytope.

The most well-known tight Bell inequality is the *Clauser-Horne-Shimony-Holt* (CHSH) inequality [60], which is formulated for the so-called *CHSH scenario* with two measurement settings A_1, A_2 and B_1, B_2 on each party, that yield outcomes ± 1. It reads

$$\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \leq 2. \tag{4.27}$$

The CHSH inequality is violated by no-signaling behaviors and the left-hand side of the CHSH inequality yields the maximal value of 4 for a PR-box behavior [68]. Obviously, other versions of the CHSH inequality Eq. (4.27), for which the parties, settings, or outcomes have been relabeled, also support facets of the local polytope.

Apart from the CHSH inequality, there is another class of facet-defining inequalities for the local polytope for the same scenario. Representatives of this class have the form

$$\langle A_1 \rangle + \langle B_1 \rangle - \langle A_1 B_1 \rangle \leq 1. \tag{4.28}$$

This inequality is an example for a *trivial* Bell inequality. A Bell inequality is called *trivial*, if it is not violated by any no-signaling behavior [1]. All different versions of the CHSH inequality and the trivial inequality Eq. (4.28) are known to completely characterize the local polytope for the CHSH scenario [17, 16].

[1] Depending on the context, also valid inequalities for other models such as hybrid models are referred to as Bell or Bell-type inequalities [72].

In the bipartite scenario with three dichotomic measurements per party there is one more type of tight Bell inequality, called the I3322 inequality [16, 75, 76]

$$\begin{aligned}\langle A_1B_3\rangle + \langle A_2B_3\rangle + \langle A_3B_1\rangle + \langle A_3B_2\rangle + \langle A_1\rangle - \langle A_2\rangle + \langle B_1\rangle - \langle B_2\rangle \\ -\langle A_1B_1\rangle + \langle A_1B_2\rangle + \langle A_2B_1\rangle - \langle A_2B_2\rangle \leq 4.\end{aligned} \tag{4.29}$$

It is worth noting that the I3322 inequality includes so-called marginal correlations, such as $\langle A_1\rangle$. This means that in every round of the experiment, every party not only has the choice between one of the three measurements but can also choose not to perform a measurement at all. Inequalities for scenarios where this option of not performing a measurement is excluded are called *correlation inequalities* [77].

The I3322 Bell inequality is part of the I_{mmnn} family of Bell inequalities for 2 parties with m settings that yield n outcomes each [76]. The inequalities in the subfamily I_{mm22} are known to be tight [78].

In general, there are trivial ways to generalize Bell inequalities to larger scenarios. Any bipartite Bell inequality can be interpreted as a multipartite Bell inequality that is only sensitive to the measurements of two parties. Similarly, any Bell inequality with two settings per party is still valid in a multi-setting scenario. Finally, one can apply an inequality that features n outcome measurements in a scenario with more than n outcomes per measurement by only distinguishing between n different equivalence classes of outcomes. This process of adapting a Bell inequality to a larger scenario is called *lifting* [67, 74]. While lifted versions of Bell inequalities still define facets of the local polytope of the larger scenario, they are not interesting for physical applications, since they are defined in a way that prohibits them from exploiting the richer structure of the more complex scenario they are lifted to [67]. We will make this notion more precise in Section 5.3.

In more complex scenarios, it is not computationally feasible to enumerate the facets of the local polytope. As a result, it is common to consider families of Bell inequalities [76, 79, 80] or to devise methods that help to generate partial lists of facets [81, 74]. For instance, a partial list of facets has been generated for the bipartite case with 10 dichotomic measurements per party by generating facets of the local polytope from the cut polytope using a method called triangular elimination [74].

In multipartite scenarios with two dichotomic measurements per party, all correlation inequalities can be constructed analytically [77]. One of these inequalities is Mermin's inequality

$$\langle A_1B_1C_2\rangle + \langle A_1B_2C_1\rangle + \langle A_2B_1C_1\rangle - \langle A_2B_2C_2\rangle \leq 2. \tag{4.30}$$

In quantum mechanics, the left-hand side can reach a value of 4 as can be seen from the GHZ argument [1], see also Section 4.8.1.

The tripartite scenario with two dichotomic measurements has been characterized completely [75]. For three parties with three dichotomic settings, correlation inequal-

ities that are symmetric under party permutations have been found [81]. Symmetric inequalities have also been found for up to 5 parties with two dichotomic settings per party [81]. Further, a family of tight multipartite Bell inequalities with many dichtomic settings has been derived in Ref. [79].

It is worth noting that also non-linear Bell inequalities have been considered. One example for this are information-theoretic Bell inequalities [82]. They based on the fact that two quantum systems can share information in a way that is prohibited classically. Specifically, consider the CHSH scenario and assume that local-realism holds. Then, one can assume that the outcomes of the measurements A_1, A_2, B_1, B_2 are well defined at all times. Hence the following estimate holds for the entropy of the measurements

$$\begin{aligned} H(A_1B_1) &\leq H(A_1B_2A_2B_1) \\ &= H(A_1 \mid B_2A_2B_1) + H(B_2 \mid A_2B_1) + H(A_2 \mid B_1) + H(B_1) \\ &\leq H(A_1 \mid B_2) + H(B_2 \mid A_2) + H(A_2 \mid B_1) + H(B_1), \end{aligned} \tag{4.31}$$

which can be simplified to

$$H(A_1B_1) \leq H(A_1 \mid B_2) + H(B_2 \mid A_2) + H(A_2 \mid B_1). \tag{4.32}$$

This inequality is violated in quantum mechanics [82].

4.5 Quantum behaviors

Consider a scenario S that contains measurements, each of which is performed on a system $k \in \Sigma = \{1, \dots n\}$. Since each measurement m only affects one party, these measurements are called *local*. Assume that we can associate a finite-dimensional Hilbert space $\mathcal{H}_k$ with each system k. We write $\mathcal{H}_\Sigma$ for the Hilbert space of the system that consists of subsystems in Σ. Each of the local measurements $m \in S$ is modeled with a POVM with sample space $O(m)$ and event space $\mathfrak{P}(O(m))$, where $\mathfrak{P}$ denotes the power set. We can then write this POVM as $P_m : \mathfrak{P}(O(m)) \to E(\mathcal{H}_{k(m)})$, where $E(\mathcal{H})$ is the set of effects on a Hilbert space $\mathcal{H}$. P_m can be defined in terms of the effects $P_m(o)$ that it assigns to the events $o \in O(m)$.

Now consider a joint measurement $J \in \mathcal{J}(S)$. This joint measurement is modeled by a POVM, too, and we denote it as $P_J : \mathfrak{P}(R(J)) \to E(\mathcal{H}_\Sigma)$. Let $r = (o_m)_{m \in J}$ be a result of J, then P_J is defined by the POVMs P_m with $m \in J$ such that

$$P_J(r) = \bigotimes_{m \in J} P_m(o_m). \tag{4.33}$$

We define $E_c = P_J(r)$ for any record $c = (r \mid J) \in C(S)$. Any quantum behavior $b_q \in \mathcal{B}$ can then be expressed as

$$b_q = \sum_{c \in C(S)} \operatorname{Tr}(E_c \varrho)\, c \tag{4.34}$$

where $\varrho \in S(\mathcal{H}_\Sigma)$ is a quantum state and the effects E_c are defined as in Eq. (4.33).

Vice versa, if it is possible for a behavior to find a state and POVMs such that Eq. (4.34) holds, then this behavior is called a *quantum behavior*. We denote the set of quantum behaviors as M_Q. M_Q forms a proper subset of the no-signaling polytope [69, 68]. One can show this by proving that the marginal probability distributions are well-defined. This follows directly from Eq. (4.33) and the fact that for any POVM the measure of the sample space is the identity operator.

The set of quantum behaviors is strictly contained in the set of no-signaling behaviors, as one can see from considering the CHSH inequality. For no-signaling behaviors, the tight bound for the CHSH expression is [68]

$$\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \leq 4 \tag{4.35}$$

while for quantum behaviors it is [66]

$$\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \leq 2\sqrt{2}. \tag{4.36}$$

Since inequality Eq. (4.36) bounds the set of quantum behaviors, it is also called a *quantum bell inequality* or *Tsirelson inequality* after its inventor [66].

The set of local behaviors is strictly included in the set of quantum behaviors [19]. This is easily understood as the set of quantum behaviors is convex and local deterministic behaviors are reproducible with deterministic POVMs. The fact that the Tsirelson bound is tight and larger than the classical bound shows that this inclusion is strict. The prediction of nonlocal behaviors in quantum mechanics shows that quantum theory is irreconcilable with local realism [2]. This result is known as Bell's theorem [1]. As a formula this statement is summarized as

$$M_{LHV} \subset M_Q \subset M_{NS} \subset \mathcal{B}. \tag{4.37}$$

This relation is illustrated in figure 4.2. Violations of Bell inequalities have also been confirmed in experiments [5, 3, 4], ruling out local realism as a physical principle.

Given an affine Bell inequality for a scenario S, that is

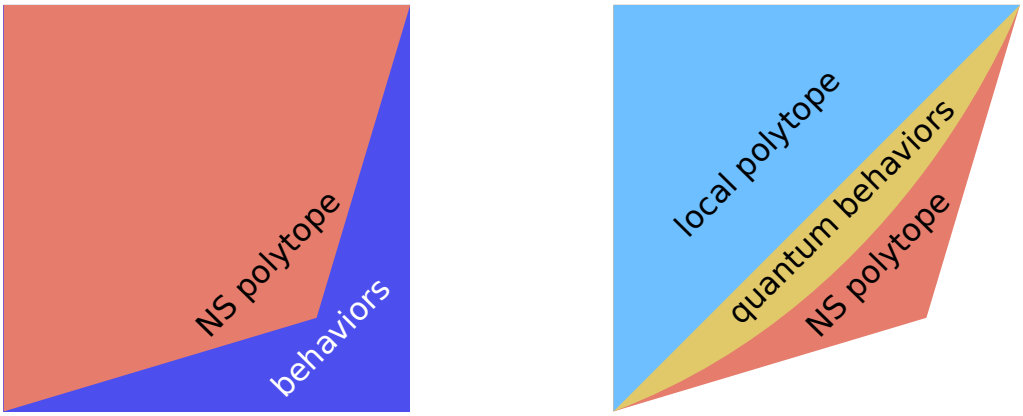
$$f \cdot b \leq \beta, \quad \text{for all } b \in M_{\text{LHV}}, \tag{4.38}$$

we can write this Bell inequality using the formalism of quantum mechanics. To this end, we first express the Bell inequality in the basis of records $c \in C(S)$. This yields

$$\sum_{c \in C(S)} f_c b_c \leq \beta. \tag{4.39}$$

As a second step, we assume that $b \in M_Q$ by using Eq. (4.34). In this way, we obtain

$$\sum_{c \in C(S)} f_c \text{Tr}\left(E_c \varrho\right) \leq \beta. \tag{4.40}$$



(a) (b)

Figure 4.2: Schematic depiction of the set of behaviors as well as the subsets of no-signaling behaviors, quantum behaviors, and local behaviors. (a) The set of behaviors $\mathcal{B} \in \mathcal{N}$ is projected into the no-signaling subspace. The image is then mapped into the expectation value behavior space using the map α. This set is a hypercube (dark blue). Its vertices are deterministic behaviors. Not all of these deterministic behaviors are no-signaling. Therefore, the no-signaling behaviors form a strict subset. (b) The set of no-signaling behaviors $\mathcal{B} \cap \mathcal{NS}$ for a scenario is mapped into the expectation value behavior space of the same scenario via the map α. Every extremal behavior of the local polytope (light blue) is also an extremal behavior of the no-signaling polytope (red). Additionally, the no-signaling polytope contains non-deterministic extremal behaviors. The set of quantum behaviors (orange) is strictly included in the no-signaling polytope and in turn strictly includes the local polytope.

One defines the *Bell operator*

$$\Xi = \sum_{c \in C(S)} f_c E_c, \tag{4.41}$$

with which one can rewrite the Bell inequality as

$$\mathrm{Tr}\,(\varrho \Xi) \leq \beta, \tag{4.42}$$

for a $\varrho \in S(\mathcal{H}_\Sigma)$. For any choice of POVMs P_J with $J \in \mathcal{J}(S)$, one can therefore compute an entanglement witness from a Bell inequality as

$$W = \beta \mathbb{1} - \Xi. \tag{4.43}$$

The set of quantum behaviors for arbitrary but finite local dimensions can be approximated using the NPA hierarchy [37]. In this case the set of quantum behaviors

can equivalently stated in a different way: Instead of imposing a tensor product structure on the measurements as in Eq. (4.33), one can instead impose that effects commute, if they are associated with measurements on different subsystems [42]. Stated in this way, polynomial optimization problems over the set of quantum behaviors are tractable with the NPA hierarchy. This includes the problem of finding upper bounds for the quantum violation of a Bell inequality [42]. The reader is referred to Section 2.3.2 for details.

In general, characterizing the set of quantum behaviors is not easy. However, in the case of fully separable quantum states the set of behaviors is simply the local polytope [48]. This result is readily obtained, if in Eq. (4.34) one assumes that ϱ is fully separable and the effects have the form of Eq. (4.33). This yields a local hidden-variable model for the resulting quantum behaviors. Nonlocality therefore implies entanglement. However, entanglement does not imply nonlocality. There are mixed entangled states that exclusively give rise to behaviors that are compatible with a local hidden-variable model [83].

4.6 Hybrid models

Bell experiments have established the existence of nonlocal behaviors [5, 3, 4]. However, these experiments do not answer the question how nonlocality can be distributed among multipartite systems. For example, consider a three party system. One may ask, whether tripartite nonlocality can be explained by a model that sets the following rule: In any round of the experiment there may be communication between any two parties. Thus, these two parties can effectively be regarded as one party. However, the correlations between these two parties on the one hand and the third party on the other have to be compatible with a LHV model. Such a model is called a *hybrid model* and it is qualitatively different from a LHV model because from round to round of the experiment, different subsets of subsystems are allowed to communicate [84, 85]. As a formula, the joint probabilities for a three party hybrid model can be expressed as

$$\begin{aligned} p(a,b,c \mid x,y,z) = &\sum_{\lambda_1} p(a,b \mid x,y,\lambda_1)p(c \mid z,\lambda_1) \\ &+ \sum_{\lambda_2} p(a,c \mid x,z,\lambda_2)p(b \mid y,\lambda_2) \\ &+ \sum_{\lambda_3} p(bc \mid y,z,\lambda_3)p(a \mid x,\lambda_3), \end{aligned} \tag{4.44}$$

where a,b,c are the outputs on the parties A,B,C and x,y,z are the respective inputs [84].

If a behavior violates this model, then it exhibits genuine tripartite nonlocality. This

can be tested with the Svetlichny inequality

$$\begin{aligned} &+ \langle A_1 B_1 C_2 \rangle + \langle A_1 B_2 C_1 \rangle + \langle A_2 B_1 C_1 \rangle - \langle A_2 B_2 C_2 \rangle \\ &+ \langle A_2 B_2 C_1 \rangle + \langle A_2 B_1 C_2 \rangle + \langle A_1 B_2 C_2 \rangle - \langle A_1 B_1 C_1 \rangle \leq 4, \end{aligned} \tag{4.45}$$

and all measurements yield outcomes ± 1. In quantum mechanics, the Svetlichny inequality is violated up to a value of roughly 5.66 [84].

In the case of four parties, three different hybrid models have been considered [85]. In the 2/1/1 model in each round two parties can cooperate, so the joint behavior of these two parties is arbitrary. In the 2/2 model, two teams of two parties can cooperate every round and finally, in the 3/1 model, three parties can cooperate [85]. A generalization of the Svetlichny inequality has been studied in Ref. [85] that detects genuine multipartite nonlocality for all of the four-body hybrid models mentioned above. Further, generalized Svetlichny inequalities have been defined for the n-partite case for $k/(n-k)$ models [85]. These generalized Svetlichny inequalities are defined using the Mermin-Klyshko polynomials, which are defined recursively as [85]

$$M_n = \frac{1}{2} M_{n-1}(A_1^n + A_2^n) + \frac{1}{2} M'_{n-1}(A_1^n - A_2^n), \tag{4.46}$$

$$M_1 = A_1^1, \tag{4.47}$$

where A_j^n is the j-th measurement setting on party n and M'_k is obtained from M_k by exchanging the settings 1 and 2 on each party. The generalized Svetlichny inequality is the same for all hybrid models for the same number of parties n and reads

$$S_n \leq \begin{cases} 2^{\frac{n-2}{2}}, \, n \text{ even} \\ 2^{\frac{n-3}{2}}, \, n \text{ odd} \end{cases}. \tag{4.48}$$

The generalized Svetlichny polynomial S_n is defined as

$$S_n = \begin{cases} M_n, \, n \text{ even}, \\ \frac{1}{2}(M_n + M'_n), \, n \text{ odd} \end{cases}. \tag{4.49}$$

The quantum bound is always by a factor of $\sqrt{2}$ larger than the hybrid model bound [85].

Svetlichny inequalities have been generalized to scenarios with an arbitrary number of outputs [86] and measurement settings [87] for $k/n-k$ hybrid models.

4.7 Guess-your-neighbors-input inequalities

This section contains text that has been published in our paper [B].

Guess-your-neighbors-input (GYNI) inequalities are Bell inequalities that are derived from the game called 'Guess your neighbors input' [9]. The game is played with

n parties that are arranged in a ring and collaboratively play against the instructor. The instructor will supply every participant i with an input bit x_i and each participant then has to guess the input of his left neighbor. The game is won if the output a_i of every player output matches the input x_{i+1} of their left neighbor. In order to achieve this goal, the players may agree on a strategy before the game and they are also provided with the probability distribution q according to which the instructor chooses the inputs. In the quantum version of the game, each party additionally possesses a part of a shared quantum state. During the game, no communication is permitted. A GYNI inequality is of the form

$$\sum_{\vec{x}} q(\vec{x}) P(a_1 = x_2, a_2 = x_3, \ldots, a_n = x_1 | x) \leq \omega_c, \tag{4.50}$$

whereby the left-hand-side expresses the average winning probability, which is bounded by the average winning probability of the best classical strategy ω_c.

As is the case with any other Bell inequality, the violation of a GYNI inequality would indicate a violation of local realism. However, GYNI inequalities are not only satisfied for local behaviors but also for quantum behaviors. Therefore, GYNI inequalities are Bell inequalities and simultaneously quantum Bell inequalities. Notably, GYNI inequalities can be violated for no-signaling behaviors [9].

One particular GYNI inequality is the facet-defining inequality listed as inequality number 10 in the paper by Śliwa [75]. This inequality reads

$$\begin{aligned} &\langle A_1 B_1 \rangle + \langle A_2 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_2 \rangle + \langle A_1 C_1 \rangle - \langle A_2 C_1 \rangle \\ &+ \langle B_1 C_1 \rangle + \langle A_1 B_1 C_1 \rangle - \langle B_2 C_1 \rangle - \langle A_2 B_2 C_1 \rangle + \langle A_1 C_2 \rangle \\ &- \langle A_2 C_2 \rangle - \langle B_1 C_2 \rangle + \langle A_2 B_1 C_2 \rangle + \langle B_2 C_2 \rangle - \langle A_1 B_2 C_2 \rangle \quad \leq \quad 4. \end{aligned} \tag{4.51}$$

4.8 Inequality-free tests of nonlocality

The perhaps most natural way to test nonlocality is to test an affine Bell inequality that supports a facet of the local polytope. However, there are other ways of testing nonlocality, two of which we are going to discuss briefly in this section.

4.8.1 Greenberger-Horne-Zeilinger argument

There are also ways to detect nonlocality entirely without inequalities. One inequality-free Bell test is the Greenberger-Horne-Zeilinger (GHZ) argument [1, 88]. Assume a three party scenario, where every party has two measurement settings – A_1, A_2 on party A, B_1, B_2 on party B, and C_1, C_2 on party C – and every measurement yields

outcomes ± 1. If the three parties observe the correlations

$$\langle A_2 B_1 C_1 \rangle = 1, \tag{4.52a}$$

$$\langle A_1 B_2 C_1 \rangle = 1, \tag{4.52b}$$

and

$$\langle A_1 B_1 C_2 \rangle = 1, \tag{4.52c}$$

and local realism holds, then every measurement yields one particular outcome with certainty. One can then associate every measurement with its outcome, which allows us to write

$$\begin{aligned} \langle A_2 B_2 C_2 \rangle &= (A_2 B_1 C_1) \cdot (A_1 B_2 C_1) \cdot (A_1 B_1 C_2) \\ &= 1. \end{aligned} \tag{4.53}$$

In quantum mechanics however, it is possible to violate the above equality Eq. (4.53) under the assumption that the conditions Eq. (4.52) are met [1]. If the measurements A_2, B_2, C_2 are described by the observable σ_x and the measurements A_1, B_1, C_1 are described by the observable σ_y, then the state

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}(|000\rangle - |111\rangle) \tag{4.54}$$

satisfies the relations Eq. (4.52). However, it holds that

$$\langle A_2 B_2 C_2 \rangle = -1. \tag{4.55}$$

4.8.2 Hardy's Paradox

A similar argument, which is known as Hardy's paradox, can be made for the CHSH setting [89]. If it holds that

$$p(1, 1 \mid A_1 B_1) = 0, \tag{4.56}$$

$$p(-1, 1 \mid A_1 B_2) = 0, \tag{4.57}$$

and

$$p(1, -1 \mid A_2 B_1) = 0, \tag{4.58}$$

then the assumption of local realism implies

$$p(1, 1 | A_2 B_2) = 0. \tag{4.59}$$

One way to see this is to consider a graph as in figure 4.3. The vertices of this graph are all pairs of measurements and outcomes, such as $(A_1, +1)$ for example. To vertices are connected by an edge, if these pairs are mutually exclusive. For example, $(A_1, +1)$ and $(B_1, +1)$ are mutually exclusive since $p(1, 1 \mid A_1 B_1) = 0$. However, in quantum mechanics the conditions can be satisfied while at the same time Eq. (4.59) can be violated up to $p(1, 1 | A_2 B_2) = \frac{1}{2}(5\sqrt{5} - 11) \approx 9\%$.

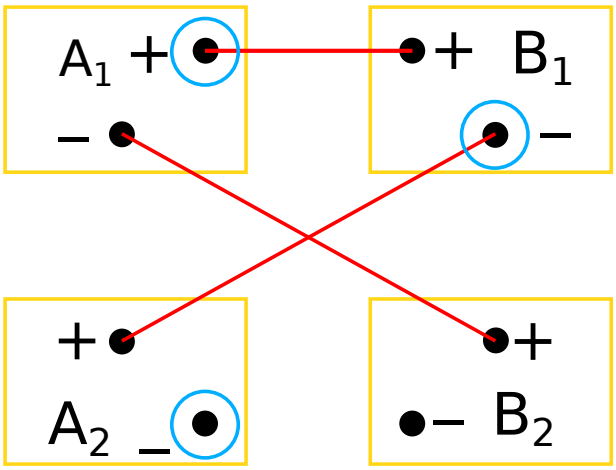


Figure 4.3: Graph for Hardy's argument – mutually exclusive outcomes according to the equations Eqs. (4.56-4.58) are connected by an edge. The task is to assign an outcome to every measurement. This is indicated by the blue circles. Since the graph is symmetric under exchange of the parties A, B, one can assume without loss of generality that A_1 yields outcome $+1$. This implies that B_1 yields -1 and hence A_2 yields -1. Therefore, we get $p(+1, +1 \mid A_2 B_2) = 0$.

4.9 Applications of Bell inequalities

Bell inequalities bound the image of functionals on the set of behaviors. Crucially, behaviors only reflect the observations that were made in an experiment; the formalism of behaviors does not presume anything regarding the description of the internal parameters of the experiment other than a consistent use of labels. Hence, functionals on behaviors yield information on the experiment that is *device independent*. Device independence is useful in a variety of situations. One application is the certification of quantum states and measurements, which is known as self-testing [11]. For instance, if a behavior maximally violates the CHSH inequality, then it can be shown that the quantum states and measurements that lead to this violation are unique up to an isomorphism [11]. In this way, a violation of a Bell inequality can provide a device independent certification of entanglement. For example, if the CHSH inequality is tested on a two-qubit system and a maximal violation is detected, this means that the quantum state is equivalent to a Bell-state up to local unitary transformations [90]. Hence in this case maximal nonlocality implies maximal entanglement [48]. For other Bell inequalities this is in general not the case [59].

Another application of Bell inequalities is device independent quantum key distribution (DIQKD) [91]. If a behavior violates a Bell inequality, this indicates that no LHV

model exists that could explain the observed behavior. Hence the local measurement outcomes cannot be predicted perfectly. Moreover, a Bell inequality violation necessitates that the local measurement outcomes are correlated. In DIQKD these correlations are exploited to share information – a raw key – between several parties, while the non-existence of a LHV model guarantees that no eavesdropper can have perfect knowledge of this shared information.

In the following, we briefly discuss the DIQKD protocol which was introduced in Ref. [92] and is similar to the protocol suggested by Ekert [91]. It relies on the violation of the CHSH inequality. Suppose that Alice and Bob share the Bell state

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}(\,|00\rangle + |11\rangle) \tag{4.60}$$

and perform the optimal measurements

$$A_1 = \frac{1}{\sqrt{2}}(\sigma_x + \sigma_z), \tag{4.61}$$

$$A_2 = \frac{1}{\sqrt{2}}(\sigma_x - \sigma_z), \tag{4.62}$$

$$B_1 = \sigma_x, \tag{4.63}$$

$$B_2 = \sigma_z, \tag{4.64}$$

to reach a maximal violation of the CHSH inequality. Additionally, Alice can perform a measurement

$$A_3 = B_1, \tag{4.65}$$

so that when Alice measures A_3 while Bob measures B_1, they obtain the same outcome. This joint outcome then becomes part of the raw key. It has been shown that in this case it is impossible for an eavesdropper to have any information about the shared key [92]. If the state and the measurements are imperfect, this is no longer the case, but the protocol can still be used and security can be ensured by using a method called *privacy amplification* [92].

5 Cone-projection technique and generalized Bell inequalities

The content of this chapter is based on our publications [A, B, C]. The first section describes the cone-projection technique and contains text and figures from publications [A, B]. The second section is taken from our manuscript [C]. The third section is yet unpublished material. In the second and third section we introduce the concept of a generalization of a Bell inequality in full generality and state a precise definition of the concept.

Bell inequalities are relevant for many problems in quantum information science, but finding them for many particles is computationally hard. Furthermore, the number of facet-defining Bell inequalities increases rapidly, making it difficult to identify the inequalities of interest. To give an example, there are only eight non-trivial Bell inequalities all of which are versions of the famous Clauser-Horne-Shimony-Holt (CHSH) inequality [60, 93] in the scenario with two parties and two different dichotomic measurements per party each. For the same number of parties, but three dichotomic measurements per party, there are already 648 non-trivial facet-defining inequalities, 72 of which are of the CHSH type and 576 are variations of the so-called I3322 inequality. The latter inequality was first identified by Froissart [16] and later independently by Śliwa [75] and Collins and Gisin [76]. If one increases the number of parties, for three parties with two dichotomic measurements each, there are 53856 facet-defining inequalities and 46 inequivalent classes of Bell inequalities [75]. Going to even more complex scenarios it is impossible so far to compute all the facets.

Luckily however, in practice we rarely need a complete characterization of the local polytope. Instead, we seek Bell inequalities with properties that are suitable for a specific purpose.

To address this problem we propose a general method for this problem, which is computationally feasible. Henceforth we refer to this method as the cone-projection technique (CPT). The CPT can be used to find all optimal Bell inequalities obeying some affine constraints.

Specifically, one important task that can be addressed with the CPT is finding generalizations of a Bell inequality. A generalization of a Bell inequality is a Bell inequality that inherits some properties from the Bell inequality it generalizes. At the same time, it may outperform the Bell inequality it generalizes in some tasks. This makes

it desirable to find such generalizations of Bell inequalities. In this chapter, we give a precise definition of our notion of a generalization of a Bell inequality and a detailled explaination of the cone-projection technique.

5.1 Description of the cone-projection technique (CPT)

In this section, we provide a detailed description of the cone-projection method.

We wish to find a method that makes it possible to compute facets of a polytope which is defined by its vertices, where the facet normal vectors obey some linear constraints. This can be interpreted geometrically as the condition that the facet normal vectors must have a fixed inner product with some vector $\boldsymbol{g}$ that represents the condition, see also Fig. 5.1.

A naive method to tackle this problem is to compute first all facets and then to find out which of them obey the constraint. For the problems we consider, however, this is not feasible, as it is already impossible to compute all the facets.

Note that constraints of the considered type include two important special cases of constraints that are important in this work: The condition that some given vertex of the polytope should lie on the desired facet and the condition that the normal vector should be symmetric under some linear transformation. In the first case, the position vector of the vertex in question takes the role of $\boldsymbol{g}$. To understand this, consider a Bell inequality $\boldsymbol{x}^T\boldsymbol{b} \leq -\beta$ with facet normal vector $\boldsymbol{b}$ that should hold for all classical behaviors $\boldsymbol{x}$. Demanding that some vertex $\boldsymbol{g}$ lies on the facet means that $\boldsymbol{g}^T\boldsymbol{b} = -\beta$, which is an affine constraint. In the second case all points that obey the symmetry lie in a plane and the normal vectors of the desired facets have to lie in this plane as well. This is illustrated in Fig. 5.2.

We consider the situation where a convex polytope P is given in its V-representation and we aim to find all facets of P that satisfy some affine conditions. Note that the CPT can be applied to any convex polyhedron, but in practice we are interested in the local polytope, the vertices of which are the local deterministic behaviors.

We consider a D-dimensional polytope P and affine conditions on the facet normal vectors. Each of these conditions can be written in the form

$$\boldsymbol{g}_k^T\boldsymbol{b}_P = \gamma_k, \tag{5.1}$$

that is, the normal vector $\boldsymbol{b}_P$ has a fixed scalar product with some vector $\boldsymbol{g}_k$, see Fig. 5.3(a). The task is to find *all* the facet normal vectors that obey these constraints.

In the first step, we construct a cone C in $(D+1)$-dimensional space that maintains a one-to-one correspondence to the polytope. One way to achieve this is to prepend one fixed coordinate to every vertex $\boldsymbol{v}_i$,

$$\boldsymbol{v}_i \mapsto \boldsymbol{w}_i = \begin{pmatrix} 1 \\ \boldsymbol{v}_i \end{pmatrix}. \tag{5.2}$$

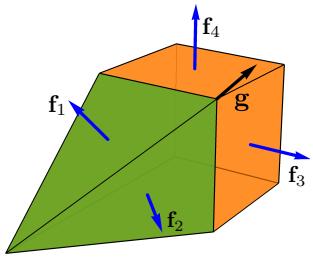


Figure 5.1: A polytope in three-dimensional space is defined by its two-dimensional surfaces, the so-called facets. Here, the normal vectors of the facets are depicted as blue (gray) vectors. The vector g (black arrow) may be used to define a condition on the facets, be requiring that the facet normal vectors f_i must enclose some absolute angle α. For instance, this angle may be chosen such that the facets drawn in green (dark gray) meet the constraints while the others drawn in orange (light gray) do not.

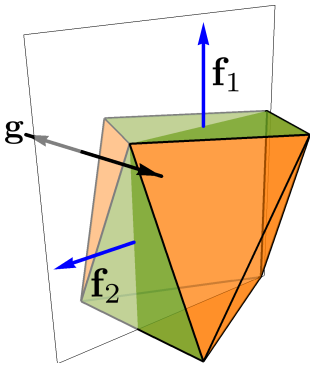


Figure 5.2: This example shows how symmetries can be formulated as affine constraints. We seek facets with normal vectors that are invariant under reflection on a plane (white, half-transparent). This is satisfied by the normal vectors f_1, f_2 (blue [dark gray]) of two of the facets (green [gray]). The condition is equivalent to demanding that the normal vectors be perpendicular to the normal vectors (black arrows) of the mirror plane.

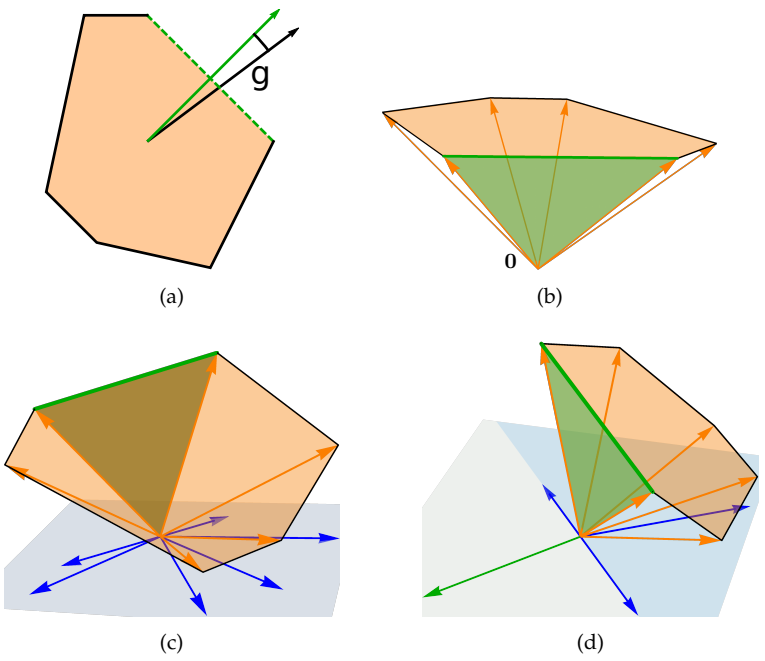


Figure 5.3: Visualization of the CPT to find facets of a polytope where the normal vector obeys some constraints. (a) We aim to find the facets of the two-dimensional polytope P where the normal vector $\boldsymbol{b}_P$ has a fixed scalar product with some vector $\boldsymbol{g}$. The facet and the normal vector that fulfill this constraint in the given example are shown in green (dark gray). (b) We embed the polytope P in a plane in three-dimensional space, where the plane does not contain the origin $\mathbf{0}$. The vertices of P define rays (orange [light gray] arrows) that define the cone C. The polytope P is then the intersection of the cone C with the plane, and each of its facets relates to a facet of C (green [dark gray]) in a unique manner. (c) The initial constraint on the facet of P can be translated to conditions on the facets of the cone. A facet of P fulfills the constraint if and only if the corresponding facet of C has a normal vector $\boldsymbol{b}_C$ that obeys a linear constraint $G\boldsymbol{b}_C = 0$, where G is some matrix. Geometrically, this means that $\boldsymbol{b}_C$ has to lie in a certain plane (light gray). Then, we project the rays of C (orange [light gray]) into that plane (blue [dark gray] arrows) to define a cone $\tilde{C}$ in the low-dimensional plane by taking projected rays as generators. By construction, facets of C which obey the constraint are also facets of the projected cone $\tilde{C}$. (d) Finally, we find the facets of $\tilde{C}$ and check which ones correspond to facets of C. From the facets of C that meet the conditions we can then compute the corresponding facets of P. In our example, $\tilde{C}$ is a half plane (light blue [light gray]) and has only one facet with the normal vector in green [gray]. It is also the normal vector of a facet of C (green [gray]). Note that in the given example $\tilde{C}$ is already generated by three rays and the other three rays are redundant.

and to define the cone C as the conic hull of the rays $\boldsymbol{w}_i$. In this way, the polytope can be seen as the intersection of the cone C with the hyperplane defined by $x_0 = 1$. This relationship is illustrated in Fig. 5.3(b). Notably, there is a one-to-one correspondence between the facets of the cone and those of the polytope, and a normal vector $\boldsymbol{b}_P$ of a polytope facet translates to a normal vector $\boldsymbol{b}_C$ of a cone facet. This is easy to formulate in the H-representation of the polytope and the corresponding cone. Let

$$\boldsymbol{x}^T \boldsymbol{b}_P \leq -\beta \tag{5.3}$$

be a facet-defining inequality of the polytope. Then,

$$(1\ \boldsymbol{x}^T)\begin{pmatrix}\beta\\ \boldsymbol{b}_P\end{pmatrix} \leq 0 \tag{5.4}$$

is the corresponding facet-defining inequality of the cone. Note that then $\binom{\beta}{\boldsymbol{b}_P}$ is the normal vector of the facet, and any normal vector of a facet can be written in this way.

The correspondence between the polytope and the cone enables us to work with the cone instead of the polytope. This construction also makes it possible to write the conditions in Eq. (5.1) in a linear form, namely

$$(g_{0,k}\ \boldsymbol{g}_k^T)\begin{pmatrix}\beta\\ \boldsymbol{b}_P\end{pmatrix} = 0, \tag{5.5}$$

where we set $g_{0,k} = -\frac{\gamma_k}{\beta}$.

Collecting all the facet conditions yields the linear matrix equation

$$G\boldsymbol{b}_C = 0, \tag{5.6}$$

where the k-th row of G is the row vector $(g_{0,k}\ \boldsymbol{g}_k^T)$ and the facet normal vector of the cone is $\boldsymbol{b}_C = \binom{\beta}{\boldsymbol{b}_P}$. Geometrically speaking, Eq. (5.6) defines a hyperplane through the origin in which the facet normal vectors of the cone must lie in order to comply with the conditions in Eq. (5.1), see also Fig. 5.3(c).

The key observation is that in this situation we can define a new cone $\tilde{C}$, such that if $\boldsymbol{b}_C$ is a facet normal vector of C that obeys the constraints, then $\boldsymbol{b}_C$ is also a facet normal vector of $\tilde{C}$. This is done by projecting the rays of C down to the subspace of vectors obeying $G\boldsymbol{v} = 0$, see Fig. 5.3(c) and Fig. 5.4.

This subspace is the kernel of G and it is spanned by a set of K vectors $\boldsymbol{t}_i$, where K is the dimension of the kernel. The advantage of $\tilde{C}$ is that its dimension K is typically considerably smaller than the dimension of C. Additionally, $\tilde{C}$ has typically much fewer rays than C. That makes it easier to find all the facets of $\tilde{C}$, compared with C, see Fig. 5.3(d).

In practice, the first step in the construction of $\tilde{C}$ is to define the $(D+1) \times K$ matrix T, whose K columns of length $D+1$ are given by the vectors $\boldsymbol{t}_j$, so we have $T_{ij} = [\boldsymbol{t}_j]_i$. With this, we define the rays $\tilde{\boldsymbol{w}}_i$ of $\tilde{C}$ as

$$\tilde{\boldsymbol{w}}_i^T = \boldsymbol{w}_i^T T, \tag{5.7}$$

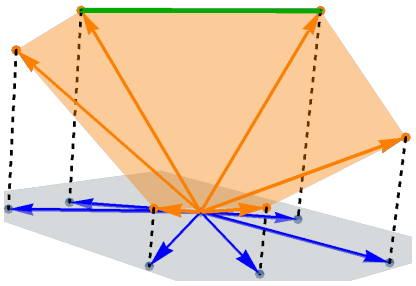


Figure 5.4: Detailed view of the of the projection of the rays of the cone C onto the plane where $G\boldsymbol{b}_C = 0$. The projected rays (in blue [dark gray]) generate a new cone $\tilde{C}$. Depending on the conditions, this new cone may have a significantly reduced dimension compared with C.

where $\boldsymbol{w}_i$ are the rays of C. Note that this is a projection if the vectors $\boldsymbol{t}_j$ form an orthonormal basis of the kernel of G. However, this is in general not necessary and in practice it can be preferable to pick vectors $\boldsymbol{t}_j$ with integer coefficients, so that the rays of $\tilde{C}$ have integer coefficients, if the rays of C have integer coefficients. In this way, one does not need to worry about the precision of the numerical calculations.

Similarly to the rays of C, we can express its facet normal vectors $\boldsymbol{b}_C$ (that satisfy the constraints) in the basis $\boldsymbol{t}_j$ of the kernel of G as

$$\boldsymbol{b}_C = T\boldsymbol{b}_{\tilde{C}}, \tag{5.8}$$

where $\boldsymbol{b}_{\tilde{C}}$ is a vector of dimension K.

The following theorem is the central result and establishes the previously claimed relation between $\tilde{C}$ and C. Namely, it states that any time $\boldsymbol{b}_C$ is a facet normal vector of C that satisfies the conditions, $\boldsymbol{b}_{\tilde{C}}$ is a facet normal vector of $\tilde{C}$. In this way, the facet normal vectors of $\tilde{C}$ are the only relevant vectors that one needs to consider.

Theorem 5.1. Let $C = \mathrm{conv}(\{\boldsymbol{w}_i\})$ be a cone and $\boldsymbol{b}_C$ a facet normal vector of C that satisfies $G\boldsymbol{b}_C = 0$ for some matrix G. With T and $\boldsymbol{b}_{\tilde{C}}$ defined as above, we define the cone $\tilde{C} = \mathrm{conv}(\{\tilde{\boldsymbol{w}}_i\})$ of dimension K with $\tilde{\boldsymbol{w}}_i^T = \boldsymbol{w}_i^T T$. Then $\boldsymbol{b}_{\tilde{C}}$ defines a facet of $\tilde{C}$.

Proof. We prove the statement in three steps. (1) The inequality $\tilde{\boldsymbol{w}}_i^T \boldsymbol{b}_{\tilde{C}} \leq 0$ holds, since Eq. (5.8) together with the definition of the $\tilde{\boldsymbol{w}}_i$ implies

$$\tilde{\boldsymbol{w}}_i^T \boldsymbol{b}_{\tilde{C}} = \boldsymbol{w}_i^T \boldsymbol{b}_C \tag{5.9}$$

and $\boldsymbol{w}_i^T \boldsymbol{b}_C \leq 0$ because $\boldsymbol{b}_C$ is facet defining.

(2) The vector $\boldsymbol{b}_{\tilde{C}}$ defines a face of $\tilde{C}$, as one can directly see from Eq. (5.9). With $K = \dim(\ker G)$, the dimension of the face is at most $K - 1$, since it is contained in the $K - 1$ dimensional subspace $\{\boldsymbol{x} \,|\, \boldsymbol{x}^T \boldsymbol{b}_{\tilde{C}} = 0\}$.

(3) The vector $\boldsymbol{b}_{\tilde{C}}$ defines a facet of $\tilde{C}$. That is, the dimension of the face is exactly $K-1$.

Let B be the $M \times (D+1)$ matrix that contains all M rays $\boldsymbol{w}_i$ as rows that fulfill $\boldsymbol{w}_i^T \boldsymbol{b}_C = 0$. Since $\boldsymbol{b}_C$ is a facet normal vector, B has rank D. Accordingly, BT is the $M \times K$ matrix that contains all rays $\tilde{\boldsymbol{w}}_i^T$ as rows that fulfill $\tilde{\boldsymbol{w}}_i^T \boldsymbol{b}_{\tilde{C}} = 0$. Showing that $\tilde{\boldsymbol{b}}$ defines a facet is equivalent to showing that $\text{rank}(BT) = K-1$. We now prove the latter by contradiction. Assume there exist two linearly independent vectors $\tilde{\boldsymbol{b}}, \tilde{\boldsymbol{c}}$ that satisfy $BT\tilde{\boldsymbol{b}} = BT\tilde{\boldsymbol{c}} = 0$. Thus, $T\tilde{\boldsymbol{b}}$ and $T\tilde{\boldsymbol{c}}$ lie in the kernel of B. Since $\text{rank}(B) = D$, the kernel is one-dimensional, so we can write $T\tilde{\boldsymbol{c}} = \ell T\tilde{\boldsymbol{b}}$ for some real number ℓ. This implies $T(\tilde{\boldsymbol{c}} - \ell\tilde{\boldsymbol{b}}) = 0$. Because $\tilde{\boldsymbol{b}}$ and $\tilde{\boldsymbol{c}}$ are linearly independent, the kernel of T has at least dimension one, which is impossible because T has full column rank. □

The facets of interest of the polytope P can now be found by finding the facets of $\tilde{C}$ first, calculating potential facets of C via Eq. (5.8), transforming these into potential facets of P and finally checking which of the found inequalities define facets of P. Note that it is computationally simple to check whether a given candidate is a facet, one just needs to compute the dimension of the surface.

5.2 Generalizations of a Bell inequality

In this section we explain generalizations of Bell inequalities and present two generalizations of Svetlichny's inequality. Consider a Bell scenario S and a Bell inequality $\mathbf{b}_1$ that is applicable to the scenario S. Additionally, consider a second scenario S', which is larger than S in the sense that it involves more parties, more measurement settings per party or more outcomes per measurement setting than S. Let $\mathbf{b}_2$ be a Bell inequality for S'. Consider the situation in which the parties perform a Bell test for the Bell inequality $\mathbf{b}_2$. However, in the number of parties involved in the Bell test and the number and kind of measurements they are allowed to perform, they restrict themselves to the rules given by scenario S. If in this case the Bell test effectively reduces to a Bell test of Bell inequality $\mathbf{b}_1$, then we call $\mathbf{b}_2$ a generalization of $\mathbf{b}_1$.

The concept is best explained by example. Consider three parties, A, B, C that perform a test of Mermin's inequality

$$+\langle A_1 B_1 C_2\rangle + \langle A_1 B_2 C_1\rangle + \langle A_2 B_1 C_1\rangle - \langle A_2 B_2 C_2\rangle \leq 2. \tag{5.10}$$

However, they restrict themselves to the resources of the CHSH scenario. This means that one of the parties is no longer allowed to contribute to the Bell test in a meaningful way. Let this party be C. If C reports the measurement outcome $+1$ in every round, then evaluating Mermins inequality effectively means evaluating the inequality

$$+\langle A_1 B_1\rangle + \langle A_1 B_2\rangle + \langle A_2 B_1\rangle - \langle A_2 B_2\rangle \leq 2, \tag{5.11}$$

which is the well-known CHSH inequality. Therefore, we call Mermin's inequality a generalization of the CHSH inequality. To be precise, we call an inequality $\mathbf{b}_2$ a generalization of an inequality $\mathbf{b}_1$ to more parties, if one can make a choice of trivial measurements for the additional parties, such that $\mathbf{b}_2$ reduces to $\mathbf{b}_1$. By trivial measurement, we mean a measurement that yields one measurement result deterministically. This has as a consequence, that each $(n+1)$-party correlation can be computed from one n-party correlation, such as

$$\langle A_1 B_2 C_1 \rangle = \langle A_1 B_2 \rangle, \tag{5.12}$$

if C_1 is set to always yield outcome $+1$.

We now turn to an example that illustrates the generalization to a scenario with more settings. Consider the $(3,3;2,2)$ scenario, that is the bipartite scenario, in which every party has three measurement settings and every setting yields outcomes ± 1. Besides the CHSH inequality, there is only one non-trivial facet-defining inequality for this scenario, the I3322 inequality [16, 76, 75]. This inequality reads

$$\begin{aligned}\langle A_1 B_3 \rangle + \langle A_2 B_3 \rangle + \langle A_3 B_1 \rangle + \langle A_3 B_2 \rangle + \langle A_1 \rangle - \langle A_2 \rangle + \langle B_1 \rangle - \langle B_2 \rangle \\ - \langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \leq 4.\end{aligned} \tag{5.13}$$

As was noted by Collins et al. [76], this inequality is strictly stronger than the CHSH inequality with regard to its ability to detect non-locality: If the measurements A_3 and B_1 are chosen trivially, so they always yield the result $+1$, then the I3322 inequality reduces to a variant of the CHSH inequality. Therefore the I3322 inequality is a generalization of the CHSH inequality. This example also illustrates an important quality of generalizing Bell inequalities: A generalization of a Bell inequality always performs at least as well as the Bell inequality itself in detecting nonlocality for a given quantum state.

In the example of the I3322 inequality, A and B choose one of their measurements trivial to comply with the CHSH scenario. In general, this is however not the only way to achieve this. Alternatively, the parties may have set two of their measurements equal up to a permutation of outcome labels. When looking for generalizations of a Bell inequality to a scenario with more settings, one must therefore take this possibility into account.

Note that the generalization of a Bell inequality to more parties is a special case of the generalization to more settings: One can first add a party and give it only a trivial measurement. This scenario is essentially still the original scenario, as the new party does not contribute to the experiment in a meaningful way. In a second step, one can then equip the new party with more settings. The parties can then comply with the original scenario, if the new party performs her measurement but always reports the outcome of her trivial measurement.

5.3 Formal definition of a generalization of a Bell inequality

After introducing the idea of a generalization of a Bell inequality in the previous section, we now formalize the notion. Recall from Chapter 4 that a measurement m is a triple consisting of the party $k(m)$ it acts on, a label $i(m)$ called input, and a set $O(m)$ that contains all the possible outcomes of m. We will also make use of the definitions of a scenario, a joint measurement and a record, all of which we introduced in Chapter 4.

Definition 5.1 (Inflation of a scenario)**.** Let S, S' be scenarios and $I = (S, S', E)$ a bipartite graph with edges $e \in E$ such that every edge $e = (m, m')$ with $m \in S, m' \in S'$, is associated with an bijective map $e : O(m) \to O(m')/\sim$ for some equivalence relation $\sim$ on $O(m')$. If for every such edge it holds that $k(m) = k(m')$ and for every $m \in S, m' \in S'$ it holds that $\deg(m) \geq 1$, and $\deg(m') = 1$, then we call I an inflation of S to S'.

The construction of an inflation is illustrated in Figure 5.5 and 5.6.

Definition 5.2 (Related record)**.** Let S, S' be scenarios and $I = (S, S', E)$ an inflation of S to S'. Let $c' \in C(S')$ be a record for scenario S' with $c' = ((o'_j)_{j\in J'} \mid J')$. We define the joint measurement $J = \{m \in S \mid (m, m') \in E \text{ for some } m' \in J'\}$ and the sequence $(o_m)_{m\in J}$ with $o_m = (m, m')^{-1}(o'_{m'})$, where $o'_{m'}$ represents the equivalence class it belongs to. Then we call the record $c = ((o_m)_{m\in J} \mid J)$ the related record of c'. We denote the related record of any $c' \in C(S')$ as $\varrho(c')$.

Definition 5.3 (Extension map)**.** Let S, S' be scenarios and $I = (S, S', E)$ an inflation of S to S'. Further, let $\mathcal{N}(S)$ and $\mathcal{N}(S')$ be the behavior spaces of S and S', respectively. We call the map $\mathcal{E} : \mathcal{N}(S) \to \mathcal{N}(S')$, $b \mapsto b'$, such that $b' \cdot c' = b \cdot \varrho(c')$ for all $c' \in C(S')$ the extension map induced by I.

Definition 5.4 (Extended behavior)**.** Let S, S' be scenarios and $I = (S, S', E)$ an inflation of S to S' and $\mathcal{E}$ the extension map induced by I. Further, let $b \in \mathcal{B} \subset \mathcal{N}(S)$ a behavior for scenario S. Then we call $\mathcal{E}(b)$ the extended behavior of b.

Definition 5.5 (Generalization of a Bell inequality)**.** Let S, S' be scenarios and

$$a'^T b' \geq a'_0 \tag{5.14}$$

an inequality for $b' \in \mathcal{B}' \subset \mathcal{N}(S')$, where $\mathcal{B}'$ is the set of behaviors for S'. Further, let

$$a^T b \geq a_0 \tag{5.15}$$

be an inequality for $b \in \mathcal{B} \subset \mathcal{N}(S)$, where $\mathcal{B}$ is the set of behaviors for S. We call Eq. (5.14) a generalization of Eq. (5.15) iff there exists an inflation I of S to S', such that

$$a'^T \mathcal{E}(b) \geq a'_0 \tag{5.16}$$

is equivalent to Eq. (5.15), where $\mathcal{E}$ is the extension map induced by I.

Note that the notion of a generalization of a Bell inequality is different from a lifting of a Bell inequality. The latter was defined in Ref. [94] in the context of cut polytopes and described in Ref. [67] specifically for Bell inequalities. We make the difference between generalizations and liftings apparent by stating a precise definition of a lifting using our terminology.

Definition 5.6 (Lifting of a Bell inequality)**.** Let S, S' be scenarios and Further, let

$$a^T b \geq a_0 \tag{5.17}$$

be an inequality for $b \in \mathcal{B} \subset \mathcal{N}(S)$, where $\mathcal{B}$ is the set of behaviors for S. Further, let $I = (S, S', E)$ be an inflation of S to S' and $\mathcal{E}$ the extension map induced by I. Let $\mathcal{P} : \mathcal{N}(S') \to \mathcal{N}(S)$ be the affine map such that $\mathcal{P}(\mathcal{B}') = \mathcal{B}$, where $\mathcal{B}'$ is the set of behaviors for scenario S' and $\mathcal{E} \circ P$ is a projection. Then the inequality

$$a^T \mathcal{P}(b') \geq a_0, \tag{5.18}$$

where $b' \in \mathcal{B}'$ is a lifting of Eq. (5.17) to the scenario S'.

While the notions of a generalization and a lifting may seem similar at the first glance, the approaches are in fact contrary. The idea of a lifting is to utilize a known Bell inequality for the simple scenario S for behaviors stemming from an experiment with a more complex scenario S'. This is achieved by mapping the behaviors of the more complex scenario to behaviors of the simpler scenario. This process in general comes with a loss of information, since the original behavior $b' \in \mathcal{B}'$ can in general not be recovered from $b = \mathcal{P}(b')$.

The approach is the opposite when generalizing a Bell inequality. The generalization Bell inequality is a Bell inequality for the more complex scenario S' that can directly test the locality of a behavior $b' \in \mathcal{B}'$.

However, if the behavior is projected into the subspace that does not utilize the advantage that comes with the larger scenario S', then the generalization is just a lifting. Writing the generalized Bell inequality Eq. (5.15) for a projected behavior yields

$$a'^T \mathcal{E} \circ \mathcal{P}(b') \geq a'_0. \tag{5.19}$$

Using matrix notation, this can be rewritten as

$$(\mathcal{E}^T a')^T \mathcal{P} b' \geq a'_0. \tag{5.20}$$

This inequality is by definition a lifting of 5.15.

Note that the condition for a Bell inequality to be a generalization of another Bell inequality can be stated as a list of affine constraints. Given a behavior b, such that $a^T b = a_0$, it must hold for any generalized inequality $a'^T b' \geq a'_0$ that $a'^T \mathcal{E} b = a'_0$. In this way, every saturating behavior of the inequality $a^T b \geq a_0$ gives rise to one affine equality condition that must be satisfied by the vector a' of the generalized Bell inequality. In the following chapter, we give a detailed description how the cone-projection technique works if applied to the problem of finding generalizations of the I3322 inequality to three parties.

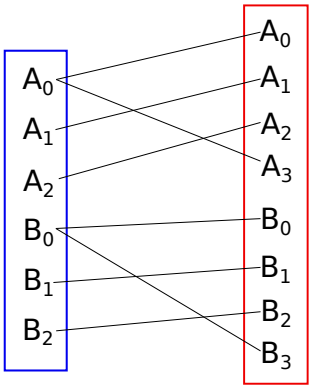


Figure 5.5: Inflation of the CHSH scenario S to the 3322 scenario S'. – On the left, the CHSH scenario is depicted as a blue box that contains measurements A_0, A_1, A_2 for party A and B_0, B_1, B_2 for B. The measurements A_0, B_0 always yield outcome $+1$. The CHSH scenario is inflated to the $(3, 3; 2, 2)$ scenario, depicted as red box. An edge (m, m') with $m \in S, m' \in S'$ means that every outcome of m' is mapped to an outcome of m. This is illustrated in figure 5.6. The outcomes of A_3 and B_3 are mapped to the outcome of the trivial measurement on their respective party.

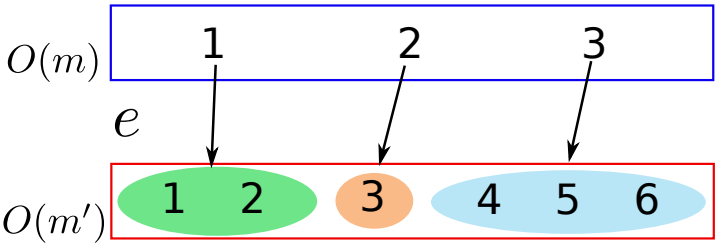


Figure 5.6: The edge $e = (m, m') \in E$ of an an inflation $I = (S, S', E)$ maps outcomes of $m \in S$ to equivalence classes of outcomes of m'. The equivalence classes are represented by the green, orange, and blue ellipses.

6 I3322 generalizations

This chapter closely follows our publication [A]. In fact, the vast majority of this chapter is drawn from the aforementioned paper. The inequalities found in this chapter are further studied in Chapter 7.

We find all symmetric facet-defining generalizations of the I3322 inequality to three parties using the cone-projection technique. For this special case, the conditions for a generalization of a I3322 inequality can be stated as follows. First, since all measurements are dichotomic, we write all Bell inequalities in terms of expectation values of observables. Observables that refer to measurements of party A (B, C) are denoted as A_i (B_j, C_k) and we define A_0, B_0, C_0 to be trivial measurements on the respective parties that always yield a measurement result $+1$. This conveniently allows to treat marginal terms such as $\langle A_1 \rangle = \langle A_1 B_0 C_0 \rangle$ and constant terms such as $1 = \langle A_0 B_0 C_0 \rangle$ on the same footing. In this notation, any Bell inequality can be written as $S = \sum_{i,j,k} b_{ijk} \langle A_i B_j C_k \rangle \geq 0$.

A three-partite Bell inequality is a generalization of the I3322 inequality if there is an assignment $C_k \to \xi_k \in \{\pm 1\}$ for the observables on the third party, such that the remaining inequality

$$\sum_{i,j,k} b_{ijk} \xi_k \langle A_i B_j \rangle \geq 0 \tag{6.1}$$

is the I3322 inequality

$$\begin{aligned} &\langle A_1 \rangle - \langle A_2 \rangle + \langle B_1 \rangle - \langle B_2 \rangle - \langle (A_1 - A_2)(B_1 - B_2) \rangle \\ &\quad + \langle (A_1 + A_2) B_3 \rangle + \langle A_3 (B_1 + B_2) \rangle \leq 4. \end{aligned} \tag{6.2}$$

Because we are only considering inequalities that are symmetric under permutation of parties, an analogous condition with the same ξ_k holds for any other reduction to two parties. One may ask why we have to consider both possibilities $\xi_k = 1$ and $\xi_k = -1$, as inequalities that arise from the condition Eq. (6.1) for $\xi_k = -1$ are just different versions of the inequalities that can be found if one chooses $\xi_k = 1$, where the outcomes of C have been relabeled. The reason for this is that this relabeling of the outcomes may destroy the party symmetry that we demand.

The critical reader may also ask at this point why we are only considering the special form of I3322 as in Eq. (6.2), and not equivalent forms arising from a relabeling of the observables or a sign flip. First, since we defined generalizations of I3322 to be

symmetric, we only have to take symmetric versions of I3322 into account. Further, as one can easily check, all symmetric versions of I3322 can be transformed into each other just by outcome relabelings on both parties. Consequently, it suffices to only consider one symmetric version of I3322.

We also note that the condition that the inequality Eq. (6.1) is the I3322 inequality can be reformulated as follows: An inequality like $\sum_{i,j,k} b_{ijk}\xi_k\langle A_iB_j\rangle \geq 0$ is the I3322 inequality, if and only if equality holds for the set of vertices of the local polytope for which also equality holds in I3322. These bipartite behaviors may be lifted to three-partite behaviors by adding the variable ξ_k, then any generalization of the I3322 inequality has to be saturated by these lifted vertices.

We can now find all generalizations of I3322 in four steps: First, we find all vertices of the local polytope of two parties which saturate I3322. Second, we choose deterministic outcomes ξ_k on C and determine the corresponding vertices in the local polytope of three parties. In the third step, we compose the matrix G for the condition $G\boldsymbol{b}_C = 0$. This matrix contains each symmetry condition and each vertex from step 2 as a row. Finally, we employ the cone-projection technique to find all facet defining inequalities that meet the criteria. We then repeat steps two to four until all possible choices for deterministic outcomes ξ_k are exhausted. In our case of three dichotomic measurements, there are eight possibilities. In this way, we find all symmetric, facet-defining generalizations of I3322, 3050 inequalities in total. We now describe the implementation in detail. The complete list of inequalities is given in the supplemental material of our publication [A].

6.1 Description of the numerical procedure

In order to find all generalizations of the I3322 inequality using our algorithm, we first need to determine its inputs, namely, the rays $\boldsymbol{w}_i$ of the cone C and the matrix G that captures the linear conditions on normal vectors of the facets we aim to find. These conditions are established through the equation $G\boldsymbol{b}_C = 0$. In the following, b_{ijk} denote the coefficients of the vector $\boldsymbol{b}_C$. We now find the rays of C by first finding all 512 local deterministic behaviours $\boldsymbol{v}_i$ of the scenario with three parties, three measurements per party and two outcomes per measurement. Then, we compute the rays of C as

$$\boldsymbol{w}_i = \begin{pmatrix} 1 \\ \boldsymbol{v}_i \end{pmatrix}. \tag{6.3}$$

By doing this, we include the trivial correlation $\langle A_0B_0C_0\rangle = 1$ as first coordinate. Now that the rays of C are found, we construct G. The first constraint G is supposed to capture is the symmetry of the Bell inequality under party permutations. Concretely,

the coefficients of the Bell inequality $S = \sum_{i,j,k} b_{ijk} \langle A_i B_j C_k \rangle$ have to fulfill

$$b_{ijk} - b_{jki} = 0, \tag{6.4}$$
$$b_{ijk} - b_{kij} = 0, \tag{6.5}$$
$$b_{ijk} - b_{ikj} = 0, \tag{6.6}$$
$$b_{ijk} - b_{kji} = 0, \tag{6.7}$$
$$b_{ijk} - b_{jik} = 0, \tag{6.8}$$

and

$$b_{iij} - b_{iji} = 0, \tag{6.9}$$
$$b_{iij} - b_{jii} = 0, \tag{6.10}$$
$$b_{jji} - b_{jij} = 0, \tag{6.11}$$
$$b_{jji} - b_{ijj} = 0, \tag{6.12}$$

with $i < j < k$. In our scenario, where the indices take values from 0 to 3 we get Eq. (6.4)-(6.8) for four different values of (i, j, k) and Eqs. (6.9)-(6.12) for six different values of (i, j). This gives us 44 equations in total. Each of them will be stated as $\boldsymbol{g}_i^T \boldsymbol{b}_C = 0$ and the vectors $\boldsymbol{g}_i$ then form the rows of the matrix G. Since $\boldsymbol{b}_C$ is a vector in dimension 64, G has 64 columns.

Next, we want to ensure that the Bell inequalities we are about to find are not only symmetric, but also generalizations of I3322. As explained in the main text, there must exist deterministic outcomes $\xi_k = \pm 1$ for the measurements on C, such that the coefficients of the I3322 inequality are obtained via $I_{ij}^{3322} = \sum_k b_{ijk} \xi_k$. So, if a bipartite behaviour $\langle A_i B_j \rangle$ saturates I3322, the extended behaviour $r_{ijk} = \langle A_i B_j \rangle \xi_k$ saturates its generalization. Conversely, if a Bell inequality is saturated by the extended behaviour r_{ijk}, then $\langle A_i B_j \rangle$ also saturates I3322. Since a facet is defined by all the vertices it contains, a symmetric and facet defining Bell inequality is a generalization of I3322 if and only if for all behaviours that saturate I3322 there is local deterministic assignment $\boldsymbol{\xi}$, such that the extended behaviours saturate this Bell inequality.

Since there are three measurements per party and we have the choice between two outcomes, we have to take eight possible deterministic assignments into account. For each of them, we find the generalized inequalities that can be reduced to I3322 by performing the assignment. Hence, we have to run our algorithm once for each local deterministic assignment of one party, totalling in eight runs in our case. In each run, we have to complete the matrix G according to the chosen assignment, only the first 44 rows that implement the symmetry conditions stay the same. For each run of the algorithm, we choose one local deterministic assignment $\boldsymbol{\xi}$ and obtain one extended behaviour $\boldsymbol{r}$ for each behaviour that saturates the I3322 inequality. The extended behaviour then has to saturate a potential generalization of I3322. This gives us the con-

dition

$$\boldsymbol{r}^T \boldsymbol{b}_C = 0. \tag{6.13}$$

There are 20 local deterministic behaviours that saturate I3322, so we get 20 conditions of type Eq. (6.13) and the extended behaviours $\boldsymbol{r}$ make up the next 20 rows of G. Therefore, G is a 64×64 matrix. However, it only has rank 53 because some of the extended behaviours are already related through the symmetry we implemented earlier.

The kernel of G can now be found using for example the sympy package in python, which returns basis vectors with integer coordinates (since G has integer entries). In our case, the kernel has dimension 11. For each of the 512 rays of C, we now obtain a ray of $\tilde{C}$ in the way described in Theorem 1. Note, that the map of a ray of C to a ray that lies in $\ker(G)$ is not necessarily a projection as the basis vectors $\boldsymbol{t}_i$ are not necessarily normalized. In fact, we want to avoid normalization to preserve the property that all the objects we deal with only have integer entries. The mapping of rays of C to rays of $\tilde{C}$ is not injective. In fact, only 88 rays remain that span the cone $\tilde{C}$. The facets of $\tilde{C}$ can be found within seconds using standard polytope software such as cdd [27]. From these facets, we keep those that correspond to facets of C.

Finally, we compose a list of all valid facets from the eight runs of the algorithm and remove Bell inequalities that are equivalent to other Bell inequalities in the list up to relabeling of the measurements or outcomes. In this way, all 3050 three-party generalizations of I3322 are found.

6.2 Properties of the generalizations of I3322

Let us now examine the three simplest ones among the generalized I3322 inequalities in some detail. For convenience, we introduce a short-hand notation for symmetric Bell inequalities. We define symmetric correlations as $(ijk) = \sum_{\pi \in \Pi} A_{\pi(i)} B_{\pi(j)} C_{\pi(k)}$, where Π denotes the set of all permutations of the indices ijk that give different terms. Note that in this notation $(112) = A_1 B_1 C_2 + A_1 B_2 C_1 + A_2 B_1 C_1$, so permutations leading to the same term are not counted multiple times. Further, as noted before, settings labeled with index zero refer to trivial measurements that always yield the result 1. Using this notation, the I3322 inequality can be written as

$$(01) - (02) - (11) - (22) + (12) + (13) + (23) \leq 4. \tag{6.14}$$

The three generalizations of I3322 that involve the least number of symmetric correlations are given by:

$$\begin{aligned} F_1 &= 8 + (110) - (210) + (211) + (220) \\ &\quad + (222) - 2(331) - 2(332) \geq 0, \end{aligned} \tag{6.15}$$

$$\begin{aligned} F_2 &= 9 + (110) + 2(220) - 2(221) - (300) \\ &\quad - (310) + (311) - 2(322) \geq 0, \end{aligned} \tag{6.16}$$

$$\begin{aligned} F_3 &= 9 - (210) + (211) + (220) + 3(222) \\ &\quad - (300) - (310) + (311) - 2(322) \geq 0. \end{aligned} \tag{6.17}$$

Let us start our analysis with the possible violations in quantum mechanics. It has recently been shown that all inequalities that exclusively utilize full correlations are violated in quantum mechanics [95], however, our inequalities also contain marginal terms.

A way to obtain bounds on possible quantum values of Bell inequalities is given by the hierarchy of Navascués, Pironio and Acín (NPA) [37, 42, 96]. For the inequality F_1, this shows that in quantum mechanics $F_1 \geq -8$ holds. The minimal value $F_1 = -8$ can indeed be reached, namely by a three-qubit GHZ state $|GHZ_3\rangle = (|000\rangle + |111\rangle)/\sqrt{2}$ and measurements $A_1 = A_2 = \sigma_x$, $A_3 = \sigma_y$, $B_1 = B_2 = -\sigma_y$, $B_3 = \sigma_x$ and $C_i = B_i$. In fact, F_1 reduces to the Mermin inequality if the first two measurement settings on each party are chosen equal.

Concerning F_2, a numerical optimization suggests that one optimal choice of settings is given by $A_1 = -A_3 = \sigma_z$, $A_2 = -\sigma_x$, $B_1 = -B_3 = -\sigma_z$, $B_2 = -\sigma_x$, and $C_i = B_i$. This leads to a quantum mechanical violation of $F_2 = 4(1 - \sqrt{7}) \approx -6.58301$ for the three-qubit state

$$|\psi_2\rangle = a\,|W_3\rangle + b\,|111\rangle, \tag{6.18}$$

with $|W_3\rangle = (|001\rangle + |010\rangle + |100\rangle)/\sqrt{3}$ being the three-qubit W state and $a = \sqrt{19 + 2\sqrt{7}}/\sqrt{74} \approx 0.57294$ and $b = \sqrt{55 - 2\sqrt{7}}/\sqrt{74} \approx 0.81960$. The violation attained by this state coincides up to numerical precision with the lower bound on F_2 for quantum states from the NPA hierarchy. It is interesting that F_2 is maximally violated by a state that does not belong to the frequently studied three-qubit states (such as the states considered in [97, 98, 99]). In this way, the Bell inequality F_2 may open an avenue for new methods of self-testing quantum states [11].

For the inequality F_3 the third level of the NPA hierarchy bounds the quantum mechanical values by $F_3 \geq -4.63097$. Within numerical precision, this can be attained using the three-qubit state

$$|\psi_3\rangle = \cos(\varphi)|W_3\rangle + \sin(\varphi)|GHZ_3\rangle \tag{6.19}$$

with $\varphi = 4.0^\circ$. The required measurement settings (for $i = 1, 2, 3$) are $A_i = \cos(\alpha_i)\sigma_z + \sin(\alpha_i)\sigma_x$ and $B_i = C_i = \cos(\alpha_i)\sigma_z - \sin(\alpha_i)\sigma_x$, where $\alpha_1 = 141.6^\circ, \alpha_2 = 22.6^\circ, \alpha_3 =$

101.6°. Again, we find a non-standard three-qubit state leading to the maximal violation of the Bell inequality. The state $|\psi_3\rangle$ is close to a three-qubit W state but the difference is significant, as for the W state one can only reach a violation of $F_3 = -4.59569$.

Now we clarify whether the new three-setting inequalities are indeed relevant, that is, whether they detect the nonlocality of some quantum states, where all two-setting inequalities fail to do so. This question can be answered positively for all *three* inequalities; moreover, all the F_i detect also entanglement that is not detected by the I3322 inequality in the reduced states. To show this, we provide a three-qubit state ϱ_{ABC} with separable two-body marginals that has a symmetric extension for each F_i, such that the respective Bell inequality F_i is violated. A symmetric extension of ϱ_{ABC} is a five-qubit state $H_{ABB'CC'}$ that is symmetric under exchange of the parties B, B' (and C, C') such that $\varrho_{ABC} = \mathrm{Tr}_{B'C'}(H_{ABB'CC'})$. A state that has a such a symmetric extension cannot violate any Bell inequality with with an arbitrary number of settings for Alice and two settings for Bob and Charlie [100], see also in appendix A.1. Note that here the number of outcomes for each setting is unrestricted. Thus this is a stronger statement than proving that the known inequalities for two settings and two outcomes [77, 101, 75] are not violated.

We find the desired state ϱ_{ABC} using a seesaw algorithm that alternates between optimizing the measurement settings (for the violation of the F_i) and the state (under some constraints). If measurement settings for all parties are fixed, finding a state with a symmetric extension that maximally violates a given F_i is a semidefinite program. On the other hand, given a state and measurement settings for two parties, finding the optimal settings for the third party is also a semidefinite program.

6.3 Conclusion

Using the cone-projection technique, we characterized all symmetric generalizations of the I3322 inequality to three particles. It turned out that already the simplest ones of these generalizations have interesting properties, making an experimental implementation of them desirable.

7 Analysing generalizations of Bell inequalities

This chapter closely follows our publication [B]. We reuse both the text and the figures of this publication for the present chapter.

We present a detailed analysis of the properties of three-particle generalizations of the I3322 inequality previously found in our work [A]. Second, we study generalizations of the so-called I4422 inequality [76] to three particles. Third, we find and investigate three-particle Bell inequalities that are generalizations of the CHSH inequality and the I3322 inequality at the same time. Finally, we study generalizations of three-particle Guess-Your-Neighbors-Input (GYNI) inequality [9] to four particles. All of these inequalities are found using the cone-projection technique.

For every inequality we find, we conduct the same numerical analysis. We compute the quantum mechanical violation for qubits, for qutrits and for the second and (if possible) third level of the NPA-hierarchy [37]. For the latter, we used the ncpol2sdpa package by Peter Wittek [96]. For the qubit and qutrit violations, we provide lower bounds using a seesaw algorithm that alternates between optimizing the measurement settings of one of the parties and optimizing the state. Every of these optimization steps is a semidefinite program. Details can be found in Appendix A.

For meaningful comparisons between the inequalities, we compute four quantities for every Bell inequality $\langle \mathcal{B} \rangle \equiv \sum_{ijk \neq 000} -b_{ijk} \langle A_i B_j C_k \rangle \leq b_{000}$ that will guide the following discussion.

The first quantity is the *relative qutrit violation* by which the maximal value of the Bell expression $\mathcal{B}$ achievable with qutrits exceeds the maximal classical value. It is defined as

$$m_Q = \frac{\max_{\text{qutrit}} \langle \mathcal{B} \rangle}{b_{000}} - 1. \tag{7.1}$$

We are interested in this quantity as a signature to identify inequalities whose quantum violation is particularly strong. In this regard, it would be more informative to consider the quantum violation with a higher-dimensional system. However, the computations become more demanding which is why we restrict our quantum mechanical analysis to qubit and qutrit systems.

The second quantity we consider is the *algebraic-classical ratio* that is defined analogously and quantifies by how much the algebraic maximum of the Bell expression

exceeds the classical bound in relation to the classical bound. It is calculated as

$$m_A = \frac{\sum_{i,j,k} |b_{ijk}| - b_{000}}{b_{000}}. \tag{7.2}$$

As a third interesting quantity, we compute the *NPA-qutrit ratio* that quantifies the relative margin between the maximal value of the Bell expression in the third level of the NPA hierarchy and the maximal value achievable with qutrits. For some of the Bell inequalities we found, it was unfeasible to compute the third level of the NPA hierarchy. In this situation we resorted to the value obtained by the second level of the NPA hierarchy. Every time this is the case it is stated explicitly. The NPA-qutrit ratio is calculated as

$$m_N = \frac{\max_{\text{NPA}} \langle \mathcal{B} \rangle}{\max_{\text{qutrit}} \langle \mathcal{B} \rangle} - 1. \tag{7.3}$$

Lastly, we consider the *qutrit-qubit ratio* m_{32} by which the maximal value of the Bell expression $\mathcal{B}$ achieved by qutrits exceeds the corresponding value for qubits. Mathematically, it is defined as

$$m_{32} = \frac{\max_{\text{qutrit}} \langle \mathcal{B} \rangle}{\max_{\text{qubit}} \langle \mathcal{B} \rangle} - 1. \tag{7.4}$$

It is natural to state these relative margins in percent, in which case $m_{32} = 10\%$ means that the maximal value of the Bell expression is 10% larger for qutrit systems than for qubit systems.

In the following, we report those inequalities that exhibit the biggest value for one of the four relative quantities. A big qutrit-qubit ratio m_{32} makes the inequality potentially interesting for experimental discrimination between qubit and qutrit states (i.e. device-independent dimension witnessing), while a large NPA-qutrit ratio m_N suggests that the maximal quantum mechanical violation of the inequality may not be achievable with qutrit states. A large relative qutrit violation m_Q hints at a particularly strong violation of the inequality in quantum mechanics, while a large algebraic-classical ratio m_A shows that observers that play by the rules of classical physics are especially limited in obtaining a large expectation value for the Bell expression in comparison with hypothetical observers who do not suffer any physical limitations.

Finally, while we defined the quantities in this subsection for three parties, all the definitions can easily be extended to four-party Bell inequalities.

7.1 Investigation of I3322 generalizations

For simplicity, we adopt a notation that takes into account that the inequalities are symmetric under arbitrary permutations of parties. We write $(ijk) := \langle A_i B_j C_k \rangle +$ permutations that yield different terms. For example $(011) = \langle A_1 B_1 \rangle + \langle B_1 C_1 \rangle + \langle A_1 C_1 \rangle$.

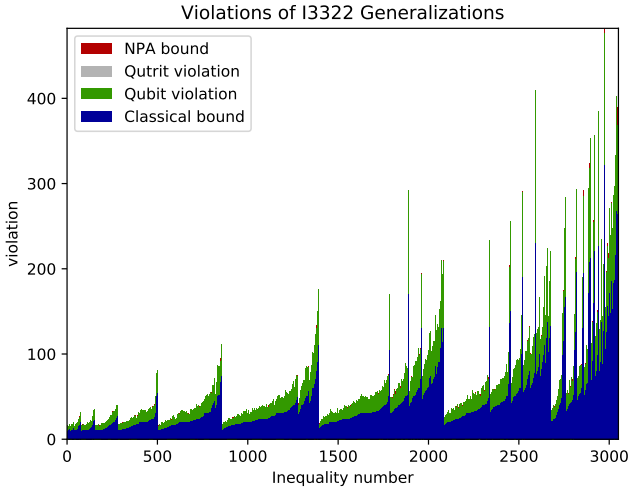


Figure 7.1: The classical bounds of the inequalities $\sum_{ijk\neq 000} -b_{ijk}\langle A_i B_j C_k\rangle \leq b_{000}$ and their violation by qubits, qutrits and the NPA hierarchy of third level. The Bell inequalities are sorted by the number of terms they incorporate. The light gray and red [dark gray] bars that represent the qutrit violations and NPA bounds can barely be seen because they are mostly hidden behind the bars representing the qubit violations.

Let us discuss these inequalities. To start with a rough overview, the classical bounds b_{000} of the Bell inequalities range from 8 to 321. Fig. 7.1 depicts the classical bound for every Bell inequality and how much this bound is exceeded by the quantum states we found as well as the upper bound thereof provided by the NPA hierarchy. We find that all of the inequalities can be violated for qubit states. The inequalities are ordered by simplicity, measured as length of their expression in the above stated notation.

Now let us consider some specific Bell inequalities. The largest qutrit-classical ratio m_Q is achieved for Bell inequality number 1 as listed in the Supplementary Material of our paper [B], which reads

$$\begin{aligned}&(110)+(210)-(211)-(220)\\&-(222)+2(331)+2(332)\leq 8.\end{aligned} \tag{7.5}$$

For this inequality, the maximal value attainable with qubits coincides with the upper bound of 16 given by the NPA-hierarchy. It is worth noting that this inequality reduces to Mermin's inequality if the first two settings of each party are chosen equal. The inequality then is maximally violated for the Greenberger-Horne-Zeilinger state $|GHZ\rangle = (|000\rangle + |111\rangle)/\sqrt{2}$ and the optimal measurement settings for Mermin's inequality. In fact, this has been already discussed in our paper [A] and more details on

states and settings can be found there.

We observe the biggest qutrit-qubit ratio m_{32} for Bell inequality number 400, which reads

$$\begin{aligned} &+5(100)+(110)-5(111)+3(200)-3(210)+3(211)\\ &-2(220)+2(221)-2(300)+(310)+(311)\\ &+2(322)-(330)\\ &-(333)\leq 18 \end{aligned} \tag{7.6}$$

For qubits, QM can lead to a violation of 20.928, whereas for qutrits a maximal violation of 21.157 can be obtained, which is less than the value 21.238 obtained using the third level of the NPA-hierarchy.

We find the biggest NPA-qutrit ratio m_N for Bell inequality number 1507, which reads

$$\begin{aligned} &8(100)-4(110)+3(111)+4(200)-3(210)\\ &+2(211)-(220)+2(221)-2(222)-(300)-(310)\\ &+3(311)+2(320)-(321)+(322)-(331)-(333)\leq 21. \end{aligned} \tag{7.7}$$

The third level of the NPA hierarchy yields the value 24.079, which is a significant improvement over the value 26.299 provided by the second level of the NPA hierarchy. However, with qutrits we could only achieve a violation of 23.603. With qubits, the violation was even lower at 23.249.

The largest algebraic-classical ratio m_A between the classical bound 12 and the algebraic bound 86 is found for Bell inequality number 532. It reads

$$\begin{aligned} &3(100)-(110)-(111)-4(200)+2(210)-(220)\\ &+(221)+(222)-3(300)+2(310)-(320)+(321)\\ &-(322)+(331)+(332)\leq 12 \end{aligned} \tag{7.8}$$

All four ratios are listed for the inequalities above in Tab. 7.2.

It is also worth mentioning that for some of the Bell inequalities, we have numerical evidence that suggests the maximal qubit violation cannot be achieved using projective measurements. For 93 inequalities the maximal qubit violation we could achieve using projective measurements falls short by more than 0.001 compared to the value we were able to obtain with POVMs. The largest percentual discrepancy occurs in Bell inequality 898, which reads

$$\begin{aligned} &(110)-3(111)-4(200)-(210)+(211)\\ &-(220)-(221)+(222)+3(300)+(311)\\ &+2(320)+(321)+(322)-(330)+2(331)\\ &-(333)\leq 15. \end{aligned} \tag{7.9}$$

Eq.	Number	m_Q	m_{32}	m_N	m_A
(7.5)	1	100.0	0.0	-0.0	250.0
(7.6)	400	17.54	1.09	0.38	433.33
(7.7)	1507	12.4	0.61	2.01	514.29
(7.8)	532	18.69	0.0	0.35	616.67

Table 7.1: This table shows the relative qutrit violation m_Q, the qutrit-qubit ratio m_{32}, the NPA-qutrit ratio m_N, and the algebraic-classical ratio m_A as well as those generalizations of the I3322 inequality for which one of these margins is the largest. The values of the margins is stated in percent. For example, for inequality number 1, the maximal value of the Bell expression is 16 for qutrit systems, exceeding the classical bound of 8 by 8, yielding $m_Q = 100\%$.

Eq.	Number	m_Q	m_{32}	m_N	m_A
(7.5)	1	100.0	0.0	-0.0	250.0
(7.6)	400	17.54	1.09	0.38	433.33
(7.7)	1507	12.4	0.61	2.01	514.29
(7.8)	532	18.69	0.0	0.35	616.67

Table 7.2: This table shows the relative qutrit violation m_Q, the qutrit-qubit ratio m_{32}, the NPA-qutrit ratio m_N, and the algebraic-classical ratio m_A as well as those generalizations of the I3322 inequality for which one of these margins is the largest. The values of the margins is stated in percent. For example, for inequality number 1, the maximal value of the Bell expression is 16 for qutrit systems, exceeding the classical bound of 8 by 8, yielding $m_Q = 100\%$.

For this inequality, we obtain a POVM violation of 19.32. This is almost 11% larger than the maximal PVM violation we were able to obtain at 17.43.

7.2 Three-party generalizations of the I4422 inequality

When Collins and Gisin discovered the I3322 inequality, they noticed that this Bell inequality is a member of a family of Bell inequalities which they called I$mm22$, where $m \geq 2$. This family contains the CHSH inequality, or I2222, as a special case. Inequalities of this family with more measurement settings generalize inequalities with fewer measurement settings in the following sense: If one substitutes $A_1 = 1$ and $B_3 = 1$ in the I3322 inequality, the inequality effectively reduces to a CHSH inequality. Essentially the same procedure also works for the I4422 inequality [76]. In the notation

introduced above, it reads

$$\begin{aligned} &2\,(10) + 2\,(20) + (30) - (11) - (21) \\ &-(31) - (41) - (22) - (32) + (42) + (33) \leq 7. \end{aligned} \tag{7.10}$$

If we substitute $A_3 = B_3 = 1$, followed by $A_2 \to -A_2$, $B_2 \to -B_2$ and $A_4 \to -A_3$, $B_4 \to -B_3$, we arrive at the I3322 inequality.

In contrast to the I3322 inequality we observe a difference between the maximal qubit-violation and the maximal qutrit-violation. For qubits, the largest value we could achieve for the Bell expression is 8, whereas for qutrits it is 8.15, which matches the upper bound obtained using the third level of the NPA hierarchy up to numerical precision. Obtaining a value of 8 is compatible with the above mentioned contraints that will reduce the inequality to the I3322 inequality. This means that the additional measurement setting of I4422 seems to lead to no advantage for achieving a large violation of the inequality, as long as only qubits are concerned. Perhaps surprisingly, the generalizations of I4422 we are going to present do not seem to inherit the feature of being able to discriminate between qubit and qutrit systems by means of their maximal violations.

Before presenting some generalizations of I4422, we should clarify that we did not tackle this problem in full generality. When looking for generalizations of the I3322 inequality to more parties, the only symmetry we demanded was invariance under permutations of parties. Finding generalizations of the I4422 inequality is, however, computationally more involved, which is why we demand a second symmetry in order to make the problem tractable. When choosing a symmetry, one needs to be careful to not impose too strong constraints. For example, a three-party inequality that is invariant under party permutations can never be the generalization of a two party inequality that is not symmetric with respect to the parties. We therefore have to take the symmetries of the inequality we seek to generalize, in our case I4422, into consideration.

Besides being invariant under exchange of the parties A and B, I4422 has a second symmetry, namely, if we swap Alice's first and second setting while simultaneously relabeling the outcomes of Bob's fourth measurement, then this leaves the inequality invariant. For convenience, we write the symmetries in the following symbolic way:

1. $A \leftrightarrow B$
2. $A_1 \leftrightarrow A_2, \quad B_4 \to -B_4$.

For three-party generalizations of I4422, the following two symmetries seem natural. The first one is

1. $A \leftrightarrow B$

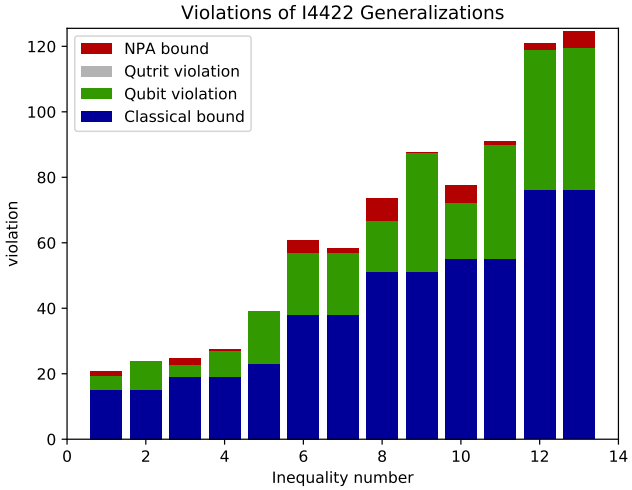


Figure 7.2: The classical bounds of the I4422 generalizations and their violation by qubits, qutrits and the NPA hierarchy (second level).

2. $A \leftrightarrow C$

3. $A_1 \leftrightarrow A_2$, $B_4 \rightarrow -B_4$, $C_4 \rightarrow -C_4$.

However, there exist no non-trivial facet inequalities that generalize I4422 and at the same time meet the above conditions.

We therefore instead look for I4422 generalizations with the following symmetry:

1. $A \leftrightarrow B$

2. $A \leftrightarrow C$

3. $A_1 \leftrightarrow A_2$, $B_1 \leftrightarrow B_2$, $C_4 \rightarrow -C_4$.

The first two symmetries together establish that the generalizations we are going to find are symmetric under arbitrary permutations of the three parties. The third symmetry resembles the second symmetry of I4422. In total, there are 13 classes of Bell inequalities that generalize I4422 and also exhibit the above symmetries. They are presented in Appendix B.

Just like some of the I3322 generalizations, also three I4422 generalizations appear to have a maximal qubit violation that can only be achieved using POVMs. These are the inequalities Eqs. (A.5 - A.7) for which the qubit violation we found using POVMs exceeded the bound we found using PVMs by 2.3%, 2.1%, and 9.1%, respectively.

Fig. 7.2 shows the classical bounds and the violations we found for the 13 inequalities. The more complex scenario of tripartite generalizations of I4422 also comes with

Eq.	Number	m_Q	m_{32}	m_N	m_A
(A.3)	1	28.87	0.0	6.81	360.0
(A.4)	2	59.05	0.0	0.24	413.33
(A.5)	3	18.39	0.0	9.49	315.79
(A.6)	4	42.11	0.0	2.18	357.89
(A.7)	5	69.57	0.0	0.0	434.78
(A.8)	6	49.83	0.0	6.82	400.0
(A.9)	7	49.56	0.0	2.75	378.95
(A.10)	8	30.39	0.0	10.85	341.18
(A.11)	9	71.33	0.0	0.33	403.92
(A.12)	10	30.87	0.0	7.56	327.27
(A.13)	11	63.34	0.0	1.39	385.45
(A.14)	12	56.21	0.0	1.76	373.68
(A.15)	13	57.12	0.0	4.26	405.26

Table 7.3: This table shows the relative qutrit violation m_Q, the qutrit-qubit ratio m_{32}, the NPA-qutrit ratio m_N, and the algebraic-classical ratio m_A for every generalization of I4422 that satisfies the symmetry described in the text. The values of the ratios are expressed in percent. Numbers refers to the number under which the corresponding inequality is listed in the Appendix A.3.

the drawback that the third level of the NPA hierarchy is already hard to compute. We therefore only compute the second level of the NPA hierarchy. Tab. 7.3 shows the relative margins as defined in Eqs. (7.1 - 7.4). For all of the inequalities the maximal qubit and qutrit violations we found were the same. However, we cannot be certain that we managed to find the maximal qutrit violation, except from the case of inequality number 5, where the quantum violations attained the upper bound given by the NPA hierarchy up to numerical precision.

7.3 Three-party hybrid CHSH-I3322 generalizations

To show that also hybrid scenarios can be studied, we find all 476 three-party facet Bell inequalities with the following properties: Alice and Bob have three measurement settings each and Charlie has only two. Also, there are deterministic assignments to the outcomes of Alice's measurements, such that the inequality effectively reduces to the CHSH inequality

$$\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \leq 2. \tag{7.11}$$

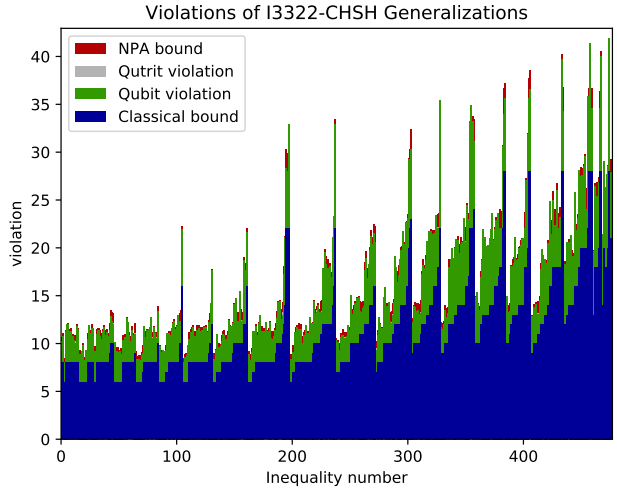


Figure 7.3: The classical bounds of the inequalities $\sum_{ijk\neq 000} -b_{ijk}\langle A_i B_j C_k\rangle \leq b_{000}$ and their violation by qubits, qutrits and the NPA hierarchy of third level. See text for further details.

Further, there are deterministic assignments to the outcomes of Charlie's measurements, such that the inequality reduces to the I3322 inequality. Lastly, we require the inequality to be symmetric under exchange of Alice and Bob. These conditions lead to 476 inequalities, these are given in the Supplementary Material of our paper [B].

As one can see in Fig. 7.3 all of the Bell inequalities are violated in quantum mechanics. The strongest violation by qutrit states is achieved for inequality 47, where with qutrits one can achieve a value of 10.154 while the classical bound is 6. It reads

$$\begin{aligned}
&\langle C_1\rangle + \langle C_2\rangle + \langle B_1\rangle \\
&- \langle B_1C_2\rangle - \langle B_2\rangle + \langle B_2C_1\rangle + \langle A_1\rangle \\
&- \langle A_1C_2\rangle + 2\langle A_1B_1C_2\rangle + \langle A_1B_2\rangle - \langle A_1B_2C_1\rangle \\
&- \langle A_1B_3C_1\rangle - \langle A_1B_3C_2\rangle - \langle A_2\rangle + \langle A_2C_1\rangle \\
&+ \langle A_2B_1\rangle - \langle A_2B_1C_1\rangle - \langle A_2B_2\rangle + 2\langle A_2B_2C_1\rangle \\
&- \langle A_2B_2C_2\rangle + \langle A_2B_3\rangle - \langle A_2B_3C_2\rangle - \langle A_3B_1C_1\rangle \\
&- \langle A_3B_1C_2\rangle + \langle A_3B_2\rangle - \langle A_3B_2C_2\rangle - \langle A_3B_3\rangle \\
&- \langle A_3B_3C_1\rangle \leq 6
\end{aligned} \tag{7.12}$$

We find the largest NPA-qutrit ratio m_N for Bell inequality number 314. It reads

$$\begin{aligned}
&5\langle C_1\rangle + \langle C_2\rangle + 2\langle B_1\rangle - 2\langle B_1C_1\rangle \\
&- 2\langle B_2\rangle + 2\langle B_2C_1\rangle + \langle B_3C_1\rangle + \langle B_3C_2\rangle \\
&+ 2\langle A_1\rangle - 2\langle A_1C_1\rangle + 2\langle A_1B_1C_1\rangle - 2\langle A_1B_1C_2\rangle \\
&+ \langle A_1B_2\rangle - 2\langle A_1B_2C_1\rangle + \langle A_1B_2C_2\rangle \\
&+ \langle A_1B_3\rangle - 2\langle A_1B_3C_1\rangle + \langle A_1B_3C_2\rangle - 2\langle A_2\rangle \\
&+ 2\langle A_2C_1\rangle + \langle A_2B_1\rangle - 2\langle A_2B_1C_1\rangle + \langle A_2B_1C_2\rangle \\
&- 2\langle A_2B_2\rangle + 3\langle A_2B_2C_1\rangle + \langle A_2B_2C_2\rangle + \langle A_2B_3\rangle \\
&- 3\langle A_2B_3C_1\rangle + \langle A_3C_1\rangle + \langle A_3C_2\rangle + \langle A_3B_1\rangle \\
&- 2\langle A_3B_1C_1\rangle + \langle A_3B_1C_2\rangle + \langle A_3B_2\rangle \\
&- 3\langle A_3B_2C_1\rangle - 2\langle A_3B_3C_1\rangle - 2\langle A_3B_3C_2\rangle \leq 12.
\end{aligned} \tag{7.13}$$

We achieve the same violation for qubit and qutrit systems at 16.339. The third level of the NPA hierarchy yields 16.488, improving the upper bound of 17.870 from the second level NPA hierarchy significantly. Because the qubit and qutrit violations are the same, one may conjecture that the upper bound for the quantum mechanical violation provided by the NPA hierarchy is not tight for the third level.

We observe the biggest gap between qubits and qutrits for Bell inequality number 1. It reads

$$\begin{aligned}
&2\langle B_1\rangle - 2\langle B_2\rangle + 2\langle A_1\rangle - \langle A_1B_1\rangle \\
&+ \langle A_1B_1C_1\rangle + \langle A_1B_2\rangle - \langle A_1B_2C_1\rangle + 2\langle A_1B_3C_2\rangle \\
&- 2\langle A_2\rangle + \langle A_2B_1\rangle - \langle A_2B_1C_1\rangle - \langle A_2B_2\rangle \\
&+ \langle A_2B_2C_1\rangle + 2\langle A_2B_3C_2\rangle + 2\langle A_3B_1C_2\rangle + 2\langle A_3B_2C_2\rangle \\
&\leq 8
\end{aligned} \tag{7.14}$$

For this inequality we find a value of 10.000 for qubits, whereas the violation of 10.286 that we find for qutrits coincides with the value of the third level of the NPA hierarchy up to numerical precision.

Lastly, the biggest gap between the classical and the algebraic bound occurs for

inequality 198, which reads

$$\begin{aligned}
&3\langle C_1\rangle + \langle C_2\rangle + 2\langle B_1\rangle \\
&- \langle B_1C_1\rangle - \langle B_1C_2\rangle - \langle B_2\rangle + 2\langle B_2C_1\rangle \\
&+ \langle B_2C_2\rangle + \langle B_3\rangle - \langle B_3C_2\rangle + 2\langle A_1\rangle \\
&- \langle A_1C_1\rangle - \langle A_1C_2\rangle - \langle A_1B_1\rangle + \langle A_1B_1C_1\rangle \\
&+ \langle A_1B_2\rangle - \langle A_1B_2C_1\rangle - \langle A_1B_3C_1\rangle + \langle A_1B_3C_2\rangle \\
&- \langle A_2\rangle + 2\langle A_2C_1\rangle + \langle A_2C_2\rangle + \langle A_2B_1\rangle \\
&- \langle A_2B_1C_1\rangle - \langle A_2B_2\rangle - \langle A_2B_2C_2\rangle + \langle A_2B_3\rangle \\
&- \langle A_2B_3C_1\rangle + \langle A_3\rangle - \langle A_3C_2\rangle - \langle A_3B_1C_1\rangle \\
&+ \langle A_3B_1C_2\rangle + \langle A_3B_2\rangle - \langle A_3B_2C_1\rangle \leq 6.
\end{aligned} \tag{7.15}$$

Compared to the classical bound of 6, the algebraic bound is almost seven times as large at 40. The numerical results are summarized in Tab. 7.4.

Once again, we were able to obtain some qubit violations only when admitting POVMs but not if the measurements were restricted to be projective. The largest discrepancy between the maximal value achieved with projective measurements and the maximal value achieved with POVMs occurs for Bell inequality number 249, where the latter exceeds the former by roughly 8%. This Bell inequality reads

$$\begin{aligned}
&2\langle C_1\rangle + 2\langle C_2\rangle + 2\langle B_1\rangle - \langle B_1C_1\rangle \\
&- \langle B_1C_2\rangle - \langle B_2\rangle + \langle B_2C_1\rangle + \langle B_3\rangle \\
&- \langle B_3C_2\rangle + 2\langle A_1\rangle - \langle A_1C_1\rangle - \langle A_1C_2\rangle \\
&- \langle A_1B_1\rangle + \langle A_1B_1C_1\rangle + \langle A_1B_2\rangle - \langle A_1B_2C_1\rangle \\
&- \langle A_1B_3C_1\rangle + \langle A_1B_3C_2\rangle - \langle A_2\rangle + \langle A_2C_1\rangle \\
&+ \langle A_2B_1\rangle - \langle A_2B_1C_1\rangle + \langle A_2B_2C_1\rangle - \langle A_2B_2C_2\rangle \\
&- \langle A_2B_3C_1\rangle + \langle A_2B_3C_2\rangle + \langle A_3\rangle - \langle A_3C_2\rangle \\
&- \langle A_3B_1C_1\rangle + \langle A_3B_1C_2\rangle - \langle A_3B_2C_1\rangle + \langle A_3B_2C_2\rangle \\
&- \langle A_3B_3\rangle + 2\langle A_3B_3C_1\rangle + 3\langle A_3B_3C_2\rangle \leq 8.
\end{aligned} \tag{7.16}$$

Eq.	Number	m_Q	m_{32}	m_N	m_A
(7.12)	47	69.23	0.0	0.0	400.0
(7.14)	1	28.57	2.86	0.0	200.0
(7.13)	314	36.16	0.0	0.91	433.33
(7.15)	198	39.71	0.0	0.0	566.67

Table 7.4: This table shows the relative qutrit violation m_Q, the qutrit-qubit ratio m_{32}, the NPA-qutrit ratio m_N, and the algebraic-classical ratio m_A as well as those hybrid generalizations of the I3322 and the CHSH inequality for which one of these margins is the largest. The values of the ratios is stated in percent.

7.4 Four-party generalizations of a Guess-Your-Neighbors-Input inequality

Here we present generalizations of the Guess-Your-Neighbors-Input (GYNI) inequality

$$\begin{aligned}&\langle A_1B_1\rangle + \langle A_2B_1\rangle + \langle A_1B_2\rangle + \langle A_2B_2\rangle\\&+ \langle A_1C_1\rangle - \langle A_2C_1\rangle + \langle B_1C_1\rangle + \langle A_1B_1C_1\rangle\\&- \langle B_2C_1\rangle - \langle A_2B_2C_1\rangle + \langle A_1C_2\rangle - \langle A_2C_2\rangle - \langle B_1C_2\rangle\\&+ \langle A_2B_1C_2\rangle + \langle B_2C_2\rangle - \langle A_1B_2C_2\rangle \leq 4. \end{aligned} \tag{7.17}$$

This inequality was found to be a GYNI inequality by Almeida et al. [9] and found by Śliwa [75]. For the properties and interpretation of such an inequality, the reader is referred to section 4.7. The inequality has the following symmetries:

1. $A_1 \leftrightarrow A_2$, $B_1 \leftrightarrow B_2$, $C_1 \to -C_1$, $C_2 \to -C_2$
2. $A_1 \leftrightarrow A_2$, $C_1 \leftrightarrow C_2$, $A_i \to -A_i \quad \forall i$, $B_i \to -B_i \quad \forall i$.

We demand exactly the same symmetries for the generalization of the above inequality. We find 23 four-partite inequalities, the detailed expressions are given in the Supplementary Material of our paper [B]. The first one is not a real four-partite Bell inequality, since it only includes a single measurement setting on the fourth party. In fact, it is of the form $B(1 + D_1) \leq 4(1 + D_1)$, where B is the left-hand-side of Eq. (7.17) after C_1 and C_2 have been swapped and all outcomes of Bob have been relabeled. All the other inequalities are violated in quantum theory. Hence, none of the generalizations shares the characteristic feature of the GYNI inequality. Fig. 7.4 shows a plot with all classical bounds and their violations for all 23 inequalities.

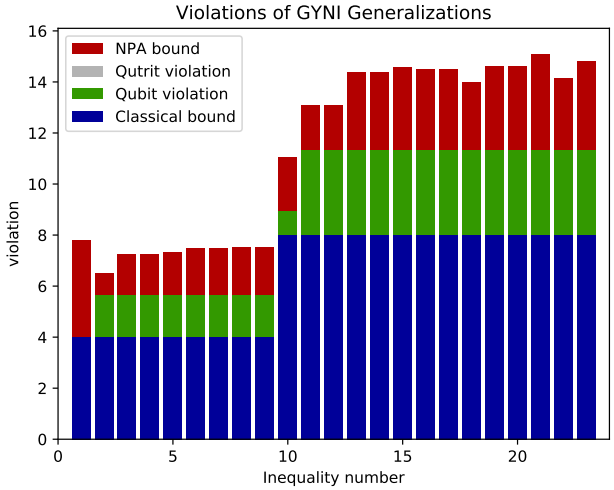


Figure 7.4: The classical bounds of the inequalities $\sum_{ijkl\neq 0000} -b_{ijkl}\langle A_i B_j C_k D_l\rangle \leq b_{0000}$ and their violation by qubits, qutrits and the NPA hierarchy of third level.

7.5 Conclusion

Using the cone-projection technique, we were able to find 3050 classes of generalizations of the I3322 inequality, 476 classes of Bell inequalities that simultaneously generalize the CHSH inequality as well as the I3322 inequality, 13 classes of Bell inequalities that generalize the I4422 inequality and 23 classes of Bell inequalities that generalize the Guess-Your-Neighbors-Input inequality first found in [75]. For all inequalities, we applied an extensive numerical analysis, providing upper bounds to their quantum violations both for qubit and qutrit systems as well as upper bounds using the NPA hierarchy.

For future research, there are several open problems. First, it would be desirable to study some of the Bell inequalities presented here in more detail. For instance, the question arises whether they can be connected to some distributed information processing task. In addition, one may study the viability of an experimental test of these inequalities. Second, it is very interesting to study the cone-projection technique also for other scenarios, such as Bell inequalities that detect genuine multi-partite nonlocality or contextuality inequalities.

8 Bell-like inequalities to distinguish hybrid models

This chapter contains results, text and figures from our manuscript [C].

Bell inequalities have been conceived as tools to test the compatibility of local hidden variable models with quantum mechanical predictions. Famously, they have later been successfully used to show that the results of Bell test experiments can in fact not be reproduced in a local hidden variable model. Given that nonlocal effects occur in the realm of quantum mechanics, one may be interested in how extensive they are. To answer this question, Svetlichny introduced so-called hybrid models. Hybrid models are a class of hidden variable models that give up on all restrictions on the correlations between a subset of parties while maintaining the restriction of local realism with respect to the remaining ones. The most famous example of a hybrid model is the one originally introduced by Svetlichny for three parties: In any round of the Bell experiment, two of the parties may collaborate to establish arbitrary correlations between themselves. However, the correlations shared between these two parties and the third party must respect a local-hidden-variable model. In this case, the correlations between the three parties satisfy the Svetlichny inequality, which reads

$$\begin{aligned} 4 - [112] - [121] - [211] + [222] \\ - [221] - [212] - [122] + [111] \geq 0, \end{aligned} \tag{8.1}$$

where $[abc] := \langle A_a B_b C_c \rangle$ and all measurements A_a, B_b, C_c yield outcomes ± 1. If one allows arbitrary correlations between Alice and Bob, this amounts to treating both together as one party. Rewriting the Svetlichny inequality in this way yields

$$4 - [12] - [21] - [31] + [42] - [41] - [32] - [22] + [11] \geq 0. \tag{8.2}$$

This inequality can be written as the sum of two CHSH inequalities

$$2 - [12] - [21] - [22] + [11] \geq 0, \tag{8.3}$$

$$2 - [31] + [42] - [41] - [32] \geq 0. \tag{8.4}$$

If both of the CHSH inequalities are satisfied, then the Svetlichny inequality is satisfied, too. Hence, Svetlichny's inequality is fulfilled if the correlations shared between the new party AB and C are local. Since the Svetlichny inequality is symmetric under

exchange of parties it is apparent that it is satisfied no matter which of the two parties collaborate. Remarkably, this inequality is violated in quantum mechanics, which means that nonlocality affects more than 2 parties at once. In the same scenario there is only one other inequivalent Bell inequality that solely involves full-body correlation terms, found by Bancal et al. [81].

In this chapter, we find Bell inequalities for hybrid models in the case of four and five parties. There is however one complication: For more than 3 parties, there is more than one hybrid model to consider. For four parties, one example of a hybrid model would be one, where two teams of two parties share local correlations while the correlations shared between parties within a team can be arbitrary. In the following we refer to this hybrid model as a $(2,2)$ model. Another example would be a hybrid model where there is one team of three parties and one party that is on its own. In our terminology, this is a $(3,1)$ model. For four and five parties, we consider all hybrid models and find all optimal, symmetric, 2-setting, 2-outcome, full-body correlation Bell inequalities. Full-body correlation Bell inequalities are those that only involve correlations where every party performs a non-trivial measurement. Requiring symmetry of the Bell inequalities imposes linear contraints, which can be supplied to the cone-projection technique (CPT) to help narrow down the search for Bell inequalities. This simplifies the task and allows us to find all of the specified Bell inequalities.

We then shift our focus back to three-partite nonlocality, where we discuss genuine-multipartite Bell inequalities with three settings that generalize Svetlichny's inequality. This means, that for a particular choices of measurements, these inequalities reduce to Svetlichny's inequality.

8.1 Hybrid models

We already briefly discussed hybrid models in Chapter 4. Here we specify the hybrid models, which are studied in this chapter.

Since all measurements considered in this chapter yield outcomes ± 1, we regard a behavior as a vector in an expectation value behavior space as defined in Section 4.2.2. If we do not impose any no-signaling constraints, the behaviors in this space form the hypercube defined by the conditions

$$-1 \leq \langle A_a B_b C_c \rangle \leq 1. \tag{8.5}$$

We refer to this hypercube as the unconstrained model $M_\square$ (see also Figure 4.2). Hybrid models are models that arise as hybrids between a LHV model and an unconstrained model [84, 85]. For a hybrid model M_H it thus holds that

$$M_{LHV} \subset M_H \subset M_\square. \tag{8.6}$$

Hybrid models can be constructed in the following way: Given an n-party system with subsystems $s \in S$, the first step is to define a partition

$$P = \{c_1, \ldots, c_k\} \tag{8.7}$$

where the cells $c_i \subset S$ are disjoint non-empty subsets of S and

$$\bigcup_{i=1}^{k} c_i = S. \tag{8.8}$$

In a second step, one defines a local hidden-variable model M^P_{LHV} for the coarse grained scenario defined by P. This works in the following way: Every cell c_i of the partition is considered as one system. The measurement settings are all combinations of measurement settings that apply to each subsystem within a cell. However, no restrictions apply to the correlations between subsystems within a cell, since the cell is regarded as one system. In particular, this allows for signaling to take place between the parties within one cell.

As a third step, two partitions are considered equivalent, if they are related by a relabeling of the parties. Each equivalence class of partitions of a partition P is then defined by what we call the cardinality tuple

$$h_P = (|c_i| \mid i \in \{1, \ldots, k\}, c_i \in P), \tag{8.9}$$

which contains the ordered cardinalities of the cells of P. Two partitions P, P' are equivalent if and only if $h_P = h_{P'}$. This equivalence relation is similar to the classification of multipartite entanglement [85]. For any ordered tuple h, one now defines a hybrid model

$$M_h = \mathrm{conv}\left(\bigcup_{P|h_P=h} M^P_{LHV} \right), \tag{8.10}$$

where conv denotes the convex hull. Note that M_h is a convex polytope, the extremal points of which are the union of the extremal points of models M^P_{LHV}.

If h is an m-tuple, we call M_h an m-local model and if m is equal to the number of parties, we call the resulting model *fully local*. A more detailled discussion on different notions of multipartite nonlocality can be found for example in Refs. [102, 103]. Some authors have for example considered hybrid models that impose a no-signaling constraint on the behaviors of each cell in a given partition [104].

8.2 Hybrid Bell inequalities

We consider the case of four and five parties that seek to perform a Bell-type experiment in order to investigate the structure of the nonlocality they might share. Every

party can choose between two measurement settings, each of which yields outcomes ± 1. Moreover, we consider the case in which every party performs one of the two measurements in every round, that is, every party performs a non-trivial measurement in every round. As far as Bell inequalities are concerned, this means that we only consider full-body correlation Bell inequalities. Further, we only consider Bell inequalities that are symmetric under relabelings of the parties. This symmetry constraint is a linear constraint on the coefficients of the Bell inequality in question, so we can employ the cone-projection technique presented in Chapter 5 to specifically find these Bell inequalities.

In order to be able to investigate the nonlocal structure, we need to consider different hybrid models. In the four-party case, these hybrid models are given by the cardinality tuples

$$h \in \{(1,1,1,1),(2,1,1),(2,2),(3,1)\}, \tag{8.11}$$

where $h = (1,1,1,1)$ corresponds to the fully local model. In the case of five parties, we consider six models given by the cardinality tuples

$$h \in \{(1,1,1,1,1),(2,1,1,1),(2,2,1), \tag{8.12}$$

$$(3,1,1),(3,2),(4,1)\}. \tag{8.13}$$

For every model, we find all optimal, symmetric, full-body correlation Bell inequalities.

To aid readability, we introduce some notation that simplifies writing down Bell inequalities which are symmetric under permutation of the parties. For example, for four parties A, B, C, D we write

$$(1122) = \langle A_1 B_1 C_2 D_2 \rangle + \text{party permutations}, \tag{8.14}$$

where 'party permutations' only includes permutations that yield different terms. Therefore, expression Eq. (8.14) consists of six terms. To give another example,

$$\begin{aligned}(1112) = {} & \langle A_1 B_1 C_1 D_2 \rangle + \langle A_1 B_1 C_2 D_1 \rangle \\ & + \langle A_1 B_2 C_1 D_1 \rangle + \langle A_2 B_1 C_1 D_1 \rangle.\end{aligned} \tag{8.15}$$

For five parties, the notation works analogously.

8.3 Numerical analysis

For each Bell inequality, we perform the same numerical analysis. We find a lower bound on the quantum violation using qubit and qutrit systems and an upper bound using the third level of the NPA-hierarchy [37, 42]. We also find the no-signaling

bound. We find that while not all inequalities are violated in quantum mechanics, all can be violated using no-signaling behaviors. We calculate the quantum violation of each Bell inequality using a seesaw algorithm that optimizes the settings of one party in every step and cycles through the parties as described in Ref. [40]. This algorithm is not guaranteed to yield the maximal quantum violation. However, comparing with the upper bound provided by the NPA-hierarchy, we can confirm that the optimum was achieved in all cases. All of bounds for the inequalities are listed together with the inequalities in the Supplemental Material [105].

Additionally, we find lower bounds on the noise robustnesses of the inequalities using the states

$$\varrho_4(p) = (1-p)|\mathrm{GHZ}_4\rangle\langle GHZ_4| + \frac{p}{16}\mathbb{1} \tag{8.16}$$

and for five qubits

$$\varrho_5(p) = (1-p)|\mathrm{GHZ}_5\rangle\langle GHZ_5| + \frac{p}{32}\mathbb{1}, \tag{8.17}$$

where the respective GHZ states are given by

$$|\mathrm{GHZ}_4\rangle = \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle), \tag{8.18}$$

$$|\mathrm{GHZ}_5\rangle = \frac{1}{\sqrt{2}}(|00000\rangle + |11111\rangle). \tag{8.19}$$

In this way, we identify regimes of the parameter p, for which some of the hybrid models can be excluded.

To estimate the noise robustness, we calculate the maximal violation of each inequality for states $\varrho_n(p_i)$, where the values $p_i \in P$ are chosen equidistantly from the interval $[0,1]$. From this, we obtain a critical interval $[p_0^c, p_1^c]$ that contains the noise robustness p^c. Specifically, p_0^c is defined as the largest possible value in P, such that the maximal quantum violation of the Bell inequality exceeds some threshold $t = 10^{-6}$. Similarly, p_1^c is defined as the smallest value in P, such that the Bell inequality is no longer violated. In a second step, we choose new, equidistant parameter values from the critical interval and repeat the procedure. This algorithm is not very efficient in the following sense. One can easily define an algorithm for which the size of the critical interval decreases more quickly as a function of the number of parameter values, for which the quantum violation of the Bell inequality is computed. However, there is an advantage. The value computed for the quantum violation is not guaranteed to be optimal. Calculating the quantum violation for more parameters allows for a sanity check: The maximal quantum violation as a function of the noise parameter p is convex. A proof of this statement can be found in appendix A.6.

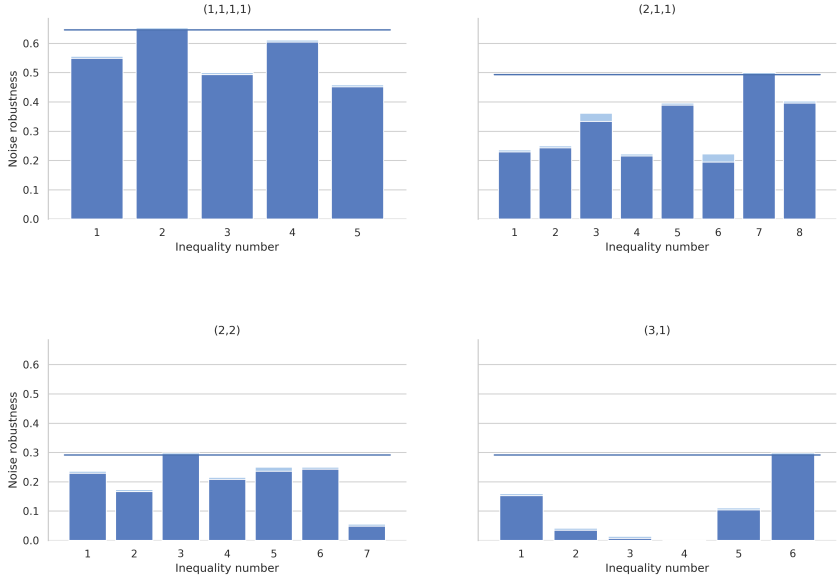


Figure 8.1: Noise robustness for each inequality for hybrid models in the four party case by the number under which it is listed in the online material. Given the family of states $\varrho_4(p)$ as defined in Eq. (8.16) the dark-blue (dark-gray) bar indicates a value of p for which the violation is larger than some threshold (here: 10^{-6}), the light-blue (light-gray) bar indicates a value of p for which the threshold is no longer exceeded. Note that inequality 4 in the 3-1 model is not violated in quantum mechanics at all.

8.4 Four party nonlocality

In the four party case, we find a total of 26 Bell inequalities, five for the fully local model, eight for the $(2,1,1)$ model, seven for the $(2,2)$ model, and six for the $(3,1)$ model. Of these inequalities, we find all but one to be violated in quantum mechanics. All inequalities that are violated in quantum mechanics, are maximally violated by the GHZ state.

For the fully local model, the inequality that exhibits the best white-noise robustness with respect to the GHZ-state is the generalized Mermin inequality

$$+4 - (1111) - (1112) + (1122) + (1222) - (2222) \geq 0, \tag{8.20}$$

which was found in Ref. [85]. With it, nonlocality can be detected with up to roughly 64.5% white noise.

One might expect that for the hybrid models considered, one would find that the

model	#ineq.	noise-robustness	ineq. number	inequality
(1,1,1,1)	5	0.645	2	$+4-(1111)-(1112)+(1122)+(1222)-(2222)$
(2,1,1)	8	0.493	7	$+4-(1112)+(1222)$
(2,2)	7	0.291	3	$+16-(1112)-2(1122)+3(1222)+4(2222)$
(3,1)	6	0.291	6	$+8+(1111)-(1112)-(1122)+(1222)+(2222)$

Table 8.1: For each hybrid model, the table shows the number of optimal symmetric full-body correlation inequalities. Further a lower bound for the white-noise robustness for the $|GHZ\rangle$ state is provided for the best inequality, which is indicated by its number in the list in the Supplementary Material [105].

Bell inequality with the best noise robustness is a generalized Svetlichny inequality. However, this is not the case. For the $(2,1,1)$ model, the most noise robust Bell inequality is

$$4-(1112)+(1222)\geq 0 \tag{8.21}$$

and it is violated up to roughly 49.3% of white noise.

If the amount of white noise is less than roughly $p = 29.1\%$, then the state $\varrho_4(p)$ violates the inequality

$$+8+(1111)-(1112)-(1122)+(1222)+(2222)\geq 0 \tag{8.22}$$

for the $(3,1)$ model and the inequality

$$+16-(1112)-2(1122)+3(1222)+4(2222)\geq 0 \tag{8.23}$$

for the $(2,2)$ model. Interestingly, the amount of white-noise required to obtain a violation of the $(2,2)$ model and the $(3,1)$ model is the same for symmetric, full-body correlation, two-setting inequalities, although the violations are established by different inequalities. We will observe the same phenomenon in the case of five parties. The findings discussed in this subsection are summarized in table 8.4. The noise robustnesses of all inequalities we found in the four party case are plotted in figure 8.1.

8.5 Five party nonlocality

In the five party scenario, we find 9 inequalities for the fully local model, 27 inequalities for the $(2,1,1,1)$ model, 38 inequalities for the $(2,2,1)$ model, 45 inequalities for the $(3,1,1)$ model, 59 inequalities for the $(3,2)$ model and 21 inequalities for the $(4,1)$ model. Among the Bell inequalities in the $(3,2)$ model there are two Bell inequalities that are not violated in quantum mechanics. For the $(4,1)$ model, there 9 such Bell inequalities. All inequalities that are violated are maximally violated by the GHZ state.

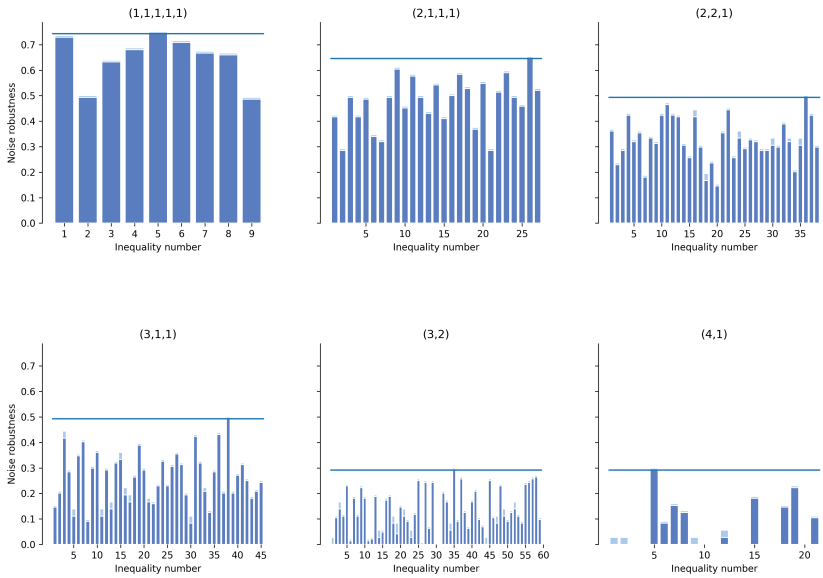


Figure 8.2: Noise robustness for each inequality for hybrid models in the five party case by the number under which it is listed in the online material. The dark-blue (dark-gray) bar indicates a value of p for which the violation is larger than some threshhold (here: 10^{-6}), the light-blue (light-gray) bar indicates a value of p for which the threshhold is no longer exceeded. Note that inequalities 30 and 31 of the 3-2 model as well as inequalities 3, 4, 10, 11, 13, 14, 16, 17 and 20 for the 4-1 local model are not violated in quantum mechanics.

model	#ineq.	noise-robustness	ineq. number	inequality
(1,1,1,1,1)	9	0.743	5	$+4-(11112)+(11222)-(22222)$
(2,1,1,1)	27	0.645	26	$+8+(11111)+(11112)-(11122)-(11222)+(12222)+(22222)$
(2,2,1)	38	0.493	36	$+8+(11111)-(11122)+(12222)$
(3,1,1)	45	0.493	38	$+8+(11111)-(11122)+(12222)$
(3,2)	59	0.291	35	$+40+(11112)+2(11122)-3(11222)-4(12222)+5(22222)$
(4,1)	21	0.291	5	$+16-(11111)+(11112)+(11122)-(11222)-(12222)+(22222)$

Table 8.2: For each hybrid model, the table shows the number of optimal symmetric full-body correlation inequalities. Further a lower bound for the white-noise robustness for the $|GHZ\rangle$ state is provided for the best inequality, which is indicated by its number in the list in the Supplemental Material [105].

The Bell inequality for the fully local model with the best white-noise robustness regarding the GHZ state is obtained for the generalized Mermin inequality

$$+4 - (11112) + (11222) - (22222) \geq 0. \tag{8.24}$$

If the percentage of white noise in the state $\varrho_5(p)$ is roughly smaller than 74.3%, then the state exhibits nonlocality.

For the $(2,1,1,1)$ model, the most sensitive Bell inequality is a generalized Svetlichny inequality

$$\begin{aligned} &+8 + (11111) + (11112) - (11122) \\ &\quad - (22222) + (12222) + (11222) \geq 0. \end{aligned} \tag{8.25}$$

It can detect a violation up to a white noise level up to roughly 64.5%. Note that this is up to numerical precision the same threshold that we obtain for a violation of the fully local model in the case of four parties.

We find numerically that the white-noise thresholds for two-local models, three-local models and five-local models of the four party scenario and the five party scenario coincide. For the three-local models given by $h = (2,2,1)$ and $h = (3,1,1)$, we find that the best inequality to detect the nonlocality of a noisy GHZ state is the generalized Mermin inequality Eq. (8.24), however with a different bound of eight instead of four in the fully local model. It can detect a violation up to approximately 49.3% of white noise.

The two-local models with $h = (3,2)$ and $h = (4,1)$ are violated up to approximately 29.1% of white noise. In the case of the $(4,1)$ model, the violation is detected by the generalized Svetlichny inequality Eq. (8.25) with an adapted bound of 16. For the $(3,2)$ model, we find that the most robust Bell inequality is

$$\begin{aligned} &+40 + (11112) + 2(11122) - 3(11222) \\ &\qquad - 4(12222) + 5(22222) \geq 0. \end{aligned} \tag{8.26}$$

8.6 Family of genuine multipartite Bell inequalities

The inequalities Eq. (8.23) and Eq. (8.26) which are useful for the $(2,2)$ model and the $(3,2)$ model can be extended to a family of Bell inequalities (F_n) for an arbitrary number of n parties. We define this family as

$$F_n = \sum_{\ell=1}^{n} (-1)^{1+\lceil \frac{\ell}{2} \rceil}\, \ell\, (\underbrace{1\ldots1}_{n-\ell}, \underbrace{2\ldots2}_{\ell}) \leq n 2^{n-2}. \tag{8.27}$$

One can show that the bound can always be achieved in any (k,m) model with $k,m > 1$. Moreover, for $m = 2$, one can show analytically that the bound holds. Proofs for both

statements can be found in appendix A.4. Moreover, we checked numerically for up to 20 parties that the bound holds for (k, m) models with $k > 1, m > 1$. We conjecture that the bound also tight for more parties. For $(n-1, 1)$ models the bound does not hold. A violation of an inequality in the family Eq. (8.27) therefore certifies that the nonlocality depth is at least $(n-1)$.

In the following, we discuss numerical findings concerning the Bell inequalities F_n for $n \leq 8$ parties. We find, that the quantum bound of the inequalities is

$$F_n = \sqrt{2}\, n 2^{n-2} \tag{8.28}$$

up to numerical precision, which is achieved by choosing $A_1 = B_1 = ... = \sigma_X$ and $A_2 = B_2 = ... = \sigma_Z$, where σ_X, σ_Z are Pauli matrices. The quantum states for which a maximal violation can be obtained using these settings are listed in appendix A.5. We find numerically that they are equivalent to the GHZ state up to the action of local unitaries.

Further, we consider the white-noise robustness. For this, we consider the states

$$\varrho_n(p) = (1-p)|\mathrm{GHZ}_n\rangle\langle +|p\, 2^{-n}\mathbb{1}_n. \tag{8.29}$$

We find that the inequalities are violated if $p \leq 1 - \frac{1}{\sqrt{2}} \approx 29.3\%$ up to numerical precision.

8.7 Generalizations of Svetlichny's inequality to more settings

Facet defining inequalities of a polytope that obey a set of affine constraints can be found using the cone-projection technique which we introduced in our publications [A, B]. Alternatively, if one is interested in finding the generalizing Bell inequality that is best suited to detect the nonlocality in an behavior r, this problem is a linear program

$$\begin{aligned} & \max_{b_2} \langle r, b_2 \rangle \\ \text{s.t} \quad & \langle \beta_2, b_2 \rangle = 1 \ \forall \text{ extended saturating behaviors } \beta_2 \\ & \langle \beta, b_2 \rangle \leq 1 \ \forall \text{ extremal behaviors of the model } \beta. \end{aligned} \tag{8.30}$$

Running the linear program Eq. (8.30) with random directions r, we find two generalizations of Svetlichny's inequality that are symmetric under party permutations for

the three party scenario with three settings per party, or $(3,3,3;2,2,2)$ for short,

$$\begin{aligned} f_1 =& (100) - (111) + (211) + (221) - (222) \\ &+ 2(300) - (310) + (330) + (331) \leq 13 \end{aligned} \tag{8.31}$$

$$\begin{aligned} f_2 =& - (122) + (123) + (133) - 3(222) \\ &- 2(223) + (233) \leq 12. \end{aligned} \tag{8.32}$$

The inequality f_1 reduces to Svetlichny's inequality, if one sets $A_3 = B_3 = C_3 = 1$. The second inequality, f_2, reduces to Svetlichny's inequality, if one sets $A_3 = A_1, B_3 = B_1, C_3 = C_1$.

By construction, the inequalities f_1, f_2 are at least as sensitive to nonlocality as Svetlichny's inequality. Moreover, since they have one more setting, one might expect that there might be an advantage of f_1 and f_2 compared to Svetlichny's inequality. Unfortunately, sampling 540 random pure three-qutrit states, we did not find a single example that shows an advantage. Rather, it appears that choosing the additional settings such that the inequalities reduce to Svetlichny's inequality is always optimal. Accordingly, f_1, f_2 share the maximally violating state, the GHZ state, and their noise-robustness with Svetlichny's inequality.

8.8 Conclusion

We presented Bell-like inequalities to detect violations of various hybrid models for four- and five-body systems. Our analysis of GHZ-states that are mixed with white noise suggests that the noise robustness of these states with regard to a k-local model only depends on k. In contrast, the particular partition of parties that defines the k-local model seems to be irrelevant. For example, we did not find a difference between the $(3,2)$-model and the $(4,1)$-model in terms of noise-robustness.

Additionally to our analysis of four- and five-party scenarios, we present a family of inequalities for an arbitrary number of parties n. The inequalities in this family are suitable to detect a nonlocality-depth of $n-1$.

Finally, we introduce the concept of a generalization of a Bell inequality to a scenario that involves more settings. We demonstrate this concept by finding two inequalities that generalize Svetlichny's inequality. Unfortunately, these inequalities do not seem to have an advantage over Svetlichny's inequality.

9 Bellpy: A toolbox for nonlocality in Python

We present the Python package Bellpy, which is designed to aid research in nonlocality. It does so in two ways: (i) It provides a framework consisting of classes that are necessary to formulate any problem regarding Bell-type black-box experiments, and (ii) it provides implementations for commonly needed functions. Bellpy is modular and easily extendable.

Any concept that is set in typewriter font is either a function, a class or abstract base class in our code.

This chapter contains text from our manuscript [D]. The software presented in this chapter was used to obtain the results presented in the Chapters 6, 7, and 8.

9.1 Framework

Consider a black-box experiment. This experiment is characterized by the settings that the parties that participate in the experiment can choose from. A `Setting` has three defining properties: (1) the party it belongs to, (2) a label that distinguishes it from other settings, (3) a set of labels for the outcomes that the party can observe after choosing this setting.

All the settings of all parties together give rise to a `Scenario`, which defines what can be done in the experiment. The CHSH scenario would for example be defined as

```
import bellpy as bell

A1 = bell.Setting(0,1,[-1,1])
A2 = bell.Setting(0,2,[-1,1])
B1 = bell.Setting(1,1,[-1,1])
B2 = bell.Setting(1,2,[-1,1])

chsh_scenario = bell.Scenario([A1, A2, B1, B2])
print(chsh_scenario.nsettings)
```

Given a scenario, one can construct the behavior space for that scenario. However, in practise one may only be interested in a subspace of the behavior space as de-

fined in Chapter 4. The specifics of this subspace may depend strongly on the research conducted. For this purpose we provide the abstract base class `Behavior_space`. We further provide an abstract base class `Model`. Using these abstract base classes has the advantage that some functions, such as obtaining the Bell inequalities from a model are already implemented in the base class. Moreover, if classes that are derived from those base classes, then these classes automatically integrate with the rest of the Bellpy library, so other functionality such as the cone-projection technique are directly available.

9.2 Concrete behavior spaces and models

In Bellpy, we provide two implementations of a behavior space: The

- `Default_expectation_behavior_space`, and the
- `Custom_expectation_behavior_space`.

Both are implementations of the abstract `Expectation_behavior_space`.

A `Default_expectation_behavior_space` is defined by listing the number of settings per party. The settings for each party are labeled with natural numbers starting from 1. Additionally, each party has a trivial setting. This setting always yields the outcome $+1$ and represents the case in which no measurement is performed by that party. The trivial setting is labeled 0. The labels and standard basis vectors of the default behavior space are the combinations of inputs of the parties. In the CHSH scenario, these labels are $A_0B_1, A_1B_0, A_1B_1, \ldots$. In the standard basis, the coefficient of a behavior is the joint expectation value of the inputs stated in the respective label.

A `Custom_expectation_behavior_space` is more flexible. It is defined by a list of tuples, called correlation list. Each of these tuples $(i_1, \ldots, i_n)$ is to be interpreted as a tuple of inputs i_k on party k. The tuples also serve as the labels of the basis vectors of the behavior space. The scenario of a custom expectation behavior space is the scenario defined by the settings with labels that occur in the correlation list. The coefficients of a behavior are the joint expectation values for the settings indicated by the label.

Apart from behavior spaces, we also provide implementations for frequently encountered models. Two examples for such models are the `Local_deterministic_model` and the `Unrestricted_model`. In the unrestricted model, the outcome of one party can depend on the inputs and outputs of all other parties in addition to being dependent on a hidden variable. Both models work for any implementation of a `Behavior_space`.

9.3 Local hidden-variable models and Bell inequalities

Models can be defined by stating their behavior space. In the case of a local hidden-variable model, this would work as follows:

```
import bellpy as bell

B = bell.Default_expectation_behavior_space(2,2)
lhv = bell.Local_deterministic_model(B)

bis = lhv.bell_inequalities()
print(bis)
```

This code yields all facet Bell inequalities in the CHSH scenario. These are different variants of the CHSH inequality and additionally some trivial Bell inequalities.

The `Local_deterministic_model` is flexible enough to handle different partitions of the parties. If a partition is passed, signaling between parties that belong to the same cell of the partition is granted. If one wants to create a model which does not restrict the correlations between the first two parties and maintains locality with respect to the bipartition $AB|C$, this model can be created as follows:

```
import bellpy as bell

B = bell.Default_expectation_behavior_space(2,2,2)
lhv = bell.Local_deterministic_model(B, [[0,1], [2]])
```

Svetlichny's hybrid model is the convex hull of three such models, namely for the bipartitions $AB|C, AC|B, A|BC$. We can create this model using the `Hybrid_model` class.

```
import bellpy as bell

B = bell.Default_expectation_behavior_space(2,2,2)
lhv_c = bell.Local_deterministic_model(B, [[0,1], [2]])
lhv_b = bell.Local_deterministic_model(B, [[0,2], [1]])
lhv_a = bell.Local_deterministic_model(B, [[0], [1,2]])

svet = bell.Hybrid_model(lhv_c, lhv_b, lhv_a)
```

If we are only interested in full-body correlations, we can alternatively specify these correlations and create a `Custom_expectation_behavior_space`.

```
import bellpy as bell
import itertools as it

correlations = [(1, 1, 1), (1, 1, 2), (1, 2, 1), (1, 2, 2),
```

```
        (2, 1, 1), (2, 1, 2), (2, 2, 1), (2, 2, 2)]
B = bell.Custom_expectation_behavior_space(correlations)
lhv_c = bell.Local_deterministic_model(B, [[0,1], [2]])
lhv_b = bell.Local_deterministic_model(B, [[0,2], [1]])
lhv_a = bell.Local_deterministic_model(B, [[0], [1,2]])

svet = bell.Hybrid_model(lhv_c, lhv_b, lhv_a)
bis = svet.bell_inequalities()
print(bis)
```

This yields three types of inequalities: trivial ones, Svetlichny inequalities, and one more non-trivial facet-defining Bell inequality different from Svetlichny's inequality, which was found by Bancal [81]. Note that the order in which the parties are listed in the partition does not matter.

9.4 Finding Bell inequalities with affine constraints

Listing all Bell inequalities of a model quickly becomes unfeasible as the scenario gets more complex. In this situation, one may consider to only consider a subset of correlations, or to only consider Bell inequalities that are symmetric under party permutations [81]. More generally, one may restrict the search to Bell inequalities

$$\langle f, b \rangle \leq \beta, \tag{9.1}$$

such that f meets a set of affine constraints. Here b is a behavior and (β, f) defines the Bell inequality. Such constraints can be given as a matrix G, which implies the condition

$$G(\beta \oplus f) = 0. \tag{9.2}$$

Apart from symmetry constraints, such affine constraints can for example be useful in order to find generalizations of Bell inequalities, as introduced in Chapter 5. We demonstrate how this works in Bellpy for an example. Assume that we aim to find generalizations of the CHSH inequality to three settings per party. The code for this task is:

```
""" Find I3322 inequalities """
from bellpy import *

bs22 = Default_expectation_behavior_space(2,2)
print(bs22.labels)
# ['A0B0', 'A0B1', 'A0B2', 'A1B0', 'A1B1',
```

```
#  'A1B2', 'A2B0', 'A2B1', 'A2B2']

bs33 = Default_expectation_behavior_space(3,3)

chsh_array = [2, 0, 0, 0, -1, -1, 0, -1, 1]
chsh = Bell_inequality(bs22, chsh_array)

lhvm2222 = Local_deterministic_model(bs22)
lhvm3322 = Local_deterministic_model(bs33)

ext_map = Canonic_extension_map(bs22, (0,0,1), (1,0,1))
eb = Extended_behaviors(lhvm2222, chsh, ext_map)

cpt = Cone_projection_technique(lhvm3322, true_facets=True)
cpt.add_extended_behaviors(eb)
bi = cpt.bell_inequalities()
bi.remove_duplicates(party=True, setting=True, outcome=True)
print(bi)
```

First, the behavior spaces of the CHSH inequality and its generalizations are generated. We then define the CHSH inequality by stating its behavior space and its coefficients in the standard basis. Afterwards, the local hidden-variable models for both behavior spaces are created. With this, we can formulate the conditions for the CHSH generalizations. These conditions are stated in terms of the extended behaviors. They are calculated from the saturating behaviors of the CHSH inequality via an extension map. One type of extension map is implemented in Bellpy, which is called `Canonic_extension_map`. It is defined by its domain and a list of tuples. Each of these tuples indicate a measurement that should be added to the scenario and how this measurement relates to the present ones. The tuple (p, s, σ) means that a setting should be added to party p. The outcome of the new setting should be equal to the one of setting s up to a sign σ. The label of the newly added setting will automatically be $1 + \max(i_p)$, where i_p is the number of settings of party p. Apart from the CHSH inequality itself and some trivial Bell inequalities, the code yields true generalizations of CHSH, which are versions of the well-known I3322 inequality [16]. All of the Bell inequalities are stored in a `Bell_inequality_collection`. Different versions of the same inquality are removed using the `remove_duplicates()` function as shown above. The arguments indicate which relabelings should be taken into account.

The `Cone_projection_technique` also allows one to impose that the inequalities to be found be symmetric under party permutations. This is done by typing

```
cpt.add_party_symmetry().
```

Similarly, arbitrary affine constraints G can be added with the function `cpt.add_constraints(G)`.

Besides finding all Bell inequalities, a `Cone_projection_technique` object can also write a file representing the projection of the model into the subspace defined by the constraints in `.ext` format. This format is commonly used in polyhedral conversion software such as CDD and LRS [25, 18, 27]. Exporting the problem to .ext format is achieved by running

```
cpt.run().
```

Apart from the `.ext` file, also the `Cone_projection_technique` object will be saved automatically. The `.ext` file can then be converted into `.ine` format by external software. Afterwards, the `Cone_projection_technique` object can be loaded using `cpt.load_cpt(filename)` and the `.ine` file is read by executing the `cpt.run()` function once again. For this to work, the `.ine` file must have the same name as the `.ext` file up to the file extension.

9.5 Analyzing two-outcome Bell inequalities

To analyse Bell inequalities in behaviors that lie in an `Expectation_behavior_space`, Bellpy provides the `Inequality` class. In contrast, the `Bell_inequality` class, which works for any affine Bell inequality for any behavior space, does not offer any functionality for analyzing Bell inequalities.

An `Inequality` can be instantiated with either a

- `Bell_inequality`, or
- with a string representation of the inequality.

We refer to the string representation of the Bell inequality as the `name` of the inequality. The name of the inequality obeys the following conventions. We define

$$[ijk] = \langle A_i B_j C_k \rangle. \tag{9.3}$$

As discussed earlier, the setting of each party with label 0 is considered trivial and always yields the outcome $+1$.

The CHSH inequality is created like this.

```
from bellpy import *

nameasy = '2 [00] - [12] - [21] - [11] + [22]'
chsh = Inequality(name=nameasy)
print(chsh.name)
# output: +2.0 (00) - (11) - (12) + (22)
```

The expression '2 [00] - [12] - [21] - [11] + [22]' is to be understood as the inequality

$$2[00] - [12] - [21] - [11] + [22] \geq 0. \tag{9.4}$$

Printing the name of the inequality shows a different string. This is a feature. Bellpy has automatically detected that the CHSH inequality is invariant, if the parties A and B are relabeled. Such symmetric inequalities can be more conveniently stated in a notation, where every correlation is symmetrized. For example,

$$(12) = [12] + [21]. \tag{9.5}$$

This notation is the same notation we used in Chapter 6.

We can find the qubit and qutrit violations of an inequality as well as bounds for the maximal violations given by the NPA hierarchy [37].

```
chsh.update_qubit_violation(verbose=True)
chsh.update_qutrit_violation()
chsh.update_npa2()
chsh.update_npa3()
```

The quantum violations are computed using the seesaw algorithm described in the paper by Pal et al. [40]. To calculate the NPA bounds, we rely on the package `ncpol2sdpa` by Peter Wittek [96]. We can also find the maximal violation possible with nonsignaling behaviors.

```
settings = chsh.behavior_space.nsettings
ns = NoSignaling(settings)
chsh.update_no_signaling_bound(ns)

print(chsh.summary())
```

The summary function yields information about the inequality. In this case, it shows:

```
---------------------------------------
Number 0 ------------------------------
 +2.0 (00) - (11) - (12) + (22) > 0
no signaling:     -2.0
npa2:             -0.828427128727788
npa3:             -0.8284271258292932
qubit violation:  (-0.8284271247461898-6.359151412848378e-17j)
qutrit violation: (-0.8284271247461903-1.3648426433378623e-16j)
algebraic bound: 4.0

Qubit violation
***************

violation: (-0.8284271247461898-6.359151412848378e-17j)
state:     [ 0.70711-0.j       -0.      -0.j
             0.      +0.j       -0.37775+0.59775j]
A1: [ 1.38777878e-16  1.11932404e-01  9.87600738e-01 -1.10072339e-01]
A2: [-2.22044605e-16  7.94258277e-01 -1.55482006e-01 -5.87349244e-01]
B1: [ 0.          0.15508262  0.85600873 -0.49315153]
B2: [2.77555756e-17 9.41026112e-01 2.39422904e-02 3.37485738e-01]
```

The settings A_1, A_2, B_1, B_2 are stated in the Pauli basis. Every inequality has a number associated to it, which is 0 by default and can be accessed through the property `line`. This is mostly important when the inequality is saved.

```
chsh.save()
```

If no further arguments are supplied, `line` will be used as filename. Inequality objects are saved as pickled dictionaries with the file extension `.ineq`.

To make working with many inequalities more convenient, Bellpy provides the `Inequalities` class. Besides functions for saving and loading inequalities from a folder, the class `Inequalities` also provides a function for removing different versions of the same inequality under relabelings of parties, settings and outcomes.

10 Absolutely maximally entangled Werner states

Throughout this chapter, we are concerned with absolutely maximally entangled (AME) Werner states $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$, where we consistently use the letter n to denote the number of parties and the letter d to denote the dimension of the local Hilbert spaces. AME-Werner states are closely related to invariant perfect tensors, which have been considered in Ref. [106], where the question was asked whether m-valent invariant perfect tensors exist for $m > 4$. We show that AME-Werner states only exist in systems of two qubits or three qutrits. The present chapter closely follows our manuscript [E].

An example for a two-qubit AME-Werner state is the singlet state [55]

$$|\psi^-\rangle = \frac{1}{\sqrt{2}}(\,|01\rangle - |10\rangle). \tag{10.1}$$

For three qutrits, one example is the state

$$|\psi\rangle = \frac{1}{\sqrt{6}} \sum_{i,j,k=1}^{3} \epsilon_{ijk} \,|i-1, j-1, k-1\rangle, \tag{10.2}$$

where ϵ_{ijk} is the Levi-Civita symbol.

10.1 Properties of pure Werner states

In this section we establish some properties of pure Werner states.

Definition 10.1 (Pure Werner state [48])**.** A pure Werner state $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$ is a state such that for any unitary $U : \mathbb{C}^d \to \mathbb{C}^d$ it holds that

$$U \otimes \ldots \otimes U \,|\psi\rangle = \xi \,|\psi\rangle, \tag{10.3}$$

where $\xi \in \mathbb{C}$ is a complex phase that depends on the unitary U and the state.

Definition 10.2 (Phase function)**.** Let $|\psi\rangle$ be a pure Werner state. Then, the function $\xi : U(d) \to \mathbb{C}$ that satisfies

$$U \otimes \ldots \otimes U \,|\psi\rangle = \xi(U) \,|\psi\rangle, \tag{10.4}$$

is called the phase function of $|\psi\rangle$.

Lemma 10.1. The phase function ξ is a group homomorphism, that is $\xi(U_1U_2) = \xi(U_1)\xi(U_2)$ and $\xi(\mathbb{1}) = 1$.

Proof. This is a trivial result that directly follows from Eq. (10.3) and the fact that the action of the unitary group on every subsystem maps Werner states to Werner states. □

Lemma 10.2. Let $(\varrho, \mathbb{C}^d)$ be a representation of the symmetric group $\mathbb{S}_d$ and $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$ a pure Werner state. Then it holds that

$$\varrho(\pi) \otimes \ldots \otimes \varrho(\pi) |\psi\rangle = \xi \circ \varrho(\pi) |\psi\rangle, \tag{10.5}$$

where ξ is the phase function of $|\psi\rangle$ and $\pi \in \mathbb{S}_d$.

Proof. The group $\varrho(\mathbb{S}_d)$ is a subgroup of the unitary group $U(d)$. □

Definition 10.3 (Canonical representation of $\mathbb{S}_d$)**.** Let $(|k\rangle)_{k\in\{1,\ldots,d\}}$ be the computational basis of $\mathbb{C}^d$. We then call the representation $(\varrho, \mathbb{C}^d)$ of the symmetric group $\mathbb{S}_d$ which is defined as

$$\varrho(\pi) |k\rangle = |\pi(k)\rangle \tag{10.6}$$

the canonical representation of $\mathbb{S}_d$.

Definition 10.4 (permutation-phase function)**.** Let $(\varrho, \mathbb{C}^d)$ be the canonical representation of S_d. Further, let ξ be the phase function of the Werner state $|\psi\rangle$. Then we call

$$f = \xi \circ \varrho \tag{10.7}$$

the permutation-phase function of $|\psi\rangle$.

Lemma 10.3. There are only two functions that are compatible with the definition of the permutation-phase function f, namely either $f(\pi) = 1$, $\forall \pi \in \mathbb{S}_d$ or $f(\pi) = \mathrm{sgn}(\pi)$, the signum of the permutation.

Proof. The same argument that shows that ξ is a group homomorphism also shows that f is a group homomorphism, too. Therefore, we only need to consider the images of the generators of S_d under f in order to characterize f.

The symmetric group $\mathbb{S}_d$ is generated by the transpositions $\theta_k = (k, k+1)$ with $k \in \{1, \ldots, d-1\}$, so f is determined by $f(\theta_k)$. Further, the transpositions fulfill the following relations

$$\theta_k^2 = \mathrm{id} \tag{10.8}$$

$$\theta_k\theta_m = \theta_m\theta_k, \text{ for } |k-m| \geq 2 \tag{10.9}$$

$$\theta_k\theta_m\theta_k = \theta_m\theta_k\theta_m, \text{ for } |k-m| = 1. \tag{10.10}$$

Since f is a homomorphism, Eq. (10.8) gives

$$f(\theta_k) = \pm 1 \tag{10.11}$$

and Eq. (10.10) gives

$$f(\theta_k) = f(\theta_m) \text{ for } |k-m| = 1. \tag{10.12}$$

Hence, we only have to distinguish two cases: (1) $f(\theta_k) = 1 \quad \forall k$ and (2) $f(\theta_k) = -1 \quad \forall k$, i.e. $f = \text{sgn}$. □

Lemma 10.3 allows us to derive a simple relation between some of the coefficients of a pure Werner state. We can state the result more elegantly by making use of the notations that are defined below.

Definition 10.5 (Multi-index)**.** We define a multi-index i as a sequence of indices $i_\alpha \in \{1,\dots,d\}$ with $\alpha \in \{1,\dots,n\}$.

Definition 10.6 (Action of the group $\mathbb{S}_d$ on a multi-index)**.** For a multi-index $i = (i_1,\dots,i_n)$ and $\pi \in \mathbb{S}_d$, we define

$$\pi(i) = (\pi(i_1),\dots,\pi(i_n)). \tag{10.13}$$

Lemma 10.4. Let $|\psi\rangle = \sum_i t_i |i\rangle \in (\mathbb{C}^d)^{\otimes n}$ be a pure Werner state and $\pi \in \mathbb{S}_d$ a permutation. Then it holds that

$$t_{\pi(i)} = f(\pi) t_i \tag{10.14}$$

where f is the permutation-phase function of $|\psi\rangle$.

Proof. According to Lemma 10.2, any pure Werner state obeys

$$\varrho(\pi') \otimes \dots \otimes \varrho(\pi') |\psi\rangle = f(\pi') |\psi\rangle, \tag{10.15}$$

where $(\varrho, \mathbb{C}^d)$ is the canonical representation of $\mathbb{S}_d$ and $\pi' \in \mathbb{S}_d$. Evaluating the left hand side yields

$$\begin{aligned} &\varrho(\pi') \otimes \dots \otimes \varrho(\pi') |\psi\rangle \\ =&\varrho(\pi') \otimes \dots \otimes \varrho(\pi') \sum_i t_{i_1,\dots,i_n} |i_1 \dots i_n\rangle \\ =&\sum_i t_{i_1\dots i_n} |\pi'(i_1)\rangle \dots |\pi'(i_n)\rangle \\ =&\sum_i t_{\pi'^{-1}(i_1)\dots\pi'^{-1}(i_n)} |i\rangle. \end{aligned} \tag{10.16}$$

Inserting this expression back into Eq. (10.15) yields

$$t_{\pi'^{-1}(i)} = f(\pi') t_i. \tag{10.17}$$

If we choose $\pi' = \pi^{-1}$, we obtain

$$t_{\pi(i)} = f(\pi^{-1})t_i. \tag{10.18}$$

However, f is a group homomorphism, so

$$1 = f(\pi^{-1}\pi) = f(\pi^{-1})f(\pi), \tag{10.19}$$

and according to Lemma 10.3 it holds that $f(\pi) = \pm 1$. Hence, we have $f(\pi) = f(\pi^{-1})$ which ends the proof. □

We now characterize the phase-function further by considering the restriction of ξ to the diagonal unitaries.

Lemma 10.5. Let $|\psi\rangle = \sum_j t_j |j\rangle$ be a pure Werner state. Then, there exists a number $N_k \in \mathbb{N}$ for every number $k \in \{1, \ldots, d\}$, such that for any non-zero coefficient $t_l \neq 0$, the multi-index l contains the value k exactly N_k times.

Proof. Since $|\psi\rangle$ is a pure Werner state, according to definition 10.1 it obeys Eq. (10.3) for any unitary matrix of suitable dimension. This in particular includes unitary matrices.

Diagonal unitary matrices are generated by the matrices

$$U_k = \mathbb{1} + (e^{i\phi_k} - 1)|k\rangle\langle k| \tag{10.20}$$

that have only diagonal entries, all of which except the k-th one are 1, while the k-th diagonal entry is $e^{i\phi_k}$. Such a diagonal matrix acts on a computation basis vector $|a\rangle$ as

$$U_k |a\rangle = \begin{cases} e^{i\phi_k} |a\rangle, \text{ if } a = k, \\ |a\rangle, \text{ else.} \end{cases} \tag{10.21}$$

We now determine $\xi(U_k)$, where ξ is the phase function of $|\psi\rangle$, by evaluating the left-hand side of Eq. (10.3). This yields

$$\begin{aligned} & U_k \otimes \ldots \otimes U_k |\psi\rangle \\ =& \sum_j t_j U_k |j_1\rangle \otimes \ldots \otimes U_k |j_n\rangle \\ =& \sum_j e^{i\phi_k K_j} t_j |j\rangle, \end{aligned} \tag{10.22}$$

where K_j is the number of times that k appears in the multi-index j. This equation holds for arbitrary ϕ_k and $\xi(U_k)$ is unique for all coefficients $t_j \neq 0$. Hence, all indices l, for which $t_l \neq 0$ must contain k the same number of times. □

From this Lemma, we can derive a corollary, where we make use of the following definition.

Definition 10.7 (Action of $\mathbb{S}_n$ on a multi-index)**.** Let i be a multi-index and $\omega \in \mathbb{S}_n$. We define

$$\omega(i) = (i_{\omega(1)}, ..., i_{\omega(n)}). \tag{10.23}$$

Corollary 10.6. Let $|\psi\rangle = \sum_i t_i |i\rangle$ be a Werner state. Further, let j, l be two multi-indices such that $t_j \neq 0, t_l \neq 0$. Then, there exists a permutation $\omega \in \mathbb{S}_n$, such that $j = \omega(l)$.

We can now make our characterization of non-zero coefficients of pure Werner states more specific.

Lemma 10.7. Let $|\psi\rangle = \sum_i t_i |i\rangle$ be a pure Werner state. Then, there exists a number $K \in \mathbb{N}$, such that for any non-zero coefficient $t_l \neq 0$, the multi-index l contains each value $k \in \{1, \dots, d\}$ exactly K times.

Proof. Consider two values $k, m \in \{1, \dots, d\}$ and let N_k, N_m be are defined as in Lemma 10.5, that is, N_k denotes how many times k is contained in the multi-index l of any coefficient $t_l \neq 0$. Now consider the permutation $(k, m) \in \mathbb{S}_d$. According to Lemma 10.4, we have

$$t_l = f((k, m)) t_{(k,m)(l)}. \tag{10.24}$$

Therefore, $t_l \neq 0$ implies $t_{(k,m)(l)} \neq 0$ due to Lemma 10.3. However, $(k, m)(l)$ contains k exactly N_m times and m exactly N_k times. Thus, we have $N_m = N_k$, which ends the proof. □

Corollary 10.8. For any pure Werner state $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$, n is always an integer multiple of d, that is $n = Kd, K \in \mathbb{N}$.

Proof. This result follows trivially from Lemma 10.7. □

10.2 Main result

With these results, we have sufficiently characterized pure Werner states for our needs and we can now consider the second property, absolute maximal entanglement. For convenience, we introduce another notational convention.

Definition 10.8 (Truncated multi-index)**.** Let $i = (i_\alpha)_{\alpha \in \{1,\dots,n\}}$ be a multi-index and $A \subset \{1, \dots, n\}$ a set of subsystems. Then, we define the truncated multi-indeces

$$i_A = (i_\alpha)_{\alpha \in A} \tag{10.25}$$

and

$$i_{\neg A} = (i_\alpha)_{\alpha \in \neg A} \tag{10.26}$$

with

$$\neg A := \{1, \ldots, n\} \setminus A. \tag{10.27}$$

Using this notation and with a set of subsystems A, we express can write the marginal state of the subsystems contained in A as

$$\begin{aligned}
&\mathrm{Tr}_{\neg A}\left(|\psi\rangle\langle\psi|\right) \\
=&\mathrm{Tr}_{\neg A}\left(\sum_{i,j} t_i t_j^* |i\rangle\langle j|\right) \\
=&\mathrm{Tr}_{\neg A}\left(\sum_{i_A, i_{\neg A}, j_A, j_{\neg A}} t_{i_A, i_{\neg A}} t^*_{j_A, j_{\neg A}} |i_A\rangle\langle j_A| \otimes |i_{\neg A}\rangle\langle j_{\neg A}|\right) \\
=&\sum_{i_A, i_{\neg A}, j_A} t_{i_A, i_{\neg A}} t^*_{j_A, i_{\neg A}} |i_A\rangle\langle j_A| \\
=&\sum_{i_A, j_A} \tau_A\left(i_A, j_A\right) |i_A\rangle\langle j_A|,
\end{aligned} \tag{10.28}$$

where we denote the coefficients of the marginal density matrix as

$$\tau_A\left(i_A, j_A\right) = \sum_{i_{\neg A}} t_{i_A, i_{\neg A}} t^*_{j_A, i_{\neg A}}. \tag{10.29}$$

Using this notation, we can state the AME condition in a very explicit form, that is useful for our means.

Definition 10.9 (AME-Werner state)**.** A state $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$ is an AME-Werner state, if it is a Werner state and for any set of parties A with $|A| = \lfloor \frac{n}{2} \rfloor$, the corresponding marginal is a maximally mixed state, that is

$$\tau_A(i_A, j_A) = \begin{cases} \frac{1}{d^{|A|}}, \text{ if } i_A = j_A \\ 0, \text{ else.} \end{cases} \tag{10.30}$$

With this notation, we are ready to proceed to the proof.

Lemma 10.9. Let $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$ be an AME-Werner state. Then,

$$\left\lfloor \frac{n}{2} \right\rfloor \leq \frac{n}{d} \tag{10.31}$$

holds.

Proof. Consider the marginal state of $|\psi\rangle = \sum_i t_i |i\rangle$ of the subsystems in $A = \{1, \ldots, \lfloor \frac{n}{2} \rfloor\}$. Then, the diagonal element

$$\tau_A(1 \ldots 1, 1 \ldots 1) = \sum_{i, i_A = 1 \ldots 1} |t_i|^2 \tag{10.32}$$

does not vanish. Therefore, there is at least one non-zero term in the sum Eq. (10.32). Hence, there exists a multi-index l with

$$l_A = (1, \ldots, 1) \tag{10.33}$$

such that $t_l \neq 0$. According to corollary 10.8, we have $n = Kd$ for some $K \in \mathbb{N}$. Further, $t_l \neq 0$ implies that the value 1 is contained in l exactly K times due to lemma 10.7. From Eq. (10.33) follows that l contains the value 1 at least $\lfloor \frac{n}{2} \rfloor$ times. This is only possible if $K \geq \lfloor \frac{n}{2} \rfloor$. □

Corollary 10.10. Let $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$ be an AME-Werner state and n an even number. Then, the local dimension is $d = 2$.

Corollary 10.11. Let $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$ be an AME-Werner state and n an odd number. If $n = 3$, then the local dimension is $d = 3$. If $n > 3$, the local dimension is $d = 2$.

Proof. From lemma 10.9 we get that $d \leq 2$ if $n > 3$ and $d \leq 3$ if $n = 3$. Then, lemma 10.7 excludes the case that $d = 2$ and $n = 3$. □

Theorem 10.12. Let $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$ be an AME-Werner state. Then either $n = d = 2$ or $n = d = 3$.

Proof. From corollaries 10.10 and 10.11 we already know that if $n = 3$ then $d = 3$ and if $n \neq 3$ then $d = 2$. Hence, the only thing that remains to be shown is that AME-Werner states do not exist for n-qubit systems with $n > 2$. We prove this statement by contradiction. First, we assume that such a state exists. Then we show that all coefficients t_i of this state vanish.

Assume there exists such a AME-Werner state $|\psi\rangle = \sum_i t_i |i\rangle$. Then there exists a non-zero coefficient t_l. According to lemma 10.7, the number K of indices in l that take the value 1 is equal to the number of indices in l that take value 2. We now consider the marginal of the state $|\psi\rangle$, that describes the subsystems α for which $l_\alpha = 1$ holds, where l_α is an index in l. We call the set of chosen subsystems A. Now consider the diagonal element of this marginal

$$\tau_A(1...1, 1...1) = |t_l|^2. \tag{10.34}$$

Since all diagonal elements of all marginals of $\frac{n}{2} = K$ systems are equal, so are the absolute values of all coefficients t_l, with l chosen as above.

We now consider a second marginal of subsystems that we choose as follows. Choose $K - 1$ subsystems α with $l_\alpha = 1$ and one subsystem with $l_\alpha = 2$. We call the set of these subsystems B. Now consider the diagonal element of the marginal state

$$\begin{aligned} \tau_B(l_B, l_B) &= \sum_{i, i_B = l_B} |t_i|^2 \\ &= K|t_l|^2. \end{aligned} \tag{10.35}$$

However, we know that

$$\tau_A(1...1, 1...1) = \tau_B(l_B, l_B) \tag{10.36}$$

and therefore

$$|t_l|^2 = K\,|t_l|^2. \tag{10.37}$$

Since by assumption we have $d = 2$ and $n > 2$, according to corollary 10.8 we have $K \geq 2$. This implies $t_l = 0$ and hence $|\psi\rangle = 0$, where we arrive at the contradiction.

□

11 Entanglement dynamics of two mesoscopic objects with gravitational interaction

The content of this chapter has appeared in our publication [F]. The author of this thesis contributed to this work, in particular to the derivation of the system's dynamics and the investigation into the entanglement and nonlocal properties.

We analyse the entanglement dynamics of the two particles interacting through gravity in the recently proposed experiments aiming at testing quantum signatures for gravity [Phys. Rev. Lett 119, 240401 & 240402 (2017)]. We consider the open dynamics of the system under decoherence due to the environmental interaction. We show that as long as the coupling between the particles is strong, the system does indeed develop entanglement, confirming the qualitative analysis in the original proposals. We show that the entanglement is also robust against stochastic fluctuations in setting up the system. The optimal interaction duration for the experiment is computed. A condition under which one can prove the entanglement in a device-independent manner is also derived.

11.1 Introduction

The unification of quantum mechanics and general relativity has been perceived as one of the most important open problems of modern physics. Although a substantial theoretical effort has been made, there is not yet an agreement on a single theory of quantum gravity [107]. One of the main difficulties of the field is the lack of experimental support [107]. Recent advances in experimentally probing quantum physics at the mesoscopic scale have now raised the hope of finding quantum effects in gravity, which might eventually be considered as a quantum signature of gravity [108]. Among these proposals, a simple one aiming at demonstrating the entanglement between mesoscopic particles interacting via gravity was proposed by Bose *et al.* [109], and independently by Marletto and Vedral [110]. The proposed experiment has been referred to as the 'BMV experiment' for short [111], which has since also inspired similar proposals in different platforms [112, 113, 114, 115, 116, 117].

We consider here a slightly different version of the BMV experiment, see Figure 11.1(a). Two mesoscopic particles of masses m_1 and m_2 are placed at distance d from each other. Each particle is then split into a superposition of two positions that are separated by a distance L orthogonally to their initial separation. Based on recent advances in setting up mesoscopic systems in superposition [112, 118, 119, 120, 113, 121, 122], the authors of Ref. [109] suggested as physically relevant quantities $m_1 \approx m_2 \approx 10^{-8}$kg, $d \approx 200^{-}$m, and we can assume $L \gg d$.

In the original proposal [109], the particles are split into superpositions of positions in the same direction as their initial separation, see Figure 11.1(b). They thus have strong gravitational interaction only when the first particle is on the right, while the second particle is on the left. In this current setup the particles interact strongly whenever they are on the same side (left or right). Experimentally, it might be more challenging to setup the system in this symmetric configuration; in particular, one may have to introduce a thin film between the particles to keep the distance constant. We note that, if realisable, a thin film however may have an additional advantage. While still allowing for gravitational interaction to permit, it helps prevent taming interactions due to the van der Waals or the Casimir effects, which has been an obstacle for the original setup [112]; as a result, the distance between the particles can also be reduced to enhance their interaction. Here we analyse this symmetric setup mainly for convenience; the analysis can be easily adapted to the original proposal as well.

Formally, the system can be modelled as a pair of spins, where states $|\uparrow\rangle$ and $|\downarrow\rangle$ can be identified with the particles being on the left and right, respectively. Due to the gravitational interaction between the particles, if both particles are on the same sides ($|\uparrow\uparrow\rangle$ or $|\downarrow\downarrow\rangle$), the energy of the system is $-Gm_1m_2/d$, with G being the gravitational constant. On the other hand, if they are on the opposite sides ($|\uparrow\downarrow\rangle$ or $|\downarrow\uparrow\rangle$), the energy is given by $-Gm_1m_2/\sqrt{L^2+d^2}$. Therefore, upto an irrelevant additive constant, the Hamiltonian can be modelled by

$$H = -\frac{\Delta}{2}\sigma_z \otimes \sigma_z, \tag{11.1}$$

where σ_z is one of the usual Pauli matrices and

$$\Delta = Gm_1m_2\left(\frac{1}{d} - \frac{1}{\sqrt{L^2+d^2}}\right). \tag{11.2}$$

Under the evolution induced by this Hamiltonian, particles that are first given in the superpositions of being left and right, $|+\rangle|+\rangle$, where $|+\rangle = (|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2}$, should evolve into an entangled state. Provided one can preserve the coherence of the system long enough, such entanglement is expected to be observable [109]. In the actual physical setting, each particle carries an additional two-level degree of freedom, which is then correlated with its positions (left or right) during the spliting [109]. The result is that, after merging their superpositions, the entanglement in the particle positions is

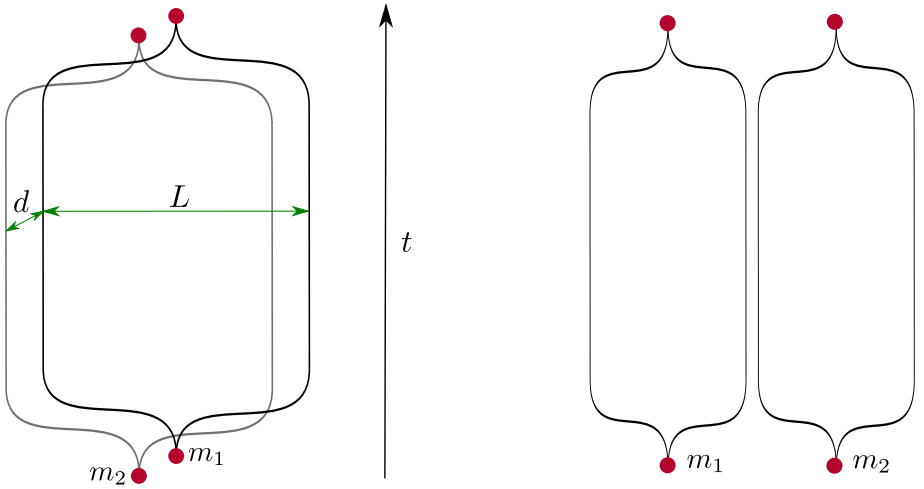


Figure 11.1: The BMV experiment. (a) The symmetric setup: two particles are initially at distance d from each other. Each of the particle is then split into a superposition of two positions at distance L from each other in the direction orthogonal to their initial separation. Due to their gravitational interaction the particles are expected to be entangled over time. (b) The original setup: the particles are split into superpositions in parallel to their initial separation.

eventually transferred to the entanglement between these additional degrees of freedom and can be directly measured.

The entanglement between the particles has been argued to be an evidence that the gravitational field is a quantum mechanical system [109, 110]. While this claim is still a subject of debate [123, 124, 125, 111, 126], we do believe that the ability to entangle particles via their gravitational interaction would greatly advance our understanding of the interface between quantum mechanics and gravity.

Let T be the decoherence time, the authors of Ref. [109] argued that the necessary condition to observe the entanglement is

$$(\Delta T)/\hbar \sim \mathcal{O}(1). \tag{11.3}$$

While this qualitative estimate is plausible, it is still important to analyse the noisy dynamics of the system in detail to pinpoint the precise condition under which entanglement between the particles can be observed. Here we analyse the details of the decoherence dynamics of the system. More importantly, we also consider fluctuations of the experimental parameters. These stochastic fluctuations in setting the parameters

in the proposed experiments imply that one has to average over the obtained entangled states from run to run of the experiment, which results in a reduction of entanglement in the averaged state. While so far this has not been considered, it is also crucial to the experiment, since entanglement can only be verified statistically through multiple runs of the experiment. Our analysis shows that the entanglement is rather robust. More precisely, we show that the entanglement indeed develops as long as $\Delta T/\hbar > 1$ if the fluctuations in setting the experimental parameters can be neglected. Moreover, we show that moderate stochastic errors in the experiment can also be tolerated. We discuss the optimal interaction duration for the particles while they are in the superposition state and find a condition under which entanglement can be detected in a device-independent manner.

11.2 The decoherence dynamics of the system

The superposition of the positions of the particles is suppressed in the long time limit because of the environmental interaction. This is known as the decoherence process, which gives rise to our classical notion of position [47]. While the details of the decoherence process depend on the details of the environment, the system under consideration is sufficiently simple that it can be analysed with some minimal assumptions of the decoherence theory. Indeed, due to decoherence the system decays into a mixed state of positions (and not any other basis), so one can assume that the environment couples only to the position operator of the particles, which is σ_z in this case [47]. The coupling Hamiltonian between one particle and the environment can be written generally as

$$H_D \propto \sigma_z \otimes R, \tag{11.4}$$

where R is an operator acting on the environment. This environmental coupling Hamiltonian Eq. (11.4) is such that if the initial reduced state of the particle is given by a (2×2) density matrix a, it will evolve in a way that its diagonal elements are constant, while its off-diagonal elements decay over time. Technically, this is a valid approximation when the particle is large enough such that recoiling effects due to scattering by the environment is negligible [47]. Assuming an exponential decay of the off-diagonal elements (known as coherence elements) for specificity, the state of the particle state at time t is given by

$$\varrho_1(t) = \begin{pmatrix} a_{11} & a_{12}e^{-\tau} \\ a_{21}e^{-\tau} & a_{22} \end{pmatrix}, \tag{11.5}$$

where we have used the dimensionless time τ, defined by the physical time divided by the decoherence time T, $\tau = t/T$. While this exponential decaying of coherence is the case for the position decoherence due to the enviromental scattering by photons or

air molecules [47], other types of decoherence dynamics can also be considered with minimal adaptation.

For the system of two particles without mutual interaction, we assume that their decoherence are independent from each other. If the system is first given in the state $a \otimes b$, where a and b are (2×2) density matrices of the first and the second particle, respectively, the density matrix of the whole system at time τ is then

$$\begin{pmatrix} a_{11} & a_{12}e^{-\tau} \\ a_{21}e^{-\tau} & a_{22} \end{pmatrix} \otimes \begin{pmatrix} b_{11} & b_{12}e^{-\tau} \\ b_{21}e^{-\tau} & b_{22} \end{pmatrix}. \tag{11.6}$$

By linearity, the two-particle system first given in a (4×4) density matrix c then evolves to

$$\begin{pmatrix} c_{11} & c_{12}e^{-\tau} & c_{13}e^{-t} & c_{14}e^{-2\tau} \\ c_{21}e^{-\tau} & c_{22} & c_{23}e^{-2t} & c_{24}e^{-\tau} \\ c_{31}e^{-\tau} & c_{32}e^{-2t} & c_{33} & c_{34}e^{-\tau} \\ c_{41}e^{-2\tau} & c_{42}e^{-\tau} & c_{43}e^{-\tau} & c_{44} \end{pmatrix}. \tag{11.7}$$

Let us now consider the interaction between the particles via the Hamiltonian Eq. (11.1). Importantly, the system Hamiltonian commutes with the environmental coupling Eq. (11.4), rendering the total dynamics also exactly solvable regardless of the details of the enviroment operator R. Indeed, as we transform to the interaction picture by substituting $\varrho = U(\tau)\varrho_I(\tau)U^\dagger(\tau)$, where $U(\tau) = e^{-iH\tau}$, we find that the interacting density matrix $\varrho_I(\tau)$ follows the dynamics of two independent particles interacting only with the environment given by equation Eq. (11.7). Assuming that at $\tau = 0$ the system is in the state $|+\rangle\langle+| \otimes |+\rangle\langle+|$, we find the density matrix of the system at time τ to be

$$\varrho = \frac{1}{4}\begin{pmatrix} 1 & e^{i\omega\tau-\tau} & e^{i\omega\tau-\tau} & e^{-2\tau} \\ e^{-i\omega\tau-\tau} & 1 & e^{-2\tau} & e^{-i\omega-\tau} \\ e^{-i\omega\tau-\tau} & e^{-2\tau} & 1 & e^{-i\omega\tau-\tau} \\ e^{-2\tau} & e^{i\omega\tau-\tau} & e^{i\omega\tau-\tau} & 1 \end{pmatrix}. \tag{11.8}$$

Recall that we are using the dimensionless time τ, and $\omega = \Delta T/\hbar$ is referred to as the (dimensionless) coupling of the system.

To analyse the entanglement dynamics in the density matrix Eq. (11.8), we use the positive partial transposition (PPT) criterion [57, 58]. The smallest eigenvalue of the partial transposition of ϱ is found to be

$$\lambda = \frac{1}{2}e^{-\tau}(\sinh\tau - |\sin\omega\tau|). \tag{11.9}$$

According to the PPT criterion, the two particles are entangled if and only if $\lambda < 0$. Figure 11.2 (a) illustrates several different evolutions of λ for different couplings ω. For $\omega < 1$, λ is positive and no entanglement develops. For $\omega > 1$, λ becomes negative for certain times, indicating that entanglement develops between the particles. This sharp

transition can be easily confirmed by analysing equation Eq. (11.9). For very large coupling parameters ω, the particles can undergo entangled-disentangled oscillations. Obviously, the particles share the highest amount of entanglement during the first phase of entanglement, where the effect of decoherence is still weak. To estimate the optimal duration τ_0 of the experiment, we find the first minimum of λ. By considering the derivative of equation Eq. (11.9), an equation for τ_0 can be found, namely

$$e^{-\tau_0} + \sin\omega\tau_0 - \omega\cos\omega\tau_0 = 0, \tag{11.10}$$

which is to be solved for the first positive time τ_0. This yields the optimal duration for the experiment as a function of the coupling parameter ω. A plot of this function is presented in Figure 11.2 (b). For $1 < \omega < 1.8$, decoherence is strong and the optimal time quickly increases with respect to the coupling strength ω toward a maximum at $\tau_0 \approx 0.4$. For $\omega > 1.8$, the internal evolution of the system dominates in the short time dynamics and the optimal time is similar to the time where the system achieves a Bell state when we ignore decoherence, $\tau_0 \approx \pi/(2\omega)$, which decreases as ω increases.

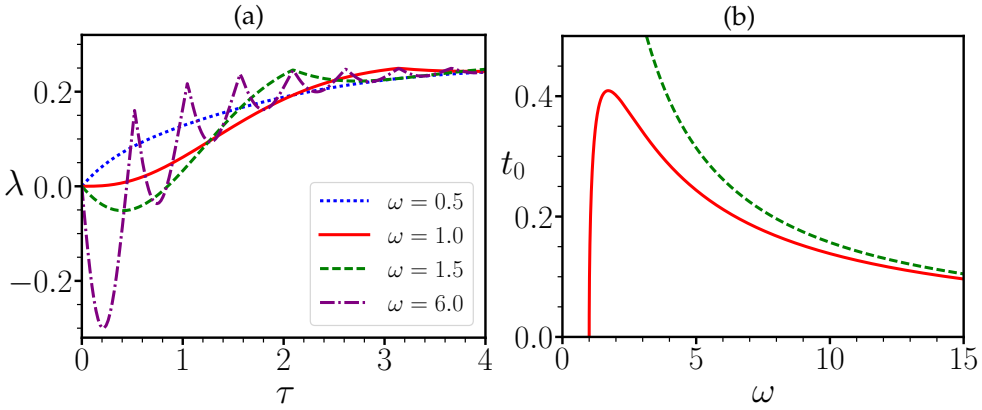


Figure 11.2: (a) The smallest eigenvalue λ of the partial transposition of ϱ as a function of time. (b) The optimal time where the two particles are most entangled. The dashed-line indicates the asymptotic optimal time in the limit of very strong couplings ω and no decoherence, given by $\pi/(2\omega)$.

11.3 Stochastic fluctuations in preparing the experiment

Let us consider now the fluctuations of the parameters of the experiment. If the separation between the two positions of a particle L is large in comparison to the typical wave length of the electromagnetic environment, we can assume that the decoherence time T is not sensitive to this separation [47]. Moreover, if L is much larger than d, fluctuations in L have only marginal effects on Δ. Thus, only fluctuations in two quantities are important: (a) fluctuations in the minimal distance between the two particles

d, which induce fluctuations in the coupling ω and (b) fluctuations in the interaction duration τ.

To model the fluctuations, one can simply replace the deterministic values of ω and τ by two gaussian random variables $\omega + \xi_\omega s_\omega$ and $\tau + \xi_\tau s_\tau$, where s_ω and s_τ are their standard deviations, and ξ_ω and ξ_τ are two standard gaussian random variables. The state of the system averaged over all runs of the experiment would then be

$$\bar{\varrho} = \langle \varrho(\tau) \rangle_{\xi_\tau, \xi_\omega}. \tag{11.11}$$

Assuming that the fluctuations are small, $\tau \gg s_\tau$, $\omega \gg s_\omega$, one can expand their contributions in the phase and the damping terms in ϱ to the first order in s_ω and s_τ. Averaging over the gaussian fluctuations in time τ and in coupling parameter ω then yields the state

$$\bar{\varrho} = \frac{1}{4} \begin{pmatrix} 1 & a & a & b \\ \bar{a} & 1 & b & \bar{a} \\ \bar{a} & b & 1 & \bar{a} \\ b & a & a & 1 \end{pmatrix}, \tag{11.12}$$

with $a = e^{i\omega\tau - \tau} e^{-\frac{1}{2}s_\omega^2 \tau^2 + \frac{1}{2}s_\tau^2 (i\omega - 1)^2}$ and $b = e^{-2\tau} e^{2s_\tau^2}$.

To analyse the entanglement in the density operator, we again compute the smallest eigenvalue $\bar{\lambda}$ of the partial transposition of $\bar{\varrho}$, which is

$$\frac{1}{2} e^{-(\tau - s_\tau^2)} [\sinh(\tau - s_\tau^2) - e^{-\frac{s_\tau^2}{2}(1+\omega^2) - \frac{s_\omega^2}{2}\tau^2} |\sin \omega(\tau - s_\tau^2)|]. \tag{11.13}$$

By sending s_ω and s_τ to zero, we can easily recover equation Eq. (11.9). Note that one should not extrapolate this formula to $\tau < s_\tau^2$, since in this regime the fluctuations in time extrapolate the decoherence dynamics backward into negative time, which is not physical. One then finds that the entanglement between the particles develops after $\tau > s_\tau^2$ if and only if

$$s_\tau^2 (1 + s_\tau^2 s_\omega^2 + \omega^2) < 2 \ln \omega. \tag{11.14}$$

To be consistent with the approximation, the higher order term $s_\tau^2 s_\omega^2$ should in fact be ignored. We thus obtain the condition

$$s_\tau^2 < \frac{2 \ln \omega}{1 + \omega^2}. \tag{11.15}$$

Remarkably, s_ω is absent in this condition. This means that if $\omega > 1$, and if the duration of interaction is well-controlled ($s_\tau \approx 0$), the entanglement persists despite arbitrary fluctuations in ω. On the other hand, equation Eq. (11.15) does pose a bound on the maximal standard deviation s_τ allowed, which is plotted in Figure 11.3. Interestingly, this indicates that if the interaction strength is strong, one has to control the time more accurately; on the other hand, in the intermediate regime, s_τ can vary to a large extent.

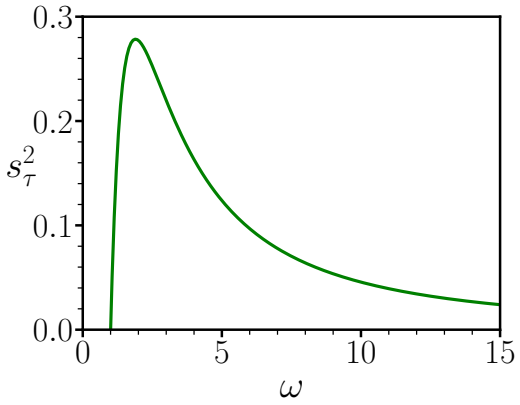


Figure 11.3: Maximum fluctuation allowed in the interaction duration s_τ^2 such that the entanglement can still be observed as a function of the coupling strength ω.

If the interaction time can be precisely controlled by an atomic clock, this accuracy can be easily achieved. In reality, the accuracy of the interaction duraction in the actual experiment could be much less than the scale of atomic clocks due to various difficulties in setting up the superposition configuration for each particle. In particular, it is known [127, 128, 129] that such a process should not be too fast, otherwise gravitational radiations would interfere with the system causing further decoherence effects. Yet, we expect that even in this case condition Eq. (11.15) can be easily achieved in reality.

11.4 Violation of the CHSH inequality

While the entanglement in the density operator can be demonstrated by state tomography or certain entanglement witnesses [48], it is generally desirable to demonstrate the entanglement in a device-independent way [72]. This can be done by demonstrating a violation of the so-called Clauser-Horne-Shimony-Holt (CHSH) inequality [72]. Suppose two parties, Alice and Bob, each of whom owns one particle of the pair. Consider a situation where Alice performs either one of two measurements A_1, A_2 on her particle while Bob performs either measurements B_1 or B_2 on his particle. Each measurement has only two outcomes ± 1. If one constrains that the system satifies the so-called assumption of local realism [61, 2], then it is easy to show that

$$|\langle A_1 B_1\rangle + \langle A_1 B_2\rangle + \langle A_2 B_1\rangle - \langle A_2 B_2\rangle| \leq 2. \tag{11.16}$$

It has been repeatedly demonstrated in experiments that the CHSH inequality is violated in quantum mechanics [72]. This shows that quantum mechanics is not compatible with the assumption of local realism, on which the CHSH inequality Eq. (11.16) is

based. What is relevant to us in the current context is that in order to violate the CHSH inequality Eq. (11.16) the state must be entangled. Notice that in order to demonstrate the violation of the CHSH inequality, we only need the statistics of the measurements A_1, A_2 and B_1, B_2 in the experiment. The details how such measurements are setup or how they are described mathematically are irrelevant [72]. In that sense, one can prove entanglement between particles in a device-independent way.

For simplicity we ignore the fluctuations in the experimental parameters in this section so that the density operator is of the simple form as in equation Eq. (11.8). To see whether ϱ violates the CHSH inequality for certain measurement settings, we make use of the criterion described in Ref. [130]. To this end, we consider the correlation matrix $T_{ij} = \mathrm{Tr}\left(\varrho\sigma_i \otimes \sigma_j\right)$, where σ_i and σ_j for $1 \leq i, j \leq 3$ are the Pauli matrices. It turns out that the correlation matrix T for ϱ is degenerate with singular values $s_1 = e^{-2\tau}$ and $s_2 = s_3 = e^{-\tau}|\sin\omega\tau|$. Then according to [130], the state ϱ violates the CHSH inequality if and only if either $s_1^2 + s_2^2 > 1$ or $2s_2^2 > 1$. Solving these two inequalities numerically, we find that the system can violate the CHSH inequality if and only if

$$\omega > 4.19135. \tag{11.17}$$

While this bound is significantly larger than the threshold of the coupling for the systems to be entangled over time ($\omega > 1$), it is still in the same order of magnitude. Thus once one can prove the entanglement of the particles, one is also close to proving it in a device-independent manner.

11.5 Conclusion

In this work we have analysed the entanglement dynamics of the two particles in the BMV experiment in detail. We showed that the system entangles as long as the coupling between the particles is strong, $\Delta T/\hbar > 1$, and the parameters are setup precisely. Fluctuations in the parameters that arise from setting up the experiment from run to run were then considered. The entanglement turns out robust against the decoherence for some time and also against stochastic fluctuations. Moreover, we discuss the optimal duration of the gravitational interaction while the particles are in a superposition state. Also, we identify a condition under which one can detect entanglement in a device-independent manner using the CHSH inequality. We would like to mention that recently a similar detailed analysis of the entanglement dynamics has been made for another setup of the experiment [117]. We hope that together these analyses provide useful inputs for a realisation of the experiment in the near future.

Summary and Outlook

We introduced generalizations of a Bell inequality as Bell inequalities that perform at least as well in any task as the Bell inequality they generalize. We further provided a method, called the cone-projection technique, that is helpful in finding such generalizations of a Bell inequality.

Using the cone-projection technique, we were able to find 3050 classes of generalizations of the I3322 inequality to three parties, 476 classes of Bell inequalities that simultaneously generalize the CHSH inequality as well as the I3322 inequality, 13 classes of Bell inequalities that generalize the I4422 inequality and 23 classes of Bell inequalities that generalize the Guess-Your-Neighbors-Input inequality first found in [75]. For all inequalities, we applied an extensive numerical analysis, providing upper bounds to their quantum violations both for qubit and qutrit systems as well as upper bounds using the NPA hierarchy. Taking a closer look at the simplest generalizations of the I3322 inequality, we found that these Bell inequalities are strong in the following sense: There are three-qubit states with separable two-body marginals that can violate the Bell inequalities for suitably chosen settings. Moreover, there are three-qubit states that can violate our inequalities but no Bell inequality with only two measurement settings per party. At the same time, these Bell inequalities are by construction as powerful as the I3322 inequality in detecting two-body nonlocality.

Furthermore, we presented symmetric Bell-like inequalities for various hybrid models as well as generalizations of Svetlichny's inequality to three settings per party. Again, we used the cone-projection technique to achieve this. Moreover, we constructed a family of inequalities for a special class of hybrid models.

Constructing and analyzing the properties of Bell inequalities was an important aspect of our research. To this end, we created the Python library Bellpy, that aims at making these tasks as easy and efficient as possible.

Beyond our research on Bell inequalities, we have explored the possibility of AME-Werner states and found that these only exist for two-qubit and three-qutrit systems. In joint work with Chau Nguyen, we have furthermore scrutinized recent suggestions for experiments to detect quantum phenomena in gravitationally interacting systems with regard to entanglement dynamics and the device independent detection thereof.

For future research there are several directions in which our work can be extended. First, the cone-projection technique is general enough to not be restricted to applications in the field of nonlocality. In fact, it is useful in any situation where facets of

polytopes are of interest, which obey affine equality constraints. In the following, we describe some directions of research, all of which rely on the cone-projection technique. We defined generalizations of Bell inequalities to more parties, more settings per party and more outcomes per setting. However, in our research we have almost exclusively explored generalizations to more parties. Concerning generalizations to more settings, we have only considered the Svetlichny inequality. Moreover, we have so far only considered scenarios with dichotomic measurements and have not investigated generalizations of Bell inequaities to more outcomes per setting at all. We believe however, that this is a promising approach to finding strong Bell inequalities. Taking a broader perspective, it would be interesting to apply our techniques to causal models and find generalizations of contextuality inequalities [64, 131] or recently discussed inequalities for testing certain views of quantum mechanics [132]. Moreover, it could be interesting to extend our approach beyond the scenario of linear optimization: While the local polytope in Bell scenarios can be characterized by the optimization method of linear programming, other forms of quantum correlations, such as quantum steering, are naturally described in terms of convex optimization and semidefinite programs [133, 134]. Thus, a generalization of our methods to this type of problems could give more insight into various problems in information processing.

Acknowledgements

First of all, I thank my PhD supervisor Professor Otfried Gühne. He sparked my interest in Bell inequalities when he introduced me to the foundations of quantum mechanics and has ever since supported my interest and research in this field.

I also thank Chau Nguyen, who was my office mate for most of my time working in Siegen, for many interesting and joyful discussions, one of which led to our paper.

I would also like to thank all of the current and former members of the TQO group I was lucky to know and learn from. My special thanks go to Yuanyuan Mao, Jonathan Steinberg, Chau Nguyen, Tristan Kraft, Pascal Höhn, Leonardo Santos, and Lina Vandré, who have proof-read large parts of this thesis.

I am deeply grateful to my wife, parents, brother, and grandparents for their care and love.

I acknowledge the financial support of the House of Young Talents Siegen.

List of Publications

(A) F. Bernards and O. Gühne.
Generalizing Optimal Bell Inequalities.
Phys. Rev. Lett. **125**(20), 200401 (2020).

(B) F. Bernards and O. Gühne.
Finding optimal Bell inequalities using the cone-projection technique.
Phys. Rev. A **104**(1), 012206 (2021).

(C) F. Bernards and O. Gühne.
Characterizing nonlocality depth with the cone-projection technique.
arXiv:2205.04250 (2022).

(D) F. Bernards.
Bellpy: A toolbox for Bell nonlocality in python.
Manuscript in preparation.

(E) F. Bernards.
Absolutely maximally entangled Werner states only exist in two-qubit and three-qutrit systems.
Manuscript in preparation.

(F) H. C. Nguyen and F. Bernards.
Entanglement dynamics of two mesoscopic objects with gravitational interaction.
The European Physical Journal D **74**(4), 69.

(X) F. Bernards, M. Kleinmann, O. Gühne, and M. Paternostro.
Daemonic Ergotropy: Generalised Measurements and Multipartite Settings.
Entropy 2019, **21**(8), 771. This publication comprises the results of the author's master's thesis.

A Appendix

A.1 Constructing LHV models from symmetric quasi-extensions

We first recall the result that a state ϱ_{ABC} does not violate any Bell inequality with two settings on B and C and an arbitrary number N of measurement settings on A, if there exists a symmetric extension, that is a positive semidefinite operator $H_{ABB'CC'} = H_{AB'BCC'}$ that fulfills $\mathrm{Tr}_{B'C'}\left(H_{ABB'CC'}\right) = \varrho_{ABC}$. The main idea stems from [100]. We focus on the case that is relevant for our paper. For convenience, we denote the symmetric extension simply as H, whenever this is possible without causing confusion.

The argument that links symmetric extensions to local hidden variable models goes as follows: Firts, if a symmetric extension exists, we can write, for any fixed measurement setting i of Alice, a joint probability for both measurements on Bob and Charlie via

$$p(a_i, b_1, b_2, c_1, c_2) = \mathrm{Tr}\left(HE_{a_i} \otimes E_{b_1} \otimes E_{b_2} \otimes E_{c_1} \otimes E_{c_2}\right) \tag{A.1}$$

where E_{a_i} is the effect that corresponds to the outcome a_i etc. Note that the non-negativity of the operator H is not required, it is sufficient if the operator is an entanglement witness, i.e. non-negative on product states. Such an operator is called symmetric quasi-extension [100].

Then, we can define a for all measuremements a joint probability distribution via [135]

$$\begin{aligned} &p(a_1, \ldots, a_N, b_1, b_2, c_1, c_2) \\ &\quad = \frac{p(a_1, b_1, b_2, c_1, c_2) \cdots p(a_N, b_1, b_2, c_1, c_2)}{p(b_1, b_2, c_1, c_2)^{N-1}}. \end{aligned} \tag{A.2}$$

Finally, it is well established that if such a joint probability distribution exists, then a LHV model exists and no Bell inequality can be violated [17, 136].

In our work, we make use of this connection to find a three-qubit state ϱ_{ABC} that violates one of the Bell inequalities F_i in Eqs. (6.15)-(6.17), but which is not violating any three-partite Bell inequality with two settings per party and, in addition, it is not violating any bipartite Bell inequality (such as I3322) with any of its two-body marginals. We achieve this by demanding that the state ϱ_{ABC} possesses a symmetric

extension and separable two-body marginals. Since the marginals are two-qubit states, the separability condition can be implemented using the criterion of the positivity of the partial transpose [58, 57].

In this way, all of the constraints are semidefinite constraints. Maximizing the violation of a Bell inequality under these constraints is a semidefinite program, if the measurement settings are given. As initially no measurement settings are given, we pick random measurement settings for the three parties before optimizing over the state. Then, keeping the state fixed, each optimization over the measurement settings of a single party is again a semidefinite programm. We then alternate between these two steps – the optimization of the state on the one hand and the optimization of the measurement settings on the other – in a seesaw algorithm. We solve the semidefinite programs with Mosek [137] through Picos [138].

Typically, after 50 iterations a convergence is reached. In this way, we find a state $\varrho^{(1)}_{ABC}$ with $F_1 \approx -0.065$, a state $\varrho^{(2)}_{ABC}$ with $F_2 \approx -0.043$ and a state $\varrho^{(3)}_{ABC}$ with $F_3 \approx -0.063$. The symmetric extensions and measurement settings that lead to these violations can be found as part of the supplemental material of our paper [A].

A.2 Details on finding lower bounds for qubit and qutrit violations

We provide lower bounds for the violations of the Bell inequalities achievable with qubits and qutrits, respectively. These lower bounds are violations achieved by specific states and measurements.

Finding good lower bounds therefore involves two optimizations: One optimization over the possible states and another optimization over the measurement settings. The objective – that is the expectation value of the Bell operator for the given state – is however not linear, if we optimize over all measurements at once. We therefore break the optimization over the measurement settings down further to optimizing over them party by party, thereby rendering the objective linear. We represent each measurement by a positive operator-valued measure (POVM). Such a optimization problem that features an objective that is linear in its arguments which are positive-semidefinite operators is called a semidefinite program (SDP) and can be solved efficiently. Likewise, the optimization over the state-space while keeping the measurement settings constant is also an SDP. However, the optimal state for given settings can more easily be obtained by calculating the eigenstate corresponding to the largest eigenvalue of the Bell operator.

This prompts a seesaw-algorithm: After initializing the measurement settings to (pseudo-)random values, we alternate between updating the quantum state and updating the measurement settings one party at a time while keeping the quantum state

constant. In this way, the value of the objective is guaranteed to increase monotonically and will eventually converge, if the objective is bounded. However, it oftentimes only converges to a local minimum. This is a problem we cannot avoid, but we can aim for a rather good local optimum by running the seesaw-algorithm many times, starting at a different initial state every time. To save computational resources, we do not wait until the objective values reach high precision. Instead, we stop after a few iterations, select the most promising instance of the optimization and only continue the seesaw-algorithm for this selected instance until the improvements of the objective value fall below some threshold.

A.3 I4422 generalizations

We find 13 classes of Bell inequalities that are symmetric under party permutations, generalize I4422 and further exhibit the symmetry $A_1 \leftrightarrow A_2, B_1 \leftrightarrow B_2, C_4 \rightarrow -C_4$. In the notation for symmetric Bell inequalities that we introduced in the main text, they

read

$$\begin{aligned} &-(100)-2(110)-(200)-2(210)-2(220)+(300)-(310)\\ &-(320)+(330)+(331)+(332)-3(333)+2(441)-2(442)\leq 15 \end{aligned} \tag{A.3}$$

$$\begin{aligned} &-(100)-2(110)-(111)-(200)-2(210)+(211)\\ &-2(220)-(221)+(222)+(300)-(310)-(320)\\ &+(330)+(331)+(332)-3(333)+2(441)-2(442)\leq 15 \end{aligned} \tag{A.4}$$

$$\begin{aligned} &-2(100)-2(110)-2(200)-2(210)-2(220)-(300)-2(310)\\ &-2(320)+(330)-(333)+2(441)-2(442)\leq 19 \end{aligned} \tag{A.5}$$

$$\begin{aligned} &-2(100)-2(110)-(111)-2(200)-2(210)+(211)\\ &-2(220)-(221)+(222)-(300)-2(310)-2(320)\\ &+(330)-(333)+2(441)-2(442)\leq 19 \end{aligned} \tag{A.6}$$

$$\begin{aligned} &+3(100)-3(110)+2(111)+3(200)-3(210)\\ &-3(220)+2(221)+(300)-(310)+(311)\\ &-(320)+(321)+(322)+(330)-(331)\\ &-(332)+(333)-4(441)+4(442)\leq 23 \end{aligned} \tag{A.7}$$

$$\begin{aligned} &+3(100)+(110)-3(111)+3(200)+(210)+(211)\\ &+(220)\\ &-3(221)+(222)+4(300)\\ &+3(310)-3(311)\\ &+3(320)-3(321)\\ &-3(322)-6(330)-(331)-(332)\\ &+12(333)+4(441)-4(442)\leq 38 \end{aligned} \tag{A.8}$$

$$\begin{aligned} &+3(100)+(110)-(111)+3(200)+(210)-(211)\\ &+(220)-(221)-\\ &(222)+4(300)+3(310)-3(311)\\ &+3(320)-3(321)-3(322)\\ &-6(330)-(331)-(332)\\ &+12(333)+4(441)-4(442)\leq 38 \end{aligned} \tag{A.9}$$

$$\begin{aligned}+2(100)-5(110)+2(200)-5(210)-5(220)+(300)\\ -4(310)+3(311)-4(320)+3(321)+3(322)+(330)\\ -2(331)-2(332)-3(333)-8(441)+8(442)\leq 51\end{aligned} \tag{A.10}$$

$$\begin{aligned}+2(100)-5(110)-4(111)+2(200)-5(210)+4(211)\\ -5(220)-4(221)+4(222)+(300)\\ -4(310)+3(311)-4(320)+3(321)+3(322)+(330)\\ -2(331)-2(332)-3(333)+8(441)-8(442)\leq 51\end{aligned} \tag{A.11}$$

$$\begin{aligned}+(100)-5(110)+(200)-5(210)-5(220)-(300)\\ -5(310)+3(311)-5(320)+3(321)+3(322)+(330)\\ -3(331)-3(332)-(333)-8(441)+8(442)\leq 55\end{aligned} \tag{A.12}$$

$$\begin{aligned}+(100)-5(110)-4(111)+(200)-5(210)+4(211)\\ -5(220)-4(221)+4(222)-(300)-5(310)+3(311)\\ -5(320)+3(321)+3(322)+(330)-3(331)-3(332)\\ -(333)+8(441)-8(442)\leq 55\end{aligned} \tag{A.13}$$

$$\begin{aligned}+7(100)+(110)-(111)+7(200)+(210)-(211)\\ +(220)-(221)-(222)+6(300)+7(310)-7(311)\\ +7(320)-7(321)-7(322)-12(330)-(331)-(332)\\ +22(333)-8(441)+8(442)\leq 76\end{aligned} \tag{A.14}$$

$$\begin{aligned}+7(100)+(110)-5(111)+7(200)+(210)+3(211)\\ +(220)-5(221)+3(222)+6(300)+7(310)-7(311)\\ +7(320)-7(321)-7(322)-12(330)-(331)-(332)\\ +22(333)+8(441)-8(442)\leq 76.\end{aligned} \tag{A.15}$$

A.4 Classical bound for family of Bell inequalities

This appendix is concerned with the family of Bell inequalities F_n presented in chapter 8.

We show that with any (k, m) model, there exists a behavior such that

$$F_n = \sum_{\ell=1}^{n} (-1)^{1+\lceil \frac{\ell}{2} \rceil}\, \ell\, (\underbrace{1\ldots1}_{n-\ell}, \underbrace{2\ldots2}_{\ell}) = n2^{n-2}. \tag{A.16}$$

Further, we show that for $m = 2$, this bound is a valid upper bound of F_n. First note that the symmetric correlation

$$(\underbrace{1\dots1}_{n-\ell},\underbrace{2\dots2}_{\ell}) \tag{A.17}$$

comprises $\binom{n}{\ell}$ terms, each of which takes values ±1. Consequently,

$$-\binom{n}{\ell} \leq (\underbrace{1\dots1}_{n-\ell},\underbrace{2\dots2}_{\ell}) \leq \binom{n}{\ell}. \tag{A.18}$$

Since the expression F_n is linear in the symmetric correlation terms and its maximum will therefore be achieved for

$$(\underbrace{1\dots1}_{n-\ell},\underbrace{2\dots2}_{\ell}) = \pm\binom{n}{\ell}. \tag{A.19}$$

We can thus treat the symmetric correlation terms as binary variables. For convenience, we define the variables

$$\gamma_i^k = (\underbrace{1\dots1}_{k-i},\underbrace{2\dots2}_{i})\binom{k}{i}^{-1}, \tag{A.20}$$

which are normalized such that they take values ±1. With this, we can write F_n as

$$F_n = \sum_{\ell=1}^{n}(-1)^{1+\lceil\frac{\ell}{2}\rceil}\,\ell\,\gamma_\ell^n\binom{n}{\ell}. \tag{A.21}$$

However, the variables γ_ℓ^n cannot be chosen independently, since they have to respect the (k,m) model under consideration. This condition is met, if we consider behaviors that stem from a local hidden variable model between the first k parties and the last m parties. For this model, we have

$$\gamma_\ell^n = \gamma_i^k\gamma_j^m \tag{A.22}$$

with

$$i + j = \ell. \tag{A.23}$$

With this, we can rewrite

$$F_n = \sum_{\ell=1}^{n}(-1)^{1+\lceil\frac{i+j}{2}\rceil}\,(i+j)\,\gamma_i^k\gamma_j^m\binom{k}{i}\binom{m}{j}. \tag{A.24}$$

Setting

$$\gamma_i^k = (-1)^{\lfloor\frac{i}{2}\rfloor} \tag{A.25}$$

$$\gamma_j^m = (-1)^{\lfloor\frac{j}{2}\rfloor} \tag{A.26}$$

yields

$$F_n = \sum_{i=0}^{k}\sum_{j=0}^{m}(-1)^{(i+1)(j+1)}\,(i+j)\binom{k}{i}\binom{m}{j} \tag{A.27}$$

$$= n\,2^{n-2}. \tag{A.28}$$

We now show that for $m = 2$, this value is a valid upper bound for F_n. For convenience, we define the matrix M with elements

$$M_{ij} = (-1)^{(i+1)(j+1)}\,(i+j)\binom{k}{i}\binom{2}{j}. \tag{A.29}$$

Note, that a different choice for $\gamma_i^k(\gamma_j^m)$ corresponds to flipping the signs of the entries in the i-th row (j-th column) of M. Hence, showing that there does not exist a subset of rows and columns, such that multiplying these columns and rows with -1 yields a larger sum $\sum_{ij} M_{ij}$ proves the claim. Since M only has three columns, we focus on the columns. For any choice of rows and columns of M, either zero, one, two, or all columns of M would be affected. However, multiplying all columns and rows with -1 leaves M invariant and therefore we only need to consider two cases: Either (1) non of the columns is affected by the sign-flip operation or (2) exactly one column is affected by the sign-flip operation. In case (1) one cannot reach a value higher than $n2^{n-2}$ since all rows have a non-negative value. For the second case, note that

$$M_{i1} = |M_{i0}| + |M_{i2}|. \tag{A.30}$$

Hence, multiplying the column $j = 1$ with -1 cannot be compensated for any choice of rows. Further, multiplying a column with $j \neq 1$ with -1 still leaves all rows non-negative. Since the sum of the entries in the columns $j = 0$ and $j = 2$ vanishes, this means, that the value $F_n = n2^{n-2}$ cannot be exceeded in an $(n-2, 2)$ model.

A.5 Optimal states for Bell inequality family

This appendix is concerned with the family of Bell inequalities F_n presented in chapter 8.

Below, we list the quantum states that lead to a maximal violation of the respective Bell inequality from the F_n family. For convenience, we define

$$(X...Z) = X \otimes ... \otimes Z + \text{permutations}, \tag{A.31}$$

where 'permutations' accounts for all party permutations of the first term and no term is present twice in the sum, that is $(XXX) = X \otimes X \otimes X$. The symbols $\mathbb{1}, X, Y, Z$ are defined as

$$\mathbb{1} = \frac{1}{2}\mathbb{1}_2,\ X = \frac{1}{2}\sigma_X,\ Y = \frac{1}{2}\sigma_Y,\ Z = \frac{1}{2}\sigma_Z. \tag{A.32}$$

The optimal states are the pure states

$$\varrho_3 = (111) + (1YY) - \frac{1}{\sqrt{2}}(XXX) + \frac{1}{\sqrt{2}}(XXZ) + \frac{1}{\sqrt{2}}(XZZ) - \frac{1}{\sqrt{2}}(ZZZ) \tag{A.33}$$

$$\begin{aligned}\varrho_4 =&(1111) + (11YY) - \frac{1}{\sqrt{2}}(XXXX) + \frac{1}{\sqrt{2}}(XXXZ) + \frac{1}{\sqrt{2}}(XXZZ)\\ &- \frac{1}{\sqrt{2}}(XZZZ) + (YYYY) - \frac{1}{\sqrt{2}}(ZZZZ)\end{aligned} \tag{A.34}$$

$$\begin{aligned}\varrho_5 =&(11111) + (111YY) + (1YYYY) - \frac{1}{\sqrt{2}}(XXXXX) + \frac{1}{\sqrt{2}}(XXXXZ)\\ &+ \frac{1}{\sqrt{2}}(XXXZZ) - \frac{1}{\sqrt{2}}(XXZZZ) - \frac{1}{\sqrt{2}}(XZZZZ) + \frac{1}{\sqrt{2}}(ZZZZZ)\end{aligned} \tag{A.35}$$

$$\begin{aligned}\varrho_6 =&(111111) + (1111YY) + (11YYYY) - \frac{1}{\sqrt{2}}(XXXXXX) + \frac{1}{\sqrt{2}}(XXXXXZ)\\ &+ \frac{1}{\sqrt{2}}(XXXXZZ) - \frac{1}{\sqrt{2}}(XXXZZZ) - \frac{1}{\sqrt{2}}(XXZZZZ) + \frac{1}{\sqrt{2}}(XZZZZZ)\\ &+ (YYYYYY) + \frac{1}{\sqrt{2}}(ZZZZZZ)\end{aligned} \tag{A.36}$$

$$\begin{aligned}\varrho_7 =&(1111111) + (11111YY) + (111YYYY) + (1YYYYYY) - \frac{1}{\sqrt{2}}(XXXXXXX)\\ &+ \frac{1}{\sqrt{2}}(XXXXXXZ) + \frac{1}{\sqrt{2}}(XXXXXZZ) - \frac{1}{\sqrt{2}}(XXXXZZZ)\\ &- \frac{1}{\sqrt{2}}(XXXZZZZ) + \frac{1}{\sqrt{2}}(XXZZZZZ) + \frac{1}{\sqrt{2}}(XZZZZZZ)\\ &- \frac{1}{\sqrt{2}}(ZZZZZZZ)\end{aligned} \tag{A.37}$$

$$\begin{aligned}\varrho_8 =&(11111111) + (111111YY) + (1111YYYY) + (11YYYYYY)\\ &- \frac{1}{\sqrt{2}}(XXXXXXXX) + \frac{1}{\sqrt{2}}(XXXXXXXZ) + \frac{1}{\sqrt{2}}(XXXXXXZZ)\\ &- \frac{1}{\sqrt{2}}(XXXXXZZZ) - \frac{1}{\sqrt{2}}(XXXXZZZZ) + \frac{1}{\sqrt{2}}(XXXZZZZZ)\\ &+ \frac{1}{\sqrt{2}}(XXZZZZZZ) - \frac{1}{\sqrt{2}}(XZZZZZZZ) + (YYYYYYYY)\\ &- \frac{1}{\sqrt{2}}(ZZZZZZZZ)\end{aligned} \tag{A.38}$$

This can be generalized to

$$\varrho_n = \sum_i (1^{n-2i}Y^{2i}) + \frac{1}{\sqrt{2}}\sum_l (-1)^l[(X^{2l}Z^{n-2l}) - (X^{2l+1}Z^{n-2l-1})], \tag{A.39}$$

which is equivalent to

$$\equiv \sum_i (1^{n-2i}Z^{2i}) + \frac{1}{\sqrt{2}}\sum_l (-1)^l[(X^{n-2l}Y^{2l}) - (X^{n-2l-1}Y^{2l+1})] \tag{A.40}$$

under the local unitary transformation

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}. \tag{A.41}$$

For comparison, the standard GHZ state written in the z-basis reads

$$\varrho_{ghz,n} = \sum_i (1^{n-2i} Z^{2i}) + \sum_l (-1)^l (X^{n-2l} Y^{2l}). \tag{A.42}$$

Numerically, we find that the optimal states are equivalent under local unitary transformations to GHZ states.

A.6 Convexity of the maximal quantum value of a Bell inequality

Consider a family of quantum states

$$\varrho(p) = p\varrho(0) + (1-p)\varrho(1) \tag{A.43}$$

with $p \in [0,1]$. Further consider a Bell inequality with Bell operator B that depends on the choice of measurement settings. We can then define a function that maps p to the maximal quantum value of the Bell inequality for that p as

$$v : p \mapsto \max_M \mathrm{Tr}\left(\varrho(p)B\right). \tag{A.44}$$

We will now show that v is a convex function. Let $p_0, p_m, p_1 \in [0,1]$ and $p_m = \lambda p_0 + (1-\lambda)p_1$ be a convex combination of p_0, p_1 for some $\lambda \in [0,1]$. Further let B_m be the Bell operator if the settings are chosen to maximize the quantum value, that is $v(p_m) = \mathrm{Tr}\left(\varrho(p_m)B_m\right)$. In this point, the maximal quantum value is equal to the value of the affine function $v_m : p \mapsto \mathrm{Tr}\left(\varrho(p)B_m\right)$. In general, $v_m(p)$ is smaller or equal than $v(p)$. This holds in particular for p_0 and p_1. This implies

$$\begin{aligned} \lambda v(p_0) + (1-\lambda)v(p_1) &\geq \lambda v_m(p_0) + (1-\lambda)v_m(p_1) \\ &= \lambda \mathrm{Tr}\left(\varrho(p_0)B_m\right) + (1-\lambda)\mathrm{Tr}\left(\varrho(p_1)B_m\right) \\ &= \mathrm{Tr}\left(\varrho(p_m)B_m\right) \\ &= v(p_m). \end{aligned} \tag{A.45}$$

Therefore, v is convex.

Bibliography

[1] V. Scarani. *Bell Nonlocality*. Oxford University Press (2019).

[2] J. S. Bell. *On the Einstein Podolsky Rosen paradox*. Physics Physique Fizika **1**, 3 195 (1964).

[3] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne et al. *Strong Loophole-Free Test of Local Realism*. Phys. Rev. Lett. **115**, 25 250402 (2015).

[4] B. Hensen, H. Bernien, A. Dréau, A. Reiserer, N. Kalb et al. *Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres*. Nature **526** 682 (2015).

[5] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer et al. *Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons*. Phys. Rev. Lett. **115**, 25 250401 (2015).

[6] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf. *Nonlocality and communication complexity*. Rev. Mod. Phys. **82**, 1 665 (2010).

[7] C. Brukner, M. Żukowski, J.-W. Pan, and A. Zeilinger. *Bell's Inequalities and Quantum Communication Complexity*. Phys. Rev. Lett. **92**, 12 127901 (2004).

[8] A. Tavakoli, M. Żukowski, and C. Brukner. *Does violation of a Bell inequality always imply quantum advantage in a communication complexity problem?* (2019).

[9] M. L. Almeida, J.-D. Bancal, N. Brunner, A. Acín, N. Gisin et al. *Guess Your Neighbor's Input: A Multipartite Nonlocal Game with No Quantum Advantage*. Phys. Rev. Lett. **104**, 23 230404 (2010).

[10] T. Holz, H. Kampermann, and D. Bruß. *Genuine multipartite Bell inequality for device-independent conference key agreement*. Phys. Rev. Research **2** 023251 (2020).

[11] I. Šupić and J. Bowles. *Self-testing of quantum systems: a review*. Quantum **4** 337 (2020).

[12] G. M. Ziegler. *Lectures on Polytopes: Updated Seventh Printing of the First Edition*. Graduate Texts in Mathematics 152. Springer-Verlag New York (1995).

[13] A. Paffenholz. *Polyhedral Geometry and Linear Optimization*. Technische Universität Darmstadt (2010).

[14] K. Fukuda. *Lecture: Polyhedral Computation*. Department of Mathematics and Institute of Theoretical Computer Science, ETH Zurich, Switzerland (2014).

[15] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press (2004).

[16] M. Froissart. *Constructive generalization of Bell's inequalities*. Il Nuovo Cimento B (1971-1996) **64**, 2 241 (1981).

[17] A. Fine. *Hidden Variables, Joint Probability, and the Bell Inequalities*. Phys. Rev. Lett. **48**, 5 291 (1982).

[18] K. Fukuda and A. Prodon. *Double description method revisited*. In M. Deza, R. Euler, and I. Manoussakis, eds., *Combinatorics and Computer Science*, 91–111. Springer Berlin Heidelberg, Berlin, Heidelberg (1996).

[19] I. Pitowsky. *The range of quantum probability*. Journal of Mathematical Physics **27**, 6 1556 (1986).

[20] M. R. Bussieck and M. E. Lübbecke. *The vertex set of a 01-polytope is strongly P-enumerable*. Computational Geometry **11**, 2 103 (1998).

[21] V. Kaibel and M. E. Pfetsch. *Some Algorithmic Problems in Polytope Theory*. In M. Joswig and N. Takayama, eds., *Algebra, Geometry and Software Systems*, 23–47. Springer, Berlin, Heidelberg (2003).

[22] D. Avis. *A Revised Implementation of the Reverse Search Vertex Enumeration Algorithm*. In G. Kalai and G. M. Ziegler, eds., *Polytopes — Combinatorics and Computation*, DMV Seminar, 177–198. Basel (2000).

[23] D. Bremner, K. Fukuda, and A. Marzetta. *Primal—Dual Methods for Vertex and Facet Enumeration*. Discrete & Computational Geometry **20**, 3 333 (1998).

[24] T. S. Motzkin, H. Raiffa, G. L. Thompson, and R. M. Thrall. *3. The Double Description Method*. In *3. The Double Description Method*, 51–74. Princeton University Press (1953).

[25] D. Avis and K. Fukuda. *A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra*. Discrete & Computational Geometry **8**, 3 295 (1992).

[26] W. Bruns, B. Ichim, and C. Söger. *The power of pyramid decomposition in Normaliz*. Journal of Symbolic Computation **74** 513 (2016).

[27] *cdd Homepage*, see www-oldurls.inf.ethz.ch/personal/fukudak/cdd_home.

[28] *PPL | BUGSENG*. See www.bugseng.com/ppl.

[29] *PORTA polyhedral conversion*. See www.porta.zib.de.

[30] *Normaliz software*. See www.normaliz.uni-osnabrueck.de.

[31] D. Avis and C. Jordan. *mplrs: A scalable parallel vertex/facet enumeration code*. arXiv:1511.06487 [cs] (2017). ArXiv: 1511.06487.

[32] T. H. Cormen, Thomas H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, 1 edition (1990).

[33] R. Diestel. *Graph Theory* (2006).

[34] A. Brüngger, A. Marzetta, K. Fukuda, and J. Nievergelt. *The parallel search bench ZRAM and its applications*. Annals of Operations Research **90**, 0 45 (1999).

[35] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. *Complete family of separability criteria*. Phys. Rev. A **69**, 2 022308 (2004).

[36] U. Herzog and J. A. Bergou. *Distinguishing mixed quantum states: Minimum-error discrimination versus optimum unambiguous discrimination*. Phys. Rev. A **70**, 2 022302 (2004).

[37] M. Navascués, S. Pironio, and A. Acín. *Bounding the Set of Quantum Correlations*. Phys. Rev. Lett. **98**, 1 010401 (2007).

[38] J. C. Spall. *Cyclic Seesaw Process for Optimization and Identification*. Journal of Optimization Theory and Applications **154**, 1 187 (2012).

[39] R. F. Werner and M. M. Wolf. *Bell inequalities and Entanglement*. arXiv:quant-ph/0107093 (2001).

[40] K. F. Pál and T. Vértesi. *Maximal violation of a bipartite three-setting, two-outcome Bell inequality using infinite-dimensional quantum systems*. Phys. Rev. A **82**, 2 022116 (2010).

[41] M. F. Anjos and J. B. Lasserre. *Handbook on Semidefinite, Conic and Polynomial Optimization*. Springer New York Dordrecht Heidelberg London (2012).

[42] M. Navascués, S. Pironio, and A. Acín. *A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations*. New J. Phys. **10**, 7 073013 (2008).

[43] M. Navascués, S. Pironio, and A. Acín. *SDP Relaxations for Non-Commutative Polynomial Optimization*. In M. F. Anjos and J. B. Lasserre, eds., *Handbook on Semidefinite, Conic and Polynomial Optimization*, 601–634. Springer US, Boston, MA (2012).

[44] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press (2010).

[45] T. Heinosaari and M. Ziman. *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*. Cambridge University Press (2011).

[46] G. N. M. Tabia and D. M. Appleby. *Exploring the geometry of qutrit state space using symmetric informationally complete probabilities*. Physical Review A **88**, 1 012131 (2013).

[47] M. Schlosshauer. *Decoherence: And the Quantum-To-Classical Transition*. The Frontiers Collection. Springer (2007).

[48] O. Gühne and G. Tóth. *Entanglement detection*. Physics Reports **474**, 1 1 (2009).

[49] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. *Quantum entanglement*. Reviews of Modern Physics **81**, 2 865 (2009).

[50] V. Scarani and N. Gisin. *Spectral decomposition of Bell's operators for qubits*. Journal of Physics A: Mathematical and General **34**, 30 6043 (2001).

[51] N. Gisin and H. Bechmann-Pasquinucci. *Bell inequality, Bell states and maximally entangled states for n qubits*. Physics Letters A **246**, 1 1 (1998).

[52] W. Helwig, W. Cui, J. I. Latorre, A. Riera, and H.-K. Lo. *Absolute maximal entanglement and quantum secret sharing*. Physical Review A **86**, 5 052335 (2012).

[53] F. Huber and N. Wyderka. *Table of AME states*.

[54] F. Huber, C. Eltschka, J. Siewert, and O. Gühne. *Bounds on absolutely maximally entangled states from shadow inequalities, and the quantum MacWilliams identity*. Journal of Physics A: Mathematical and Theoretical **51**, 17 175301 (2018).

[55] R. F. Werner. *Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model*. Physical Review A **40**, 8 4277 (1989).

[56] T. Eggeling and R. F. Werner. *Separability properties of tripartite states with $U \otimes U \otimes U$ symmetry*. Physical Review A **63**, 4 042111 (2001).

[57] A. Peres. *Separability Criterion for Density Matrices*. Phys. Rev. Lett. **77**, 8 1413 (1996).

[58] M. Horodecki, P. Horodecki, and R. Horodecki. *Separability of mixed states: necessary and sufficient conditions*. Phys. Lett. A **223**, 1 1 (1996).

[59] T. Vidick and S. Wehner. *More nonlocality with less entanglement*. Physical Review A **83**, 5 052310 (2011).

[60] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. *Proposed Experiment to Test Local Hidden-Variable Theories*. Phys. Rev. Lett. **23**, 15 880 (1969).

[61] A. Einstein, B. Podolsky, and N. Rosen. *Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?* Physical Review **47**, 10 777 (1935).

[62] D. Bohm and Y. Aharonov. *Discussion of Experimental Proof for the Paradox of Einstein, Rosen, and Podolsky*. Physical Review **108**, 4 1070 (1957).

[63] A. B. Einstein, M. Born, and H. Born. *The Born-Einstein letters: Correspondence between Albert Einstein and Max and Hedwig Born from 1916-1955, with commentaries by Max Born;* (1971).

[64] J. C. Pearl and E. G. Cavalcanti. *Classical causal models cannot faithfully explain Bell nonlocality or Kochen-Specker contextuality in arbitrary scenarios*. Quantum **5** 518 (2021).

[65] B. S. Tsirelson. *Some results and problems on quantum Bell-type inequalities*. Hadronic Journal Supplement **8**, 4 329 (1993).

[66] B. S. Cirel'son. *Quantum generalizations of Bell's inequality*. Letters in Mathematical Physics **4**, 2 93 (1980).

[67] S. Pironio. *Lifting Bell inequalities*. Journal of Mathematical Physics **46**, 6 062112 (2005).

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

[69] L. Khalfin and B. Tsirelson. *Quantum and quasi-classical analogs of Bell inequalities*. Symposium on the Foundations of Modern Physics 1985 (ed. Lahti et al.; World Sci. Publ.) 441–460 (1985).

[70] N. S. Jones and L. Masanes. *Interconversion of nonlocal correlations*. Physical Review A **72**, 5 052312 (2005).

[71] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu et al. *Nonlocal correlations as an information-theoretic resource*. Physical Review A **71**, 2 022101 (2005).

[72] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. *Bell nonlocality*. Rev. Mod. Phys. **86**, 2 419 (2014).

[73] I. Pitowsky. *Correlation polytopes: Their geometry and complexity*. Mathematical Programming **50**, 1 395 (1991).

[74] D. Avis, H. Imai, T. Ito, and Y. Sasaki. *Deriving Tight Bell Inequalities for 2 Parties with Many 2-valued Observables from Facets of Cut Polytopes* (2004).

[75] C. Śliwa. *Symmetries of the Bell correlation inequalities*. Phys. Lett. A **317**, 3 165 (2003).

[76] D. Collins and N. Gisin. *A relevant two qubit Bell inequality inequivalent to the CHSH inequality*. J. Phys. A: Math. Gen. **37**, 5 1775 (2004).

[77] R. F. Werner and M. M. Wolf. *All multipartite Bell-correlation inequalities for two dichotomic observables per site*. Phys. Rev. A **64**, 3 032112 (2001).

[78] D. Avis and T. Ito. *New classes of facets of the cut polytope and tightness of Imm22 Bell inequalities*. Discrete Applied Mathematics **155**, 13 1689 (2007).

[79] W. Laskowski, T. Paterek, M. Żukowski, and Č. Brukner. *Tight Multipartite Bell's Inequalities Involving Many Measurement Settings*. Physical Review Letters **93**, 20 200401 (2004).

[80] S. L. Braunstein and C. M. Caves. *Wringing out better Bell inequalities*. Annals of Physics **202**, 1 22 (1990).

[81] J.-D. Bancal, N. Gisin, and S. Pironio. *Looking for symmetric Bell inequalities*. J. Phys. A: Math. Theor. **43**, 38 385303 (2010).

[82] S. L. Braunstein and C. M. Caves. *Information-Theoretic Bell Inequalities*. Physical Review Letters **61**, 6 662 (1988).

[83] J. Barrett. *Nonsequential positive-operator-valued measurements on entangled mixed states do not always violate a Bell inequality*. Physical Review A **65**, 4 042302 (2002).

[84] G. Svetlichny. *Distinguishing three-body from two-body nonseparability by a Bell-type inequality*. Physical Review D **35**, 10 3066 (1987).

[85] D. Collins, N. Gisin, S. Popescu, D. Roberts, and V. Scarani. *Bell-Type Inequalities to Detect True n-Body Nonseparability*. Physical Review Letters **88**, 17 170405 (2002).

[86] J.-D. Bancal, N. Brunner, N. Gisin, and Y.-C. Liang. *Detecting Genuine Multipartite Quantum Nonlocality: A Simple Approach and Generalization to Arbitrary Dimensions*. Physical Review Letters **106**, 2 020405 (2011).

[87] J.-D. Bancal, C. Branciard, N. Brunner, N. Gisin, and Y.-C. Liang. *A framework for the study of symmetric full-correlation Bell-like inequalities*. J. Phys. A: Math. Theor. **45**, 12 125301 (2012).

[88] D. M. Greenberger, M. A. Horne, and A. Zeilinger. *Going Beyond Bell's Theorem*. In M. Kafatos, ed., *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, Fundamental Theories of Physics, 69–72. Springer Netherlands, Dordrecht (1989).

[89] L. Hardy. *Nonlocality for two particles without inequalities for almost all entangled states*. Physical Review Letters **71**, 11 1665 (1993).

[90] S. Popescu and D. Rohrlich. *Which states violate Bell's inequality maximally?* Physics Letters A **169**, 6 411 (1992).

[91] A. K. Ekert. *Quantum cryptography based on Bell's theorem*. Physical Review Letters **67**, 6 661 (1991).

[92] A. Acín, S. Massar, and S. Pironio. *Efficient quantum key distribution secure against no-signalling eavesdroppers*. New Journal of Physics **8**, 8 126 (2006).

[93] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. *Proposed Experiment to Test Local Hidden Variable Theories.* Phys. Rev. Lett. **24** 549 (1970).

[94] C. De Simone. *Lifting facets of the cut polytope*. Operations Research Letters **9**, 5 341 (1990).

[95] L. Escolà, J. Calsamiglia, and A. Winter. *All tight correlation Bell inequalities have quantum violations*. Phys. Rev. Research **2**, 1 012044 (2020).

[96] P. Wittek. *Algorithm 950: Ncpol2sdpa—Sparse Semidefinite Programming Relaxations for Polynomial Optimization Problems of Noncommuting Variables*. ACM Trans. Math. Softw. **41**, 3 (2015).

[97] D. Bruß, D. P. DiVincenzo, A. Ekert, C. A. Fuchs, C. Macchiavello et al. *Optimal universal and state-dependent quantum cloning*. Phys. Rev. A **57**, 4 2368 (1998).

[98] M. Hein, J. Eisert, and H. J. Briegel. *Multiparty entanglement in graph states*. Phys. Rev. A **69**, 6 062311 (2004).

[99] O. Gühne, M. Cuquet, F. E. S. Steinhoff, T. Moroder, M. Rossi et al. *Entanglement and nonclassical properties of hypergraph states*. J. Phys. A: Math. Theor. **47**, 33 335303 (2014).

[100] B. M. Terhal, A. C. Doherty, and D. Schwab. *Symmetric Extensions of Quantum States and Local Hidden Variable Theories*. Phys. Rev. Lett. **90**, 15 157903 (2003).

[101] M. Żukowski and C. Brukner. *Bell's Theorem for General N-Qubit States*. Phys. Rev. Lett. **88**, 21 210401 (2002).

[102] S. Szalay. *k-stretchability of entanglement, and the duality of k-separability and k-producibility*. Quantum **3** 204 (2019).

[103] F. Baccari, J. Tura, M. Fadel, A. Aloy, J.-D. Bancal et al. *Bell correlation depth in many-body systems*. Physical Review A **100**, 2 022121 (2019).

[104] M. L. Almeida, D. Cavalcanti, V. Scarani, and A. Acín. *Multipartite fully nonlocal quantum states*. Physical Review A **81**, 5 052111 (2010).

[105] The Supplemental Material of our manuscript [C] will be available with the publication.

[106] Y. Li, M. Han, M. Grassl, and B. Zeng. *Invariant perfect tensors*. New Journal of Physics **19**, 6 063029 (2017).

[107] C. Kiefer. *Quantum gravity*. Cambridge University Press (2012).

[108] S. Hossenfelder, ed. *Experimental search for quantum gravity*. Springer Berlin (2018).

[109] S. Bose, A. Mazumdar, G. W. Morley, H. Ulbricht, M. Toroš et al. *Spin entanglement witness for quantum gravity*. Phys. Rev. Lett. **119** 240401 (2017).

[110] C. Marletto and V. Vedral. *Gravitational induced entanglement between two massive particles is sufficient evidence of quantum effects in gravity*. Phys. Rev. Lett. **119** 240402 (2017).

[111] M. Christodoulou and C. Rovelli. *On the possibility of laboratory evidence for quantum superposition of geometries*. arXiv:1808.05842v1 (2018).

[112] S. Bose and G. V. Morley. *Matter and spin superposition in vacuum experiment*. arXiv:1810.07045v1 (2018).

[113] S. Qvarfort, S. Bose, and A. Serafini. *Mesoscopic entanglement from central potential interactions*. arXiv:1812.09776v1 (2018).

[114] A. A. Balushi, W. Cong, and R. B. Mann. *Optomechanical quantum Cavendish experiment*. Phys. Rev. A **98** 043811 (2018).

[115] M. Carlesso, A. Bassi, M. Parernostro, and H. Ulbricht. *Testing the gravitational field generated by a quantum superposition*. arXiv:1906.04513 (2019).

[116] H. Miao, D. Martynov, and H. Yang. *Quantum correlation of light mediated by gravity*. arXiv:1901.05827 (2019).

[117] T. Kristanda, G. Y. Tham, M. Paternostro, and T. Paterek. *Observable quantum entanglement due to gravity*. arXiv:1906.08808 (2019).

[118] D. Carney, P. C. Stamp, and J. M. Taylor. *Tabletop experiments for quantum gravity: a user's maunal*. arXiv:1807.11494v2 (2018).

[119] R. Howl, L. Hackermüller, D. E. Bruschi, and I. Fuentes. *Gravity in the quantum lab*. Adv. Phys. **3** 138184 (2017).

[120] R. Marshman, A. Mazumdar, G. W. Morley, P. F. Barker, S. Hoekstra et al. *Mesoscopic interference for metric and curvature (MIMAC) and gravitational waves*. arXiv:1807.10830v1 (2018).

[121] H. Pino, J. Prat-Camps, K. Sinha, B. P. Venkatesh, and O. Romero-Isart. *On-chip quantum interference of a superconducting microsphere*. Quant. Sci. Tech. **3**, 10 25001 (2018).

[122] J. S. Pedernales, G. W. Morley, and M. B. Plenio. *Motional dynamical decoupling for matter-wave interferometry*. arXiv:1906.00835 (2019).

[123] M. J. W. Hall and M. Reginatto. *On two recent proposals for witnessing nonclassical gravity*. J. Phys. A: Math. Theor. **51** 085303 (2018).

[124] C. Anastopoulos and B. L. Hu. *Comment on "A spin entanglement witness for quantum gravity" and on "Gravitationally induced entanglement between two massive particles is sufficient evidence of quantum effects in gravity"*. arXiv:1804.11315v1 (2018).

[125] M. Reginatto and M. J. W. Hall. *Entanglemetn of quantum fields via classical gravity*. arXiv:1809.04989v1 (2018).

[126] M. Christodoulou and C. Rovelli. *On the possibility of experimental detection of the discreteness of time*. arXiv:1812.01542v2 (2018).

[127] G. Baym and T. Ozawa. *Two-slit diffraction with highly charged particles: Niels Bohr's consistency argument that the electromagnetic field must be quantized*. Proc. Natl. Asoc. Sci. **106** 3035 (2008).

[128] A. Mari, G. De Palma, and V. Giovannetti. *Experiments testing macroscopic quantum superpositions must be slow*. Sci. Rep. **6** 22777 (2016).

[129] A. Belenchia, R. Wald, F. Giacomini, E. Castro–Ruiz, C. Brukner et al. *Quantum superposition of massive objects and the quantization of gravity*. arXiv:1807.07015v1 (2018).

[130] R. Horodecki, R. Horodecki, and M. Horodecki. *Violating Bell inequality by mixed spin-$\frac{1}{2}$ states: necessary and sufficient condition*. Phys. Lett. A **200** 340 (1995).

[131] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-Å. Larsson. *Quantum contextuality* (2020).

[132] K.-W. Bong, A. Utreras-Alarcón, F. Ghafari, Y.-C. Liang, N. Tischler et al. *Testing the reality of Wigner's friend's observations* (2019). ArXiv: 1907.05607.

[133] D. Cavalcanti and P. Skrzypczyk. *Quantum steering: a review with focus on semidefinite programming*. Rep. Prog. Phys. **80**, 2 024001 (2017).

[134] R. Uola, A. C. S. Costa, H. C. Nguyen, and O. Gühne. *Quantum steering*. Rev. Mod. Phys. **92** 015001 (2020).

[135] M. M. Wolf, D. Perez-Garcia, and C. Fernandez. *Measurements Incompatible in Quantum Theory Cannot Be Measured Jointly in Any Other No-Signaling Theory*. Phys. Rev. Lett. **103**, 23 230402 (2009).

[136] S. Abramsky and A. Brandenburger. *The sheaf-theoretic structure of non-locality and contextuality*. New J. Phys. **13**, 11 113036 (2011).

[137] The Mosek optimization software, see www.mosek.com.

[138] G. Sagnol and M. Stahlberg. *PICOS Homepage*. PICOS, the Python API to conic and integer programming solvers, see picos-api.gitlab.io/picos/index.html.