Characterizing quantum correlations: entanglement, uncertainty relations and exponential families

DISSERTATION
zur Erlangung des Grades eines Doktors
der Naturwissenschaften

vorgelegt von
Dipl.-Phys. Sönke Niekamp
geb. am 17. September 1982 in Hannover

eingereicht
bei der Naturwissenschaftlich-Technischen Fakultät
der Universität Siegen
Siegen 2012

Gutachter der Dissertation: Prof. Dr. Otfried Gühne und Prof. Dr. Dagmar Bruß

Datum der mündlichen Prüfung: 20. April 2012

gedruckt auf alterungsbeständigem holz- und säurefreiem Papier

Abstract

This thesis is concerned with different characterizations of multi-particle quantum correlations and with entropic uncertainty relations.

The effect of statistical errors on the detection of entanglement is investigated. First, general results on the statistical significance of entanglement witnesses are obtained. Then, using an error model for experiments with polarization-entangled photons, it is demonstrated that Bell inequalities with lower violation can have higher significance.

The question for the best observables to discriminate between a state and the equivalence class of another state is addressed. Two measures for the discrimination strength of an observable are defined, and optimal families of observables are constructed for several examples.

A property of stabilizer bases is shown which is a natural generalization of mutual unbiasedness. For sets of several dichotomic, pairwise anticommuting observables, uncertainty relations using different entropies are constructed in a systematic way.

Exponential families provide a classification of states according to their correlations. In this classification scheme, a state is considered as k-correlated if it can be written as thermal state of a k-body Hamiltonian. Witness operators for the detection of higher-order interactions are constructed, and an algorithm for the computation of the nearest k-correlated state is developed.

Zusammenfassung

Diese Arbeit befasst sich mit Charakterisierungen von Mehrteilchen-Quantenkorrelationen und mit entropischen Unschärferelationen.

Der Einfluss statistischer Fehler auf die Detektion von Verschränkung wird untersucht. Zuerst werden allgemeine Resultate zur statistischen Signifikanz von Verschränkungszeugen erzielt, dann wird unter Verwendung eines Fehlermodells für polarisationsverschränkte Photonen gezeigt, dass Bellsche Ungleichungen mit niedrigerer Verletzung höhere Signifikanz haben können.

Die Frage nach den besten Observablen zur Unterscheidung eines Zustands von der Äquivalenzklasse eines anderen wird behandelt. Zwei Maße für die Unterscheidungskraft werden definiert, und für mehrere Beispiele werden optimale Familien von Observablen gefunden.

Es wird eine Eigenschaft von Stabilisatorbasen gezeigt, die eine natürliche Verallgemeinerung der *mutual unbiasedness* darstellt. Für Familien aus mehreren dichotomen, paarweise antikommutierenden Observablen werden Unschärferelationen mit verschiedenen Entropien systematisch konstruiert.

Exponentielle Familien ermöglichen eine Klassifikation von Zuständen nach den enthaltenen Korrelationen. Hierbei wird ein Zustand als k-korreliert angesehen, wenn er sich als thermischer Zustand eines k-Teilchen-Hamiltonoperators schreiben lässt. Es werden Zeugenoperatoren zur Detektion von Wechselwirkungen höherer Ordnung konstruiert, und ein Algorithmus zur Berechnung des nächsten k-korrelierten Zustands wird entwickelt.

Contents

1 Introduction

Perhaps the greatest success of quantum information theory has not been in computation or cryptography, but in the contributions it has made to our better understanding of the fundamental concepts of quantum mechanics. The focus on states and measurements as objects of investigation, detached from their concrete physical implementation, and the introduction of information-theoretic concepts have helped to determine more clearly the differences between quantum and classical physics. The impressive progress in control of individual quantum systems during the last decades has made it possible to test many of the predicted phenomena experimentally.

A central role is played by the concept of entanglement [32, 100, 101, 102]. The presence of entanglement is necessary to show nonlocality in the sense of Bell's theorem, which establishes most dramatically the fundamental difference between quantum and classical physics [13]. It has become popular to refer to entanglement as the key resource for quantum information processing [86], though this is an oversimplification: At least, one has to distinguish between different kinds of entanglement, some of which are useful for a given task, while others are not. Naturally, the situation becomes more involved if more than two degrees of freedom are entangled. At a more pragmatic level, entanglement is useful as a benchmark for the experimental control of individual quantum systems. This shows the particular relevance of methods that allow to verify the presence of entanglement in an experiment and to determine its type. While entanglement detection is a highly developed field [44], the question of the statistical significance which is provided by an entanglement test has received little attention. The terminology which is used to describe entanglement witnesses provides an example: It is customary to call a witness *optimal* if there is no other witness which detects more entangled states [74]. However, for experimental applications one is rather more interested in a witness that detects a given target state with the highest possible significance. In Chapter 3 of this thesis, the optimization of witnesses in the latter sense is considered. Similarly, it is shown that Bell inequalities with a lower violation can have a higher significance. In the multipartite case, where different types of entanglement can be distinguished, one is interested in finding observables to discriminate inequivalent states. In Chapter 4 the question for families of observables with the highest discrimination strength is addressed.

A prototypical example of the ways in which quantum mechanics departs from classical physics is given by the uncertainty principle [52]. This term refers to the fact that one cannot prepare a quantum system in such a way that for all possible measurements the outcome is certain. A quantitative formulation of the principle is provided by uncertainty relations. For continuous variables, such as position and momentum, uncertainty is usually quantified by the standard deviation. In the case of finitely many measurement outcomes, the entropy of the outcome probabilities is a more natural measure.

This has lead to the study of entropic uncertainty relations [30, 69, 80].

While entanglement describes a form of correlation that is inherently quantum mechanical, one is also interested in classifying all correlations that are contained in a quantum state, whether they can be described classically or not. A very natural way of characterizing correlations is by asking: How much information is contained in a given state, but is not contained in its k-party reduced density matrices? It turns out that the answer to this question can be understood geometrically as a distance from the state to the class of thermal states of Hamiltonians with at most k-body interactions. The term "exponential families" in the title of this thesis refers to such classes of thermal states. In the framework of information geometry, a theory of exponential families of classical probability distributions has been developed [2,3]. Even though the concept of thermal states is very natural in quantum mechanics, the quantum version of the theory has been developed only recently [132, 133, 134].

This thesis is structured as follows:

In Chapter 2 those aspects of quantum information theory are reviewed which are prerequisites for the main part of the thesis. These topics include entanglement and its detection, Bell inequalities, classical and quantum entropies and uncertainty relations. An introduction is given to the stabilizer formalism, which will be used as an important tool in all subsequent chapters. The theory of classical exponential families of interaction spaces is outlined.

Chapters 3–6 constitute the main part of the thesis. The subject of Chapter 3 is the statistical significance of experimental entanglement tests, and Chapter 4 is concerned with finding optimal sets of observables for discriminating classes of multipartite states with different entanglement properties.

Chapter 5 describes the results on entropic uncertainty relations. These results fall into two categories: the characterization of measurement bases that admit strong uncertainty relations and the generalization of the theory to the case of more than two observables.

In Chapter 6 the theory of exponential families of interaction spaces is applied to the study of quantum correlations. The first section aims at establishing connections between different notions of genuine three-party correlations by studying the measurement probabilities of three-party Bell experiments. The other sections concern exponential families of quantum states. The focus is on the computation of the information projection and the detection of higher-order interactions.

The results reported in Chapters 3, 4 and 5 have been published in Refs.[1] B, C and D, respectively. The results in Chapter 6 are as of now unpublished.

Finally, in Chapter 7 the thesis ends with a conclusion.

[1] References in the form of letters refer to the publication list on p. 131.

2 Basic concepts

2.1 Entanglement and its detection

2.1.1 Entanglement

The notion of entanglement pertains to quantum systems that are composed of identifiable subsystems. Like many concepts of quantum information theory, entanglement is best introduced while imagining that the system under consideration consists of a number of distinguishable particles which are situated in different, spatially separated laboratories.[1] The persons experimenting on the individual particles are customarily called Alice, Bob, Charlie, ...

We begin with the case of two parties. If the two subsystems are prepared independently and they are adequately described by two pure states $|\psi^A\rangle$ and $|\psi^B\rangle$, then the composite system is in a pure *separable or product state*

$$|\psi\rangle = |\psi^A\rangle \otimes |\psi^B\rangle. \tag{2.1}$$

A pure state that cannot be written as a product is called *entangled*. Examples of entangled states are the four *Bell states*[2]

$$|\phi^\pm\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle), \qquad |\psi^\pm\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle). \tag{2.2}$$

The defining property of separable states is that they can be prepared locally, i. e., by two devices which act independently on either subsystem. Consequently, a mixed state ρ is called separable if one can find states $|\psi_i^A\rangle$ of system A and $|\psi_i^B\rangle$ of system B and convex weights $p_i \geq 0$ with $\sum_i p_i = 1$ such that [126]

$$\rho = \sum_i p_i \, |\psi_i^A\rangle\langle\psi_i^A| \otimes |\psi_i^B\rangle\langle\psi_i^B|. \tag{2.3}$$

Such a state can be prepared locally in the following way: A random generator produces numbers with probabilities p_i. Given the random number i, the systems A and B are prepared in states $|\psi_i^A\rangle$ and $|\psi_i^B\rangle$, respectively. The set of separable states is a convex and compact subset of all states.[3] Again, states which are not separable are called entangled.

[1] The concept of entanglement is applicable more generally, though. *Hyperentanglement* specifically refers to entanglement between different degrees of freedom of the same particle.

[2] As usual in the quantum information literature, $|0\rangle$ and $|1\rangle$ are the eigenvectors of the Pauli matrix σ_z with eigenvalue $+1$ and -1, respectively, and $|01\rangle = |0\rangle|1\rangle = |0\rangle \otimes |1\rangle$.

[3] Throughout this thesis all Hilbert spaces are finite-dimensional. In this case, Carathéodory's theorem implies that any separable state can be written as a convex combination of finitely many pure product states [44, Sec. 2.2].

Separable mixed states are not necessarily product states, i. e., of the form $\rho = \rho^A \otimes \rho^B$. The results of local measurements will in general be correlated[4].

In the case of more than two parties, different classes of entangled states can be distinguished. We call an n-party pure state *fully separable* if it is a product of n factors,

$$|\psi_{\text{fs}}\rangle = |\psi^1\rangle \otimes \cdots \otimes |\psi^n\rangle. \tag{2.4}$$

A pure state is called *biseparable* if it can be written as a product of two factors,

$$|\psi_{\text{bs}}\rangle = |\psi^1\rangle_{i_1,\ldots,i_m} \otimes |\psi^2\rangle_{i_{m+1},\ldots,i_n}. \tag{2.5}$$

The states $|\psi^1\rangle$ and $|\psi^2\rangle$ need not be entangled. For example, a pure biseparable state of three parties has one of the following forms:

$$|\psi_{AB|C}\rangle = |\psi^1\rangle_{AB} \otimes |\psi^2\rangle_C, \quad |\psi_{AC|B}\rangle = |\psi^1\rangle_{AC} \otimes |\psi^2\rangle_B, \quad |\psi_{A|BC}\rangle = |\psi^1\rangle_A \otimes |\psi^2\rangle_{BC}. \tag{2.6}$$

(There are three *bipartitions* of three parties.) States which are not biseparable are called *genuinely n-partite* (or *genuinely multipartite*) entangled. For $n > 3$ parties, the biseparable states can be differentiated further [44, Sec. 3.3].

Mixed states which can be written as convex combinations of pure fully separable (biseparable) states are called fully separable (biseparable) [1]. The important point here is that the states in the convex decomposition of a biseparable mixed state may be biseparable with respect to different bipartitions of the n parties. This definition ensures that only those states count as genuinely n-partite entangled which require for their preparation a quantum operation on all parties.

2.1.2 Local operations

Thinking again in the picture of spatially separated laboratories, it is natural to ask what can (or cannot) be done with local operations. Answering this question also helps to understand multipartite entanglement.

Two states certainly have the same entanglement properties if they differ only by the choice of local bases. This is the case if one state can be obtained from the other by a local unitary (LU) transformation,

$$\rho' = U_1 \otimes \cdots \otimes U_n \, \rho \, U_1^\dagger \otimes \cdots \otimes U_n^\dagger. \tag{2.7}$$

We call these states *LU-equivalent*. For a necessary and sufficient condition for the LU-equivalence of pure n-qubit states see Ref. 68. For the study of stabilizer states, a certain subgroup of local unitaries, namely, local Clifford operations, are particularly relevant (see Section 2.5.3).

The most general local operation consists of the use of an additional quantum system (a so-called *ancilla*) and arbitrary unitary operations and measurements on the local

[4] It seems to be common practice in quantum information theory to use the word *correlation* somewhat loosely. This is in contrast to mathematical statistics, where one carefully distinguishes between *uncorrelated* (vanishing correlation coefficient) and *independent* (factorizing probability distribution).

system combined with the ancilla.[5] Furthermore, we allow the parties to communicate classically. Thus, the operation of one party may depend on a previous measurement result of another party. We can now ask if it is possible in this way to transform a single copy of a multipartite state ρ into another state ρ'. If we do not demand that the conversion always succeeds, but only that the success probability is nonzero, we refer to this scheme as *stochastic local operations and classical communication (SLOCC)*. If the conversion is possible in either direction (with nonvanishing probability), we call the states *SLOCC-equivalent*. It can be shown that two pure states are SLOCC-equivalent if and only if they are related by an invertible local operation [31],

$$|\phi\rangle = A_1 \otimes \cdots \otimes A_n |\psi\rangle. \tag{2.8}$$

Here the A_i are arbitrary invertible operators.

The concept of SLOCC equivalence is relevant for quantum information processing because SLOCC-equivalent states can in principle be used for the same applications (by first applying SLOCC transformations, if necessary). Note, however, that any quantum speedup might be lost due to the overhead resulting from the conversion.

For the case of pure three-qubit states there, are six SLOCC equivalence classes with the following canonical representatives [31]:

$$\begin{aligned} |\psi_{\mathrm{fs}}\rangle &= |000\rangle, & |\psi_{A|BC}\rangle &= |0\rangle_A \otimes |\phi^+\rangle_{BC}, \\ |\psi_{B|AC}\rangle &= |0\rangle_B \otimes |\phi^+\rangle_{AC}, & |\psi_{AB|C}\rangle &= |\phi^+\rangle_{AB} \otimes |0\rangle_C, \\ |\mathrm{W}_3\rangle &= \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle), & |\mathrm{GHZ}_3\rangle &= \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle). \end{aligned} \tag{2.9}$$

Here $|\phi^+\rangle$ is the Bell state from Eq. (2.2). (Any other Bell state would do as well, since all Bell states are LU-equivalent.) The SLOCC equivalence classes of the first four states consist respectively of the fully separable states and those states which are biseparable with respect to a specific bipartition, but not fully separable. There are two SLOCC classes of genuinely threepartite entangled states, which are represented by the three-qubit W state[6] and the three-qubit Greenberger-Horne-Zeilinger (GHZ) state. It is easy to see that the W and the GHZ state have different entanglement properties: If one party of the GHZ state is traced out, the remaining two parties are in the maximally mixed state, while the reduced two-party density matrix of the W state is entangled. A parameter counting argument shows that for more than three qubits there are continuous families of SLOCC-inequivalent pure states [31].

For three qubits, the class of mixed W states is defined as the convex combinations of biseparable states and pure states which are SLOCC-equivalent to the W state [1]. In this classification scheme, the class of mixed GHZ states is the set of all states. (One might ask if the roles of the W and the GHZ state in these definitions could be swapped. This is not possible for the following reason: The closure of the set of pure GHZ-type state contains all pure W-type states. Therefore the set of convex combinations of GHZ-type pure states and biseparable states cannot be closed [1].)

[5] Such operations can be described in the *Kraus operator or operator sum representation* [86, Sec. 8.2.3].

[6] This state is named after Wolfgang Dür.

2.1.3 Entanglement criteria

The field of entanglement detection is concerned with two closely related questions: *Is the state described by a given density matrix separable or entangled?*, and: *Does a certain experiment produce entangled or separable states?*

For pure bipartite states, the *Schmidt decomposition* provides a simple answer to the first question. It can be shown that for any such state $|\psi\rangle$ there exists an orthonormal basis $|\psi_i^A\rangle$ of the first subsystem and an orthonormal basis $|\psi_i^B\rangle$ of the second subsystem such that [86, Thm. 2.7]

$$|\psi\rangle = \sum_i \lambda_i |\psi_i^A\rangle \otimes |\psi_i^B\rangle \tag{2.10}$$

with nonnegative real numbers λ_i, which are called *Schmidt coefficients*. This decomposition can be found by first expanding the state $|\psi\rangle$ in a product basis and then computing the the singular value decomposition of the coefficient matrix. As a consequence of the uniqueness of the Schmidt coefficients, the state $|\psi\rangle$ is separable if and only if the number of nonzero coefficients (the *Schmidt number*) is one.

In general the separability problem is hard, both in the complexity-theoretic and in the practical sense. In fact, the separability problem for bipartite mixed states is proven to be strongly NP-hard [36, 45, 57]. Practical approaches fall into two broad categories: On the one hand, a number of algorithms have been developed, which are usually based on convex optimization or semidefinite programming [44, Sec. 2.3.3]. On the other hand, a large variety of theorems characterizing separable and entangled states have been proven. It is the latter approach which is considered in this thesis.

Usually one wants to prove that some state is entangled, rather than that it is separable.[7] Therefore, theorems in entanglement detection typically take the form of entanglement criteria: An *entanglement criterion* consists of a condition that is satisfied by all separable states. Thus, violation of the criterion for some state proves that it is entangled. The state is then said to be *detected* by the criterion. For reviews see Refs. 44, 55. As an example, consider the *positive partial transpose (PPT) or Peres-Horodecki criterion* [54, 88]: The partial transpose of the state

$$\rho = \sum_{i,j,k,\ell} \rho_{ijk\ell} \, |i\rangle\langle j| \otimes |k\rangle\langle \ell| \tag{2.11}$$

with respect to the first subsystem is defined as

$$\rho^{T_A} = \sum_{i,j,k,\ell} \rho_{jik\ell} \, |i\rangle\langle j| \otimes |k\rangle\langle \ell| \tag{2.12}$$

(note that the indices i and j of the coefficient matrix have been swapped), and analogously for the second subsystem and in the multipartite case. The criterion now states that separable states have positive[8] partial transpose. (Though the partial transpose of a state depends on the choice of local bases its spectrum does not.)

[7] For an algorithm to show separability see the supplementary information to Ref. 11.

[8] As customary we call a Hermitian matrix P positive and write $P \geq 0$ if it is positive semidefinite.

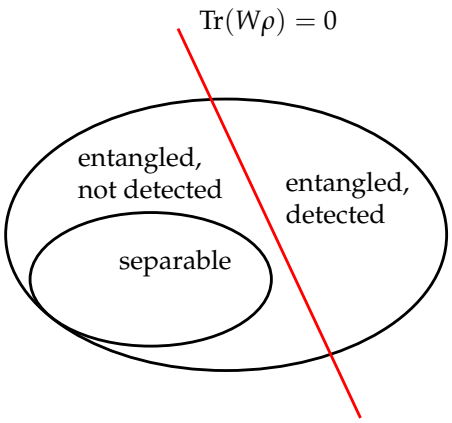

Figure 2.1: Illustration of an entanglement witness. Shown are the convex set of all states (large ellipse), the convex set of separable states (small ellipse) and the hyperplane defined by $\mathrm{Tr}(W\rho) = 0$ (red line). The hyperplane separates the entangled states which are detected by the witness operator from those that are not.

For the purpose of experimental entanglement detection one is particularly interested in criteria which do not require knowledge of the complete density matrix. Witness operators are the most important example. An *entanglement witness* (*witness* for short) is an observable W which has nonnegative expectation value on all separable states and negative expectation value on at least one entangled state [54, 108],

$$\begin{aligned} \mathrm{Tr}(W\rho) &\geq 0 \quad \text{for all separable } \rho, \\ \mathrm{Tr}(W\rho) &< 0 \quad \text{for at least one entangled } \rho. \end{aligned} \tag{2.13}$$

Entanglement witnesses have a useful geometrical interpretation, which is illustrated in Fig. 2.1: The equation $\mathrm{Tr}(W\rho) = 0$ defines a hyperplane in the space of all states, which separates the states with positive from those with negative expectation value. The set of separable states is by definition convex and compact. For any entangled state there exists a hyperplane that separates it from this set [54]. (This is a consequence of the Hahn-Banach theorem.) This shows that for every entangled state there is a witness which detects it. Since the sets of biseparable and fully separable states in the multipartite case are also convex and compact, witnesses are equally useful in that scenario.

Let us consider two methods to construct entanglement witnesses [44, Sec. 2.5.1]. Suppose we know a state ρ_{NPT} which does not have positive partial transpose.[9] Let $|\eta\rangle$ be an eigenvector of $\rho_{\mathrm{NPT}}^{T_A}$ corresponding to a negative eigenvalue. Then

$$W = |\eta\rangle\langle\eta|^{T_A} \tag{2.14}$$

is an entanglement witness detecting ρ_{NPT}, among other states.[10] As a second example,

[9] States are called *PPT* if they have positive partial transpose, and *NPT* if they do not.

[10] The PPT criterion is an example of an entanglement criterion based on a positive, but not completely

note that for any entangled pure state $|\psi\rangle$ we can construct the *projector witness*

$$W = \alpha \mathbb{1} - |\psi\rangle\langle\psi| \qquad \text{where} \qquad 1 > \alpha \geq \max_{|\phi\rangle = |\phi^A\rangle \otimes |\phi^B\rangle} |\langle\phi|\psi\rangle|^2. \tag{2.15}$$

Here the the maximum is given by the square of the largest Schmidt coefficient of $|\psi\rangle$. Projector witnesses can also be constructed in the multipartite case.

We say that the witness W' is *finer* than the witness W if it detects all states that are detected by W and possibly more [74],

$$\mathrm{Tr}(W\rho) < 0 \quad \Rightarrow \quad \mathrm{Tr}(W'\rho) < 0. \tag{2.16}$$

One can show that in this case [74, Lemma 2]

$$W' = \alpha(W - P) \tag{2.17}$$

for a positive operator P and a positive coefficient α. Conversely, any W' of the form Eq. (2.17) which is a valid witness is obviously finer than W. We call a witness *optimal* if there is no other witness which is finer. In other words, the witness W is optimal if for any positive operator P the expectation value $\mathrm{Tr}[(W - P)\rho]$ is negative for some separable state ρ. It is a necessary, but not a sufficient condition for the optimality of a witness W that the hyperplane defined by $\mathrm{Tr}(W\rho) = 0$ touches the set of separable states.[11] A method has been developed to *optimize* a witness, i. e., to find a finer witness by subtracting a suitable positive operator [74]. Note that subtracting a positive operator from a witness can only increase the violation for any fixed detected state. However, a major result of Chapter 3 will be that in general there is a trade-off between optimizing a witness in the sense discussed here and maximizing the statistical significance of entanglement detection for a given target state.

Let us conclude our discussion of witnesses by commenting on their implementation in experiments. Although any witness corresponds to a valid measurement, in practice only local measurements are feasible. Thus the witness has to be decomposed into projectors onto product vectors. The required number of measurement settings determines the experimental effort [44, Sec. 6.1.2].

2.2 Bell inequalities

2.2.1 Bell's theorem and the CHSH inequality

Bell's theorem states that the measurement probabilities predicted by quantum mechanics are incompatible with local realism [13]. In the form of Bell inequalities the theorem can be tested experimentally, showing that Nature cannot be described by a local

positive map. Any such criterion gives rise to a witness [44, Sec. 2.5.1]. Conversely, with the Choi-Jamiołkowski isomorphism it can be shown that any witness originates from such a map [44, Sec. 2.5.3].

[11] A simple counterexample shows that touching the set of separable states (this is sometimes called *weak optimality*) is not sufficient for optimality: Consider a 4×4-system and the states $|\psi_{01}^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ and $|\psi_{23}^-\rangle = (|23\rangle - |32\rangle)/\sqrt{2}$. Then $W = \mathbb{1}/2 - |\psi_{01}^-\rangle\langle\psi_{01}^-|$ is a witness touching the set of separable states, since the maximal fidelity of a Bell state with a product state is $1/2$. But a similar argument shows that $W' = W - |\psi_{23}^-\rangle\langle\psi_{23}^-|$ is a finer witness.

realistic theory. The theorem and its subsequent experimental verification [5,96,125] (if we forget about the infamous loopholes) constitute one of the deepest results of quantum physics.

We begin by recapitulating the prototypical Clauser-Horne-Shimony-Holt inequality. Consider two parties (Alice and Bob), each of whom can choose between two measurements, and suppose that each measurement has two possible outcomes. Denote by $P_{AB}(a,b)$ the probability that Alice and Bob obtain results a and b, respectively, if they measure A and B (where $A \in \{A_1, A_2\}$, $B \in \{B_1, B_2\}$ and $a, b \in \{-1, +1\}$). We say that the probability table allows a *deterministic local hidden variable model* if it can be represented in the form

$$P_{AB}(a,b) = \int \mathrm{d}\lambda\, \mu(\lambda)\chi_A(a,\lambda)\chi_B(b,\lambda), \tag{2.18}$$

where λ stands for the hidden variable, μ is a probability density, and the response functions χ_A and χ_B [which are normalized probability distributions, that is, $\chi_A(1,\lambda) + \chi_A(-1,\lambda) = 1$ and similarly for χ_B] take only the values 0 and 1. The assumptions leading to such a model are discussed below. One can easily check that probabilities of this form obey the *Clauser-Horne inequality* [22]

$$P_{A_1B_1}(-1,-1) + P_{A_1B_2}(+1,-1) + P_{A_2B_1}(-1,+1) - P_{A_2B_2}(-1,-1) \geq 0. \tag{2.19}$$

Rewriting this in terms of expectation values, we obtain the *Clauser-Horne-Shimony-Holt (CHSH) inequality* [23]

$$\langle A_1B_1\rangle + \langle A_1B_2\rangle + \langle A_2B_1\rangle - \langle A_2B_2\rangle \leq 2. \tag{2.20}$$

For its derivation none of the laws of quantum mechanics were used. The expectation values are not defined by observables, but are simply given as

$$\langle AB\rangle = P_{AB}(+1,+1) + P_{AB}(-1,-1) - P_{AB}(+1,-1) - P_{AB}(-1,+1). \tag{2.21}$$

Consider now a quantum mechanical two-qubit system with local observables $A_1 = -\sigma_x$ and $A_2 = -\sigma_y$ on Alice's and $B_1 = (\sigma_x + \sigma_y)/\sqrt{2}$ and $B_2 = (\sigma_x - \sigma_y)/\sqrt{2}$ on Bob's side. For the Bell state $|\psi^-\rangle$, quantum mechanics predicts

$$\langle A_1B_1\rangle + \langle A_1B_2\rangle + \langle A_2B_1\rangle - \langle A_2B_2\rangle = 2\sqrt{2}, \tag{2.22}$$

violating the CHSH inequality.

It is worth discussing the assumptions that lead to Eq. (2.18). First note that a general theory [not necessarily of the form Eq. (2.18)] is called *nonsignalling* if the probabilities for Alice's results do not depend on Bob's choice of measurement and vice versa,

$$P_A(a) = \sum_b P_{AB}(a,b) \qquad \text{is independent of } B \tag{2.23}$$

and analogously for Bob's results. We call a hidden variable model *local* if it is nonsignalling already for a fixed value of the variable. The idea behind the locality assumption is that the hidden variable could in principle be "uncovered" and the theory should still be nonsignalling in this case. A hidden variable model

$$P_{AB}(a,b) = \int \mathrm{d}\lambda\, \mu(\lambda)\chi_{AB}(a,b,\lambda) \tag{2.24}$$

is local in this sense precisely if it satisfies Eq. (2.23) with the response function χ_{AB} in place of P_{AB},

$$\sum_b \chi_{AB}(a,b,\lambda) \qquad \text{is independent of } B \tag{2.25}$$

and analogously for $\sum_a$. Each of the following sets of assumptions is sufficient to show Eq. (2.18):

1. There is a deterministic hidden variable model of the form

   $$P_{AB}(a,b) = \int \mathrm{d}\lambda\, \mu(\lambda)\chi_{AB}(a,b,\lambda), \tag{2.26}$$

   where the (a priori not necessarily factorizing) response function χ_{AB} takes only the values 0 and 1, and the model is local in the sense of Eq. (2.25).

2. There is a factorizable stochastic model of the form [34, 128]

   $$P_{AB}(a,b) = \int \mathrm{d}\lambda\, \mu(\lambda)\chi_{AB}(a,\lambda)\chi_{AB}(b,\lambda), \tag{2.27}$$

   where the response functions can take any value in $[0,1]$, and the model is local in the sense of Eq. (2.25).

3. There is a joint probability function $P(a_1,a_2,b_1,b_2)$ for the results of all measurements [34] whose marginals give the probabilities $P_{AB}(a,b)$,

   $$P_{A_1B_1}(a,b) = \sum_{a_2,b_2} P(a,a_2,b,b_2) \qquad \text{etc.} \tag{2.28}$$

The proof is also instructive:

Proof. To see that the first set of assumptions implies Eq. (2.18), note that any deterministic (i. e., taking only the values 0 and 1) response function factorizes,

$$\chi_{AB}(a,b,\lambda) = \delta\big((a,b)-(a_0,b_0)\big) = \delta(a-a_0)\delta(b-b_0), \tag{2.29}$$

where a_0 can depend on A, B and λ and similarly for b_0. For a factorizing response function of the form $\chi_{AB}(a,\lambda)\chi_{AB}(b,\lambda)$ locality implies that $\chi_A(a,\lambda) = \chi_{AB}(a,\lambda)$ is independent of B and $\chi_B(b,\lambda) = \chi_{AB}(b,\lambda)$ is independent of A. For the second set of assumptions argue again that the response function for a does not depend on B etc. and then use the fact that we can always replace a non-deterministic response function

with a deterministic one at the cost of introducing an additional hidden variable [128]: Let $\widetilde{\lambda} = (\lambda, \xi_A, \xi_B)$, where ξ_A and ξ_B are independently uniformly distributed on $[0, 1]$, and define

$$\widetilde{\chi}_A(a, \widetilde{\lambda}) = \widetilde{\chi}_A\big(a, (\lambda, \xi_A, \xi_B)\big) = \begin{cases} 1 & \text{if } \xi_A \leq \chi_A(a, \lambda), \\ 0 & \text{otherwise,} \end{cases} \tag{2.30}$$

and similarly for $\widetilde{\chi}_B$. This gives the same probability table. Lastly show that the third set of assumptions is equivalent to Eq. (2.18): Given Eq. (2.18), we can define a joint probability distribution simply as

$$P(a_1, a_2, b_1, b_2) = \int \mathrm{d}\lambda\, \mu(\lambda) \chi_{A_1}(a_1, \lambda) \chi_{A_2}(a_2, \lambda) \chi_{B_1}(b_1, \lambda) \chi_{B_2}(b_2, \lambda). \tag{2.31}$$

Conversely, given the joint probability distribution let $\lambda = (a_1, a_2, b_1, b_2)$,

$$\mu(\lambda) = \mu\big((a_1, a_2, b_1, b_2)\big) = P(a_1, a_2, b_1, b_2), \tag{2.32}$$

and define $\chi_A(a, \lambda)$ by $\chi_{A_1}(a, \lambda) = \delta(a - a_1)$ etc. [34]. □

Allowing for P_{AB} arbitrary probability distributions on (a, b), the left-hand side of the CHSH inequality can reach the value 4. This is known as the *algebraic bound*. The maximal value allowed by quantum mechanics is given by the *Tsirelson bound*[12] of $2\sqrt{2}$ (see Ref. 21). One might presume that this discrepancy could be explained by the fact that quantum mechanics is nonsignalling. This is not the case, however: A model assigning the following probabilities to the measurement results is easily seen to be nonsignalling:

$$P_{A_1B_1}(a, b) = P_{A_1B_2}(a, b) = P_{A_2B_1}(a, b) = \begin{cases} \frac{1}{2} & \text{if } a = b, \\ 0 & \text{if } a \neq b, \end{cases} \tag{2.33}$$

$$P_{A_2B_2}(a, b) = \begin{cases} 0 & \text{if } a = b, \\ \frac{1}{2} & \text{if } a \neq b. \end{cases} \tag{2.34}$$

Yet it reaches the algebraic bound of 4, which is thus equal to the *nonsignalling bound*. (For other Bell inequalities these two bounds are in general not equal.) The model in Eqs. (2.33) and (2.34) is known as a *nonlocal box or Popescu-Rohrlich (PR) box* [90]. In Section 6.1 of this thesis possible conditions on nonsignalling probabilistic theories that restrict the violation of Bell inequalities will be discussed in a multipartite setting.

In a successful Bell experiment the left-hand side of the inequality takes a value higher than the local hidden variable bound. To show convincingly that Nature cannot be explained by local hidden variable model, one has to rule out the possibility that this value is only an effect of statistical fluctuations. One way to do so is to put an error bar on the experimental violation. This strategy will be pursued in Chapter 3 of this thesis. We will see that different Bell inequalities require the experiment to be repeated for different numbers of times to reach the same confidence level. In this sense, Bell inequalities differ in statistical strength. A rigorous definition of the statistical strength

[12] The name *Tsirelson* is also spelt *Cirel'son*.

of Bell inequalities, based on the theory of statistical hypothesis testing, was given in Ref. 27. That article inspired the work on strategies for the discrimination of different classes of states in Chapter 4.

2.2.2 Genuine multiparty nonlocality

In a multipartite setting Bell inequalities can also be used to rule out models that contain a certain type of nonlocality. For three parties, in addition to local hidden variable models, which have probability tables of the form

$$P_{ABC}(a,b,c) = \int \mathrm{d}\lambda\, \mu(\lambda)\chi_A(a,\lambda)\chi_B(b,\lambda)\chi_C(c,\lambda), \tag{2.35}$$

we can also define *hybrid local-nonlocal hidden variable models*, for which [24]

$$\begin{aligned} P_{ABC}(a,b,c) = q_{12} &\int \mathrm{d}\lambda\, \mu_{12}(\lambda)\chi_{AB}(a,b,\lambda)\chi_C(c,\lambda) \\ &+ q_{13} \int \mathrm{d}\lambda\, \mu_{13}(\lambda)\chi_{AC}(a,c,\lambda)\chi_B(b,\lambda) \\ &+ q_{23} \int \mathrm{d}\lambda\, \mu_{23}(\lambda)\chi_{BC}(b,c,\lambda)\chi_A(a,\lambda). \end{aligned} \tag{2.36}$$

Here the q_{ij} are convex weights, $0 \leq q_{ij} \leq 1$ and $q_{12} + q_{13} + q_{23} = 1$. Note how this parallels the definition of biseparability (see Section 2.1.1). For the threeparty *Svetlichny inequality* [105]

$$\begin{aligned} \langle A_1B_1C_2\rangle + \langle A_1B_2C_1\rangle + \langle A_2B_1C_1\rangle - \langle A_2B_2C_2\rangle \\ + \langle A_1B_2C_2\rangle + \langle A_2B_1C_2\rangle + \langle A_2B_2C_1\rangle - \langle A_1B_1C_1\rangle \leq 4 \end{aligned} \tag{2.37}$$

the bound of 4 holds not only for local, but also for hybrid models. The proof of this results is short:

Proof. First we prove the inequality for a specific bipartite split. For example, assume $q_{12} = 1$ and $q_{13} = q_{23} = 0$ in Eq. (2.36). We consider the parties A and B as one party D with measurements $D_1 = A_1B_1$, $D_2 = A_2B_1$, $D_1' = A_1B_1$ and $D_2' = A_2B_2$. We can then write the Svetlichny inequality as the sum of two CHSH inequalities [9]

$$\langle D_1C_2\rangle + \langle D_2C_2\rangle + \langle D_2C_1\rangle - \langle D_1C_1\rangle \leq 2, \tag{2.38}$$

$$\langle D_1'C_2\rangle + \langle D_1'C_1\rangle + \langle D_2'C_1\rangle - \langle D_2'C_2\rangle \leq 2. \tag{2.39}$$

This proves the inequality for the chosen bipartite split. We treat the remaining bipartite splits analogously and then take convex combinations. □

The quantum mechanical bound is $4\sqrt{2}$, it can be attained with [24, 84]

$$A_1 = B_1 = \sigma_x, \quad A_2 = B_2 = \sigma_y, \quad C_1 = -\frac{1}{\sqrt{2}}(\sigma_x + \sigma_y), \quad C_2 = \frac{1}{\sqrt{2}}(\sigma_x - \sigma_y) \tag{2.40}$$

and the GHZ state. The nonsignalling bound is equal to the algebraic bound of 8. An example of a nonsignalling model reaching this bound will be given later in Eqs. (6.10) and (6.11).

Just as any Bell inequality requires entanglement for its violation, the Svetlichny inequality requires genuine threepartite entanglement.[13] The inequality has been generalized to arbitrary numbers of parties and measurement outcomes [9, 24, 103]. In Ref. 73, an experiment demonstrating violation of the threepartite Svetlichny inequality is described. Finally, Ref. 8 proposes two measures to quantify multipartite nonlocality, which are also based on Svetlichny's idea.

2.3 Entropies

2.3.1 Shannon entropy and classical relative entropy

The *Shannon entropy* of a discrete classical random variable with probability distribution $P = (p_1, \ldots, p_m)$ is (see e. g. Ch. 2 in Ref. 25)

$$S(P) = -\sum_{i=1}^{m} p_i \log(p_i), \tag{2.41}$$

where we define $0 \log(0) := \lim_{p \to 0^+} x \log(x) = 0$. Throughout this thesis, log denotes the logarithm to base 2. Instead of "entropy of the random variable" we usually say "entropy of the probability distribution". The entropy quantifies the information that is gained on average if one learns the value of the random variable, of equivalently, the uncertainty about the random variable if its value is still unknown [86, Ch. 11.1]. This interpretation can be justified in various ways: By observing that the entropy gives the expected number of bits per symbol required to store a message (*Shannon's source coding theorem* [25, Thm. 5.4.2]), by postulating axioms for a measure of information and showing that they uniquely define the entropy [25, Problem 2.46] or by arguing that $\log(1/p_i)$ is a reasonable measure for the surprise upon observing the event i (see our discussion of the relative entropy below).

We recall the most important properties of the Shannon entropy (see e. g. Ch. 2 in Ref. 25):

1. $0 \leq S(P) \leq \log(m)$ with $S(P) = 0 \Leftrightarrow p_i = \delta_{i,i_0}$ and $S(P) = \log(m) \Leftrightarrow p_i = 1/m$.
2. *Strict concavity:*

$$S\big(\lambda P + (1-\lambda)Q\big) \geq \lambda S(P) + (1-\lambda) S(Q) \quad \text{for} \quad 0 \leq \lambda \leq 1, \tag{2.42}$$

for $0 < \lambda < 1$ equality holds if and only if $P = Q$.

[13] Interestingly, the Svetlichny inequality, published in 1987, pre-dates the idea of genuine multipartite entanglement. The three-qubit GHZ state also appears in Svetlichny's article.

3. *Subadditivity:* Let $P = (p_{ij})$ be the joint probability distribution of two random variables with marginals $P^{(1)} = (\sum_j p_{ij})$ and $P^{(2)} = (\sum_i p_{ij})$. Then

$$S(P) \leq S(P^{(1)}) + S(P^{(2)}) \tag{2.43}$$

with equality if and only if the variables are independent, $p_{ij} = p_i^{(1)} p_j^{(2)}$.

4. *Grouping rule:* If we decide to consider events 1 and 2 as one event, which leaves us with the probability distribution $\widetilde{P} = (p_1 + p_2, p_3, \ldots, p_m)$, then

$$S(P) = S(\widetilde{P}) + (p_1 + p_2) S(P'), \tag{2.44}$$

where $P' = \left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$. Note that $(p_1 + p_2)S(P')$ is the conditional entropy for the condition that p_1 or p_2 occurs.

The *relative entropy*, or *Kullback-Leibler divergence*, from the probability distribution $P = (p_1, \ldots, p_m)$ to the probability distribution $Q = (q_1, \ldots, q_m)$ is defined as (see e. g. Ch. 2 in Ref. 25)

$$D(P\|Q) = \sum_{i=1}^{m} p_i \log\left(\frac{p_i}{q_i}\right), \tag{2.45}$$

where $0 \log(0/q) := 0$, $0 \log(0/0) := 0$ and $p \log(p/0) := \infty$ for $p > 0$. We list its most important properties (see e. g. Ch. 2 in Ref. 25):

1. *Positive definiteness:* $D(P\|Q) \geq 0$ with $D(P\|Q) = 0 \Leftrightarrow P = Q$.

2. $D(P\|Q) = \infty$ if and only if $q_i = 0$, but $p_i > 0$ for some index i, that is, if there is an event which is impossible according to the second probability distribution, but occurs with nonvanishing probability according to the first one.

3. *Joint convexity* in the arguments:

$$D\big(\lambda P + (1-\lambda)P' \big\| \lambda Q + (1-\lambda)Q'\big) \geq \lambda D(P\|Q) + (1-\lambda) D(P'\|Q') \tag{2.46}$$

for $0 \leq \lambda \leq 1$.

4. *Additivity for independent variables:* If P and Q are joint probability distributions of two independent random variables, $p_{ij} = p_i^{(1)} p_j^{(2)}$ and $q_{ij} = q_i^{(1)} q_j^{(2)}$, then

$$D(P\|Q) = D(P^{(1)}\|Q^{(1)}) + D(P^{(2)}\|Q^{(2)}). \tag{2.47}$$

5. *Grouping rule:* For $\widetilde{P} = (p_1 + p_2, p_3, \ldots, p_m)$ and $\widetilde{Q} = (q_1 + q_2, q_3, \ldots, q_m)$,

$$D(P\|Q) = D(\widetilde{P}\|\widetilde{Q}) + (p_1 + p_2) D(P'\|Q'), \tag{2.48}$$

where $P' = \left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$ and $Q' = \left(\frac{q_1}{q_1+q_2}, \frac{q_2}{q_1+q_2}\right)$.

It is tempting (and often even helpful) to interpret the relative entropy as a distance between probability distributions. Note, however, that it is not a metric, since it is not symmetric and does not satisfy the triangle inequality.[14]

The relative entropy gives an answer to the question [27]: *Given a sample from the probability distribution P, how strongly does it indicate on average that it was indeed drawn from P rather than from some other distribution Q?* Here, only an intuitive explanation of this interpretation will be given (following Sec. II in Ref. 118), a more rigorous justification is postponed until Section 4.2.2.

We begin by defining a measure for *surprise*. If a random event occurs with probability p, our surprise upon observing the event increases monotonically with decreasing p: The less likely an event is, the more surprised we are if we actually see it happen. It is reasonable to postulate that the surprise for an event composed of two independent events is the sum of the surprises. This fixes our measure of surprise as $\log(1/p)$, up to a constant. The average surprise is nothing but the Shannon entropy. Suppose now that the probability distribution describing a collection of events is $P = (p_i)$, but we mistakenly believe it to be $Q = (q_i)$. For example, we may think that a die is fair, when it is actually loaded. Our average surprise is then $\sum_i p_i \log(1/q_i)$, where the surprise depends on the probabilities in which we believe, but the average has to be taken with respect to the correct distribution. We subtract from this the average surprise inherent in the process, which is given by the entropy, and obtain the amount of surprise which is due to our erroneous assumption: $\sum_i p_i \log(1/q_i) - \sum_i p_i \log(1/p_i) = D(P\|Q)$. This line of thought also helps to explain why the relative entropy is not symmetric: If $P = (1/3, 1/3, 1/3)$ and $Q = (1/2, 1/2, 0)$, a single observation of the third event is enough to shatter our belief in the distribution Q, which is reflected by $D(P\|Q) = \infty$. If we swap the roles of the distributions, a larger number of observations is necessary to show with some certainty that the real distribution is Q, and correspondingly $D(Q\|P) = \log(3)$ is finite.

2.3.2 Other classical entropies

By relaxing the postulates which characterize the Shannon entropy other entropies can be defined.

The *Rényi entropy* of the probability distribution $P = (p_1, \ldots, p_m)$ is defined as [92,93]

$$S_q^R(P) = \frac{\log\left[\sum_{i=1}^m (p_i)^q\right]}{1-q}, \qquad q > 0, \quad q \neq 1. \tag{2.49}$$

In the limit $q \to 1$ it gives the Shannon entropy and for $q \to \infty$ the *min-entropy*

$$S_\infty(P) = -\log\left(\max_i p_i\right). \tag{2.50}$$

The Rényi entropy of order $q = 2$ is also called *collision entropy*. Like the Shannon entropy the Rényi entropy satisfies $0 \leq S_q^R(P) \leq \log(m)$ with equality only for the δ-

[14] A simple counterexample to the triangle inequality: Let $P = (0, 1)$, $Q = (1/2, 1/2)$ and $R = (2/3, 1/3)$. Then $D(P\|R) = \log(3)$, but $D(P\|Q) + D(Q\|R) = \log(3) - 1/2$.

and the uniform distribution. It is additive for independent random variables,

$$S_q^R(P) = S_q^R(P^{(1)}) + S_q^R(P^{(2)}) \qquad \text{for} \qquad p_{ij} = p_i^{(1)} p_j^{(2)}. \tag{2.51}$$

Unlike the Shannon entropy it is not concave. As a function of the parameter q it is monotonically decreasing.

The *Tsallis entropy* is defined as[15] [49, 111]

$$S_q^T(P) = \frac{1 - \sum_{i=1}^m (p_i)^q}{q-1}, \qquad q > 1. \tag{2.52}$$

Again, the limit $q \to 1$ gives the Shannon entropy, up to a factor $\ln(2)$. The Tsallis entropy satisfies $0 \leq S_q^T(P) \leq (1 - m^{1-q})/(q-1)$ with equality only for the δ- and the uniform distribution. Like the Shannon entropy it is concave. However, it is not additive for independent random variables. Like the Rényi entropy it is monotonically decreasing as a function of the parameter.

Both the Rényi and the Tsallis entropy are functions of the q-norm of the probability vector,

$$S_q^R(P) = \frac{q}{1-q} \log\big(\|P\|_q\big) \qquad \text{and} \qquad S_q^T(P) = \frac{1 - \|P\|_q^q}{q-1}, \tag{2.53}$$

where

$$\|P\|_q = \Big[\sum_{i=1}^m (p_i)^q\Big]^{1/q}. \tag{2.54}$$

2.3.3 Von Neumann entropy and quantum relative entropy

The *von Neumann entropy* of a quantum state is defined as (see e. g. Ch. 11.3 in Ref. 86)

$$S(\rho) = -\operatorname{Tr}[\rho \log(\rho)]. \tag{2.55}$$

For vanishing eigenvalues of the density matrix the convention $0 \log(0) := 0$ applies. We list the most important properties [86, Ch. 11.3]:

1. $0 \leq S(\rho) \leq \log(d)$ where d is the dimension of the Hilbert space, $S(\rho) = 0$ if and only if ρ is pure, $S(\rho) = \log(d)$ if and only if $\rho = \mathbb{1}/d$.
2. *Strict concavity:* $S(\sum_i p_i \rho_i) \geq \sum_i p_i S(\rho_i)$, where $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$, equality holds if and only if all ρ_i with $p_i > 0$ are equal.
3. *Subadditivity:* $S(\rho) \leq S(\rho^A) + S(\rho^B)$ where $\rho^A = \operatorname{Tr}_B(\rho)$ and $\rho^B = \operatorname{Tr}_A(\rho)$, equality holds if and only if $\rho = \rho^A \otimes \rho^B$.
4. *Triangle or Araki-Lieb inequality:* $S(\rho) \geq |S(\rho^A) - S(\rho^B)|$.

[15] Sometimes the Tsallis entropy is defined for all nonnegative $q \neq 1$ or even all real $q \neq 1$.

In contrast to the classical case, $S(\rho) \geq S(\rho^A)$ where $\rho^A = \mathrm{Tr}_B(\rho)$ does not always hold. If $\rho = |\psi\rangle\langle\psi|$ is pure, $S(|\psi\rangle\langle\psi|) < S(\rho^A)$ if and only if $|\psi\rangle$ is entangled.

We are interested in the behaviour of the von Neumann entropy under measurements. More specifically, suppose we perform a projective measurement, but ignore the result. This corresponds to the map

$$\rho \to \rho' = \sum_i \Pi_i \rho \Pi_i \tag{2.56}$$

where the Π_i are orthogonal projectors with $\sum_i \Pi_i = \mathbb{1}$. Once can show that this process can only increase the entropy of the state [86, Thm. 11.9],

$$S(\rho') \geq S(\rho) \tag{2.57}$$

with equality if and only if $\rho = \rho'$. For generalized measurements the result no longer holds [86, Exercise 11.15]. If the measurement corresponds to a nondegenerate observable, in other words, if the Π_i are one-dimensional projectors, $S(\rho')$ is equal to the Shannon entropy of the measurement probabilities. The above result thus shows

$$S(\rho) \leq S\big((p_1, \ldots, p_d)\big) \quad \text{where} \quad p_i = \langle\psi_i|\rho|\psi_i\rangle \tag{2.58}$$

for any orthonormal basis $|\psi_i\rangle$.

The *quantum relative entropy* is defined as (see e. g. Ch. 11.3.1 in Ref. 86)

$$D(\rho\|\sigma) = \mathrm{Tr}\{\rho[\log(\rho) - \log(\sigma)]\} \tag{2.59}$$

where again $0\log(0) \coloneqq 0$. We list its most important properties [86, Ch. 11.3]:

1. *Klein's inequality* [86, Thm. 11.7]: The quantum relative entropy is positive definite, $D(\rho\|\sigma) \geq 0$ and $D(\rho\|\sigma) = 0 \Leftrightarrow \rho = \sigma$.

2. $D(\rho\|\sigma) = \infty$ if and only if $\ker(\sigma)$ has nontrivial intersection with $\mathrm{supp}(\rho)$.

3. *Joint convexity* in the arguments [86, Thm. 11.12]:

$$D\big(\lambda\rho + (1-\lambda)\rho' \big\| \lambda\sigma + (1-\lambda)\sigma'\big) \geq \lambda D(\rho\|\sigma) + (1-\lambda) D(\rho'\|\sigma') \tag{2.60}$$

   for $0 \leq \lambda \leq 1$.

4. *Additivity for product states:*

$$D(\rho^A \otimes \rho^B \| \sigma^A \otimes \sigma^B) = D(\rho^A\|\sigma^A) + D(\rho^B\|\sigma^B). \tag{2.61}$$

5. *Monotonicity* [86, Thm. 11.17]: $S(\rho^A\|\sigma^A) \leq S(\rho\|\sigma)$ where $\rho^A = \mathrm{Tr}_B(\rho)$ and $\sigma^A = \mathrm{Tr}_B(\sigma)$.

Let

$$D_N(\rho\|\sigma) = \sup_{\{E_i\}} \sum_i \mathrm{Tr}(E_i\rho^{\otimes N})\Big\{\log\big[\mathrm{Tr}(E_i\rho^{\otimes N})\big] - \log\big[\mathrm{Tr}(E_i\sigma^{\otimes N})\big]\Big\} \tag{2.62}$$

be the classical relative entropy of the measurement probabilities maximized over all positive operator-valued measurements acting on N copies of two states. One can show that [118, Sec. II.E]

$$\lim_{N\to\infty} \frac{D_N(\rho\|\sigma)}{N} = D(\rho\|\sigma). \tag{2.63}$$

In this sense, the quantum relative entropy is a measure for the distinguishability of two states.

2.4 Uncertainty relations

2.4.1 Uncertainty principle

Quantum mechanics does not allow to predict a measurement outcome with certainty unless the system is in an eigenstate of the observable being measured. It follows that if two or more observables have no common eigenstate it is not possible to prepare the system such that for each observable only one measurement outcome can occur. We will refer to this fact as the *uncertainty principle*. It can be formulated quantitatively in terms of uncertainty relations.

The first and still the most celebrated uncertainty relation was given by Heisenberg [52] and formulated rigorously by Kennard [64]. It applies to canonically conjugate observables such as position and momentum and states that the product of their variances $\Delta^2(q) = \langle q^2\rangle - \langle q\rangle^2$ and $\Delta^2(p) = \langle p^2\rangle - \langle p\rangle^2$ is lower bounded by a constant,

$$\Delta^2(q)\Delta^2(p) \geq \frac{\hbar}{4}. \tag{2.64}$$

It is usually generalized to arbitrary observables in the form of the *Heisenberg-Robertson uncertainty relation* [94]

$$\Delta^2(A)\Delta^2(B) \geq \frac{1}{4}\big|\langle[A,B]\rangle\big|^2. \tag{2.65}$$

The Heisenberg-Robertson relation can be generalized to the case of more than two observables [95, 110].

There is another aspect of uncertainty, which is discussed in Heisenberg's article [52] as well (see also Refs. 7,17,127): Observing the position a of particle, such as an electron, requires interaction with a photon. By the Compton effect, during this interaction an uncontrolled momentum Δp is transferred to the electron. The spatial resolution Δq of this measurement depends on the wavelength of the photon and is related in this way to the momentum transfer by $\Delta q\Delta p \approx h$. In this sense, the determination of a particle's position with an accuracy Δq results in an uncontrolled change in momentum of the order $\Delta p \approx h/\Delta q$. This thought experiment is known as the *Heisenberg microscope*.

While the conclusion of the thought experiment is of course correct, it is not the immediate content of Eq. (2.64). Unfortunately, this distinction is often ignored.[16] In this thesis, the words *uncertainty principle* and *uncertainty relation* will always be used as in the beginning of this section and never in the sense of the Heisenberg microscope. This may be called *preparation uncertainty*, as opposed to *measurement uncertainty* [127]. Ballentine [7] has proposed the term *statistical dispersion principle*.

Uncertainty relations not only describe a fundamental property of quantum mechanics, but they have also found application in quantum information tasks, for example in quantum cryptography [28, 66, 67] and entanglement detection [42, 43, 53]. Entropic uncertainty relations (see below) have turned out to be particularly useful. More recently, the uncertainty principle has also been formulated in terms of majorization relations [87]. In Ref. 14, Berta et al. derived an entropic uncertainty relation for a system which is entangled to a quantum memory. The access to this memory can then be used to lower the uncertainties of measurements on the system. This relation has been the subject of recent experiments [75, 91].

2.4.2 Entropic uncertainty relations

Uncertainty can also be quantified by the entropy of the measurement probabilities. This approach leads to *entropic uncertainty relations*. For a review see Ref. 124. To fix our notation, let A be an observable with spectral decomposition $A = \sum_i a_i \Pi_i$ with mutually distinct a_i, and let S_X be a classical entropy function. Most often S_X stands for the Shannon entropy S, but the Tsallis entropy S_q^T and the min-entropy S_∞ [see Eqs. (2.52) and (2.50)] will also be used in this thesis. We denote by $S_X(A|\rho)$ the entropy of the measurement probabilities,

$$S_X(A|\rho) = S_X\big((p_1, \ldots, p_m)\big) \quad \text{where} \quad p_i = \mathrm{Tr}(\Pi_i \rho). \tag{2.66}$$

We are interested in uncertainty relations for a family of observables $\{A_1, \ldots, A_L\}$ of the form

$$\frac{1}{L} \sum_{k=1}^{L} S_X(A_k|\rho) \geq c_{\{A_k\}}. \tag{2.67}$$

The lower bound $c_{\{A_k\}}$ may depend on the observables, but preferable is independent of the state. For a given set of observables an uncertainty relation is called *tight* if a state ρ_0 exists that attains the lower bound, $1/L \sum_{k=1}^{L} S_X(A_k|\rho_0) = c_{\{A_k\}}$.

The entropy $S_X(A|\rho)$ depends only on the eigenstates, but not on the eigenvalues of the observable A, as long as they are nondegenerate. It is thus independent of the

[16] In the words of R. F. Werner [127]: "Heisenberg's Uncertainty Relation $(\Delta Q)(\Delta P) \geq \hbar/2$ is one of the most fundamental features of quantum theory, and is taught in even the most basic course on the subject. All too often, however, teachers succumb to the persistent bad habit of proving the relations as an inequality on variances for arbitrary state preparations, but then to go on to explain their 'physical meaning' in terms of a perturbation of the momentum of a particle caused by an approximate position measurement. Since the usual proof contains nothing of that sort, attentive students quickly get the impression that quantum uncertainty rubs off on their teachers as some kind of conceptual fuzziness."

labelling of the measurement results, which in finite dimensions is to some extent arbitrary. Therefore the entropy can be regarded as a more natural measure of uncertainty than the variance, at least in the case of finitely many measurement outcomes, which we consider here.[17] For the same reason, we will not distinguish between a nondegenerate observable A and its eigenbasis $\mathcal{A}$.

The use of entropies also allows uncertainty bounds $c_{\{A_k\}}$ which hold for all states. For Deutsch [30] this is the main argument in favour of entropic uncertainty relations. By contrast, any uncertainty relation whose left-hand side is the product of variances will always be trivial for any eigenstate of any of the observables (assuming a finite-dimensional Hilbert space).

For any measurement basis $\mathcal{A}$ of length m the Shannon entropy satisfies $S(\mathcal{A}|\rho) \leq \log(m)$. If we choose for ρ one of the states of the bases $\mathcal{A}_k$, we have $1/L \sum_{k=1}^{L} S(\mathcal{A}_k|\rho) \leq \log(m)(L-1)/L$, because in this case the entropy is zero for one basis and upper bounded by $\log(m)$ for the remaining $L-1$ bases. This implies that for the Shannon entropy the right-hand side of Eq. (2.67) cannot exceed $\log(m)(L-1)/L$. An uncertainty relation that reaches this limit is called *maximally strong*, and the corresponding measurements are called *maximally incompatible* [124]. In more physical terms, maximal incompatibility means that if the outcome of one measurement is certain, for any of the remaining measurements all outcomes are equally likely.

A classic result on entropic uncertainty relations is the connection between maximal incompatibility and mutual unbiasedness. Two orthonormal bases $|a_i\rangle$ and $|b_i\rangle$, $i = 1, \ldots, d$, are called *mutually unbiased* if

$$|\langle a_i|b_j\rangle| = \frac{1}{\sqrt{d}} \qquad \forall\, i, j. \tag{2.68}$$

An almost trivial example is given by the eigenbases of the three Pauli matrices. Pairwise mutual unbiasedness of the eigenbases is a necessary condition for maximal incompatibility. Perhaps surprisingly, for more than two observables this condition is not sufficient [124].

Mutually unbiased bases have become a subject of research in their own right. A central question is the maximal number of such bases for a given Hilbert space dimension. It is known that in a d-dimensional space there are at most $d+1$ mutually unbiased bases, and for the case that d is a prime power an explicit construction has been found [10, 131]. In general, though, the problem is still unsolved [124].

The most prominent example of an entropic uncertainty relation is the *Maassen-Uffink relation:* For any two measurement bases $\mathcal{A} = \{|a_i\rangle\}$ and $\mathcal{B} = \{|b_i\rangle\}$,

$$\frac{1}{2}\big[S(\mathcal{A}|\rho) + S(\mathcal{B}|\rho)\big] \geq -\log\big(\max_{i,j} |\langle a_i|b_j\rangle|\big). \tag{2.69}$$

This relation was shown by Maassen and Uffink [79,80]; for the special case of mutually unbiased bases it had been conjectured before by Kraus [69]. It follows from a similar result for Rényi entropies, which in turn is a consequence of the Riesz-Thorin theorem.

[17] Entropic uncertainty relations for continuous variables also exist (see e. g. Ref. 15).

In the case of mutually unbiased bases the Maassen-Uffink relation is maximally strong and thus tight. Equality holds for any of the basis states $|a_i\rangle$ and $|b_i\rangle$. (Note that for two arbitrary observables A and B the entropy sum $S(A|\rho) + S(B|\rho)$ is in general not minimized by an eigenstate of either of them [37].) If the bases are not mutually unbiased, the Maassen-Uffink relation is in general not tight [119]. While most uncertainty relations in the literature apply to projective measurements only, the Maassen-Uffink relation has been generalized to arbitrary positive operator-valued measurements; for the result see Ref. 70.

Sections 5.1 and 5.2 of this thesis are concerned with measurement bases that are not mutually unbiased, but for which the Maassen-Uffink relation is still tight.

Results on entropic uncertainty relations for more than two measurements are comparatively rare [124]. Exceptions are Refs. 81 and 123. The latter reference deals with pairwise anticommuting observables with two distinct eigenvalues. In Section 5.3 those results will be generalized, leading to a systematic construction of uncertainty relations for these observables using variances and different entropy functions. In this setting the relative strengths of entropic and variance-based uncertainty relations will be compared.

2.5 Stabilizer formalism

2.5.1 Stabilizer states

The stabilizer formalism allows to describe certain many-qubit states in an efficient way. It will be used in all chapters of this thesis. Originally developed for quantum error correction [39, 40], it has proven useful in many areas if quantum information theory, including quantum computation, where it is used to prove the Gottesman-Knill theorem [86, Thm. 10.7], and entanglement detection [109].

This section is based on Ch. 10.5 in Ref. 86 and on Refs. 50, 112.

We define a *Pauli operator* on n qubits as any n-fold tensor product of Pauli matrices, including the identity. The n-qubit Pauli operators generate (under the usual matrix multiplication) a group, which consists of all Pauli operators multiplied with phase factors $\{+1, -1, +\mathrm{i}, -\mathrm{i}\}$. We call this group the *Pauli group* and denote it by $\mathcal{G}_n$. For example, for one qubit we have

$$\mathcal{G}_1 = \{\pm\mathbb{1}, \pm\mathrm{i}\mathbb{1}, \pm\sigma_x, \pm\mathrm{i}\sigma_x, \pm\sigma_y, \pm\mathrm{i}\sigma_y, \pm\sigma_z, \pm\mathrm{i}\sigma_z\}. \tag{2.70}$$

Obviously, $\mathcal{G}_n$ has 4^{n+1} elements, which are Hermitian or anti-Hermitian. Any two elements either commute or anticommute.

The idea of the stabilizer formalism is to describe a pure state, rather than by its state vector, by a subgroup of the Pauli group such that the state is an eigenvector of all group elements. One can show [112, Prop. 2.4] that the elements of a subgroup S of $\mathcal{G}_n$ have a nontrivial simultaneous eigenspace V_S with eigenvalue $+1$ if and only if S is Abelian and $-\mathbb{1} \notin S$. In particular, these conditions are sufficient to show that all group elements are Hermitian. We call such a subgroup a *stabilizer group* (or *stabilizer*

for short) and say that the space V_S is *stabilized* by it. The group elements are called the *stabilizing operators* of the space.

A stabilizer group can be described by a set of generators $g_1, \ldots, g_m$. We choose the generators to be independent in the sense that removing any of them makes the group smaller. Any group element M has a unique representation as a product of generators,

$$M = g_1^{x_1} \cdots g_m^{x_m}, \qquad x_i \in \{0, 1\}. \tag{2.71}$$

This shows that $|S| = 2^m$ and $1 \leq m \leq n$. The stabilized space has dimension 2^{n-m} (see Prop. 10.5 in Ref. 86 or Prop. 2.11 in Ref. 112). Adding a generator divides the dimension by two.

After normalizing it, we identify the projector onto the stabilized space V_S with a *stabilizer state* ρ_S. (Most often only pure states are considered, but we will be dealing with mixed stabilizer states in Chapter 6.) The state can be expressed by the generators g_i or by the stabilizing operators $M_i \in S$ as

$$\rho_S = \frac{1}{2^n} \prod_{i=1}^{m} (\mathbb{1} + g_i) = \frac{1}{2^n} \sum_{i=1}^{2^m} M_i. \tag{2.72}$$

The rank of the state is $\operatorname{rank}(\rho_S) = 2^{n-m}$. Pure states correspond to $m = n$,

$$|\psi_S\rangle\langle\psi_S| = \frac{1}{2^n} \prod_{i=1}^{n} (\mathbb{1} + g_i) = \frac{1}{2^n} \sum_{i=1}^{2^n} M_i. \tag{2.73}$$

The stabilizer state $|\psi_S\rangle$ is uniquely defined (up to a phase, of course) by the condition

$$g_i |\psi_S\rangle = |\psi_S\rangle, \qquad i = 1, \ldots, n. \tag{2.74}$$

Equation (2.72) gives the expansion of the density matrix in the Pauli basis

$$\rho_S = \frac{1}{2^n} \sum_{i_1, \ldots, i_n = 0}^{3} t_{i_1, \ldots, i_n} \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}. \tag{2.75}$$

The expansion coefficient comprise what is sometimes called the *correlation tensor* $t_{i_1,\ldots,i_n}$. As they are given by the expectation values of the Pauli operators,

$$t_{i_1,\ldots,i_n} = \operatorname{Tr}(\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n} \rho_S) = \begin{cases} +1 & \text{if } \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n} \in S, \\ -1 & \text{if } -\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n} \in S, \\ 0 & \text{otherwise,} \end{cases} \tag{2.76}$$

the stabilizing operators can be understood as a description of the correlations of the state.

Like any set of Hermitian, pairwise commuting operators, the elements of a stabilizer group have a basis of common eigenstates. If the group has the maximal cardinality of 2^n, this basis is uniquely determined. The stabilizer state is one of the basis states. We refer to the basis as *stabilizer basis*.

2.5.2 Graph states

Graph states are a special class of stabilizer states, whose underlying interaction pattern can be described by a simple undirected graph. It turns out that any pure stabilizer state is equivalent under local Clifford operations (a subgroup of local unitaries) to a graph state. For a review on graph states see Ref. 50.

A finite undirected *graph* consists of a finite number of vertices and finitely many edges connecting pairs of vertices. We require all our graphs to be simple, i. e., we do not allow loops (edges connecting a vertex to itself) or multiple edges. Examples are shown in Fig. 2.2. Every vertex represents a qubit. Edges represent pair interactions that have created the graph state from a product state in the following sense [50, Sec. 2.1]: With the empty or completely unconnected graph we associate the product state

$$|G_0\rangle = |+\rangle^{\otimes n} \qquad \text{where} \qquad |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle). \tag{2.77}$$

For every edge we apply a controlled phase gate on the corresponding pair of qubits,

$$|G\rangle = \prod_{(i,j)\in E} C_{ij}\, |+\rangle^{\otimes n}, \qquad \text{where} \qquad C_{ij} = |0\rangle_i\langle 0| \, \mathbb{1} + |1\rangle_i\langle 1| \, \sigma_z^{(j)} \tag{2.78}$$

and the product is over the set all edges E. As usual

$$\sigma_z^{(j)} = \mathbb{1}^{\otimes(j-1)} \otimes \sigma_z \otimes \mathbb{1}^{\otimes(n-j)} \tag{2.79}$$

is the Pauli matrix σ_z acting on qubit j, and the projectors $|0\rangle_i\langle 0|$ and $|1\rangle_i\langle 1|$ are defined similarly. Note that for the controlled phase gate it does not matter which is the control qubit and which the target. Up to local unitaries the controlled phase gate is an Ising interaction,

$$C_{ij} = \mathrm{e}^{-\mathrm{i}\pi/4} \exp\bigl[\mathrm{i}\frac{\pi}{4}\sigma_z^{(i)}\bigr] \exp\bigl[\mathrm{i}\frac{\pi}{4}\sigma_z^{(j)}\bigr] \exp\bigl[-\mathrm{i}\frac{\pi}{4}H_{ij}\bigr] \tag{2.80}$$

where $H_{ij} = \sigma_z^{(i)}\sigma_z^{(j)}$.

The generators of the stabilizer group can immediately be read off the graph: With each vertex i we associate the operator

$$g_i = \sigma_x^{(i)} \prod_{j\in N(i)} \sigma_z^{(j)}, \tag{2.81}$$

where the product is over the neighbourhood $N(i)$ of the vertex i, that is, over all vertices directly connected to it by an edge. The g_i defined in this way commute pairwise and generate the stabilizer group [50, Prop. 2].

As an example we consider the graph Fig. 2.2 (b). The corresponding generators are

$$g_1 = XZZ, \quad g_2 = ZX\mathbb{1}, \quad g_3 = Z\mathbb{1}X, \tag{2.82}$$

where X, Y and Z denote the Pauli matrices σ_x, σ_y and σ_z and tensor product signs have been omitted. Up to a local Clifford operation the corresponding graph state is the three-qubit GHZ state.

Occasionally we need an explicit expression for the coefficients of a graph state with respect to the standard basis. The *adjacency matrix* of an n-vertex graph is the symmetric $n \times n$-matrix Γ whose entry Γ_{ij} is 1 if there is an edge between vertices i and j and 0 if there is not. It thus provides a complete description of the graph state. We denote the basis vectors of the standard basis as $|\mathbf{x}\rangle = |x_1\rangle \otimes \cdots \otimes |x_n\rangle$, where $\mathbf{x} = (x_1, \ldots, x_n)$ with $x_i \in \{0, 1\}$ is an n-bit string. The representation of the graph state in the standard basis is [112, Prop. 2.14]

$$|G\rangle = \frac{1}{2^{n/2}} \sum_{\mathbf{x}} (-1)^{\sum_{i<j} x_i \Gamma_{ij} x_j} |\mathbf{x}\rangle, \tag{2.83}$$

where the sum is over all n-bit strings. (Linear algebra and quadratic forms over the finite field $\mathbb{F}_2$ are powerful tools to study stabilizer states and Clifford operations [29, 112], which will not be used much in this thesis, though.)

The generators defined by a graph have a complete basis of common eigenstates, namely, the stabilizer basis, which in this case is called *graph state basis* [50, Prop. 3]. We introduce the notation $|G, \mathbf{x}\rangle$ for the basis states, where $\mathbf{x} = (x_1, \ldots, x_n)$ with $x_i \in \{0, 1\}$ is an n-bit string encoding the eigenvalues,

$$g_i |G, \mathbf{x}\rangle = (-1)^{x_i} |G, \mathbf{x}\rangle \tag{2.84}$$

and

$$|G, \mathbf{x}\rangle\langle G, \mathbf{x}| = \frac{1}{2^n} \prod_{i=1}^{n} \big[\mathbb{1} + (-1)^{x_i} g_i\big]. \tag{2.85}$$

The graph state itself corresponds to $\mathbf{x} = (0, \ldots, 0)$. Later we will also use the fact the remaining basis states can be obtained from the graph state by applying σ_z on a subset of the qubits,

$$|G, \mathbf{x}\rangle = \prod_{i=1}^{n} \big(\sigma_z^{(i)}\big)^{x_i} |G\rangle = \prod_{(i,j) \in E} C_{ij} \, |\mathbf{x}\rangle_x \tag{2.86}$$

where

$$|\mathbf{x}\rangle_x = \frac{1}{2^{n/2}} \bigotimes_{i=1}^{n} \big[|0\rangle + (-1)^{x_i} |1\rangle\big]. \tag{2.87}$$

2.5.3 Local equivalence of stabilizer states

To study the equivalence of stabilizer states under different local operations let us first introduce a special class of local unitaries: The *local Clifford group* $C_1^{\otimes n}$ on n qubits is defined as the group of local unitaries that map the Pauli group onto itself under conjugation,

$$C_1^{\otimes n} = \{U \in U(2)^{\otimes n} \mid U \mathcal{G}_n U^\dagger = \mathcal{G}_n\}. \tag{2.88}$$

As our notation suggests, it is the n-fold tensor product of the Clifford group of one qubit. Up to global phase factors, the latter is generated by the Hadamard gate $H = 2^{-1/2} \left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$ and the phase gate $S = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{smallmatrix}\right)$. (For the proof see Thm. 10.6 in Ref. 86 or

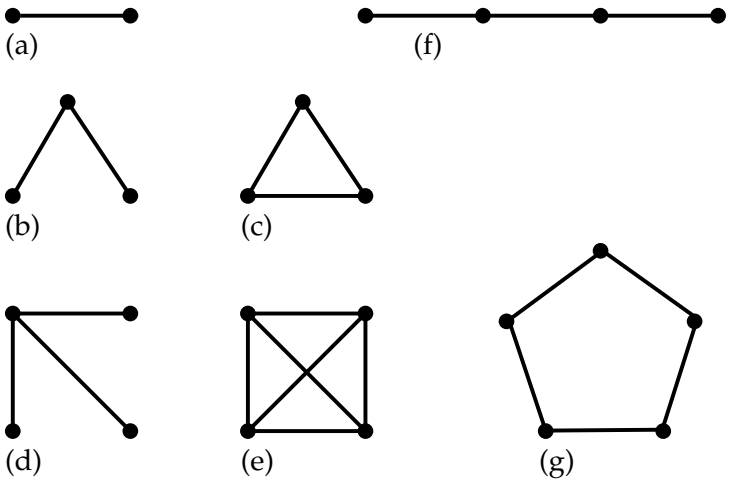

Figure 2.2: Examples of graphs. Up to local unitaries, the corresponding graph states are: **(a)** Bell state, **(b), (c)** three-qubit GHZ state, **(d), (e)** four-qubit GHZ state, **(f)** four-qubit linear cluster state, **(g)** five-qubit ring cluster state.

Props. 3.5 and 3.6 in Ref. 112). By definition local Clifford operations map stabilizer states onto stabilizer states. We call two states *LC-equivalent* if they can be mapped onto each other by a local Clifford operation. It turns out that any pure stabilizer state is LC-equivalent to a graph state. (This was first shown in Refs. 41, 98. Proofs can also be found in Ref. 114 and in Corr. 3.16 of Ref. 112.) Therefore, for many purposes it suffices to consider only graph states.

There is a one-to-one correspondence between graphs and graph states. However, the states defined by two different graphs can be equivalent under local operations. It has been shown that two graph states are SLOCC-equivalent if and only if they are LU-equivalent (see Ref. 115 or Corr. 3.4 in Ref. 112). A conjecture stating that LU equivalence also equals LC equivalence has been disproven [59].

Two graph states are LC-equivalent if and only if their graphs can be transformed into each other by a sequence of local complementations (see Ref. 114 or Prop. 5 in Ref. 50). As graph states are mostly studied for their entanglement properties, LC-equivalent states are often identified. However, in Chapter 4 we will not make this identification.

2.5.4 Examples of graph states

To illustrate the theory outlined above, we give several examples of graph states with small numbers of qubits.

For two qubits there is only one nonempty graph [see Fig. 2.2 (a)]. The corresponding generating operators are

$$g_1 = XZ, \quad g_2 = ZX; \tag{2.89}$$

they define the stabilizer group

$$S_{\text{Bell}} = \{\mathbb{1}\mathbb{1},\ XZ,\ ZX,\ YY\}. \tag{2.90}$$

The corresponding graph state is LC-equivalent to any Bell state.

For three qubits there are two different connected graphs, though they are related by a local complementation. We call Fig. 2.2 (b) the star graph and Fig. 2.2 (c) the complete or fully connected graph. They have generating operators

$$g_1 = XZZ, \quad g_2 = ZX\mathbb{1}, \quad g_3 = Z\mathbb{1}X \quad \text{(star graph)} \tag{2.91}$$

and

$$g_1 = XZZ, \quad g_2 = ZXZ, \quad g_3 = ZZX \quad \text{(complete graph)}, \tag{2.92}$$

respectively. The groups defined by these sets of generators are both LC-equivalent to the group

$$S_{\text{GHZ}_3} = \{\mathbb{1}\mathbb{1}\mathbb{1},\ \mathbb{1}ZZ,\ Z\mathbb{1}Z,\ ZZ\mathbb{1},\ XXX,\ -XYY,\ -YXY,\ -YYX\}, \tag{2.93}$$

which is the stabilizer group of the three-qubit GHZ state

$$|\text{GHZ}_3\rangle = \frac{1}{\sqrt{2}}\big(|000\rangle + |111\rangle\big). \tag{2.94}$$

For four qubits there are two LC-inequivalent graph states. Again, the graph states defined by the star graph Fig. 2.2 (d) and the complete graph Fig. 2.2 (e) are LC-equivalent to the GHZ state

$$|\text{GHZ}_4\rangle = \frac{1}{\sqrt{2}}\big(|0000\rangle + |1111\rangle\big), \tag{2.95}$$

which has the stabilizer group

$$\begin{aligned} S_{\text{GHZ}_4} = \{&\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1},\ \mathbb{1}\mathbb{1}ZZ \text{ and permutations},\ ZZZZ,\\ &XXXX,\ -XXYY \text{ and permutations},\ YYYY\}, \end{aligned} \tag{2.96}$$

where "and permutations" stands for all permutations of qubits that give distinct operators, for example

$$\mathbb{1}\mathbb{1}ZZ \text{ and permutations} = \mathbb{1}\mathbb{1}ZZ,\ \mathbb{1}Z\mathbb{1}Z,\ \mathbb{1}ZZ\mathbb{1},\ Z\mathbb{1}\mathbb{1}Z,\ Z\mathbb{1}Z\mathbb{1},\ ZZ\mathbb{1}\mathbb{1}. \tag{2.97}$$

The state defined by the four-qubit linear graph Fig. 2.2 (f) is LC-equivalent to the *four-qubit linear cluster state*

$$|\text{C}_4\rangle = \frac{1}{2}\big(|0000\rangle + |0011\rangle + |1100\rangle - |1111\rangle\big) \tag{2.98}$$

with stabilizer group

$$\begin{aligned} S_{C_4} = \{&\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1},\ \mathbb{1}\mathbb{1}ZZ,\ ZZ\mathbb{1}\mathbb{1},\ ZZZZ,\\ &XYXY,\ XYYX,\ YXXY,\ YXYX,\\ &\mathbb{1}ZXX,\ Z\mathbb{1}XX,\ XX\mathbb{1}Z,\ XXZ\mathbb{1},\\ &-\mathbb{1}ZYY,\ -Z\mathbb{1}YY,\ -YY\mathbb{1}Z,\ -YYZ\mathbb{1}\}. \end{aligned} \tag{2.99}$$

This state is not LC-equivalent to the GHZ state.

For five qubits there are four LC-inequivalent graph states. In this thesis only the *five-qubit ring cluster state*, which is defined by the graph Fig. 2.2 (g), will be needed. After application of a suitable LC operation its state vector is

$$\begin{aligned} |\mathrm{R}_5\rangle = \frac{1}{2\sqrt{2}}\big(&|00000\rangle + |00110\rangle - |01011\rangle + |01101\rangle\\ &+ |10001\rangle - |10111\rangle + |11010\rangle + |11100\rangle\big) \end{aligned} \tag{2.100}$$

and its stabilizer group is

$$\begin{aligned} S_{\mathrm{R}_5} = \{&\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1},\ ZZ\mathbb{1}\mathbb{1}Z,\ XXX\mathbb{1}\mathbb{1},\ \mathbb{1}ZZZ\mathbb{1},\\ &\mathbb{1}\mathbb{1}XXZ, X\mathbb{1}\mathbb{1}ZX,\ -YYX\mathbb{1}Z,\ Z\mathbb{1}ZZZ,\\ &ZZXX\mathbb{1},\ -YZ\mathbb{1}ZY,\ -XYYZ\mathbb{1},\ XX\mathbb{1}XZ,\\ &\mathbb{1}XXZX,\ -\mathbb{1}ZYYZ,\ XZZ\mathbb{1}X,\ X\mathbb{1}XYY,\\ &-YXYZZ,\ -YY\mathbb{1}X\mathbb{1},\ -ZYXZY,\ -Z\mathbb{1}YY\mathbb{1},\\ &-Y\mathbb{1}Z\mathbb{1}Y,\ YZXYX,\ -XYZYZ,\ -\mathbb{1}YY\mathbb{1}X,\\ &\mathbb{1}X\mathbb{1}YY,\ XZYXY,\ -YXZY\mathbb{1},\ -ZXY\mathbb{1}Y,\\ &ZY\mathbb{1}YX,\ Y\mathbb{1}YXX,\ \mathbb{1}YZXY,\ -ZXZXX\}. \end{aligned} \tag{2.101}$$

For up to eight qubits, all LC-inequivalent graph states have been classified [18, 51].

2.6 Classical exponential families of interaction spaces

2.6.1 Hierarchy of exponential families

Exponential families provide a classification of probability distributions based on the interactions between parts of a system. The exponential families considered here consist of all probability distributions that can be written as thermal distributions of classical Hamiltonians containing at most k-party interactions. The notion of exponential families naturally leads to a quantification of degrees of interaction. These measures are motivated by the questions: *How far is a probability distribution from a thermal distribution of a k-party Hamiltonian?*, and: *How much information is contained in its k-party, but not in its* $(k-1)$*-party marginals?*

In this section we will review the definitions and results that will be needed later in this thesis. Much of the previous work on this subject has been done within the

framework of information geometry, where methods of differential geometry are used to study families of probability distributions [3]. However, from the point of view of information geometry we will always be dealing with very special cases. As it turns out, the results we need can be obtained on a more elementary level using methods from statistical physics. This thesis therefore makes little use of the language of differential geometry. Only Section 2.6.4 is devoted to putting the results in the larger context of information geometry.

We follow Refs. 2, 62. We consider probability distributions on a configuration space which has a product structure,

$$\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n. \tag{2.102}$$

The $\mathcal{X}_i$ are finite sets; for our purposes it suffices to consider $\mathcal{X}_i = \{0, 1\}$. The elements of the space $\mathcal{X}$ will be denoted as $\mathbf{x} = (x, \ldots, x_n)$ where $x_i \in \mathcal{X}_i$. It is helpful to think of the $\mathbf{x}$ as the configurations of a classical dynamical system of n particles.

Given a probability distribution P on $\mathcal{X}$, one can ask if it can be written as a *thermal or Gibbs distribution* of a "Hamiltonian" H,

$$P(\mathbf{x}) = \frac{1}{Z} \mathrm{e}^{H(\mathbf{x})} \qquad \text{where} \qquad Z = \sum_{\mathbf{x}} \mathrm{e}^{H(\mathbf{x})}. \tag{2.103}$$

The Hamiltonian can be any real-valued function on $\mathcal{X}$. It should not be construed as describing an actual physical system. We therefore could take the liberty of assuming the inverse temperature to be unity and reversing the sign in the exponent. If we put no constraints on the Hamiltonian, any probability distribution with full support[18] is thermal. A Hamiltonian can be found by taking the component-wise logarithm of the probability vector. In the following the set of probability distributions on $\mathcal{X}$ with full support will be denoted as $\mathcal{P}(\mathcal{X})$.

Let $V = \{1, \ldots, n\}$ be the set of all subsystems (or parties). Consider the class of Hamiltonians acting only on a subset $A \subseteq V$ of the parties

$$\mathcal{I}_A = \big\{ H \colon \mathcal{X} \to \mathbb{R} \;\big|\; H(\mathbf{x}) = H(\mathbf{x}') \text{ if } x_i = x_i' \text{ for all } i \in A \big\}. \tag{2.104}$$

We define the class of *k-party Hamiltonians* as

$$\mathcal{I}_k = \Big\{ H \colon \mathcal{X} \to \mathbb{R} \;\Big|\; H = \sum_{|A|=k} H_A \text{ where } H_A \in \mathcal{I}_A \Big\}. \tag{2.105}$$

We are interested in the set of all possible thermal distribution of k-party Hamiltonians

$$\mathcal{E}_k = \Big\{ P \;\Big|\; P(\mathbf{x}) = \frac{\mathrm{e}^{H(\mathbf{x})}}{\sum_{\mathbf{x}'} \mathrm{e}^{H(\mathbf{x}')}} \text{ where } H \in \mathcal{I}_k \Big\}. \tag{2.106}$$

We call the sets $\mathcal{E}_k$ *exponential families of interaction spaces*. The general form of an exponential family will be given later in Eq. (2.143). The exponential families considered in this thesis will always be of the form of Eq. (2.106).

[18] The support of a probability distribution P on $\mathcal{X}$ is defined as $\operatorname{supp}(P) = \{\mathbf{x} \in \mathcal{X} \mid P(\mathbf{x}) > 0\}$.

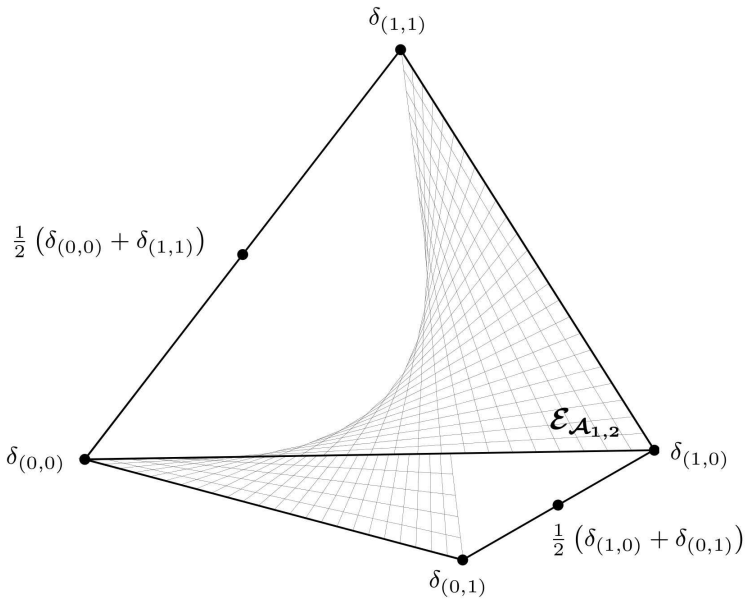

Figure 2.3: The exponential family $\mathcal{E}_1$ (here labelled $\mathcal{E}_{\mathcal{A}_{1,2}}$) in the simplex of probability distributions on $\mathcal{X} = \{0,1\}^2$. The symbol δ stands for the Kronecker delta, for example $\delta_{(0,0)} = \delta_{x_1,0}\delta_{x_2,0}$. (Figure taken from Ref. 63.)

In this way, we obtain a series of nested families (or *hierarchy*)

$$\mathcal{E}_1 \subset \mathcal{E}_2 \subset \cdots \subset \mathcal{E}_n. \tag{2.107}$$

Note that $\mathcal{E}_n = \mathcal{P}(\mathcal{X})$ is the set of all distributions with full support, and $\mathcal{E}_1 = \mathcal{P}(\mathcal{X}_1) \times \cdots \times \mathcal{P}(\mathcal{X}_n)$ is the set of all product distributions with full support. If we want to include distributions without full support, we work with the compactified exponential families $\overline{\mathcal{E}}_k$. Figure 2.3 illustrates the case of two parties, $\mathcal{X} = \{0,1\}^2$. Here, the set of all probability distributions is a three-dimensional simplex. The only nontrivial exponential family $\mathcal{E}_1$ is a hyperplane in this simplex.

The remainder of this subsection is based on Ref. 134. We parametrize Hamiltonians by expanding them into an orthogonal basis of the space of functions on $\mathcal{X}$,

$$H(\mathbf{x},\boldsymbol{\theta}) = \sum_i \theta_i k_i(\mathbf{x}) + \sum_{i<j} \theta_{ij} k_i(\mathbf{x}) k_j(\mathbf{x}) + \cdots + \theta_{1,\ldots,n} k_1(\mathbf{x}) \cdots k_n(\mathbf{x}) \tag{2.108}$$

where

$$k_i(\mathbf{x}) = \begin{cases} +1 & \text{if } x_i = 0, \\ -1 & \text{if } x_i = 1. \end{cases} \tag{2.109}$$

The constant term in the Hamiltonian has been omitted, because it only changes the normalization. Note that this parametrization is completely analogous to the expansion

of a diagonal quantum mechanical Hamiltonian into tensor products of σ_z,

$$H(\boldsymbol{\theta}) = \sum_i \theta_i \sigma_z^{(i)} + \sum_{i<j} \theta_{ij} \sigma_z^{(i)} \sigma_z^{(j)} + \cdots + \theta_{1,\ldots,n} \sigma_z^{(1)} \cdots \sigma_z^{(n)}. \tag{2.110}$$

We can thus parametrize any probability distribution with full support as

$$P_\theta(\mathbf{x}, \boldsymbol{\theta}) = \exp\big[H(\mathbf{x}, \boldsymbol{\theta}) - \psi(\boldsymbol{\theta})\big], \tag{2.111}$$

where

$$\psi(\boldsymbol{\theta}) = \ln\big[\sum_{\mathbf{x}} \mathrm{e}^{H(\mathbf{x},\boldsymbol{\theta})}\big] \tag{2.112}$$

ensures normalization. Alternatively, we can expand the distribution directly,

$$P_\eta(\mathbf{x}, \boldsymbol{\eta}) = \frac{1}{2^n} + \sum_i \eta_i k_i(\mathbf{x}) + \sum_{i<j} \eta_{ij} k_i(\mathbf{x}) k_j(\mathbf{x}) + \cdots + \eta_{1,\ldots,n} k_1(\mathbf{x}) \cdots k_n(\mathbf{x}). \tag{2.113}$$

This is just the classical version of the Bloch representation.

The relative entropy of two distributions $P(\mathbf{x}) = P_\eta(\mathbf{x}, \boldsymbol{\eta})$ and $P'(\mathbf{x}) = P_\theta(\mathbf{x}, \boldsymbol{\theta}')$ is

$$\ln(2) D(P\|P') = \phi(\boldsymbol{\eta}) + \psi(\boldsymbol{\theta}') - \boldsymbol{\eta} \cdot \boldsymbol{\theta}'. \tag{2.114}$$

Here the function ϕ is given by the Shannon entropy as[19]

$$\phi(\boldsymbol{\eta}) = -\ln(2) S(P_\eta(\boldsymbol{\eta})), \tag{2.115}$$

and the scalar product $\boldsymbol{\eta} \cdot \boldsymbol{\theta}'$ is defined as

$$\boldsymbol{\eta} \cdot \boldsymbol{\theta}' = \sum_i \eta_i \theta_i' + \sum_{i<j} \eta_{ij} \theta_{ij}' + \cdots + \eta_{1,\ldots,n} \theta_{1,\ldots,n}'. \tag{2.116}$$

For $P = P'$ we obtain

$$\phi(\boldsymbol{\eta}) + \psi(\boldsymbol{\theta}) - \boldsymbol{\eta} \cdot \boldsymbol{\theta} = 0. \tag{2.117}$$

The last equation shows that the θ- and the η-parametrization are related by a Legendre transformation,

$$\eta_{i_1,\ldots,i_k} = \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta_{i_1,\ldots,i_k}} \qquad \text{and} \qquad \theta_{i_1,\ldots,i_k} = \frac{\partial \phi(\boldsymbol{\eta})}{\partial \eta_{i_1,\ldots,i_k}}. \tag{2.118}$$

This relation is well-known in statistical physics. Note that ψ corresponds to the minus the free energy. For three distributions we have (as can easily be verified)

$$D(P\|P'') = D(P\|P') + D(P'\|P'') + \frac{1}{\ln(2)} (\boldsymbol{\eta} - \boldsymbol{\eta}') \cdot (\boldsymbol{\theta}' - \boldsymbol{\theta}''). \tag{2.119}$$

If the scalar product vanishes, we call this equation the *generalized Pythagoras theorem*.

[19] The reason for the appearance of a factor $\ln(2)$ in various places is that in this thesis the entropy $S(P)$ and the relative entropy $D(P\|Q)$ are defined with the binary logarithm, but thermal distributions are defined as $P(\mathbf{x}) = \mathrm{e}^{H(\mathbf{x})}/Z$ and not $P(\mathbf{x}) = 2^{H(\mathbf{x})}/Z$.

2.6.2 Information projection

The definition of the information projection captures the notion of the "closest distribution with at most k-party interaction". In this section three equivalent definitions will be given. This section is based on Refs. 2, 26, 133, 134.

Definition 2.1. The *information projection*[20] $\widetilde{P}_k$ of a probability distribution P is the element of the compactified exponential family $\overline{\mathcal{E}}_k$ which is closest to P in terms of the relative entropy,

$$\widetilde{P}_k = \operatorname*{argmin}_{P' \in \overline{\mathcal{E}}_k} D(P\|P'). \tag{2.120}$$

This definition has an interpretation as a maximum-likelihood estimate: Suppose a sample is drawn according to the probability distribution P, resulting in a vector of relative frequencies or empirical probability distribution F. Then the likelihood function is

$$L(P|F) = \prod_{\mathbf{x}} P(\mathbf{x})^{F(\mathbf{x})}. \tag{2.121}$$

The log-likelihood is, up to a P-independent term, equal to minus the relative entropy:

$$\log[L(P|F)] = \sum_{\mathbf{x}} F(\mathbf{x}) \log[P(\mathbf{x})] = -D(F\|P) - S(F). \tag{2.122}$$

The information projection $\widetilde{P}_k$ is thus the maximum-likelihood estimate for P among all $P' \in \overline{\mathcal{E}}_k$.

For a subset $A \subseteq V$ of the parties one can compute the marginal P_A of the distribution P,

$$P_A(\mathbf{x}) = \sum_{\substack{\mathbf{x}' \text{ with} \\ x'_i = x_i \ \forall\, i \in A}} P(\mathbf{x}'). \tag{2.123}$$

We define the set of all distributions $M_k(P)$ with the same k-party marginals as P by

$$M_k(P) = \left\{ P' \;\middle|\; P'_A = P_A \text{ for all } |A| = k \right\}. \tag{2.124}$$

This is an example of a *linear family* [cf. Eq. (2.144)]. Working in the Bloch representation Eq. (2.113),

$$M_k(P_\eta) = \left\{ P_{\eta'} \;\middle|\; \eta'_{i_1,\ldots,i_\ell} = \eta_{i_1,\ldots,i_\ell} \text{ for all } 1 \le \ell \le k \right\}. \tag{2.125}$$

The following lemma contains the second definition of the information projection:

Lemma 2.2. *The information projection $\widetilde{P}_k$ of a probability distribution P is the maximizer of the entropy among all distributions with the same k-party marginals as P,*

$$\widetilde{P}_k = \operatorname*{argmax}_{P' \in M_k(P)} S(P'). \tag{2.126}$$

[20] Note that this is sometimes called the *reverse I-projection or rI-projection*, e. g. in Ref. 26.

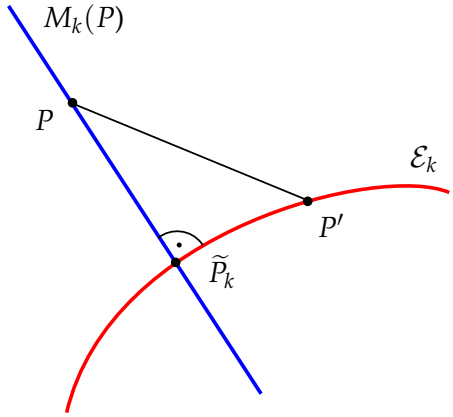

Figure 2.4: Illustration of the information projection onto an exponential family. Shown are the linear family $M_k(P)$ of distributions with the same k-party marginals as P (blue line), the exponential family $\mathcal{E}_k$ of thermal distributions of k-party Hamiltonians (red curve) and the information projection $\widetilde{P}_k$ of P onto $\mathcal{E}_k$; and P' represents an arbitrary distribution in $\mathcal{E}_k$.

The proof of the equivalence of the definitions, as well as all other proofs in this section, are omitted here. These proofs are contained as special cases in the quantum version of the theory, which is treated in Section 6.2. Alternative proofs for the classical theory, which are mathematically rigorous even for the case that the information projection does not have full support, can be found in Ch. 3 of Ref. 26.

By combining the constraints in Eq. (2.120) and Eq. (2.126) one arrives at a definition of the information projection that does not involve any optimization:

Lemma 2.3. *The information projection $\widetilde{P}_k$ of a probability distribution P is the uniquely defined element of the compactified exponential family $\overline{\mathcal{E}}_k$ with the same k-party marginals as P,*

$$\{\widetilde{P}_k\} = \overline{\mathcal{E}}_k \cap M_k(P). \tag{2.127}$$

Let $\widetilde{P}_k$ be the projection of P onto $\overline{\mathcal{E}}_k$ and $P' \in \overline{\mathcal{E}}_k$ arbitrary. Then the generalized Pythagoras theorem Eq. (2.119) holds in the form

$$D(P\|P') = D(P\|\widetilde{P}_k) + D(\widetilde{P}_k\|P'). \tag{2.128}$$

In this sense $\widetilde{P}_k$ is really an orthogonal projection. Figure 2.4 illustrates the situation.

The information projection $\widetilde{P}_1$ onto the exponential family of product distributions $\overline{\mathcal{E}}_k$ is simply given by the product of the one-party marginals,

$$\widetilde{P}_1(\mathbf{x}) = P_{\{1\}}(x_1)\cdots P_{\{n\}}(x_n) \qquad \text{where} \qquad P_{\{i\}}(x_i) = \sum_{\substack{\mathbf{x}' \text{ with} \\ x'_i = x_i}} P(\mathbf{x}). \tag{2.129}$$

For the projections $\widetilde{P}_k$ with $k > 1$ there is no explicit formula. For an iterative algorithm to compute these projections numerically see the next subsection.

We denote the distance from a probability distribution P to its projection $\widetilde{P}_k$ in terms of the relative entropy by

$$D_k(P) = D(P\|\widetilde{P}_k), \qquad k = 1, \ldots, n-1. \tag{2.130}$$

It can be shown that

$$D_k(P) = S(\widetilde{P}_k) - S(P), \qquad k = 1, \ldots, n-1. \tag{2.131}$$

For $k = 1$ this is the *multi-information* [62]

$$D_1(P) = D(P\|\widetilde{P}_1) = S(\widetilde{P}_1) - S(P) = \sum_{i=1}^{n} S(P_{\{i\}}) - S(P) \tag{2.132}$$

[cf. Eq. (2.129)]. We also define the *degree of irreducible k-party interaction*[21] as

$$C_k(P) = D_{k-1}(P) - D_k(P), \qquad k = 2, \ldots, n \tag{2.133}$$

(where $D_n \equiv 0$). By the generalized Pythagoras theorem,

$$C_k(P) = D(\widetilde{P}_k\|\widetilde{P}_{k-1}), \qquad k = 2, \ldots, n-1, \tag{2.134}$$

or equivalently

$$C_k(P) = S(\widetilde{P}_{k-1}) - S(\widetilde{P}_k), \qquad k = 2, \ldots, n-1. \tag{2.135}$$

The multi-information [Eq. (2.132)], which in the following will be called *degree of total interaction* and denoted by C_{tot}, has a decomposition

$$D_1(P) = C_{\text{tot}}(P) = \sum_{k=2}^{n} C_k(P). \tag{2.136}$$

The terms in this decomposition are orthogonal in the sense of the generalized Pythagoras theorem. The distance $D_k(P)$ is a measure for the information that is contained in the distribution P, but not in its k-party marginals. Similarly, $C_k(P)$ measures the information contained in the k-party, but not in the $(k-1)$-party marginals of P.

In Ref. 62 the C_k were called *complexity measures* and applied to the study of complex dynamical systems.[22] This leads to the question which properties should be expected from complexity measures, and if the C_k possess these properties. It has been pointed out [35] that $C_k(P)$ can increase under local transformations of P, and in particular under tracing out of parties.[23] This may indicate that the C_k are no good complexity measures. (By comparison, entanglement measures are by definition invariant under local unitaries and non-increasing under LOCCs [44, Sec. 4.1.1].) The reason for this

[21] In Ref. 2 the degree of irreducible k-party interaction is called *amount of kth-order effect*.

[22] In that reference also C_1 is considered, which is the distance from $\widetilde{P}_1$ to the uniform probability distribution.

[23] The increase of C_k under local transformations was observed before in the quantum case in Ref. 134 and suspected to be a quantum feature.

behaviour of the C_k is that the exponential families are not invariant under local transformations; the relative entropy itself is non-increasing even under nonlocal transformations of both arguments. (However, the exponential family $\overline{\mathcal{E}}_1$, which consists of all product distributions, is invariant, and consequently C_{tot} is non-increasing.) Consequently, it has been suggested [35] to replace D_k by the distance to the local orbit of $\overline{\mathcal{E}}_k$. In this thesis, the term *complexity measure* (and similarly *correlation measure*) will be avoided; the C_k will always be referred to as *interaction measures*.

2.6.3 Iterative scaling

The information projection onto an exponential family can be computed numerically by iterative scaling. This algorithm is based on the following dual formulation of the minimization problem in Eq. (2.120):

Recall the definition of the linear family $M_k(P)$ of probability distributions with the same k-party marginals as P [see Eq. (2.124)]. We define the *dual information projection* of a distribution Q onto $M_k(P)$ as[24]

$$Q^* = \operatorname*{argmin}_{Q' \in M_k(P)} D(Q'\|Q) \tag{2.137}$$

[note that the minimization is over the first argument of the relative entropy, while in Eq. (2.120) it is over the second argument]. Theorem 3.3 in Ref. 26 states that the information projection of P onto $\overline{\mathcal{E}}_k$ is given by the dual information projection of the uniform distribution P_0 onto $M_k(P)$,

$$\operatorname*{argmin}_{P' \in \overline{\mathcal{E}}_k} D(P\|P') = \operatorname*{argmin}_{Q' \in M_k(P)} D(Q'\|P_0) \qquad \text{where} \qquad P_0 \equiv \frac{1}{2^n}. \tag{2.138}$$

In the *iterative scaling algorithm*, which will now be described, the dual projection is computed by scaling a distribution to adjust its marginals (see Ch. 5 in Ref. 26 and Refs. 35, 104). The iteration can be proven to converge [26, Thm. 5.1].

Algorithm 2.4 (Iterative scaling).
Problem: *Given a probability distribution P of n parties, compute its information projection $\widetilde{P}_k$ onto the exponential family $\overline{\mathcal{E}}_k$.*

1. *For each k-element subset $A \subseteq \{1, \ldots, n\}$ compute the marginal P_A of P,*

$$P_A(\mathbf{x}) = \sum_{\substack{\mathbf{x}' \text{ with} \\ x'_i = x_i \,\forall\, i \in A}} P(\mathbf{x}'). \tag{2.139}$$

2. *Initialize Q as the uniform probability distribution, $Q \equiv 1/2^n$.*

[24] In Ref. 26 the dual information projection is called *I-projection*, and the information projection is called *reverse I-projection or rI-projection*. Reference 2 calls them *m-projection* and *e-projection*, respectively.

3. *Looping through all k-element subsets A of the parties, update Q according to*

$$Q(\mathbf{x}) \to Q'(\mathbf{x}) = \frac{P_A(\mathbf{x})}{Q_A(\mathbf{x})} Q(\mathbf{x}), \tag{2.140}$$

where Q_A is the marginal of Q,

$$Q_A(\mathbf{x}) = \sum_{\substack{\mathbf{x}' \text{ with} \\ x'_i = x_i \,\forall\, i \in A}} Q(\mathbf{x}'). \tag{2.141}$$

4. *Repeat the last step.*

An algorithm for the corresponding quantum mechanical problem is developed in Section 6.4.

2.6.4 Information geometry

This subsection aims at putting the topic of exponential families of interaction spaces into the larger context of information geometry. It is based on Refs. 2, 3.

In information geometry, families of probability distributions which depend on continuous parameters are viewed as manifolds. For example, the family of normal distributions with mean μ and standard deviation σ

$$P(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{ -\frac{(x-\mu)^2}{2\sigma^2} \Big\} \tag{2.142}$$

is a two-dimensional manifold with coordinates (μ, σ). Every point in the manifold is a probability distribution.

We will now give the general definitions of exponential and linear families. They are defined by the way in which the probability distributions depend on the parameters (or coordinates). An *exponential family* has the form

$$P(\mathbf{x}, \boldsymbol{\theta}) = \exp\big\{ k_0(\mathbf{x}) + \sum_i \theta_i k_i(\mathbf{x}) - \psi(\boldsymbol{\theta}) \big\}, \tag{2.143}$$

where the k_i are given functions [not necessarily of product form as in Eq. (2.108)] and ψ ensures normalization. The θ_i are called *e-affine coordinates*. Any curve in the space of coordinates of the form $\boldsymbol{\theta}(t) = t\mathbf{a} + \mathbf{b}$ is called an *e-geodesic*.

A *linear family*[25] has the form

$$P(\mathbf{x}, \boldsymbol{\eta}) = q_0(\mathbf{x}) + \sum_i \eta_i q_i(\mathbf{x}), \tag{2.144}$$

where the q_i are given functions. Analogously to exponential families we speak of *m-affine coordinates* and *m-geodesics*.

[25] Linear families are called *mixture families* in Refs. 2, 3.

The use of the terms *e-geodesic* and *m-geodesic* is justified by the existence of two corresponding affine connections. These two connections are related to each other by a property called *duality* with respect to the Riemannian metric defined by the Fisher information matrix. It is a consequence of this duality that any *e*-flat manifold, such as an exponential family, is also *m*-flat; and any *m*-flat manifold, such as a linear family, also *e*-flat. We call such manifolds dually flat. For any dually flat manifold, the *e*- and *m*-affine coordinates are related by a Legendre transformation, and the coordinate curves are orthogonal at any point. If now the point (or distribution) P is connected by an *m*-geodesic to P', and P' is connected by an *e*-geodesic to P'', and these geodesics are orthogonal at P', the Pythagoras theorem

$$D(P\|P'') = D(P\|P') + D(P'\|P'') \tag{2.145}$$

holds [cf. Fig. 2.4]. One can re-formulate this as the dual Pythagoras theorem, where the *e*- and *m*-geodesics trade places and $D(P\|Q)$ is replaced by the dual divergence $D^*(P\|Q) = D(Q\|P)$.

Any linear subspace of the space of *e*-affine coordinates of a dually flat manifold defines an *e*-flat submanifold, and analogously for the *m*-affine coordinates. This defines orthogonal foliations of the manifold.

3 Increasing the statistical significance of entanglement detection in experiments

This chapter deals with the effect of finite measurement statistics on entanglement detection. Two detection methods are considered: witness operators and Bell inequalities. In Section 3.1 the significance of an entanglement test is defined, and two particular Bell inequalities, namely, the Mermin and the Ardehali inequality, are introduced. Section 3.2 describes a general method for maximizing the significance of a witness operator under the assumption that the statistical error can be estimated by the standard deviation. In Section 3.3 the error estimation for multiphoton experiments is discussed, and the Mermin and the Ardehali inequality are compared with respect to their statistical significances. The result of this comparison has been tested in an experiment carried out by the group of Jian-Wei Pan; a description of the experiment is given in Section 3.4.

The results of this chapter have been published in Ref. B, but are presented here in greater detail.

3.1 Statement of the problem

Our first aim is to give a definition of the statistical significance of an entanglement test. Two kinds of entanglement tests are considered, namely, witness operators and Bell inequalities. An introduction to these concepts was given in Sections 2.1.3 and 2.2.1, respectively. Recall that a witness operator W satisfies $\mathrm{Tr}(W\rho) \geq 0$ for all separable states ρ. For any state ρ, we define the *violation* of the witness as

$$V_W = -\mathrm{Tr}(W\rho). \tag{3.1}$$

A Bell inequality takes the form $\mathrm{Tr}(B\rho) \leq C_{\mathrm{lhv}}$. The observable B is called *Bell operator*. For the CHSH inequality, $B = A_1 \otimes B_1 + A_1 \otimes B_2 + A_2 \otimes B_1 - A_2 \otimes B_2$. The number C_{lhv} is the local hidden variable bound. We define the violation as

$$V_B = \mathrm{Tr}(B\rho) - C_{\mathrm{lhv}}. \tag{3.2}$$

In either case a positive violation proves entanglement. The experimental entanglement test consists in determining the value of V.

Quantum mechanics predicts in general only probabilities for measurement results. Thus, even in the absence of any experimental imperfections, expectation values such as $\mathrm{Tr}(W\rho)$ and $\mathrm{Tr}(B\rho)$ can be determined perfectly only in the limit of an infinite number of repetitions of the experiment. When evaluating the data of an experiment implementing a witness or a Bell operator, one has to rule out the possibility that a positive violation is only an effect of statistical fluctuations.[1]

[1] Clearly, this also holds for the so-called "nonlocality without inequality" proofs, such as the GHZ argument [27, 89].

The effect of finite statistics is particularly relevant for experiments generating entangled photons by spontaneous parametric downconversion. (For a description of such an experiment see Section 3.4 below.) In this type of experiment, the rate at which entangled states are generated scales very unfavourably with the number of parties. As a consequence, the effect of statistical fluctuations is by far larger than all systematic errors, such as misalignment of the axes of polarization measurements. In this chapter we will always assume an experiment where the statistical error is dominant. (In fact, we shall neglect systematic errors completely.) In Section 3.3 we discuss the error estimation explicitly for experiments with entangled photons.

A way to account for statistical errors, which is obvious to any experimental physicist, is to equip the measured value of V with an "error bar", $V \pm \mathcal{E}$. The proper definition of the error $\mathcal{E}$ is the first main subject of this chapter.

It should be mentioned that this is not the only possible way to deal with statistical fluctuations. At a more fundamental level, we are concerned with the task of discriminating classes of probability distributions for measurement results: Those which can result from separable states and those which require entanglement. This is a problem from the realm of statistical hypothesis testing. A rigorous analysis of Bell inequalities from this point of view has been carried out in Ref. 27. In Chapter 4 this idea will be applied to the discrimination of different classes of entangled states.

The experiments that we have in mind aim at preparing a certain target state. Thus it makes sense to optimize the entanglement test by maximizing the violation V or minimizing the error $\mathcal{E}$ for this state. In general there will be a trade-off between these optimizations. We define the *significance* of the entanglement test as the ratio

$$S = \frac{V}{\mathcal{E}}\,, \tag{3.3}$$

in other words, as the inverse relative error. If we use the standard deviation $\Delta(V)$ as an estimate for the error $\mathcal{E}$, reporting the value of the significance is nothing but reporting "violation by S standard deviations". The optimization of witness operators and Bell inequalities with respect to their significance S is the second main subject of this chapter.

Recall that for witness operators the word *optimization* is conventionally used in a different sense (see Section 2.1.3): It refers to enlarging the set of detected states by subtracting a positive operator from the witness. This has the additional effect of increasing (or at least not decreasing) the violation for any fixed detected state. However, when trying to prove entanglement in an experiment with a given target state, increasing the set of detected states is not the primary concern, and increasing the violation for the target state is desirable only if it leads to a higher significance. The relation between optimization of a witness in the conventional sense and maximization of its significance will be studied in the next section.

Furthermore we will compare the significances of two Bell inequalities. The *four-party Mermin inequality* is given by [83]

$$\langle A_1 B_1 C_1 D_1\rangle - \sum_{\pi(1,1,2,2)} \langle A_{\pi_1} B_{\pi_2} C_{\pi_3} D_{\pi_4}\rangle + \langle A_2 B_2 C_2 D_2\rangle \leq 4. \tag{3.4}$$

The symbol $\sum_{\pi(i_1,\ldots,i_k)}$ stands for all distinct permutations of the indices, in this case

$$\sum_{\pi(1,1,2,2)} \langle A_{\pi_1} B_{\pi_2} C_{\pi_3} D_{\pi_4} \rangle = \langle A_1 B_1 C_2 D_2 \rangle + \langle A_1 B_2 C_1 D_2 \rangle + \langle A_1 B_2 C_2 D_1 \rangle + \langle A_2 B_1 C_1 D_2 \rangle + \langle A_2 B_1 C_2 D_1 \rangle + \langle A_2 B_2 C_1 D_1 \rangle. \quad (3.5)$$

We choose $A_1 = B_1 = C_1 = D_1 = \sigma_x$ and $A_2 = B_2 = C_2 = D_2 = \sigma_y$. For this choice the eight terms of the Mermin inequality are expectation values of stabilizing operators of the GHZ state [see Eq. (2.96)]. Thus the quantum mechanical bound is 8 and it is attained by the GHZ state.

The *four-party Ardehali inequality* is [4]

$$\frac{1}{\sqrt{2}}\Big[\langle A_1 B_1 C_1 D_1 \rangle + \langle A_1 B_1 C_1 D_2 \rangle - \sum_{\pi(1,2,2)} \langle A_{\pi_1} B_{\pi_2} C_{\pi_3} D_1 \rangle + \sum_{\pi(1,2,2)} \langle A_{\pi_1} B_{\pi_2} C_{\pi_3} D_2 \rangle - \sum_{\pi(1,1,2)} \langle A_{\pi_1} B_{\pi_2} C_{\pi_3} D_1 \rangle + \sum_{\pi(1,1,2)} \langle A_{\pi_1} B_{\pi_2} C_{\pi_3} D_2 \rangle + \langle A_2 B_2 C_2 D_1 \rangle - \langle A_2 B_2 C_2 D_2 \rangle\Big] \leq 2\sqrt{2}. \quad (3.6)$$

We choose

$$A_1 = B_1 = C_1 = \sigma_x, \qquad A_2 = B_2 = C_2 = \sigma_y, \quad (3.7)$$

$$D_1 = \frac{1}{\sqrt{2}}(\sigma_x + \sigma_y), \qquad D_2 = \frac{1}{\sqrt{2}}(\sigma_x - \sigma_y). \quad (3.8)$$

Then the Bell operator is equal to the Bell operator of the Mermin inequality. Thus the quantum mechanical bound is again 8 and is again attained with the GHZ state. From the local realistic point of view the inequalities are not equivalent, of course. The maximal quantum mechanical violation is higher for the Ardehali inequality,

$$V_{\text{Mermin}} = 4, \qquad V_{\text{Ardehali}} = 8 - 2\sqrt{2} \approx 5.1716. \quad (3.9)$$

Unlike the Svetlichny inequality (2.37), both the Mermin and the Ardehali inequality hold for local realistic models only, but not for the more general hybrid models.

3.2 Optimizing a witness with respect to its variance

In a naive error model for a witness operator, the error is estimated by the standard deviation of the witness

$$\mathcal{E}(W) = \Delta(W) = \sqrt{\langle W^2 \rangle - \langle W \rangle^2}. \quad (3.10)$$

In this model the fact is neglected that in a typical experiment the observable W is not implemented directly, but rather decomposed into a sum of product observables [44, Sec. 6.1.2]. Also, even for an observable that is directly implemented the use of the

standard deviation as an error estimate needs to be justified (cf. the next section). Nevertheless, even with this simple model one can demonstrate the difference between maximization of the violation and maximization of the significance.

Clearly the significance

$$S(W) = \frac{-\langle W \rangle}{\Delta(W)} \tag{3.11}$$

diverges if and only if the state being measured is an eigenstate of the witness. We will show the following result:

Lemma 3.1. *Let $\rho = |\psi\rangle\langle\psi|$ be a pure state detected by the witness W with finite significance. Then there exists another witness W' which is less fine than W and which detects ρ with lower violation, but arbitrary high significance. In other words, there is a positive operator P such that $|\psi\rangle$ is an eigenstate of $W' = W + P$.*

Before proving the lemma, we describe an iterative procedure to construct a witness with a higher violation. This procedure motivates the ansatz for P made in the proof and may also be of independent interest.

Let ρ be a state detected by the witness W. For the time being we do not require ρ to be pure. We want to increase the significance iteratively by adding only a small positive operator at a time. So let $W' = W + \varepsilon P$, where P is a positive operator with unit trace and $\varepsilon > 0$. The violation and the standard deviation of W' for the state ρ are given by

$$V_{W'} = -\langle W \rangle - \varepsilon \langle P \rangle \tag{3.12}$$

and

$$\Delta(W') = \Big[\Delta^2(W) + \varepsilon\big(\langle WP \rangle + \langle PW \rangle - 2\langle W \rangle \langle P \rangle\big) + \varepsilon^2 \Delta^2(P)\Big]^{1/2}. \tag{3.13}$$

We expand the significance for small ε,

$$S(W') = \frac{-\langle W \rangle}{\Delta(W)} + \varepsilon \frac{1}{2} \frac{\langle W \rangle}{\Delta^3(W)} \Big(\langle WP \rangle + \langle PW \rangle + 2 \frac{\langle W^2 \rangle}{-\langle W \rangle} \langle P \rangle \Big) + \mathcal{O}(\varepsilon^2). \tag{3.14}$$

Introducing the Hermitian operator

$$Q = \rho W + W \rho + 2 \frac{\langle W^2 \rangle}{-\langle W \rangle} \rho, \tag{3.15}$$

we rewrite this expression as

$$S(W') = S(W) + \varepsilon \frac{1}{2} \frac{\langle W \rangle}{\Delta^3(W)} \operatorname{Tr}(QP) + \mathcal{O}(\varepsilon^2). \tag{3.16}$$

We neglect the terms of quadratic and higher order and maximize the significance over all positive P with unit trace. Since the prefactor $\langle W \rangle / \Delta^3(W)$ is negative, this is equivalent to minimizing $\operatorname{Tr}(QP)$. The optimal P is thus a projector onto the eigenspace of Q corresponding to the minimal eigenvalue. If this eigenvalue is negative, we succeeded in improving the significance.

We will now show that for a pure state $\rho = |\psi\rangle\langle\psi|$ the operator Q always has a negative eigenvalue.[2] As an ansatz for a corresponding eigenvector we use

$$|\eta\rangle = \cos\left(\tfrac{\theta}{2}\right)|\psi\rangle + \mathrm{e}^{\mathrm{i}\phi}\sin\left(\tfrac{\theta}{2}\right)|\psi^\perp\rangle, \quad 0 \le \theta < \pi, \quad 0 \le \phi < 2\pi, \tag{3.17}$$

where $|\psi^\perp\rangle$ is an as of yet undetermined normalized vector orthogonal to $|\psi\rangle$. For $P = |\eta\rangle\langle\eta|$ we obtain

$$\mathrm{Tr}(QP) = 2\cos\left(\tfrac{\theta}{2}\right)\sin\left(\tfrac{\theta}{2}\right)\mathrm{Re}\left(\mathrm{e}^{\mathrm{i}\phi}\langle\psi|W|\psi^\perp\rangle\right) + 2\cos^2\left(\tfrac{\theta}{2}\right)\frac{\Delta_\psi^2(W)}{-\langle W\rangle_\psi}. \tag{3.18}$$

This expression has to be minimized over $|\psi^\perp\rangle$ and the angles θ and ϕ. By choosing $\phi = \pi - \arg(\langle\psi|W|\psi^\perp\rangle)$ we can always make $\mathrm{Re}(\mathrm{e}^{\mathrm{i}\phi}\langle\psi|W|\psi^\perp\rangle)$ negative. Thus the optimal $|\psi^\perp\rangle$ is the normalized vector orthogonal to $|\psi\rangle$ maximizing $|\langle\psi|W|\psi^\perp\rangle|$, which is

$$|\psi^\perp_{\mathrm{opt}}\rangle = \frac{1}{\Delta_\psi(W)}\left(\mathbb{1} - |\psi\rangle\langle\psi|\right)W|\psi\rangle, \tag{3.19}$$

where the phase has been chosen arbitrarily. Observing $\langle\psi|W|\psi^\perp_{\mathrm{opt}}\rangle = \Delta_\psi(W)$ and choosing $\phi = \pi$ we arrive at

$$\mathrm{Tr}(QP) = 2\cos\left(\tfrac{\theta}{2}\right)\Delta_\psi(W)\left[-\sin\left(\tfrac{\theta}{2}\right) + \cos\left(\tfrac{\theta}{2}\right)\frac{\Delta_\psi(W)}{-\langle W\rangle_\psi}\right]. \tag{3.20}$$

The second term in the square brackets is always positive. We can now choose θ so close to π that

$$\tan\left(\tfrac{\theta}{2}\right) > \frac{\Delta_\psi(W)}{-\langle W\rangle_\psi} \tag{3.21}$$

and thus $\mathrm{Tr}(QP) < 0$. This shows that for pure $\rho = |\psi\rangle\langle\psi|$ the minimal eigenvalue of Q is negative.

We can now repeat the procedure with $W' = W + \varepsilon P$ for sufficiently small ε in place of W. The new optimal $|\psi^\perp\rangle$ is given by [cf. Eq. (3.19)]

$$|\psi^\perp_{\mathrm{opt}}\rangle_{\mathrm{new}} = \frac{1}{\Delta_\psi(W')}\left(\mathbb{1} - |\psi\rangle\langle\psi|\right)\left(W + \varepsilon|\eta\rangle\langle\eta|\right)|\psi\rangle \tag{3.22}$$

with $|\eta\rangle = \cos(\theta/2)|\psi\rangle - \sin(\theta/2)|\psi^\perp_{\mathrm{opt}}\rangle$. The last equation can be simplified:

$$\begin{aligned}
|\psi^\perp_{\mathrm{opt}}\rangle_{\mathrm{new}} &= \frac{1}{\Delta_\psi(W')}\left(\mathbb{1} - |\psi\rangle\langle\psi|\right)\left[W|\psi\rangle + \varepsilon\cos\left(\tfrac{\theta}{2}\right)|\eta\rangle\right] \\
&= \frac{1}{\Delta_\psi(W')}\left(\mathbb{1} - |\psi\rangle\langle\psi|\right)W|\psi\rangle - \varepsilon\cos\left(\tfrac{\theta}{2}\right)\sin\left(\tfrac{\theta}{2}\right)|\psi^\perp_{\mathrm{opt}}\rangle \\
&= \frac{\Delta_\psi(W) - \varepsilon\cos(\theta/2)\sin(\theta/2)}{\Delta_\psi(W')}|\psi^\perp_{\mathrm{opt}}\rangle \\
&= |\psi^\perp_{\mathrm{opt}}\rangle.
\end{aligned} \tag{3.23}$$

[2] For mixed ρ this is in general not true. A counterexample is given by $W = \mathbb{1}/2 - |\psi^-\rangle\langle\psi^-|$ and $\rho = \mathbb{1}/8 + 1/2|\psi^-\rangle\langle\psi^-|$. The state is detected with $\mathrm{Tr}(W\rho) = -1/8$, but the minimal eigenvalue of Q is $19/32$.

This shows that $|\psi_{\text{opt}}^{\perp}\rangle$ will be the same in every iteration step. The angle θ will in general be different each time. In other words, in each iteration step P is a different one-dimensional projector, but all these projectors are supported on the same two-dimensional subspace spanned by $|\psi\rangle$ and $|\psi_{\text{opt}}^{\perp}\rangle$, or equivalently, by $|\psi\rangle$ and $W|\psi\rangle$. This observation motivates the following proof of the lemma:

Proof of Lemma 3.1. Let $\rho = |\psi\rangle\langle\psi|$ be a pure state detected by the witness W. We need to find a positive operator P such that $|\psi\rangle$ is an eigenvector with negative eigenvalue of $W' = W + P$. We make the ansatz

$$P = a|\psi\rangle\langle\psi| + b|\psi_{\text{opt}}^{\perp}\rangle\langle\psi_{\text{opt}}^{\perp}| + c|\psi\rangle\langle\psi_{\text{opt}}^{\perp}| + c^*|\psi_{\text{opt}}^{\perp}\rangle\langle\psi|, \tag{3.24}$$

where a and b are real and c is a complex coefficient and

$$|\psi_{\text{opt}}^{\perp}\rangle = \frac{1}{\Delta_\psi(W)}\big(\mathbb{1} - |\psi\rangle\langle\psi|\big)W|\psi\rangle. \tag{3.25}$$

For this ansatz,

$$W'|\psi\rangle = \big(\langle W\rangle_\psi + a\big)|\psi\rangle + \big(\Delta_\psi(W) + c^*\big)|\psi_{\text{opt}}^{\perp}\rangle. \tag{3.26}$$

So $|\psi\rangle$ is an eigenvector of W' for $c = c^* = -\Delta_\psi(W)$. The corresponding eigenvalue is negative for $a < -\langle W\rangle_\psi$. The operator P is positive if $a, b \geq 0$ and $\det(P) = ab - |c|^2 \geq 0$. Since these conditions can always be satisfied, the lemma is proven. □

3.3 Error estimation for multiphoton experiments

In this section the error estimation for multiphoton experiments is discussed and applied to the Mermin and the Ardehali inequality. The setup of such an experiment is described in the next section. The part of the setup implementing the measurement consists of a combination of wave plates, a polarizing beam splitter and two photon detectors for each spatial mode [see Fig. 3.2 (c)]. By setting the angles of the wave plates, local measurement bases are chosen. Any product basis thus corresponds to a measurement setting. Only those events are counted where exactly one of the two detectors for each mode gives a signal. The immediate result of the measurement is thus a set of 2^n coincidence count numbers, each corresponding to a vector of the measurement basis. Any observable is decomposed into a sum of terms, each being diagonal in a product basis (cf. Sec. 6.1.2 in Ref. 44). The two assumptions of the standard error model [58] are that the count numbers may be treated as statistically independent Poisson-distributed random variables and that Gaussian error propagation may be applied to calculate the error of the observable's expectation value. We will discuss this model now in more detail.

Let Π_{ki} be the projector onto the vector i of the product basis (or setting) k, and let $p_{ki} = \text{Tr}(\Pi_{ki}\rho)$ be the probability that we wish to estimate. Measuring in the setting k for a certain amount of time results in coincidence count numbers, which the model describes as Poisson-distributed random variables N_{ki},

$$\text{Prob}(N_{ki} = n) = \frac{\lambda^n \mathrm{e}^{-\lambda}}{n!} \qquad \text{where} \qquad \lambda = n_0 p_{ki} \tag{3.27}$$

for some unknown n_0. In particular this implies that the expectation value and variance are equal,

$$\langle N_{ki} \rangle = \Delta^2(N_{ki}) = n_0 p_{ki}. \tag{3.28}$$

Denote the observed values of the count numbers by n_{ki}, and let $n_k = \sum_i n_{ki}$ be the total count number for the setting. We estimate the probabilities p_{ki} by the relative count numbers or frequencies,

$$p_{ki} \approx f_{ki} = \frac{n_{ki}}{n_k}. \tag{3.29}$$

For an observable diagonal in the product basis k,

$$A_k = \sum_i \lambda_i \Pi_{ki}, \tag{3.30}$$

the expectation value is estimated by

$$\langle A_k \rangle \approx \sum_i \lambda_i f_{ki}. \tag{3.31}$$

The statistical error of the count number n_{ki} is estimated by the standard deviation,

$$\mathcal{E}(n_{ki}) = \Delta(n_{ki}) = \sqrt{n_{ki}}. \tag{3.32}$$

We use Gaussian error propagation to determine the error of $\langle A_k \rangle$,

$$\begin{aligned} \mathcal{E}^2(\langle A_k \rangle) &\approx \sum_i \Big[\frac{\partial \langle A_k \rangle}{\partial n_{ki}} \Big]^2 \mathcal{E}^2(n_{ki}) \\ &= \sum_i \Big[\frac{\lambda_i}{n_k} - \frac{\sum_j \lambda_j n_{kj}}{n_k^2} \Big]^2 n_{ki} \\ &= \sum_i \Big[\frac{\lambda_i}{n_k} - \frac{\langle A_k \rangle}{n_k} \Big]^2 n_{ki} \\ &= \frac{\Delta^2(A_k)}{n_k}. \end{aligned} \tag{3.33}$$

The form of the derivative given in the second line of the previous equation led the authors of Ref. 58 to believe that the second term $\sum_j \lambda_j n_{kj} / n_k^2$ decreased faster with increasing total count number n_k than the first term λ_i / n_k. Consequently, they dropped the second term. This is a mistake, because the numerator $\sum_j \lambda_j n_{kj}$ scales linearly with n_k. For the experimental data given in that reference [in the paragraph below Eq. (3.20)] the neglected terms are not significantly smaller than the retained terms. More precisely: Both the retained and the neglected term in Eq. (5.6) of Ref. 58 take values up to the order of 10^{-6}. (The situation considered in that reference is not completely the same as ours: There, only one photon detector per spatial mode is used, and the total count number is determined by measuring the identity operator.)

Let us comment on the assumptions we made. The assumption of Poissonian probability distributions is justified by the low overall efficiency of the experiment. Gaussian

error propagation involves a linearization of the function whose error is calculated and assumes that the errors of a sum of terms add quadratically. Adding errors quadratically gives the exact result if they are the standard deviations of uncorrelated variables. Finally, when interpreting a standard deviation as half of the length of a 68% confidence interval, as is frequently done, one tacitly assumes a normal distribution. However, the Poisson distribution can be well approximated by a normal distribution for sufficiently large numbers.

We now apply the theory outlined above to the Mermin and Ardehali inequality (3.4) and (3.6). For the perfect GHZ state the error of the Mermin inequality is zero. This is due to the fact that we chose the operators in the Mermin inequality as stabilizing operators of the GHZ state. In particular, the GHZ state is an eigenstate of all those observables. The operators in the Ardehali inequality are not stabilizing operators, and the error does not vanish. This shows that for states sufficiently close to the GHZ state the Mermin inequality has a higher significance than the Ardehali inequality. This behaviour of the significance should be compared to the violation, which is higher for the Ardehali inequality [see Eq. (3.9)].

For experimental applications it is relevant to know if the Mermin inequality has a higher significance even for states with a realistic amount of noise. To answer this question, we repeat the error analysis for GHZ states with different noise levels. As our noise model we choose bit-flip noise,

$$\rho \to \big(\mathcal{E}_1 \circ \mathcal{E}_2 \circ \mathcal{E}_3 \circ \mathcal{E}_4\big)(\rho) \qquad \text{where} \qquad \mathcal{E}_i(\rho) = (1-p)\rho + p\sigma_x^{(i)}\rho\sigma_x^{(i)} \tag{3.34}$$

and $0 \leq p \leq 1$ is the bit-flip probability. In a photonic experiment this type of noise can easily be introduced on purpose to test the theory for different noise levels (see the next section).

Figure 3.1 shows the calculated significance of the Mermin and the Ardehali inequality for the GHZ state with bit-flip noise. The noise can be quantified by the bit-flip probability p or the fidelity with the perfect GHZ state $F = \langle \mathrm{GHZ}_4 | \rho_{\mathrm{exp}} | \mathrm{GHZ}_4 \rangle$. It is assumed that per data point 8000 copies of the GHZ state are prepared, such that that either each of the eight terms of the Mermin inequality is measured 1000 times or each of the sixteen terms of the Ardehali inequality 500 times. One observes that the Mermin inequality has a higher significance for fidelities $F \geq 0.70$. Such fidelities are within reach of current experiments (again, see the next section). For the six-party versions of the inequalities the threshold value changes to $F \geq 0.40$. The calculations have been repeated for white noise, yielding $F \geq 0.72$ for four parties and $F \geq 0.41$ for six parties. This suggests that the effect does not depend strongly on the details of the noise channel. One can show that for white noise the threshold value decreases exponentially with the number of parties. (For the derivation of the results on white noise and larger numbers of parties see Ref. 61.)

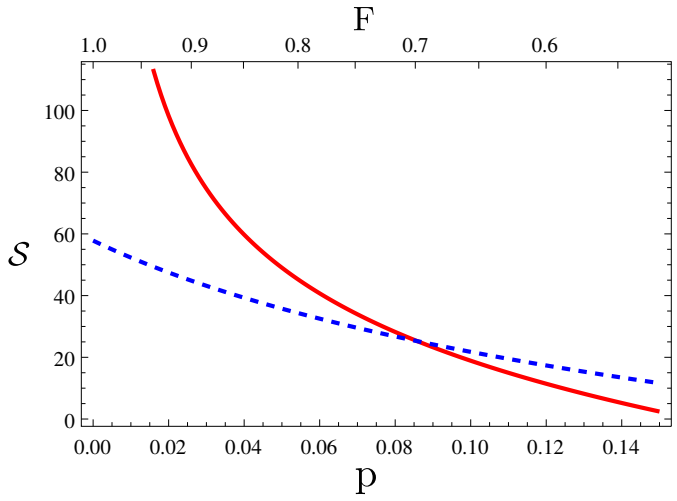

Figure 3.1: Theoretically predicted values of the significance S of the four-party Mermin (red, solid) and Ardehali inequality (blue, dashed) for the GHZ state with different levels of bit-flip noise. The noise can be quantified by the bit-flip probability p (lower horizontal axis) or the fidelity F (upper horizontal axis). It is assumed that 8000 instances of the GHZ state are prepared and that either each of the eight terms of the Mermin inequality is measured 1000 times or each of the sixteen terms of the Ardehali inequality 500 times. (Figure taken from Ref. B.)

3.4 Description of the experiment and results

In this section an experiment testing the above predictions for the significance of the Mermin and the Ardehali inequality is described. The experiment was carried out by He Lu, Wei-Bo Gao, Yu-Ao Chen, Zeng-Bing Chen and Jian-Wei Pan at the University of Science and Technology of China in Hefei. A four-photon GHZ state was prepared, and different levels of simulated (or engineered) bit-flip noise were applied.

The experimental setup is shown in Fig. 3.2. In the first step, which is not shown in the figure, femtosecond laser pulses (pulse length $\approx$ 200 fs) with a repetition rate of 76 MHz and a wavelength of 788 nm are produced and frequency-doubled with an LiB_3O_5 (LBO) crystal, thus converting them to ultraviolet. With these pulses two β-barium borate (BBO) crystals with a length of 2 mm are pumped. By spontaneous parametric downconversion [72], polarization-entangled photon pairs of the form

$$|\psi_{\mathrm{SPDC}}\rangle = \frac{1}{\sqrt{2}}\left(|H_1\rangle|V_2\rangle + \mathrm{e}^{\mathrm{i}\alpha}|V_1\rangle|H_2\rangle\right) \tag{3.35}$$

are produced in those crystals. Here H and V denote horizontal and vertical polarization and the indices 1 and 2 label spatial modes. The phase α can be compensated. For each BBO crystal the observed two-fold coincidence count rates are about $1.6 \times 10^4\,\mathrm{s}^{-1}$. The visibilities are 96% in the H/V and 94% in the $+/-$ basis. Modes 2 and 3 (see figure) each go through a half-wave plate (HWP). Following the conventions of the Jones

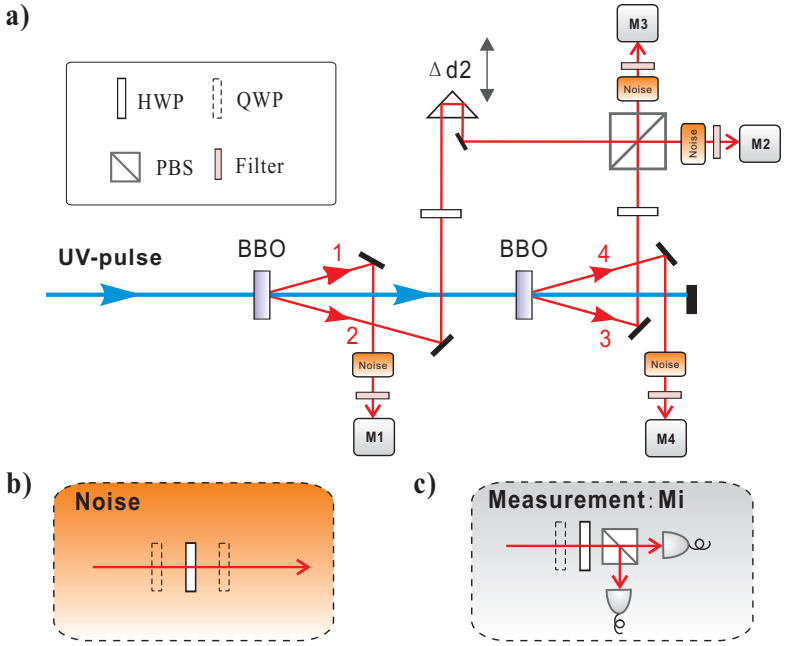

Figure 3.2: Scheme of the experimental setup. **(a)** The setup to generate the required four-photon GHZ state. Femtosecond laser pulses ($\approx$ 200 fs, 76 MHz, 788 nm) are converted to ultraviolet through a frequency-doubling LiB_3O_5 (LBO) crystal (not shown). The pulses go through two main β-barium borate (BBO) crystals (2 mm), generating two pairs of photons. For each crystal, the observed two-fold coincidence count rates are about $1.6 \times 10^4\,s^{-1}$ with a visibility of 96% (94%) in the H/V $(+/-)$ basis. **(b)** Setup for engineering the bit-flip noise. **(c)** The measurement setup. (Figure taken from Ref. B.)

calculus, we write the polarization states as $|H\rangle = |0\rangle = \binom{1}{0}$ and $|V\rangle = |1\rangle = \binom{0}{1}$. With this convention the HWP, set at an angle of 45° between the fast axis and the vertical, implements [58] the operation $-\sigma_x$. In this way the biseparable state

$$|\psi\rangle = \frac{1}{2}\left(|H_1\rangle|H_1\rangle + |V_1\rangle|V_2\rangle\right) \otimes \left(|H_3\rangle|H_4\rangle + |V_3\rangle|V_4\rangle\right) \tag{3.36}$$

is produced. Modes 2 and 3 then enter a polarizing beam splitter (PBS). By moving the prism, which in the figure is labelled Δ d2, the path length of mode 2 is adjusted such that the photons arrive simultaneously at the PBS. The PBS transmits horizontally and reflects vertically polarized photons. Each reflected photon acquires a phase shift of π. Thus the PBS acts like

$$|H_2\rangle|H_3\rangle \to |H_2\rangle|H_3\rangle, \qquad |H_2\rangle|V_3\rangle \to \mathrm{i}|H_2\rangle|V_2\rangle, \tag{3.37}$$

$$|V_2\rangle|H_3\rangle \to \mathrm{i}|V_3\rangle|H_3\rangle, \qquad |V_2\rangle|V_3\rangle \to -|V_2\rangle|V_3\rangle. \tag{3.38}$$

The final state is

$$|\psi'\rangle = \frac{1}{2}\big(|H_1\rangle|H_2\rangle|H_3\rangle|H_4\rangle + \mathrm{i}|H_1\rangle|H_2\rangle|V_2\rangle|V_4\rangle + \mathrm{i}|V_1\rangle|V_3\rangle|H_3\rangle|H_4\rangle - |V_1\rangle|V_2\rangle|V_3\rangle|V_4\rangle\big). \tag{3.39}$$

Each mode then goes through a noise channel (described below) and is finally measured with a setup consisting of a HWP, a quarter-wave plate (QWP), a PBS and two detectors. Postselection is employed in such a way that only those events are retained where a signal is detected in each of the four modes. This effectively projects the state in Eq. (3.39) onto the GHZ state

$$|\mathrm{GHZ}_4\rangle = \frac{1}{2}\big(|H_1\rangle|H_2\rangle|H_3\rangle|H_4\rangle + |V_1\rangle|V_2\rangle|V_3\rangle|V_4\rangle\big). \tag{3.40}$$

We tacitly assumed that a phase has been compensated.

The bit-flip noise channel is implemented by a HWP that is sandwiched between two QWPs. (The same implementation was used in Ref. 20.) Both QWPs are set at an angle of 0° between the fast axis and the vertical. The HWP is set at a variable angle θ. With these settings, the wave plates are described by the matrices [58]

$$U_{\mathrm{QWP}} = \frac{1}{\sqrt{2}}\begin{pmatrix} \mathrm{i}-1 & 0 \\ 0 & \mathrm{i}+1 \end{pmatrix} \qquad \text{and} \qquad U_{\mathrm{HWP}}(\theta) = \begin{pmatrix} \cos(2\theta) & -\sin(2\theta) \\ -\sin(2\theta) & -\cos(2\theta) \end{pmatrix}. \tag{3.41}$$

The HWP is now flipped randomly between $+\theta$ and $-\theta$. Then the three wave plates implement the channel

$$\begin{aligned} \rho \to & \frac{1}{2}\sum_{\pm} U_{\mathrm{QWP}}\, U_{\mathrm{HWP}}(\pm\theta)\, U_{\mathrm{QWP}}\, \rho\, U_{\mathrm{QWP}}^\dagger\, U_{\mathrm{HWP}}^\dagger(\pm\theta)\, U_{\mathrm{QWP}}^\dagger \\ & = \cos^2(2\theta)\,\rho + \sin^2(2\theta)\,\sigma_x\,\rho\,\sigma_x, \end{aligned} \tag{3.42}$$

which is the bit-flip noise channel with bit-flip probability $p = \sin^2(2\theta)$.

The fidelity of the prepared GHZ state is determined via

$$\begin{aligned} F &= \langle \mathrm{GHZ}_4|\rho_{\mathrm{exp}}|\mathrm{GHZ}_4\rangle \\ &= \frac{1}{2}\big\langle |0000\rangle\langle 0000| + |1111\rangle\langle 1111|\big\rangle + \frac{1}{16}\langle B_M\rangle, \end{aligned} \tag{3.43}$$

where B_M is the Bell operator of the Mermin inequality and we again used the fact that this operator is a sum of GHZ stabilizing operators. In this experiment, the fidelity without simulated bit-flip noise is $F = 0.84 \pm 0.01$.

The results of the Bell experiment are shown in Table 3.1. The angle determining the level of engineered noise is varied from $\theta = 0^\circ$ to $\theta = \pm 8^\circ$, which corresponds to bit-flip probabilities from $p = 0$ to $p = 0.076$. Each term in the Mermin inequality is measured for 800 s and each term in the Ardehali inequality for 400 s. This results in an average total count number of about 7500 for each inequality. In particular, the total count number is approximately the same for both inequalities. Without engineered noise, the

Table 3.1: Experimental values of the violation V, the statistical error $\mathcal{E}$ and the significance S for different values of the angle θ, which determines the level p of engineered bit-flip noise. Each setting in the Mermin inequality is measured for 800 s, while each setting in the Ardehali inequality is measured for 400 s. The average total count number for each inequality is about 7500. (Table taken from Ref. B.)

		Mermin			Ardehali		
θ	p	V	$\mathcal{E}$	S	V	$\mathcal{E}$	S
$\pm 0^\circ$	0	2.37	0.05	44.3	3.65	0.10	35.0
$\pm 2^\circ$	0.005	2.00	0.06	33.4	3.14	0.11	29.2
$\pm 4^\circ$	0.019	1.57	0.07	23.7	2.48	0.11	21.8
$\pm 6^\circ$	0.043	1.13	0.07	16.2	2.05	0.11	17.8
$\pm 8^\circ$	0.076	0.67	0.08	8.8	1.63	0.12	13.7

violation of the Mermin inequality is smaller than that of the Ardehali inequality. The significance, however, is larger. With increasing noise level the significance of the Mermin inequality decreases more quickly. For $\theta = \pm 6^\circ$ (corresponding to $p = 0.043$) the significance of the Ardehali inequality is already larger. When plotting the significance with respect to the fidelity the points do not lie on the theoretically predicted curves in Fig. 3.1. This is not to be expected, though, because the experimental state for $\theta = 0$ is not the perfect GHZ state with some amount of bit-flip noise. Rather, the type of noise present in this state is unknown. In conclusion, the experiment clearly confirms our prediction that the Mermin inequality has a higher significance that the Ardehali inequality for high fidelities.

4 Discrimination strategies for inequivalent classes of multipartite entangled states

This chapter, based on Ref. C, aims at answering the question for the best observables for discriminating classes of multipartite states with different entanglement properties. It is organized as follows: In Section 4.1 the discrimination problem is formulated, and a previous approach is reviewed. In Section 4.2 two measures for the discrimination strength of an observable are defined, and their interpretations are discussed. In Sections 4.3, 4.4 and 4.5 the stabilizer formalism is employed to compute these quantities for certain graph states and to find sets of observables that result in the strongest discrimination. Finally, Section 4.6 contains a conclusion and some open questions.

4.1 Statement of the problem

Multipartite entangled states differ in their entanglement properties. Two important classification schemes are based on the equivalence under local unitary (LU) operations and stochastic local operations and classical communication (SLOCC), respectively. (See Section 2.1.2 for the definitions.) Here an approach is presented for the discrimination of an experimentally prepared state from the equivalence class of another state. This is a relevant problem, since it has been shown that different classes of entangled states are suited for different applications. For example, cluster states are useful for measurement-based quantum computation, whereas GHZ states are not [117]. Conversely, for sub shot-noise interferometry, GHZ states are optimally suited, while cluster states are useless [56].

For the experimental verification of entanglement a number of tools exist – the most prominent example are witness operators (see Section 2.1.3). As experiments no longer aim only at the creation of entanglement but also at creating specific classes of entangled states, tools are needed for the experimental discrimination of these classes. In the context of entanglement detection, it is well-known that a given Bell inequality or witness operator detects only a part of all entangled states and fails to detect others. Thus the violation (or non-violation) of a Bell inequality can provide information not only about the entanglement present in a state but also about its type [65, 97].

Consequently, in Ref. 99 Schmid et al. constructed Bell operators and implemented them experimentally for discriminating different classes of entangled states. For an experiment aiming at the creation of a particular state, a Bell operator characteristic for this state was designed, that is, a Bell operator that has the desired state as eigenstate with maximal eigenvalue. The maximal expectation value of this Bell operator for various other classes of states (defined as all LU equivalents or all SLOCC equivalents of some prominent entangled state) was determined. Measuring the Bell operator then proved that the prepared state was not in those classes with maximal expectation value

lower than the experimentally obtained value. In this approach, the characteristic operator is far from unique. Neither is it necessary to use a Bell operator, as has already been remarked in Ref. 99. As entangled states with increasingly large numbers of qubits are being prepared, analysis tools that give strong results in spite of a limited number of measurement events are needed.

In the next section two measures for the discrimination strength of an observable will be defined. The first measure is based of the difference of expectation values and coincides with the one implicitly used in Ref. 99. It is shown to have an interpretation as a noise tolerance. The second measure is based on the relative entropy (see Section 2.3.1). It is directly related to the probability that measuring another state reproduces the observed measurement outcomes in a given number of measurement runs. This use of the relative entropy is motivated by a work of van Dam, Gill and Grünwald [27], where it was used to assess the statistical strength of nonlocality proofs.

It should be mentioned that the approach presented here is not directly related to the task of state discrimination as it is often discussed in the literature [19]. In particular, it does not rely on the promise that the state is always in either of two families; such an assumption cannot be justified in an experiment aiming for the verification of entanglement properties.

4.2 Distance measures

We consider the following situation: In an experiment aiming at the preparation of a state ρ the experimenter wants to verify that the prepared state is not in a certain class of undesired states, given by all local unitaries (and maybe permutations of qubits) of a pure state $|\phi\rangle$. For that purpose he or she can measure an observable A (or several observables A_k) and one has to define to what extent such a measurement can exclude the undesired states.

In this section two quantities will be defined that measure how well an observable A discriminates a state ρ from all local unitaries of another state $|\phi\rangle$. We are restricting our attention to LU classes only to simplify the calculations; the same quantities can equally well be defined for SLOCC classes.

4.2.1 A measure based on the fidelity

In analogy to the approach taken in Ref. 99, we define the fidelity-based measure as

$$\begin{aligned} \mathcal{F}_A(\rho\|\phi) &= \min_{\{U_i\}} \left|\mathrm{Tr}(A\rho) - \langle\phi|U_1^\dagger \otimes \cdots \otimes U_n^\dagger A U_1 \otimes \cdots \otimes U_n|\phi\rangle\right| \\ &= \min_{U\in\mathrm{LU}} \left|\mathrm{Tr}(A\rho) - \langle\phi|U^\dagger A U|\phi\rangle\right|. \end{aligned} \tag{4.1}$$

In the following we will always use the shorthand notation $\min_{U\in\mathrm{LU}}$ for the minimization over all local unitaries of a state. Later we will consider also the minimization over all permutations of qubits.

Adding or subtracting a multiple of the identity matrix to A does not change the value of $\mathcal{F}$ in Eq. (4.1). So without loss of generality we can assume $\mathrm{Tr}(A) = 0$. Also without loss of generality we assume $\mathrm{Tr}(A\rho) \geq 0$.

In the following we will always assume that $\mathrm{Tr}(A\rho) \geq \max_{U\in\mathrm{LU}} \langle\phi|U^\dagger A U|\phi\rangle$. This is no restriction for our purposes due to the following reasoning: Let us assume that for the pure n-qubit state $|\phi\rangle$ there exist 2^n local unitaries U_i such that $\{U_i|\phi\rangle\}$ forms an orthonormal basis. States with this property are called locally encodable [106]. It has been conjectured that all pure states are locally encodable, and the conjecture has been proven for a variety of states, including all stabilizer states and the W state [106]. Then we can write $\mathbb{1} = \sum_i U_i|\phi\rangle\langle\phi|U_i^\dagger$, and since A is traceless, $\mathrm{Tr}(A\sum_i U_i|\phi\rangle\langle\phi|U_i^\dagger) = 0$. So, if $\langle\phi|A|\phi\rangle < 0$ there exists a local unitary U such that $\langle\phi|U^\dagger A U|\phi\rangle > 0$ and vice versa. If $\max_{U\in\mathrm{LU}}\langle\phi|U^\dagger A U|\phi\rangle \geq \mathrm{Tr}(A\rho)$, by local encodability there exists a local unitary U such that $\mathrm{Tr}(A\rho) \geq 0 \geq \langle\phi|U^\dagger A U|\phi\rangle$. By continuity there exists another local unitary such that $\mathrm{Tr}(A\rho) = \langle\phi|U^\dagger A U|\phi\rangle$, that is, $\mathcal{F}_A(\rho\|\phi) = 0$ and the observable A is not suitable for a discrimination procedure based on $\mathcal{F}$. In this chapter we shall be concerned with graph states and sometimes with the W state, so local encodability is proven for our purposes and we have

$$\mathcal{F}_A(\rho\|\phi) = \mathrm{Tr}(A\rho) - \max_{U\in\mathrm{LU}} \langle\phi|U^\dagger A U|\phi\rangle. \tag{4.2}$$

This quantity, however, is not invariant under rescaling of A. To be able to compare different observables, we have to agree on a normalization. We choose $\mathrm{Tr}(A\rho) = 1$, which is the same normalization as in Ref. 99, and obtain

$$\mathcal{F}_A(\rho\|\phi) = 1 - \max_{U\in\mathrm{LU}} \langle\phi|U^\dagger A U|\phi\rangle. \tag{4.3}$$

This is the first quantity that will serve us as a measure for the strength with which A discriminates ρ from all local unitaries of $|\phi\rangle$.

For more than one observable we define

$$\mathcal{F}_{A_1,\ldots,A_L}(\rho\|\phi) = \mathcal{F}_{\frac{1}{L}\sum_{k=1}^L A_k}(\rho\|\phi). \tag{4.4}$$

In the remainder of the chapter we will discuss how to find optimal families of observables $A_1, \ldots, A_L$ for given states ρ and $|\phi\rangle$.

Our definition has a direct physical interpretation as a noise tolerance. To see this, we exploit the similarity of our problem to the task of entanglement detection by virtue of witness operators and consider the robustness of $\mathcal{F}$ in Eq. (4.2) against white noise: Let

$$\rho_{\mathrm{wn}}(p) = (1-p)\frac{\mathbb{1}}{d} + p\rho \tag{4.5}$$

be the state ρ affected by white noise. The maximal noise level $(1-p)$ such that

$$\mathrm{Tr}[A\rho_{\mathrm{wn}}(p)] - \max_{U\in\mathrm{LU}} \langle\phi|U^\dagger A U|\phi\rangle \geq 0 \tag{4.6}$$

is given by [using $\mathrm{Tr}(A) = 0$]

$$1 - p = 1 - \max_{U \in \mathrm{LU}} \langle \phi | U^\dagger A U | \phi \rangle = \mathcal{F}_A(\rho \| \phi). \tag{4.7}$$

The interpretation of $\mathcal{F}$ as a noise tolerance will be discussed in more detail in Section 4.3.3.

4.2.2 A measure based on the relative entropy

From a statistical point of view, the task of discriminating a state ρ and a state $\sigma = |\phi\rangle\langle\phi|$ by virtue of an observable A is the task of discriminating the corresponding probability distributions for the measurement outcomes of A.

The relative entropy (see Section 2.3.1) is a well-established information-theoretic measure for the discrepancy between two classical probability distributions. Note that, since we are dealing with the discrimination of classical probability distributions, we are not using the quantum (or von Neumann) relative entropy (see Section 2.3.3). The relative entropy $D(P\|Q)$ can be used to answer the question: How strongly does a sample (of a fixed length) from the distribution P on average indicate that it was indeed drawn from P rather than from Q? This interpretation was justified heuristically in Section 2.3.1. It can be made precise with the theory of statistical hypothesis testing [27]: Suppose that a sample of length N has been drawn from Q. We consider the empirical probability distribution P defined by the observed frequencies. Then the probability $Q^N[T(P)]$ of drawing a sample from Q with the same frequencies as P [i. e., within the type class $T(P)$ of P] decays exponentially for large N (see Thm. 11.1.4 in Ref. 25),

$$Q^N[T(P)] \sim 2^{-ND(P\|Q)}. \tag{4.8}$$

Consequently, if one observes a probability distribution P yielding a large value for the relative entropy $D(P\|Q)$, the assumption that it was rather drawn from the probability distribution Q is very questionable (see below for a quantitative statement).

Let us now return to our original problem: For an experiment aiming at the preparation of the state ρ, we define a measure for how well the observable A can exclude the state σ as the relative entropy of the corresponding measurement outcomes for A,

$$D_A(\rho\|\sigma) = \sum_{i=1}^{m} \mathrm{Tr}(\rho\Pi_i) \log\Big[\frac{\mathrm{Tr}(\rho\Pi_i)}{\mathrm{Tr}(\sigma\Pi_i)}\Big], \tag{4.9}$$

where $A = \sum_{i=1}^m a_i \Pi_i$ is the spectral decomposition of A. From the above discussion, this is a measure for how strongly the measurement results of the observable A on the state ρ on average show that they are due to the state ρ rather than the state σ. (Note that in this interpretation we assume that the experimental precision in implementing the observable A outperforms the precision that can be achieved for the preparation of the state ρ.)

Let us discuss the interpretation of this quantity. Suppose that the measurement has been carried out, resulting in an observed probability distribution $\widetilde{P} = (\widetilde{p}_1, \ldots, \widetilde{p}_m)$

of the outcomes $a_1, \ldots, a_m$, and let $\widetilde{N}$ be the number of measurement runs. Then, by Eq. (4.8), the probability that a measurement on the state σ, after $\widetilde{N}$ measurement runs, results in the same frequencies is given by

$$Q^{\widetilde{N}}[T(\widetilde{P})] \sim 2^{-\widetilde{N}D(\widetilde{P}\|Q)}, \tag{4.10}$$

where the distribution Q is given by $Q = (\mathrm{Tr}(\sigma\Pi_1), \ldots, \mathrm{Tr}(\sigma\Pi_m))$. If the experimentally prepared state is close enough to the intended state ρ, the relative entropy $D(\widetilde{P}\|Q)$ will attain a large value only if this is already the case for $D_A(\rho\|\sigma)$.

For comparison, when tossing a fair coin N times, the probability of the outcome always being "tails" is

$$2^{-ND((1,0)\|(\frac{1}{2},\frac{1}{2}))} = 2^{-N}, \tag{4.11}$$

since Eq. (4.8) is exact in this example. Thus, the probability Eq. (4.10) of obtaining the frequencies $\widetilde{P}$ after measuring $\widetilde{N}$ times the state σ is equal to the probability of always obtaining "tails" in $N = \widetilde{N}D(\widetilde{P}\|Q)$ tosses of a fair coin [27]. In other words, the likelihood after $\widetilde{N}$ measurement runs that the prepared state is σ is the same as that of a coin to be fair after $N = \widetilde{N}D(\widetilde{P}\|Q)$ tosses resulting in "tails". This gives our results for the measure D in Eq. (4.9) a quantitative interpretation.

When measuring several observables $A_1, \ldots, A_L$ independently of each other, the relative entropy of the joint probability distributions is given by the sum of the relative entropies for the individual observables [see Eq. (2.47)]. However, we renormalize the relative entropy in this case and define

$$D_{A_1,\ldots,A_L}(\rho\|\sigma) = \frac{1}{L}\sum_{k=1}^{L} D_{A_k}(\rho\|\sigma), \tag{4.12}$$

where the prefactor $1/L$ corresponds to keeping the overall number of measurement runs constant, independent of the number of observables, i. e., each observable A_k will be measured in $\widetilde{N}/L$ runs. We choose this definition because in experiments the rate at which entangled states are being created is typically low, so the total number of measurement runs is the critical resource.

Finally, we consider the minimum of D over all local unitaries of σ,

$$\mathcal{D}_{A_1,\ldots,A_L}(\rho\|\sigma) = \min_{U\in\mathrm{LU}} D_{A_1,\ldots,A_L}(\rho\|U\sigma U^\dagger). \tag{4.13}$$

In the following we will discuss how to find families of observables A_k which maximize this quantity.

4.3 Discriminating four-qubit states

As our first example we will calculate the quantities $\mathcal{F}$ and $\mathcal{D}$ for the discrimination of the four-qubit GHZ state [Eq. (2.95)] from the four-qubit linear cluster state [Eq. (2.98)] and vice versa. Recall that both these states are graph states (cf. Section 4.5). Their stabilizer groups were given in Eqs. (2.96) and (2.99), respectively.

The stabilizing operators of a graph state $|\psi\rangle$ provide a natural choice of observables for the discrimination of $\rho = |\psi\rangle\langle\psi|$ from other states. In the language of Ref. 99, they are characteristic operators for the graph state. In the following we will restrict our analysis to these observables.

4.3.1 Discriminating the GHZ state from the cluster state

We first consider the discrimination of the GHZ state from all LU equivalents of the cluster state, using *all* stabilizing operators of the former, excluding only the identity, as it is useless for any discrimination task. We introduce the notation $S^* = S \setminus \{\mathbb{1}\}$ for the stabilizer group S minus the identity. Later we will discuss which *subset* of the stabilizer group gives the strongest discrimination. We start with the calculation of the quantity $\mathcal{D}$ in Eq. (4.13), which is based on the relative entropy.

It is useful to think of the projector onto the cluster state as the sum of its stabilizing operators, $|\mathrm{C}_4\rangle\langle\mathrm{C}_4| = \frac{1}{16}\sum_{N\in S_{\mathrm{C}_4}} N$. For any GHZ stabilizing operator $M \in S^*_{\mathrm{GHZ}_4}$, the term D_M is a function of the overlap of M with the stabilizing operators of the cluster state,

$$D_M(\mathrm{GHZ}_4\|\mathrm{C}_4) = -\log\Big\{\frac{1}{2}\Big[1+\frac{1}{16}\sum_{N\in S_{\mathrm{C}}}\mathrm{Tr}(MN)\Big]\Big\}. \tag{4.14}$$

If we do not consider local unitaries, $\mathrm{Tr}(MN)$ is zero unless $M = \pm N$. For the minimization over local unitaries in Eq. (4.13) we classify stabilizing operators by the qubits on which they act nontrivially. For any GHZ stabilizing operator M, only those stabilizing operators of $|\mathrm{C}_4\rangle$ which act nontrivially on the exactly the same qubits as M can have a nonvanishing overlap with M. This still holds if arbitrary local unitary operations are applied to $|\mathrm{C}_4\rangle$. We can thus identify those stabilizing operators of $|\mathrm{C}_4\rangle$ that can contribute to D_M.

If no local unitary is applied, the GHZ stabilizing operators $\mathbb{1}\mathbb{1}ZZ$, $ZZ\mathbb{1}\mathbb{1}$ and $ZZZZ$ have maximal overlap with cluster stabilizing operators and thus give the minimal relative entropy of zero, while $\mathbb{1}Z\mathbb{1}Z$, $\mathbb{1}ZZ\mathbb{1}$, $Z\mathbb{1}\mathbb{1}Z$ and $Z\mathbb{1}Z\mathbb{1}$ each have zero overlap and thus give relative entropy of 1. A minimization over local unitaries cannot improve this result, as the cluster state has no stabilizing operators acting nontrivially on the same qubits.

All of the remaining stabilizing operators of $|\mathrm{GHZ}_4\rangle$,

$$\begin{aligned}\Sigma = \{&XXXX,\ -XXYY,\ -YYXX,\ YYYY,\\ &-XYXY,\ -XYYX,\ -YXXY,\ -YXYX\},\end{aligned} \tag{4.15}$$

act on all four qubits (such stabilizing operators describing four-point correlations we call four-point stabilizing operators for short). We note that both these and the four-point stabilizing operators of $|\mathrm{C}_4\rangle$ except $ZZZZ$ are products of local operators X and Y. It is therefore reasonable to assume that for the minimization of D_Σ it suffices to consider rotations about the z axes. The rotated cluster state is

$$|\mathrm{C}_4(\gamma,\delta)\rangle = \frac{1}{2}\big(|0000\rangle + \mathrm{e}^{-\mathrm{i}\delta}|0011\rangle + \mathrm{e}^{-\mathrm{i}\gamma}|1100\rangle - \mathrm{e}^{-\mathrm{i}(\gamma+\delta)}|1111\rangle\big), \tag{4.16}$$

where $\gamma = \varphi_1 + \varphi_2$, $\delta = \varphi_3 + \varphi_4$, and the φ_i are the rotation angles about the local z axes, and we obtain

$$D_\Sigma(\mathrm{GHZ}_4\|\mathrm{C}_4(\gamma,\delta)) = -\frac{1}{2}\Big\{\log\Big(\frac{1}{2}\big[1+\sin(\gamma)\sin(\delta)\big]\Big)+\log\Big(\frac{1}{2}\big[1-\cos(\gamma)\cos(\delta)\big]\Big)\Big\}. \quad (4.17)$$

The minimum of this expression is $-\log(3/4)$. In conclusion, we have found that

$$\mathcal{D}_{S^*_{\mathrm{GHZ}_4}}(\mathrm{GHZ}_4\|\mathrm{C}_4) = \frac{1}{15}\big(4-8\log\frac{3}{4}\big) \approx 0.4880. \quad (4.18)$$

Since the analytic optimization required an assumption, it should be mentioned that the same result is also obtained via numerical minimization over all local unitaries. The 15 GHZ stabilizing operators do not contribute equally to $\mathcal{D}$, rather, $D = 0$ for $\mathbb{1}\mathbb{1}ZZ$, $ZZ\mathbb{1}\mathbb{1}$ and $ZZZZ$; $D = 1$ for $\mathbb{1}Z\mathbb{1}Z$, $\mathbb{1}ZZ\mathbb{1}$, $Z\mathbb{1}\mathbb{1}Z$ and $Z\mathbb{1}Z\mathbb{1}$; and $D = -\log(3/4) \approx 0.4150$ for all others.

Let us now turn to the fidelity-based measure $\mathcal{F}$ for the same observables and states. If we use all stabilizing operators of $|\psi\rangle$, excluding again only the identity, $\mathcal{F}(\psi\|\phi)$ is a function of the fidelity

$$\mathcal{F}_{S^*_\psi}(\psi\|\phi) = \frac{2^n}{2^n-1}\Big(1-\max_{U\in\mathrm{LU}}|\langle\psi|U|\phi\rangle|^2\Big), \quad (4.19)$$

where n is the number of qubits.

For our example we note that for an arbitrary local unitary U we have

$$|\langle\mathrm{GHZ}_4|U|\mathrm{C}_4\rangle|^2 \leq \frac{1}{2}. \quad (4.20)$$

This follows from the known fact [82] that the maximal overlap of the cluster state with any product state, and thus with $|0000\rangle$ and $|1111\rangle$, is given by $1/4$. This bound is attained for example with $U = Z\mathbb{1}\mathbb{1}\mathbb{1}$. So we have

$$\mathcal{F}_{S^*_{\mathrm{GHZ}_4}}(\mathrm{GHZ}_4\|\mathrm{C}_4) = \frac{8}{15} \quad (4.21)$$

as the fidelity-based measure for the discrimination.

Let us now discuss subsets of the stabilizer group as observables for the discrimination. In our previous analysis, it turned out that not all stabilizing operators contribute equally to the discrimination, in fact, some of them do not contribute at all. We therefore ask for families of stabilizing operators of $|\mathrm{GHZ}_4\rangle$ that discriminate $|\mathrm{GHZ}_4\rangle$ from the local unitaries of $|\mathrm{C}_4\rangle$ most strongly, that is, families for which $\mathcal{F}$ (or $\mathcal{D}$) is maximal.

From the previous discussion, candidates are

$$\mathbb{1}Z\mathbb{1}Z,\ \mathbb{1}ZZ\mathbb{1},\ Z\mathbb{1}\mathbb{1}Z,\ Z\mathbb{1}Z\mathbb{1}, \quad (4.22)$$

since any of them gives $\mathcal{F} = \mathcal{D} = 1$. But if we want to exclude not only all LU equivalents, but also all permutations of qubits of the cluster state, we still have to minimize

Table 4.1: Stabilizing operators of the GHZ state and stabilizing operators of the three permutations of the cluster state acting on the same qubits. (Table taken from Ref. C.)

$\lvert \mathrm{GHZ}_4 \rangle$	$\lvert \mathrm{C}_4^{(1)} \rangle$	$\lvert \mathrm{C}_4^{(2)} \rangle$	$\lvert \mathrm{C}_4^{(3)} \rangle$
$\mathbb{1}\mathbb{1}ZZ$	$\mathbb{1}\mathbb{1}ZZ$		
$\mathbb{1}Z\mathbb{1}Z$			$\mathbb{1}Z\mathbb{1}Z$
$\mathbb{1}ZZ\mathbb{1}$		$\mathbb{1}ZZ\mathbb{1}$	
$Z\mathbb{1}\mathbb{1}Z$		$Z\mathbb{1}\mathbb{1}Z$	
$Z\mathbb{1}Z\mathbb{1}$			$Z\mathbb{1}Z\mathbb{1}$
$ZZ\mathbb{1}\mathbb{1}$	$ZZ\mathbb{1}\mathbb{1}$		

$\mathcal{F}$ and $\mathcal{D}$ over all permutations, because the set of observables is no longer necessarily permutation-invariant. There are three distinct permutations of the cluster state, namely

$$\lvert \mathrm{C}_4^{(1)} \rangle = \lvert \mathrm{C}_4 \rangle = \tfrac{1}{2}\big(\lvert 0000 \rangle + \lvert 0011 \rangle + \lvert 1100 \rangle - \lvert 1111 \rangle\big), \tag{4.23}$$

$$\lvert \mathrm{C}_4^{(2)} \rangle = \tfrac{1}{2}\big(\lvert 0000 \rangle + \lvert 0110 \rangle + \lvert 1001 \rangle - \lvert 1111 \rangle\big), \tag{4.24}$$

$$\lvert \mathrm{C}_4^{(3)} \rangle = \tfrac{1}{2}\big(\lvert 0000 \rangle + \lvert 0101 \rangle + \lvert 1010 \rangle - \lvert 1111 \rangle\big). \tag{4.25}$$

Table 4.1 shows the GHZ stabilizing operators from Eq. (4.22) along with all stabilizing operators of the permutations of $\lvert \mathrm{C}_4 \rangle$ that act on the same qubits. We see that any single one of these six stabilizing operators gives a relative entropy of zero, if the entropy is minimized over all permutations. Any pair of stabilizing operators gives an entropy of either zero or $1/2$. The three-element family $\{\mathbb{1}\mathbb{1}ZZ, \mathbb{1}Z\mathbb{1}Z, \mathbb{1}ZZ\mathbb{1}\}$ gives $2/3$, in total there are eight such families giving the same value.

It is easy to see that these families of stabilizing operators are optimal: It is clear that they are optimal among all subsets of the six stabilizing operators in the table. Furthermore, we recall that we found a local unitary transformation such that all remaining stabilizing operators contribute either 0 or $-\log(3/4)$ to the entropy. Because of the permutation invariance of the set of these remaining stabilizing operators, this holds for all permutations of $\lvert \mathrm{C}_4 \rangle$. As $-\log(3/4) < 2/3$, adding some of the remaining observables cannot improve the discrimination. This shows the optimality of our three-element families. These families are also optimal when using $\mathcal{F}$ instead of $\mathcal{D}$.

We summarize the main results of this subsection in the following observation:

Observation 4.1. *For the discrimination of the GHZ state from all local unitaries and permutations of qubits of the cluster state, using all GHZ stabilizing operators except the identity, the measures $\mathcal{D}$ and $\mathcal{F}$ are given by Eqs.* (4.18) *and* (4.21)*. When considering subsets of the stabilizer group, $\{\mathbb{1}\mathbb{1}ZZ,\ \mathbb{1}Z\mathbb{1}Z,\ \mathbb{1}ZZ\mathbb{1}\}$ is an example of an optimal family of observables, giving $\mathcal{F} = \mathcal{D} = 2/3$.*

Finally, let us add that until now we assumed that all observables are measured independently. However, as the observables in Eq. (4.22), from which we constructed the

optimal families, have a common eigenbasis of product states (the computational basis), they can also be measured jointly in one experiment with more than two outcomes. In this case, the relative entropy is no longer given by Eq. (4.12). When measuring the computational basis in one experiment with 16 outcomes, we obtain

$$\mathcal{D} = \min_{U\in \mathrm{LU}} \sum_{i=1}^{16} |\langle e_i|\mathrm{GHZ}_4\rangle|^2 \log\Big(\frac{|\langle e_i|\mathrm{GHZ}_4\rangle|^2}{|\langle e_i|U|\mathrm{C}_4\rangle|^2}\Big) = 1. \tag{4.26}$$

Consequently, considering measurements with more outcomes can give a stronger discrimination. This is a consequence of a general feature of the relative entropy: For each of the observables in Eq. (4.22), the probability distribution for the measurement outcomes is obtained from the one for the measurement of the computational basis by considering several events as one (in other words, by "forgetting" information). The relative entropy satisfies a grouping rule similar to the Shannon entropy [see Eq. (2.48)], which implies that this process can only decrease the relative entropy.

4.3.2 Discriminating the cluster state from the GHZ state

Let us now consider the reverse discrimination $\mathcal{D}_{S^*_{\mathrm{C}_4}}(\mathrm{C}_4\|\mathrm{GHZ}_4)$. This will turn out to be relatively simple.

First, note that the eight three-point stabilizing operators of $|\mathrm{C}_4\rangle$ will for any local unitary operation have zero overlap with $|\mathrm{GHZ}_4\rangle$, as the GHZ state has no three-point stabilizing operators. For the remaining eight stabilizing operators, however, the overlap with the GHZ state can brought to 1 by an appropriate rotation, such as $U = Z\mathbb{1}\mathbb{1}\mathbb{1}$. Thus

$$\mathcal{D}_{S^*_{\mathrm{C}_4}}(\mathrm{C}_4\|\mathrm{GHZ}_4) = \frac{8}{15}. \tag{4.27}$$

As a function of the fidelity, $\mathcal{F}$ is the same as for the reverse discrimination,

$$\mathcal{F}_{S^*_{\mathrm{C}_4}}(\mathrm{C}_4\|\mathrm{GHZ}_4) = \frac{8}{15}. \tag{4.28}$$

Considering the optimal subsets of the stabilizer group, it is clear that any set of three-point stabilizing operators of $|\mathrm{C}_4\rangle$ is an optimal family of observables, resulting in $\mathcal{D} = \mathcal{F} = 1$. Note that the GHZ state is permutation-invariant, so the optimization over permutations does not play a role.

4.3.3 Application to a four-photon experiment

To study the noise tolerance of the quantities $\mathcal{F}$ and $\mathcal{D}$ and their performance for experimental data, we use the measurement results for the stabilizer correlations of the cluster state obtained by Kiesel et al. in a photonic experiment [65]. When using all cluster stabilizing operators for the discrimination (excluding the identity), these data

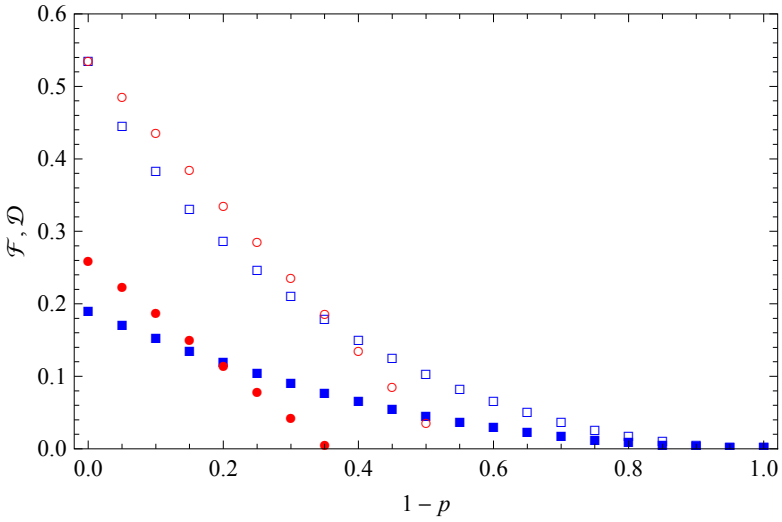

Figure 4.1: Discriminating the four-qubit linear cluster state with noise from all local unitaries of the GHZ state, using all stabilizing operators of the former. Shown are $\mathcal{F}$ (red circles) and $\mathcal{D}$ (blue squares) versus the level of white noise $(1-p)$ for the perfect (empty symbols) and the experimental (filled symbols) state. (Figure taken from Ref. C.)

give[1]

$$\mathcal{F}_{S^*_{C_4}}(\rho_{\text{exp}}\|\text{GHZ}_4) = 0.257 \pm 0.014, \tag{4.29}$$

$$\mathcal{D}_{S^*_{C_4}}(\rho_{\text{exp}}\|\text{GHZ}_4) = 0.189 \pm 0.012. \tag{4.30}$$

When using only the three-point stabilizing operators, which form an optimal family Q, we get

$$\mathcal{F}_Q(\rho_{\text{exp}}\|\text{GHZ}_4) = 0.668 \pm 0.019, \tag{4.31}$$

$$\mathcal{D}_Q(\rho_{\text{exp}}\|\text{GHZ}_4) = 0.353 \pm 0.021. \tag{4.32}$$

Note that in all cases the observables are normalized with respect to the perfect cluster state as $\langle C_4|A_k|C_4\rangle = 1$, while for the experimental data we have $\text{Tr}(\rho_{\text{exp}}A_k) < 1$. Also, it should be noted that the subsets of observables we use were chosen to be optimal for the perfect cluster state but not necessarily for the experimental one. This, however, is similar to the implementation of entanglement witnesses in experiments: There, one typically considers some optimal witness for the pure state that one aims to prepare and applies it to the experimental data in order to obtain a significant entanglement test (see [44], cf. also Chapter 3).

To investigate the power of our discrimination methods, we calculate $\mathcal{F}$ and $\mathcal{D}$ for both the perfect and the experimental cluster state under the influence of white noise.

[1] Error estimates are obtained by Gaussian error propagation from the Poissonian counting statistics and the errors in the independently determined detector efficiencies [65].

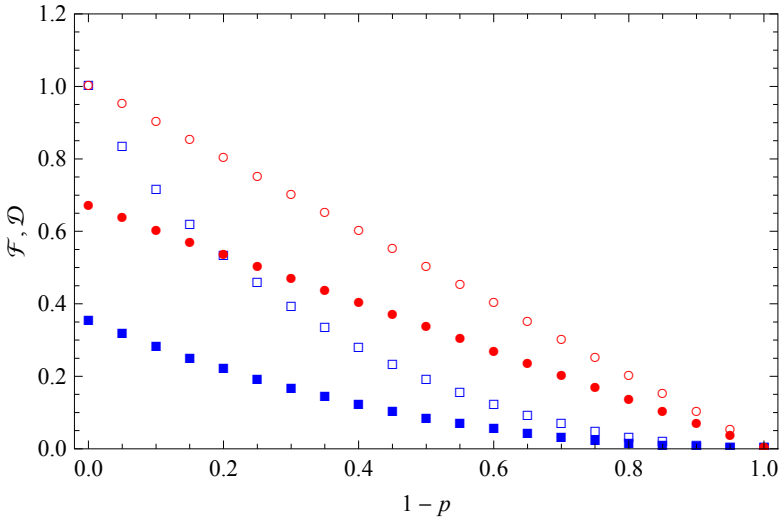

Figure 4.2: Discriminating the four-qubit linear cluster state with noise from all local unitaries of the GHZ state, using all three-point stabilizing operators of the former. Shown are $\mathcal{F}$ (red circles) and $\mathcal{D}$ (blue squares) versus the level of white noise $(1-p)$ for the perfect (empty symbols) and the experimental (filled symbols) state. (Figure taken from Ref. C.)

Figures 4.1 and 4.2 show $\mathcal{F}$ and $\mathcal{D}$ as functions of the noise level. We make a number of observations:

For the perfect cluster state with additional white noise, the quantity $\mathcal{F}$ decreases with increasing noise level until it reaches zero at the noise level given by $(1-p) = \mathcal{F}(\mathrm{C}_4\|\mathrm{GHZ}_4)$. This interpretation of $\mathcal{F}$ as a noise tolerance was already mentioned in Section 4.2.1. Note, however, that the in the case of the experimental state the noise tolerance is no longer given by $\mathcal{F}(\rho_{\mathrm{exp}}\|\mathrm{GHZ}_4)$ but is larger due to $\mathrm{Tr}(\rho_{\mathrm{exp}}A_k) < 1$.

For the same observables, the maximal noise level at which $\mathcal{D} > 0$ is at least as high as the maximal noise level at which $\mathcal{F} > 0$. This is a general feature: As a consequence of the positive definiteness of the relative entropy, D is nonzero whenever F is. In this particular example, $\mathcal{D}$ is nonzero for noise levels arbitrarily close to 1, though this is not a general feature.

For the three-point stabilizing operators of $|\mathrm{C}_4\rangle$, also the noise tolerance of $\mathcal{F}$ is 1 (Fig. 4.2). It is instructive to compare this to the case of witness operators: The set of separable states contains a ball around the completely mixed state [16], which implies that for any witness W and entangled state ρ detected by W the noise tolerance is strictly less than 1. In our case the situation is different: The reason for the noise tolerance of one is that no local unitaries of the GHZ state have any three-point correlations. This implies that the set of states LU equivalent to $|\mathrm{GHZ}_4\rangle$ does not contain a ball around $\mathbb{1}/16$. For a fair comparison, one may therefore consider the three-point correlations in experiments *aiming* at the generation of GHZ states (for instance, in Ref. 129 they were maximally 0.097) and ask whether the measured three-point stabilizer correlations in a

cluster state experiment significantly exceed these values.

Comparing the figures shows another difference between the measures $\mathcal{F}$ and $\mathcal{D}$: For the measure $\mathcal{F}$, the noise tolerance is higher in the case of the three-point stabilizing operators (Fig. 4.2) than in the case of all stabilizing operators (Fig. 4.1). This is remarkable because the former set of observables is contained in the latter. In other words, adding an observable can reduce the noise tolerance of $\mathcal{F}$. From the definition of the quantity $\mathcal{D}$ it is clear that its noise tolerance of $\mathcal{D}_{A_1,\ldots,A_L}$ is lower bounded by the noise tolerance of $\mathcal{D}$ for any subset of $A_1,\ldots,A_L$. In Ref. 6, a quantity similar to $\mathcal{F}$ was constructed from the correlations of an entangled state and used for entanglement detection. The same phenomenon of a decreasing noise tolerance when including more correlations was observed.

The preceding observations concerning the comparison of the measures $\mathcal{F}$ and $\mathcal{D}$ can be understood by noting that the relative entropy uses all information contained in the probability distributions for the measurement outcomes, whereas the quantity $\mathcal{F}$ effectively reduces each probability distribution to one parameter.

We recall that the value of the relative entropy $\mathcal{D}$ has an interpretation in terms of probabilities (Section 4.2.2). This is in contrast to the quantity $\mathcal{F}$, whose numerical value is fixed only by a normalization condition on the observables (Section 4.2.1). While in certain cases it can be interpreted as a noise tolerance and it is useful for comparing different observables, it gives no quantitative statement about the discrepancy between the prepared state and states which we want to exclude. Finally, we note that in experimental applications also the error estimates for either quantity must be taken into account. Though this has been done in Eqs. (4.29)–(4.32), a systematic analysis of this point is beyond the scope of this work.

4.4 Discriminating three-qubit states

Now we consider the three-qubit case, aiming at the discrimination of the three-qubit GHZ state and the three-qubit W state. These two states are relevant as representatives of the two different entanglement classes of genuine three-qubit entanglement [31].

The three-qubit GHZ state has the stabilizer group Eq. (2.93). The three-qubit W state

$$|\mathrm{W}_3\rangle = \frac{1}{\sqrt{3}}\big(|001\rangle + |010\rangle + |100\rangle\big) \tag{4.33}$$

is not a stabilizer state. If we expand its density matrix into Pauli matrices, we arrive at

$$\begin{aligned} |\mathrm{W}_3\rangle\langle\mathrm{W}_3| = \frac{1}{24}\big[&3\cdot\mathbb{1}\mathbb{1}\mathbb{1} + (\mathbb{1}\mathbb{1}Z + \text{perm.}) \\ &+ 2(\mathbb{1}XX + \text{perm.}) + 2(\mathbb{1}YY + \text{perm.}) - (\mathbb{1}ZZ + \text{perm.}) \\ &+ 2(XXZ + \text{perm.}) + 2(YYZ + \text{perm.}) - 3\cdot ZZZ\big]. \end{aligned} \tag{4.34}$$

4.4.1 Discriminating the GHZ state from the W state

Again, we will first compute $\mathcal{F}$ and $\mathcal{D}$ for the case that all stabilizing operators (except for $\mathbb{1}$) of the GHZ state are used, and afterwards look for optimal families of observ-

ables.

Parametrizing local unitaries as

$$U(\varphi,\theta,\psi) = \exp\big[\mathrm{i}\frac{\psi}{2}\sigma_z\big]\exp\big[\mathrm{i}\frac{\theta}{2}\sigma_y\big]\exp\big[\mathrm{i}\frac{\varphi}{2}\sigma_z\big], \tag{4.35}$$

we obtain

$$\langle \mathrm{W}_3|U^\dagger(\mathbb{1}ZZ)U|\mathrm{W}_3\rangle = \frac{1}{3}\big[-\cos(\theta_2)\cos(\theta_3) + 2\cos(\varphi_2-\varphi_3)\sin(\theta_2)\sin(\theta_3)\big]. \tag{4.36}$$

The expectation values of $Z\mathbb{1}Z$ and $ZZ\mathbb{1}$ can be obtained by cyclically permuting the indices of the angles, as the W state is permutationally invariant. We find

$$\max_{U\in\mathrm{LU}} \langle \mathrm{W}_3|U^\dagger(\mathbb{1}ZZ)U|\mathrm{W}_3\rangle = \frac{2}{3}, \tag{4.37}$$

where the maximum is attained when $\cos(\varphi_2-\varphi_3)\sin(\theta_2)\sin(\theta_3) = 1$. Thus the expectation values of $\mathbb{1}ZZ$, $Z\mathbb{1}Z$ and $ZZ\mathbb{1}$ can be maximized simultaneously. We choose the solution $\varphi_1 = \varphi_2 = \varphi_3 = 0$ and $\theta_1 = \theta_2 = \theta_3 = \pi/2$.

Let us now consider the remaining stabilizing operators. Assuming the above choice for the angles φ_i and θ_i and making the symmetry assumption $\psi_1 = \psi_2 = \psi_3 = \psi$, we have

$$\langle \mathrm{W}_3|U^\dagger(XXX)U|\mathrm{W}_3\rangle = 3\cos^3(\psi) - 2\cos(\psi), \tag{4.38}$$

$$\langle \mathrm{W}_3|U^\dagger(-XYY)U|\mathrm{W}_3\rangle = 3\cos^3(\psi) - \frac{7}{3}\cos(\psi) \tag{4.39}$$

and finally

$$\begin{aligned} &D_{\substack{XXX,-XYY,\\ -YXY,-YYX}}\big(\mathrm{GHZ}_3\big\|U|\mathrm{W}_3\rangle\big) \\ &\quad= -\log\Big\{\frac{1}{8}\big[1+3\cos^3(\psi)-2\cos(\psi)\big]\big[1+3\cos^3(\psi)-\frac{7}{3}\cos(\psi)\big]^3\Big\}. \end{aligned} \tag{4.40}$$

This function is minimal at $\cos(\psi) = -1/2$.

In conclusion, we found that

$$\mathcal{D}_{S^*_{\mathrm{GHZ}_3}}(\mathrm{GHZ}_3\|\mathrm{W}_3) = \frac{1}{7}\big(-3\log\frac{5}{6} - \log\frac{13}{16} - 3\log\frac{43}{48}\big) \approx 0.2235. \tag{4.41}$$

While the analytical calculation required some symmetry assumptions, this value is also obtained by numerical minimization over all Euler angles.

For the fidelity-based measure $\mathcal{F}$ we note that the rotation we found when minimizing $\mathcal{D}$ gives $|\langle \mathrm{GHZ}_3|U|\mathrm{W}_3\rangle|^2 = 3/4$. This is known to be the highest possible overlap of the GHZ state with local unitaries (or, indeed, SLOCCs) of the W state [1]. With Eq. (4.19) we obtain

$$\mathcal{F}_{S^*_{\mathrm{GHZ}_3}}(\mathrm{GHZ}_3\|\mathrm{W}_3) = \frac{2}{7}. \tag{4.42}$$

In Ref. 1 the state

$$|\widehat{\mathrm{W}}_3\rangle = \frac{1}{2\sqrt{6}}(3,-1,-1,-1,-1,-1,-1,3) \tag{4.43}$$

was found to maximize the overlap of $|\mathrm{GHZ}_3\rangle$ with the SLOCC class of the W state. This state is in fact LU-equivalent to $|\mathrm{W}_3\rangle$. Though $|\widehat{\mathrm{W}}_3\rangle$ minimizes F, it does not minimize D, as it gives the value $D_{S^*_{\mathrm{GHZ}_3}}(\mathrm{GHZ}_3\|\widehat{\mathrm{W}}_3) = -(6/7)\log(5/6) \approx 0.2255$.

We want to find families of observables that give the highest values of $\mathcal{F}$ and $\mathcal{D}$. We claim that any combination of

$$\mathbb{1}ZZ,\ Z\mathbb{1}Z,\ ZZ\mathbb{1} \tag{4.44}$$

is optimal in this sense among all GHZ stabilizing operators. As we have seen above, any element of this family gives $\mathcal{F} = 1/3$ and $\mathcal{D} = -\log(5/6)$. We only have to show that for any other stabilizing operator, or combination of stabilizing operators, there exists a local unitary such that $F \leq 1/3$ and $D \leq -\log(5/6)$. Taking in the above calculations $\cos(\psi) = 1$ we have $\langle \mathrm{W}_3|U^\dagger(XXX)U|\mathrm{W}_3\rangle = 1$ and $\langle \mathrm{W}_3|U^\dagger(-XYY)U|\mathrm{W}_3\rangle = 2/3$. This proves our claim.

One might ask if these optimal families of observables can be used for the construction of a witness operator for the GHZ entanglement class (see Section 2.1.2 for the definition of this class). This is, however, not the case, as

$$\max_{\text{SLOCC of } \mathrm{W}_3} \langle \mathrm{W}_3|(\mathbb{1}ZZ + Z\mathbb{1}Z + ZZ\mathbb{1})|\mathrm{W}_3\rangle = 3 \tag{4.45}$$

holds, which can be seen from the fact that the fully separable state $|000\rangle$ gives this value.

We summarize the main results of this subsection:

Observation 4.2. *If all GHZ stabilizing operators except the identity are used for the discrimination of the GHZ state from all LU equivalents of the W state, the measures $\mathcal{D}$ and $\mathcal{F}$ are given by Eqs.* (4.41) *and* (4.42)*. If we consider subsets of the stabilizer group, any combination of $\mathbb{1}ZZ$, $Z\mathbb{1}Z$ and $ZZ\mathbb{1}$ is an optimal family of observables, yielding $\mathcal{F} = 1/3$ and $\mathcal{D} = -\log(5/6)$.*

4.4.2 Discriminating the W state from the GHZ state

For the reverse problem, the discrimination of $|\mathrm{W}_3\rangle$ from the local unitaries of $|\mathrm{GHZ}_3\rangle$, there is no obvious choice of observables, as the W state is not a stabilizer state. However, each of the observables

$$\mathbb{1}\mathbb{1}Z,\ \mathbb{1}Z\mathbb{1},\ Z\mathbb{1}\mathbb{1} \tag{4.46}$$

has an expectation value of $1/3$ for the W state and zero expectation value for all local unitaries of $|\mathrm{GHZ}_3\rangle$, as for the GHZ state all reduced one-qubit density matrices are maximally mixed. Thus any appropriately normalized combination of the above observables gives $\mathcal{F} = 1$, which is the optimal value.

All of these combinations give

$$\mathcal{D}(\mathrm{W}_3\|\mathrm{GHZ}_3) = \frac{2}{3}\log\frac{4}{3} - \frac{1}{3}\log\frac{3}{2} \approx 0.0817. \tag{4.47}$$

This is the best possible value among all operators occurring in the decomposition of $|W_3\rangle$ in Eq. (4.34). To see this, we choose the rotation $\varphi = 2\pi/3$, $\theta = 3\pi/2$ and $\psi = 5\pi/4$ on all qubits and observe that all these operators give a value of D less or equal to the one in Eq. (4.47).

4.5 General graph states

In the previous sections we observed that the stabilizing operators of a given state are natural candidates for observables that discriminate this state from other states. In this section we will derive some general statements about the discrimination of graph states.

In our discussion on the discrimination of the GHZ from the cluster state we learned that the optimal families of observables consist of two-point stabilizing operators. The reason for their optimality is that the cluster state has fewer two-point stabilizing operators than the GHZ state. Hence one may try do derive general results depending only on the numbers of two-point (or higher-order) stabilizing operators.

The number of two-point correlations of a graph state can easily be obtained from its graph [38, 56]. Restricting ourselves to connected graphs with three or more vertices, there are three possibilities to obtain two-point stabilizing operators:

1. Vertices connected to the rest of the graph by only one edge. The generator associated to such a vertex is a two-point operator of the form XZ.

2. Pairs of unconnected vertices whose neighbourhoods are equal: $N(i) = N(j)$. The product of the two generators associated to such a pair has the form XX.

3. Pairs of connected vertices for which $N(i) \cup \{i\} = N(j) \cup \{j\}$. This means that their neighbourhoods apart from i and j are the same. The product of their generators has the form YY.

The product of the generators associated to vertices i and j is never equal to the identity at positions i and j. Therefore it is a two-point stabilizing operator if and only if it is equal to the identity at all other positions, which leaves only the possibilities 2 and 3. For the same reason the product of three or more generators is never a two-point stabilizing operator. This shows that the above list exhausts all possibilities to obtain a two-point stabilizing operator.

Now, let $|G_1\rangle$ and $|G_2\rangle$ be two graph states, k_1 and k_2 the numbers of two-point correlations of these states, and let P_{G_1} be the set of two-point stabilizing operators of $|G_1\rangle$. We assume $k_1 > k_2$ (our result will be trivial otherwise). We can then derive a lower bound on $\mathcal{F}_{P_{G_1}}(G_1\|G_2)$ and $\mathcal{D}_{P_{G_1}}(G_1\|G_2)$ that depends only on the numbers k_1 and k_2. Namely, we have

$$\mathcal{F}_{P_{G_1}}(G_1\|G_2) \geq \frac{k_1 - k_2}{k_1}, \tag{4.48}$$

$$\mathcal{D}_{P_{G_1}}(G_1\|G_2) \geq \frac{k_1 - k_2}{k_1}. \tag{4.49}$$

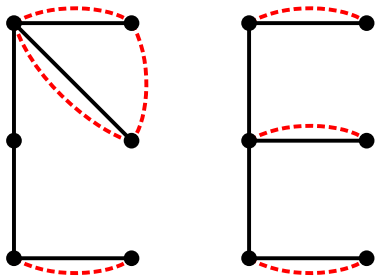

Figure 4.3: Graphs of the graph states discussed in Section 4.5. Dashed red lines denote the presence of two-point correlations. (Figure taken from Ref. C.)

To see this, note that from the above discussion it is clear that all two-point stabilizing operators of the same graph state act on different pairs of qubits [38]. It follows that any two-point stabilizing operator of $|G_2\rangle$ can have nonzero overlap with at most one stabilizing operator of $|G_1\rangle$. This shows that at least $k_1 - k_2$ two-point stabilizing operators of $|G_1\rangle$ will have zero overlap with $|G_2\rangle$, and thus give $\mathcal{F} = \mathcal{D} = 1$. Note that this still holds if local unitaries and permutations of qubits are considered.

For $|G_1\rangle = |\mathrm{GHZ}_4\rangle$ and $|G_2\rangle = |\mathrm{C}_4\rangle$ the bounds give the exact results. This is, however, not always the case, as the example of Fig. 4.3 shows: here, $k_1 = 4$ and $k_2 = 3$. Of the two-point correlations of $|G_1\rangle$, three connect three qubits in a triangle, while for $|G_2\rangle$ the connected pairs are all disjoint. This shows that at most two of the two-point stabilizing operators of $|G_1\rangle$ can have nonzero overlap with $|G_2\rangle$. In this example, $\mathcal{F}_{P_{G_1}}(G_1\|G_2) = \mathcal{D}_{P_{G_1}}(G_1\|G_2) = 1/2 > 1/4 = (k_1 - k_2)/k_1$.

While one can also use higher-order stabilizing operators for the discrimination, it is more difficult to derive general results for them. Nevertheless, for two given graph states the number of three-point (or higher-order) correlations can directly be computed by writing down the whole stabilizer group; one may then compare the different numbers of higher order stabilizing operators.

4.6 Conclusion and outlook

Concerning the comparison of the two measures $\mathcal{F}$ and $\mathcal{D}$, the observation that the measure $\mathcal{D}$ is more robust against noise than $\mathcal{F}$ could be explained by the fact that the relative entropy uses all information contained in the probability distribution, while the measure $\mathcal{F}$ effectively reduces each probability distribution to one parameter (namely, the sum of the expectation values). This leads to the following conclusion: Either measure can be used to compare the suitability of different observables for a given discrimination task. For the evaluation of experimental results the relative entropy $\mathcal{D}$ is to be preferred, because it uses all available information and allows a clear statistical interpretation.

There are several interesting open questions and possible generalizations: First, one

could extend the analysis to the measurement of non-dichotomic observables or arbitrary product bases [see the discussion of Eq. (4.26)]. Second, one could consider SLOCC equivalence classes instead of LU classes. The measures $\mathcal{F}$ and $\mathcal{D}$ are equally applicable to that case, only the optimization is different. Third, one could connect our results to other methods for characterizing multipartite entanglement classes. For instance, there exist witnesses distinguishing the class of mixed three-qubit GHZ states from the class of mixed W states [1] (see Section 2.1.2 for the definition of these classes). The discrimination of such classes of mixed states is a different problem than the one considered in this chapter; nevertheless, it would be interesting to understand possible connections.

5 Entropic uncertainty relations and the stabilizer formalism

This chapter, based on Ref. D, is concerned with entropic uncertainty relations. A short introduction to this field was given in Section 2.4.2.

Central questions in the theory of entropic uncertainty relations include the derivation of lower bounds on the total uncertainty for given observables, the characterization of observables that admit strong uncertainty relations and the construction of such relations for the case of several observables. In this chapter it is demonstrated how the stabilizer formalism (see Section 2.5) can be applied to these questions.

The chapter is organized as follows: In Section 5.1 conditions are studied which guarantee that the Maassen-Uffink relation is tight. Generalizing a well-known result on mutually unbiased bases, it is shown that this is the case for the measurement in any pair of stabilizer bases, which is a consequence of a deeper geometric property of these bases. In Section 5.2 this result is applied to the special case of graph state bases, demonstrating how the graph formalism helps to compute the maximal overlap of the basis vectors, which determines the lower bound on the entropy sum. Section 5.3 is devoted to entropic uncertainty relations for several dichotomic, pairwise anticommuting observables. Generalizing a result by Wehner and Winter [123], a systematic construction of such relations is presented. The family of uncertainty relations which is obtained contains both entropic and variance-based ones, and their relative strengths are compared. Finally, in Section 5.4 the relations are applied to the stabilizing operators of two stabilizer states.

5.1 A generalization of mutual unbiasedness

Our starting point is the Maassen-Uffink relation [69, 79, 80], which was given before in Eq. (2.69), but is repeated here for convenience: for any two measurement bases $\mathcal{A} = \{|a_i\rangle\}$ and $\mathcal{B} = \{|b_i\rangle\}$, $i = 1, \ldots, d$,

$$\frac{1}{2}\big[S(\mathcal{A}|\rho) + S(\mathcal{B}|\rho)\big] \geq -\log\big(\max_{i,j} |\langle a_i|b_j\rangle|\big). \tag{5.1}$$

It is well-known that this relation is maximally strong [meaning that the right-hand side attains the maximal possible value of $\log(d)/2$] and tight [meaning that there exists a state for which equality holds] if the measurement bases are mutually unbiased,

$$|\langle a_i|b_j\rangle| = \frac{1}{\sqrt{d}} \qquad \forall\, i, j. \tag{5.2}$$

It is not difficult to find a more general condition which is sufficient for tightness:

Lemma 5.1. *If a pair of bases $\mathcal{A} = \{|a_i\rangle\}$ and $\mathcal{B} = \{|b_i\rangle\}$ satisfies*

$$|\langle a_i|b_j\rangle| \in \{0, r\} \qquad \forall\, i, j \tag{5.3}$$

for some r, then the Maassen-Uffink relation Eq. (5.1) *for the measurement in these bases is tight. Equality holds for any of the basis states, $\rho = |a_i\rangle\langle a_i|$ or $\rho = |b_i\rangle\langle b_i|$.*

Proof. For $\rho = |b_{j_0}\rangle\langle b_{j_0}|$ the Maassen-Uffink relation reads

$$-\sum_i p_i \log(p_i) \geq -\log(\max_i p_i) \qquad \text{where} \qquad p_i = |\langle a_i|b_{j_0}\rangle|^2. \tag{5.4}$$

Note that the right-hand side is the min-entropy of the probability distribution p_i. By assumption $p_i \in \{0, r^2\}$ for all i. It follows that equality holds. □

We now assume an n-qubit system. Recall that the elements of a stabilizer group (see Section 2.5.1 for the definition) have a basis of common eigenvectors. If the group has the maximal cardinality of 2^n, the eigenbasis is unique. We call this basis a stabilizer basis. In this chapter all stabilizer groups are assumed to have cardinality 2^n. The main result of this section is the following theorem:

Theorem 5.2. *Any pair of stabilizer bases $\mathcal{A} = \{|a_i\rangle\}$ and $\mathcal{B} = \{|b_i\rangle\}$ satisfies*

$$|\langle a_i|b_j\rangle| \in \{0, r\} \qquad \forall\, i, j \tag{5.5}$$

for some r. As a consequence, the Maassen-Uffink uncertainty relation Eq. (5.1) *is tight for the measurement in these bases. The bound is attained with any of the basis states.*

Two proofs of this theorem will be given. The first proof is based on the theory of mutually unbiased bases; the second proof is more elementary, requiring only basic results of stabilizer theory. For the special case of graph state bases, the next section contains a third proof.

The first proof makes use of the following construction method for mutually unbiased bases, which is due to Bandyopadhyay et al. (see the proof of Thm. 3.2 in Ref. 10):

Theorem 5.3 (Ref. 10). *Let C_1 and C_2 each be a set of d commuting unitary $d \times d$-matrices. Furthermore, assume that $C_1 \cap C_2 = \{\mathbb{1}\}$ and that all matrices in $C_1 \cup C_2$ are pairwise orthogonal with respect to the Hilbert-Schmidt scalar product. Then the eigenbases defined by either set of matrices are mutually unbiased.*

Two Pauli operators (tensor products of Pauli matrices and the identity with prefactors ± 1 or $\pm \mathrm{i}$) are either equal up to a phase or orthogonal in the Hilbert-Schmidt sense. For stabilizing operators the phase factor can only be ± 1, because these operators are Hermitian. If we consider two stabilizing operators that differ only by a factor -1 as equal, the above theorem shows that two stabilizer groups whose intersection contains only the identity define mutually unbiased bases. In the first proof of Theorem 5.2 this result is generalized in the following way: It is shown that two stabilizer groups (which may have nontrivial intersection) define bases which are mutually unbiased on a factor space of the Hilbert space and equal on the other factor. Such a pair of bases satisfies Eq. (5.5).

Proof of Theorem 5.2. Throughout the proof we consider two stabilizing operators that differ only by a minus sign as equal. Let S_1 and S_2 be the two stabilizer groups and define $C_0 = S_1 \cap S_2$. Then C_0 is a subgroup of both S_1 and S_2. We consider the factor groups $C_1 = S_1/C_0$ and $C_2 = S_2/C_0$. The groups C_0, C_1 and C_2 are all stabilizer groups. The dimension of the space stabilized by the group C_k is 2^{n-m_k}, where m_k is the cardinality of the group. This gives us a decomposition of the Hilbert space $\mathcal{H} = \mathcal{H}_0 \otimes \mathcal{H}_{12}$, where C_0 acts trivially on $\mathcal{H}_{12}$ and C_1 and C_2 act trivially on $\mathcal{H}_0$. By the previous theorem, C_1 and C_2 define mutually unbiased bases $|{c_i}^{(1)}\rangle$ and $|{c_i}^{(2)}\rangle$ of $\mathcal{H}_{12}$. It follows that the stabilizer bases can be written as

$$|s_{ij}^{(1)}\rangle = |c_i^{(0)}\rangle \otimes |c_j^{(1)}\rangle \qquad \text{and} \qquad |s_{ij}^{(2)}\rangle = |c_i^{(0)}\rangle \otimes |c_j^{(2)}\rangle, \tag{5.6}$$

where $|{c_i}^{(0)}\rangle$ is the basis of $\mathcal{H}_0$ defined by C_0. The stabilizer bases thus satisfy Eq. (5.5) with $r = (\dim H_{12})^{-1/2}$. □

Alternative proof of Theorem 5.2. Let $S = \{M_k\}$ and $T = \{N_k\}$ be the two stabilizer groups, and let $|S\rangle$ and $|T\rangle$ be the corresponding stabilizer states. Define $S^+ = S \cap T$, where, unlike in the first proof, we consider two operators as distinct if they differ by a minus sign. Also define $S^- = S \cap -T$. Both S^+ and $S^+ \cup S^-$ are easily seen to be subgroups of S. By Lagrange's theorem, they have cardinalities $|S^+| = 2^p$ and $|S^+ \cup S^-| = 2^q$ with some $p \in \{1, 2, \ldots, n\}$ and $q \in \{p, p+1, \ldots, n\}$. The projectors onto the stabilizer states are given by their stabilizing operators as $|S\rangle\langle S| = \frac{1}{2^n} \sum_{k=1}^{2^n} M_k$ and similarly for $|T\rangle$ [see Eq. (2.72)]. Thus

$$\begin{aligned} |\langle S|T\rangle|^2 &= \frac{1}{2^{2n}} \sum_{k,\ell=1}^{2^n} \operatorname{Tr}(M_k N_\ell) \\ &= \frac{1}{2^n}\left(|S^+| - |S^-|\right) \\ &= \frac{1}{2^n}(2^{p+1} - 2^q) \\ &= \begin{cases} 2^{q-n} & \text{for } p = q, \\ 0 & \text{for } p = q-1. \end{cases} \end{aligned} \tag{5.7}$$

The case $p < q - 1$ cannot occur, because it would give a negative value of $|\langle S|T\rangle|^2$ and thus lead to a contradiction.

Consider now another vector $|T'\rangle$ of the stabilizer basis defined by the group T. This vector is again a stabilizer state, and its stabilizing operators are equal to those of $|T\rangle$ up to some minus signs. In particular, $S^+ \cup S^-$ and thus q are the same for $|T\rangle$ and for $|T'\rangle$. □

The converse of the theorem on mutually unbiased bases (Theorem 5.3) also holds:

Theorem 5.4 (Ref. 10). *If $\mathcal{A}_1, \ldots, \mathcal{A}_L$ is a set of L mutually unbiased bases in $\mathbb{C}^d$, there are L sets $C_1, \ldots, C_L$, each consisting of d commuting unitary $d \times d$-matrices, such that all matrices in $C_1 \cup \cdots \cup C_L$ are pairwise orthogonal.*

Though it is not stated explicitly in that reference, the sets of unitary matrices C_k that are constructed in the proof of the theorem are groups. This raises the question if the converse of Theorem 5.2 also holds in the sense that any pair of bases satisfying Eq. (5.5) originates from two groups of unitary matrices (cf. the first proof of Theorem 5.2).

5.2 Application to graph state bases

The aim of this section is to determine the right-hand side of the Maassen-Uffink relation explicitly for pairs of graph state bases. The main result takes the form of a recurrence relation for the maximal overlap $\max_{i,j}|\langle a_i|b_j\rangle|$ of the basis vectors. This relation connects the overlap of two n-qubit bases to the overlap of the two $(n-1)$-qubit bases which are obtained from them by deleting one vertex of the graphs.

An introduction to graph states and graph state bases was given in Section 2.5.2. Recall that we denoted the basis vectors as $|G, \mathbf{x}\rangle$, where G stands for the graph and the bitstring $\mathbf{x} = (x_1, \ldots, x_n)$ with $x_i \in \{0, 1\}$ labels the individual vectors. Furthermore, recall that the vector $|G, \mathbf{x}\rangle$ is obtained from $|+\rangle^{\otimes n}$ by applying first a controlled phase gate C_{ij} for each edge $(i, j) \in E$ of the graph and then a local phase σ_z for each vertex i with $x_i = 1$,

$$|G, \mathbf{x}\rangle = \prod_{i=1}^{n} \left(\sigma_z^{(i)}\right)^{x_i} \prod_{(k,\ell)\in E} C_{k\ell}\, |+\rangle^{\otimes n}. \tag{5.8}$$

Since all these operations commute, we can move all phase gates in the scalar product $\langle G_1, \mathbf{x}|G_2, \mathbf{y}\rangle$ to the right and all local phases to the left,

$$\begin{aligned} \langle G_1, \mathbf{x}|G_2, \mathbf{y}\rangle &= \langle +|^{\otimes n} \prod_{(k,\ell)\in E_1} C_{k\ell} \prod_{i=1}^{n} \left(\sigma_z^{(i)}\right)^{x_i} \prod_{j=1}^{n} \left(\sigma_z^{(j)}\right)^{y_j} \prod_{(m,p)\in E_2} C_{mp}\, |+\rangle^{\otimes n} \\ &= \langle +|^{\otimes n} \prod_{i=1}^{n} \left(\sigma_z^{(i)}\right)^{x_i+y_i} \prod_{(k,\ell)\in E_1 \oplus E_2} C_{k\ell}\, |+\rangle^{\otimes n} \\ &= \langle G_0, \mathbf{x} \oplus \mathbf{y}|G_1 \oplus G_2, \mathbf{0}\rangle. \end{aligned} \tag{5.9}$$

Because of $(C_{k\ell})^2 = \mathbb{1}$ this corresponds to replacing G_1 by the empty or completely unconnected graph G_0 and G_2 by the "sum modulo 2" of the graphs, denoted here as $G_1 \oplus G_2$. Similarly, $\mathbf{x} \oplus \mathbf{y}$ denotes the sum modulo 2 of the bitstrings. The graph state basis of the empty graph is given by $|G_0, \mathbf{x}\rangle = H^{\otimes n}|\mathbf{x}\rangle$, where $H = 2^{-1/2}\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$ is the Hadamard gate and $|\mathbf{x}\rangle$ is a vector of the standard basis in binary notation.

As the Hadamard gate is a local Clifford operation, the vector $H^{\otimes n}|G_1 \oplus G_2, \mathbf{0}\rangle$ is a stabilizer state. This shows that the scalar products $\langle G_1, \mathbf{x}|G_2, \mathbf{y}\rangle = \langle \mathbf{x} \oplus \mathbf{y}|H^{\otimes n}|G_1 \oplus G_2, \mathbf{0}\rangle$ can be understood as the coefficients of a stabilizer state with respect to the standard basis. It has been shown that for any stabilizer state these coefficients are 0, ± 1 and $\pm \mathrm{i}$, up to a global normalization. (See Thm. 5 and the paragraph below in Ref. 29. For an alternative proof see Ref. 113. I thank Maarten Van den Nest for pointing out these references.) This constitutes yet another proof of Theorem 5.2 for the case of graph state bases.

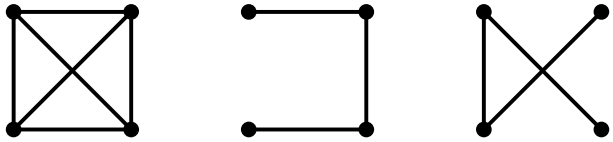

Figure 5.1: Graphs of the graph states discussed in Section 5.2. (Figure taken from Ref. D.)

As an example consider the graphs in Fig. 5.1 (a) and (b). Up to local unitary operations, the corresponding graph states are the four-qubit GHZ state and the four-qubit linear cluster state, but the uncertainty relation is not invariant under local unitaries. The above line of thoughts shows that the Maassen-Uffink bound for the corresponding bases is equal to the bound for the empty graph and the sum modulo 2 of the graphs, which in this case is given by Fig. 5.1 (c). The latter is again equal to the graph (b), up to a permutation of vertices.

Let us return to the explicit calculation of the overlaps. We have seen that it suffices to consider scalar products of the form $\langle G_0, \mathbf{y}|G, \mathbf{0}\rangle$. The relation between the graph state basis of the empty graph and the standard basis is given by

$$|G_0, \mathbf{y}\rangle = \frac{1}{2^{n/2}} \sum_{\mathbf{x}} (-1)^{\sum_i y_i x_i} |\mathbf{x}\rangle. \tag{5.10}$$

The coefficients of the graph state $|G, \mathbf{0}\rangle$ with respect to the standard basis are determined by its adjacency matrix Γ as [see Eq. (2.83)]

$$|G, \mathbf{0}\rangle = \frac{1}{2^{n/2}} \sum_{\mathbf{x}} (-1)^{\sum_{i<j} x_i \Gamma_{ij} x_j} |\mathbf{x}\rangle. \tag{5.11}$$

The scalar products are thus given by

$$\langle G_0, \mathbf{y}|G, \mathbf{0}\rangle = \frac{1}{2^n} \sum_{\mathbf{x}} (-1)^{\sum_i y_i x_i + \sum_{i<j} x_i \Gamma_{ij} x_j}. \tag{5.12}$$

To derive a recurrence relation, we write

$$\mathbf{x} = \begin{pmatrix} \xi \\ \mathbf{x}' \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} v \\ \mathbf{y}' \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 0 & \gamma'^t \\ \gamma' & \Gamma' \end{pmatrix}. \tag{5.13}$$

We obtain

$$\begin{aligned} \langle G_0, \mathbf{y}|G, \mathbf{0}\rangle &= \frac{1}{2^n} \sum_{\xi \in \{0,1\}} \sum_{\mathbf{x}'} (-1)^{v\xi + \sum_i y'_i x'_i + \xi \sum_i \gamma'_i x'_i + \sum_{i<j} x'_i \Gamma'_{ij} x'_j} \\ &= \frac{1}{2^n} \sum_{\mathbf{x}'} (-1)^{\sum_i y'_i x'_i + \sum_{i<j} x'_i \Gamma'_{ij} x'_j} + \frac{1}{2^n} (-1)^v \sum_{\mathbf{x}'} (-1)^{\sum_i (y'_i + \gamma'_i) x'_i + \sum_{i<j} x'_i \Gamma'_{ij} x'_j} \\ &= \frac{1}{2} \langle G_0, \mathbf{y}'|G', \mathbf{0}\rangle + (-1)^v \frac{1}{2} \langle G_0, \mathbf{y}' + \gamma'|G', \mathbf{0}\rangle, \end{aligned} \tag{5.14}$$

where G' is the graph defined by the adjacency matrix Γ'. This is the desired recurrence relation.

From Theorem 5.2 we already know that $|\langle G_0, \mathbf{y}'|G', \mathbf{0}\rangle| \in \{0, r'\}$ for all $\mathbf{y}'$ for some r', which depends on the graph G'. Since the scalar products are real, this means that $\langle G_0, \mathbf{y}'|G', \mathbf{0}\rangle \in \{0, \pm r'\}$. From the recurrence relation we learn that $\langle G_0, \mathbf{y}|G, \mathbf{0}\rangle \in \{0, \pm r'/2, \pm r'\}$. On the other hand, $\langle G_0, \mathbf{y}|G, \mathbf{0}\rangle \in \{0, \pm r\}$ for some r. This shows that either $r = r'/2$ or $r = r'$, depending on γ' and G'.

The graph state bases are mutually unbiased and the Maassen-Uffink relation is maximally strong precisely if $r = 2^{-n/2}$. On the other hand, r is an integer multiple of 2^{-n} (even 2^{1-n}), as one can see by induction with the recurrence relation. This shows that a pair of graph state bases is never mutually unbiased if the number of qubits is odd.

We will now use the recurrence relation to compute r for certain classes of graph states. As above we shall assume one graph to be empty and vary only the other one. The application of the recurrence relation is particularly easy if $\langle G_0, \mathbf{y}'|G', \mathbf{0}\rangle = \pm r'$ for all $\mathbf{y}'$, which is the case if G_0 and G' define mutually unbiased bases.

Consider the fully connected graph G, defined by the adjacency matrix $\Gamma_{ij} = 1 - \delta_{ij}$, whose graph state is LC-equivalent to the GHZ state. First we show by induction that for an even number n of qubits the fully connected graph has $\langle G_0, \mathbf{y}|G, \mathbf{0}\rangle = \pm 2^{-n/2}$ for all $\mathbf{y}$: For $n = 2$ one finds $r = 1/2$. Assume the assertion to be true for $n - 2$. Let G' and G'' be the fully connected graphs with $n - 1$ and $n - 2$ qubits, respectively, and define the $(n-1)$-vector $\gamma' = (1, 1, \ldots, 1)$ and the $(n-2)$-vector $\gamma'' = (1, 1, \ldots, 1)$. For $\mathbf{y} = \mathbf{0}$, application of the recurrence relation gives

$$\begin{aligned}
\langle G_0, \mathbf{0}|G, \mathbf{0}\rangle &= \frac{1}{2}\langle G_0, \mathbf{0}|G', \mathbf{0}\rangle + \frac{1}{2}\langle G_0, \gamma'|G', \mathbf{0}\rangle \\
&= \frac{1}{2}\Big(\frac{1}{2}\langle G_0, \mathbf{0}|G'', \mathbf{0}\rangle + \frac{1}{2}\langle G_0, \gamma''|G'', \mathbf{0}\rangle\Big) \\
&\qquad + \frac{1}{2}\Big(\frac{1}{2}\langle G_0, \gamma''|G'', \mathbf{0}\rangle - \frac{1}{2}\langle G_0, \mathbf{0}|G'', \mathbf{0}\rangle\Big) \\
&= \frac{1}{2}\langle G_0, \gamma''|G'', \mathbf{0}\rangle \\
&= \pm\frac{1}{2}\frac{1}{2^{(n-2)/2}} \\
&= \pm\frac{1}{2^{n/2}}.
\end{aligned} \tag{5.15}$$

This implies that $\langle G_0, \mathbf{y}|G, \mathbf{0}\rangle \in \{0, \pm 2^{-n/2}\}$ for all $\mathbf{y}$. But because of normalization $\langle G_0, \mathbf{y}|G, \mathbf{0}\rangle = 0$ is not possible. This shows the assertion. For the fully connected graph with an odd number of qubits we have

$$\langle G_0, \mathbf{y}|G, \mathbf{0}\rangle = \frac{1}{2}\langle G_0, \mathbf{y}'|G, \mathbf{0}\rangle \pm \frac{1}{2}\langle G_0, \mathbf{y}' + \gamma'|G, \mathbf{0}\rangle \in \Big\{0, \pm\frac{1}{2^{(n-1)/2}}\Big\}. \tag{5.16}$$

The generalization of the state in Fig. 5.1 (b), which is equivalent under local unitary operations to the linear cluster state, can be treated in exactly the same way, yielding the same values of r_n.

In Sec. 4 of Ref. 10 a general method for the construction of sets of mutually unbiased bases in prime power dimensions $d = p^n$ was given. Though the term "graph" is never used in that article, the construction uses a form of the graph formalism. (The matrices $A_1, \ldots, A_\ell$ in Thm. 4.4 play the role of adjacency matrices.) It should be noted that the method of that reference is more general than the graph state formalism as it is used in this thesis: Firstly, the method is not restricted to qubits, but the Hilbert space dimension of the constituent systems can be any prime number. Secondly, the diagonal elements of the adjacency matrices are not required to be zero. In graph theoretic terms this means that loops are allowed. The focus of that article is however different to this thesis: There, the authors' main interest is in finding a general construction of the maximal number of $p^n + 1$ mutually unbiased bases. They are not interested in explicit calculations for given graphs. They also do not consider the overlap of graph state bases that are not mutually unbiased.

5.3 Uncertainty relations for several dichotomic anticommuting observables

Little is known about uncertainty relations for more than two measurements [124] (see, however, Ref. 81). Following Wehner and Winter [123], this section concentrates on dichotomic anticommuting observables. An observable is called dichotomic if it has exactly two distinct eigenvalues. We will always normalize dichotomic observables such that their eigenvalues are ± 1. In other words, these observables square to the identity.

The following result has been called a meta-uncertainty relation [123,124], for reasons that soon will become apparent.

Lemma 5.5. *Let $A_1, \ldots, A_L$ be observables which anticommute pairwise, $\{A_k, A_\ell\} = 0$ for $k \neq \ell$, and which have eigenvalues ± 1. Then $\sum_{k=1}^{L} \langle A_k \rangle^2 \leq 1$, or equivalently,*

$$\sum_{k=1}^{L} \Delta^2(A_k) \geq L - 1, \tag{5.17}$$

where $\Delta^2(A) = \langle (A - \langle A \rangle)^2 \rangle$ is the variance of A.

The following proof of this lemma was given in Ref. 109. For an alternative proof, based on the Clifford algebra, see Ref. 123.

Proof. Choose real coefficients $\lambda_1, \ldots, \lambda_L$ with $\sum_{k=1}^{L} \lambda_k^2 = 1$. Because of the anticommutativity of the observables and $A_k^2 = \mathbb{1}$ we have $(\sum_{k=1}^{L} \lambda_k A_k)^2 = \sum_{k=1}^{L} \lambda_k^2 A_k^2 = \sum_{k=1}^{L} \lambda_k^2 \mathbb{1} = \mathbb{1}$ and thus $|\sum_{k=1}^{L} \lambda_k \langle A_k \rangle| = |\langle \sum_{k=1}^{L} \lambda_k A_k \rangle| \leq 1$ for all states, because for any observable $\langle X \rangle^2 \leq \langle X^2 \rangle$. Interpreting the expression $\sum_{k=1}^{L} \lambda_k \langle A_k \rangle$ as the Euclidian scalar product of the vector of coefficients λ_k and the vector of expectation values $\langle A_k \rangle$, and noting that the vector of coefficients λ_k is an arbitrary unit vector, we see that the vector $\langle A_k \rangle$ has a length less than or equal to 1. Observing $\sum_{k=1}^{L} \langle A_k^2 \rangle = L$, we obtain the lemma. □

The converse implication is also true in the following sense, as was already shown in Ref. 123:

Lemma 5.6. *Let $A_1, \ldots, A_L$ be dichotomic anticommuting observables as above, and choose arbitrary real numbers $a_1, \ldots, a_L$ with $\sum_{k=1}^{L} a_k^2 \leq 1$. Then there exists a quantum state ρ such that the numbers a_k are the expectation values of the observables, $a_k = \mathrm{Tr}(A_k \rho)$.*

Proof. Consider the state $\rho = \frac{1}{d}(\mathbb{1} + \sum_{k=1}^{L} a_k A_k)$, where d is the dimension of the Hilbert space. Because of the properties of the observables, $\mathrm{Tr}(A_k A_\ell) = d\delta_{k\ell}$. Furthermore, the observables A_k are traceless: $\mathrm{Tr}(A_k) = \mathrm{Tr}(A_k A_\ell A_\ell) = \mathrm{Tr}(A_\ell A_k A_\ell) = -\mathrm{Tr}(A_k A_\ell A_\ell) = -\mathrm{Tr}(A_k)$. This shows that the state ρ has the desired expectation values. It remains to show that $\rho \geq 0$. But in the proof of the previous lemma we have already seen that $|\sum_{k=1}^{L} a_k \langle A_k \rangle| \leq 1$. □

The meta-uncertainty relation is thus the best possible bound on the expectation values of the observables. Note that in the case of one qubit and the three Pauli matrices it reduces to the Bloch sphere picture. The relation has also been used to study monogamy relations for Bell inequalities [71]. Generalizing Wehner and Winter's result for the Shannon entropy [123], we can derive from it entropic uncertainty relations for various entropies.

For the purposes of this section an entropy is any nonnegative function of probability distributions $P = (p, \ldots, p_m)$ which is invariant under all permutations of the probabilities p_i and which has the value zero for the δ-distribution. Since we reserved the symbol S for the Shannon entropy, general entropies will be denoted as S_X.

Let A be an observable with eigenvalues ± 1 and $x = [\mathrm{Tr}(A\rho)]^2$ its squared expectation value. Then the probability distribution for the measurement outcomes of A is given by $P = (\frac{1+\sqrt{x}}{2}, \frac{1-\sqrt{x}}{2})$ or $P = (\frac{1-\sqrt{x}}{2}, \frac{1+\sqrt{x}}{2})$. Any entropy S_X, being invariant under permutation of P, is thus a function of x, which we denote by $\widetilde{S}_X$,

$$\widetilde{S}_X(x) = S_X(A|\rho) = S_X\big((\frac{1 \pm \sqrt{x}}{2}, \frac{1 \mp \sqrt{x}}{2})\big). \tag{5.18}$$

We say that the entropy S_X is *concave in the squared expectation value* if the function $\widetilde{S}_X$ is concave. This property is the crucial condition for the following entropic uncertainty relation:

Theorem 5.7. *Let $A_1, \ldots, A_L$ be observables which anticommute pairwise, $\{A_k, A_\ell\} = 0$ for $k \neq \ell$, and which have eigenvalues ± 1, and let S_X be an entropy which is concave in the squared expectation value (that is, an entropy for which the function $\widetilde{S}_X$ defined in Eq.* (5.18) *is concave). Then*

$$\frac{1}{L} \sum_{k=1}^{L} S_X(A_k|\rho) \geq \frac{L-1}{L} S_0, \tag{5.19}$$

where $S_0 = S_X\big((\frac{1}{2}, \frac{1}{2})\big)$ is the entropy value of the uniform probability distribution. This relation is tight.

Proof. For the case of the Shannon entropy the proof was given in Ref. 123. Let $x_k = [\mathrm{Tr}(A_k\rho)]^2$. Lemma 5.5 states that $\mathbf{x}$ lies in the simplex defined by $\sum_{k=1}^L x_k \leq 1$ and $x_k \geq 0$. As the function $\widetilde{S}_X$ is concave on the interval $[0,1]$, the function $\mathbf{x} \mapsto \sum_k \widetilde{S}_X(x_k)$ is concave on the simplex. Thus it attains its minimum at an extremal point of the simplex, that is, when $x_k = 1$ for one k and $x_\ell = 0$ for $\ell \neq k$. At an extremal point, $1/L \sum_{k=1}^L \widetilde{S}_X(x_k) = S_0(L-1)/L$. □

Before discussing the implications of this theorem, it makes sense to ask which of the commonly used entropies satisfy the requirement of being concave in the squared expectation value. For the Shannon entropy this was already shown in Ref. 123. The Tsallis entropy S_q^T can be treated analogously, though one has to distinguish between different values of the parameter q.

Lemma 5.8 (Ref. 123). *The Shannon entropy S of a dichotomic observable is concave in the squared expectation value (that is, the function $\widetilde{S}$ defined as in Eq.* (5.18) *is concave on the interval* $[0,1]$*).*

Lemma 5.9. *The Tsallis entropy S_q^T of a dichotomic observable is concave in the squared expectation value (meaning that the function $\widetilde{S}_q^T$ defined as in Eq.* (5.18) *is concave on the interval* $[0,1]$*) for parameter values $1 < q < 2$ and $3 < q$, but convex for $2 < q < 3$.*

Proof. [1] Explicitly,

$$\widetilde{S}_q^T(x) = \frac{1}{q-1}\Big[1 - \Big(\frac{1+\sqrt{x}}{2}\Big)^q - \Big(\frac{1-\sqrt{x}}{2}\Big)^q\Big]. \tag{5.20}$$

For $q = 2$ and $q = 3$ this function is easily seen to be linear. For the second derivative we obtain

$$\partial_x^2 \widetilde{S}_q^T(x) = \frac{q}{q-1}\frac{1}{2^{q+2}}\frac{1}{x^{3/2}}\Big\{(1+\sqrt{x})^{q-2}\big[1-\sqrt{x}(q-2)\big] - (1-\sqrt{x})^{q-2}\big[1+\sqrt{x}(q-2)\big]\Big\}. \tag{5.21}$$

Substituting $y = \sqrt{x}$ and omitting the prefactor (which is always positive), we arrive at the function

$$f_q(y) = (1+y)^{q-2}[1-y(q-2)] - (1-y)^{q-2}[1+y(q-2)]. \tag{5.22}$$

Observing that $f_q(0) = 0$, we note that $f_q(y)$ is positive (negative) for all $0 < y \leq 1$ if its derivative $f_q'(y)$ is positive (negative) for all $0 < y \leq 1$. The derivative is given by

$$f_q'(y) = -(q-2)(q-1)y\big[(1+y)^{q-3} - (1-y)^{q-3}\big]. \tag{5.23}$$

For $1 < q < 2$, the prefactor $-(q-2)(q-1)$ is positive and the term in the square brackets is negative; for $2 < q < 3$, the prefactor is negative and the term in the brackets is still negative; for $q > 3$, the prefactor is negative and the term in the brackets positive. This proves the lemma. □

[1] The first version of this work contained a much longer and rather technical proof of this lemma. I am indebted to Mary Beth Ruskai for this simple proof.

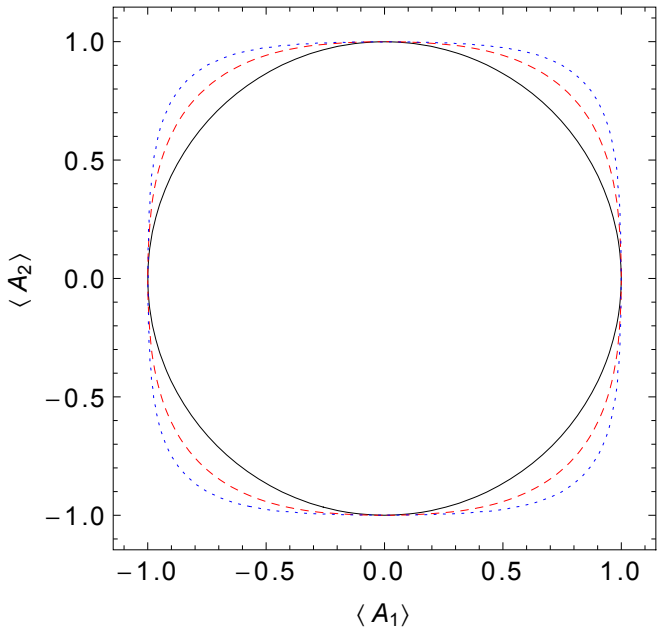

Figure 5.2: Bounds on the expectation values $(\langle A_1\rangle, \langle A_2\rangle)$ for two dichotomic anticommuting observables provided by different uncertainty relations. The black solid line corresponds to the meta-uncertainty relation Lemma 5.5, which can also be understood as an entropic uncertainty relation for the Tsallis entropy with parameter value $q = 2$ or $q = 3$. The red dashed line and the blue dotted line correspond to the entropic uncertainty relation Theorem 5.7 for the Shannon entropy and the Tsallis entropy with $q = 8$, respectively. (Figure taken from Ref. D.)

A few remarks can be made on the above theorem. The Shannon entropy has the required concavity in the squared expectation value (Lemma 5.8), and the resulting uncertainty relation is the one found by Wehner and Winter [123]. For the Tsallis entropy S_q^T we have to distinguish between different parameter ranges: For parameter values $q = 2$ and $q = 3$ this entropy is, up to a constant factor, equal to the variance, $S_2^T(A|\rho) = 1/2\Delta^2(A)$ and $S_3^T(A|\rho) = 3/8\Delta^2(A)$, and the uncertainty relation is equivalent to the meta-uncertainty relation itself. Thus it is the optimal uncertainty relation for these observables; the relation based on the Shannon entropy is strictly weaker.

In Lemma 5.9 it has been shown that the Tsallis entropy satisfies the condition of Theorem 5.7 for parameter values $1 < q \leq 2$ and $3 \leq q$. The entropy value for the uniform probability distribution, which determines the bound, is $S_0 = (1 - 2^{1-q})/(q - 1)$. In the special case of the observables σ_x and σ_y and parameter $q \in [2n - 1, 2n]$ with $n \in \mathbb{N}$, the uncertainty relation was derived before in Footnote 32 of Ref. 43.

As remarked above, Lemmas 5.5 and 5.6 provide a complete characterization of the set of expectation values of dichotomic anticommuting observables which can originate from valid quantum states. Deriving uncertainty relations from them means approxi-

mating this set from the outside. This is illustrated in Fig. 5.2.

In the parameter range $2 < q < 3$ the Tsallis entropy does not satisfy the condition for the theorem (see Lemma 5.9). An exceptional behaviour of the Tsallis entropy in this parameter range was also reported in Ref. 43. The collision entropy or Rényi entropy of order 2 and the min-entropy do not satisfy the condition either. The uncertainty relations for these entropies given in Ref. 123 also follow from the meta-uncertainty relation, but do not fit into this scheme.

In the following section Theorem 5.7 will be applied to the stabilizing operators of two stabilizer states.

Uncertainty relations for several observables can also be constructed without requiring anticommutativity: Applying a result by Mandayam et al. [81] to a set of stabilizer bases $\mathcal{A}_1, \ldots, \mathcal{A}_L$ with basis vectors denoted by $\mathcal{A}_k = \{|a_i^{(k)}\rangle\}_i$, one finds for their min-entropies

$$\frac{1}{L}\sum_{k=1}^{L} S^{\min}(\mathcal{A}_k|\rho) \geq -\log\Big[\frac{1+r(L-1)}{L}\Big], \tag{5.24}$$

where

$$r = \max_{k\neq\ell}\max_{i,j}\big|\langle a_i^{(k)}|a_j^{(\ell)}\rangle\big| \tag{5.25}$$

is the maximal overlap of the basis vectors. The proof is omitted, since this relation readily follows from Appendix C.1 of Ref. 81.

5.4 An uncertainty relation for stabilizing operators

In this section the uncertainty relation for anticommuting dichotomic observables (Theorem 5.7) is applied to the elements of a pair of stabilizer groups.

Let S and T be two n-qubit stabilizer groups, both with maximal cardinality $|S| = |T| = 2^n$. Throughout this section two operators will be considered as equal if they differ only by a minus sign. Define $\widetilde{S} = S \setminus T$ and $\widetilde{T} = T \setminus S$. Let $R = \widetilde{S} \cup \widetilde{T}$ be the symmetric difference of S and T, and denote its elements by $R = \{A_1, \ldots, A_L\}$. In Theorem 5.13 a lower bound on $1/L\sum_{k=1}^{L} S(A_k|\rho)$ will be given.

We begin by proving the following lemma:

Lemma 5.10. *Let S be a stabilizer group and N a Pauli operator (that is, a tensor product of Pauli matrices and the identity matrix) which anticommutes with an element $M_0 \in S$. Then N anticommutes with exactly half of the elements of S.*

Proof. Choose any $M_1 \in S$ with $M_1 \neq M_0$ and let $M_2 = M_0M_1$. We will now show that N anticommutes with M_2 if it commutes with M_1 and vice versa. Consider first the case $[M_1, N] = 0$. The identity $\{AB, C\} = A[B, C] + \{A, C\}B$ shows that $\{M_2, N\} = \{M_0M_1, N\} = M_0[M_1, N] + \{M_0, N\}M_1 = 0$. Consider now the case $\{M_1, N\} = 0$. Using the same identity, we obtain $M_0[M_2, N] = \{M_0M_2, N\} - \{M_0, N\}M_2 = \{M_1, N\} - \{M_0, N\}M_2 = 0$ and thus $[M_2, N] = 0$. We now iterate this procedure by choosing $M_3 \in S \setminus \{M_0, M_1, M_2\}$ and using it in place of M_1. (Note that $M_0M_3 \notin \{M_1, M_2\}$.)

The operator M_0 is kept fixed during the whole iteration. In this way S can be divided into pairs of observables, each consisting of one element commuting with N and one anticommuting with N. Note that M_0 forms a pair with the identity. □

Excluding the trivial case $S = T$, any element of T anticommutes with at least one element of S, because at most 2^n orthogonal (with respect to the Hilbert-Schmidt scalar product) unitary $2^n \times 2^n$-matrices can commute pairwise (see e. g. Ref. 10). Thus $L = |R|$ varies from 2^n to $2(2^n - 1)$.

We will now see, using only combinatorical reasoning, that this lemma implies that R can be divided into anticommuting pairs. We shall need the following combinatorical result:

Lemma 5.11 (Marriage lemma). *Consider a bipartite graph, that is, two disjoint sets of vertices U and V and a collection of edges, each connecting a vertex in U with a vertex in V. We consider the case of $|U| = |V|$. Then the graph contains a perfect matching, that is, the vertices can be divided into disjoint pairs of connected vertices, if and only if the following "marriage condition" is satisfied: For each subset U' of U, the set V' of vertices in V connected to vertices in U' is at least as large as U'.*

This theorem was first proven in Ref. 46.

Lemma 5.12. *The symmetric difference R of any two stabilizer groups S and T with $|S| = |T| = 2^n$ can be divided into anticommuting pairs of operators.*

Proof. We show that the marriage condition is fulfilled. Let S' be any subset of $\widetilde{S}$. Consider first the case $|S'| > 2^{n-1}$. Then any $N \in T$ anticommutes with at least one $M \in S'$, because any such N anticommutes with exactly 2^{n-1} elements of $\widetilde{S}$. Thus the number of $N \in \widetilde{T}$ anticommuting with at least one $M \in S'$ is $|\widetilde{T}| = |\widetilde{S}| \geq |S'|$. Consider now the case $|S'| \leq 2^{n-1}$. For any $M \in S'$ we then find 2^{n-1} elements of $\widetilde{T}$ anticommuting with M. Thus the number of $N \in \widetilde{T}$ anticommuting with at least one $M \in S'$ is $2^{n-1} \geq |S'|$. □

We are now ready to state the main result of this section:

Theorem 5.13. *Let $R = \{A_1, \ldots, A_L\}$ be the symmetric difference of two stabilizer groups S and T with $|S| = |T| = 2^n$. Then*

$$\frac{1}{L} \sum_{k=1}^{L} S_X(A_k|\rho) \geq \frac{1}{2} S_0 \tag{5.26}$$

holds, where S_X is an entropy which is concave in the squared expectation value (that is, an entropy for which the function $\widetilde{S}_X$ defined in Eq. (5.18) is concave) and S_0 is the entropy value of the uniform probability distribution. For any state of either stabilizer basis the lower bound is attained.

Proof. Due to Theorem 5.7 the uncertainty relation $S_X(A_k|\rho) + S_X(A_\ell|\rho) \geq S_0$ holds for any anticommuting pair A_k, A_ℓ. Lemma 5.12 states that R consists of $L/2$ such pairs. This shows the uncertainty relation. In remains to show that the bound is attained. The density matrix of the stabilizer state defined by the group $T = \{N_k\}$ is given by $\rho_T = \frac{1}{2^n} \sum_{k=1}^{2^n} N_k$ [see Eq. (2.72)]. Thus

$$\mathrm{Tr}(A_k \rho_T) = \frac{1}{2^n} \sum_{\ell=1}^{2^n} \mathrm{Tr}(A_k N_\ell) = 0 \quad \text{for all} \quad A_k \notin T, \tag{5.27}$$

showing $S_X(A_k|\rho_T) = S_0$ for $L/2$ observables A_k. □

This relation is not maximally strong. This is due to the fact that some of the observables commute.

6 Exponential families of interaction spaces in quantum theory

In this chapter the theory of exponential families of interaction spaces, as outlined in Section 2.6, is applied to the study of quantum correlations. Two rather different approaches are pursued: In Section 6.1 the theory is used to characterize the probability distributions of the measurement outcomes of threepartite Bell experiments. The exponential families under consideration are thus still classical. In the remaining sections, exponential families of quantum states are studied. Section 6.2 presents the quantum versions of basic definitions and results such as exponential and linear families, the generalized Pythagoras theorem, the information projection and interaction measures. In contrast to the other sections, it contains no original research, but rather aims at giving a coherent presentation of the results of Refs. 132, 133, 134. The aim of Section 6.3 is the construction of witness operators for the detection of higher-order interactions. In Section 6.4 an algorithm for the numerical computation of the quantum information projection is developed. Finally, in Section 6.5 the chapter ends with several open questions.

The contents of this chapter (with the exception of those in Section 6.2, of course) have not been published before.

6.1 Exponential families of measurement probabilities

The aim of this section is to establish connections between different notions of genuine three-party correlations. Our starting point is the Svetlichny inequality (2.37), repeated here for convenience

$$\begin{aligned}\langle A_1B_1C_2\rangle + \langle A_1B_2C_1\rangle + \langle A_2B_1C_1\rangle - \langle A_2B_2C_2\rangle \\ + \langle A_1B_2C_2\rangle + \langle A_2B_1C_2\rangle + \langle A_2B_2C_1\rangle - \langle A_1B_1C_1\rangle \leq 4. \quad (6.1)\end{aligned}$$

Recall that this Bell inequality cannot be violated by hybrid local-nonlocal hidden variable models as defined in Eq. (2.36). One can thus say that a probabilistic model contains genuine three-party nonlocality if it violates this inequality. The predictions of a probabilistic model for the outcomes of a Svetlichny experiment are summarized in a set of eight probability distributions

$$P_{ABC}(a,b,c), \quad (6.2)$$

where the indices $A \in \{A_1, A_2\}$ etc. denote the choice of local measurements and the arguments $a, b, c \in \{-1, +1\}$ denote the measurement results. Accordingly, such a model is characterized by 56 real parameters. For the purposes of this section, any choice of eight probability distributions of length eight constitutes a valid model.

Exponential families provide us with another notion of genuine three-party correlations, which originated in a different context. In that classification scheme, a probability distribution contains irreducible three-party interactions if it cannot be written as a thermal distribution of a two-party Hamiltonian, or as a limit of such distributions.

In this section we will address the question: *If we restrict the probability distributions that constitute a probabilistic model to the exponential family given by two-party Hamiltonians, what does this mean for the model's ability to violate the Svetlichny inequality?*

To make this constraint more explicit, let $\mathcal{E}_2$ be the exponential family of probability distributions on $\{-1,1\}^3$ that can be written as $P(a,b,c) = \exp[H(a,b,c) - \psi]$ with a classical two-party Hamiltonian H. (The constant ψ ensures normalization.) If the Hamiltonian is parametrized as

$$\begin{aligned} H(a,b,c;\boldsymbol{\theta}) = \theta_A k(a) + \theta_B k(b) + \theta_C k(c) \\ + \theta_{AB} k(a)k(b) + \theta_{AC} k(a)k(c) + \theta_{BC} k(b)k(c) \\ + \theta_{ABC} k(a)k(b)k(c) \end{aligned} \tag{6.3}$$

[where for $a,b,c \in \{-1,1\}$ one can choose $k(x) = x$], it contains three-party interactions precisely if θ_{ABC} is nonzero. Vanishing θ_{ABC} is equivalent to

$$\begin{aligned} &\log[P(+1,+1,+1)] + \log[P(+1,-1,-1)] \\ &+ \log[P(-1,+1,-1)] + \log[P(-1,-1,+1)] \\ &\qquad = \log[P(+1,+1,-1)] + \log[P(+1,-1,+1)] \\ &\qquad\quad + \log[P(-1,+1,+1)] + \log[P(-1,-1,-1)] \end{aligned} \tag{6.4}$$

and hence to

$$\begin{aligned} &P(+1,+1,+1)P(+1,-1,-1)P(-1,+1,-1)P(-1,-1,+1) \\ &\qquad = P(+1,+1,-1)P(+1,-1,+1)P(-1,+1,+1)P(-1,-1,-1). \end{aligned} \tag{6.5}$$

[If we write the probability distribution as a diagonal matrix as if it was a quantum mechanical density operator, Eq. (6.4) reads $\mathrm{Tr}[\sigma_z \otimes \sigma_z \otimes \sigma_z \log(P)] = 0$.] In this section we also allow probability distributions without full support, meaning that the set of allowed distributions is the compactified exponential family $\overline{\mathcal{E}}_2$.

It is not difficult at all to write down a model satisfying the $\overline{\mathcal{E}}_2$constraint and still violating the Svetlichny inequality maximally: Let

$$P_{A_1B_1C_1}(a,b,c) = P_{A_2B_2C_2}(a,b,c) = \begin{cases} 1 & \text{if } a=b=c=-1, \\ 0 & \text{otherwise,} \end{cases} \tag{6.6}$$

$$\text{all remaining distributions} = \begin{cases} 1 & \text{if } a=b=c=+1, \\ 0 & \text{otherwise.} \end{cases} \tag{6.7}$$

Then the left-hand side of Eq. (6.1) takes the value 8. The probability distributions of this model are not only in $\overline{\mathcal{E}}_2$, but even in $\overline{\mathcal{E}}_1$, that is, they are product distributions.

The negative answer to the question raised at the beginning should not come as a surprise: The $\overline{\mathcal{E}}_2$-condition is a constraint only on the individual probability distributions. It does not restrict the way in which the model's predictions depend on the choice of local measurements. Indeed, the above example is highly signalling. Signalling models can be dismissed for being unphysical. We modify our question by allowing only nonsignalling models with probability distributions in $\overline{\mathcal{E}}_2$.

So let us discuss the no-signalling conditions for Bell experiments with three parties. A model is called nonsignalling if [12]

1. Alice cannot signal to the combined systems of Bob and Charlie,

$$\sum_a P_{A_1BC}(a,b,c) = \sum_a P_{A_2BC}(a,b,c) \qquad \forall\, B,C,b,c, \tag{6.8}$$

   and analogously for all permutations of the parties;

2. the combined systems of Alice and Bob cannot signal to Charlie,

$$\begin{aligned} \sum_{a,b} P_{A_1B_1C}(a,b,c) &= \sum_{a,b} P_{A_1B_2C}(a,b,c) \\ &= \sum_{a,b} P_{A_2B_1C}(a,b,c) \\ &= \sum_{a,b} P_{A_2B_2C}(a,b,c) \qquad \forall\, C,c, \end{aligned} \tag{6.9}$$

   and analogously for all permutations of the parties.

Obviously, the first set of conditions implies the second. It also includes the weaker condition that Alice cannot signal to Bob or to Charlie.

Equation (6.8) and its permutations impose 48 linear constraints on the probability distributions, but these constraints are not all linearly independent: If we collect the probability distributions in a vector $\mathbf{P} \in \mathbb{R}^{64}$, the 48 no-signalling constraints together with the 8 normalization conditions take the form of a system of affine linear equations $\hat{T}\mathbf{P} = \mathbf{t}$. The 56×64-matrix $\hat{T}$ turns out to have rank 38. Choosing a basis in the nullspace of $\hat{T}$, one can parametrize all nonsignalling models with $64 - 38 = 26$ parameters. However, in this parametrization the probabilities are not automatically guaranteed to be nonnegative.

The no-signalling conditions alone do not restrict the violation of the Svetlichny inequality either: The model

$$P_{A_1B_1C_1}(a,b,c) = P_{A_2B_2C_2}(a,b,c) = \begin{cases} 0 & \text{if } abc = +1, \\ \frac{1}{4} & \text{if } abc = -1, \end{cases} \tag{6.10}$$

$$\text{all remaining distributions} = \begin{cases} \frac{1}{4} & \text{if } abc = +1, \\ 0 & \text{if } abc = -1 \end{cases} \tag{6.11}$$

is nonsignalling, but the left-hand side of Eq. (6.1) takes the value 8. In other words, the nonsignalling bound of the Svetlichny inequality is equal to its algebraic bound.

The probability distributions in Eqs. (6.10) and (6.11) are prototypical examples for distributions containing irreducible three-party interactions [35, 62]. They can be written as limits of thermal distributions as follows:

$$P_{A_1B_1C_1}(a,b,c) = P_{A_2B_2C_2}(a,b,c) = \lim_{\theta\to\infty} \exp\{-\theta abc - \ln[8\cosh(\theta)]\}, \tag{6.12}$$

$$\text{all remaining distributions} = \lim_{\theta\to\infty} \exp\{+\theta abc - \ln[8\cosh(\theta)]\}. \tag{6.13}$$

Note that the Hamiltonians $H(a,b,c;\theta) = \pm\theta abc$ contain only three-party and no lower-order interactions. Intuitively, this is because the distributions are essentially given by the parity function, which is $+1$ (-1) if an even (odd) number of the arguments a, b, c has the value -1. For the parity, knowledge of the values of one or two arguments gives no information about the value of the function. The interaction measures defined in Eqs. (2.133) and (2.136) take the values $C_1 = C_2 = 0$ and $C_{\text{tot}} = C_3 = 1$.

It is instructive to see how the quantum mechanical measurement probabilities which violate the inequality maximally fit into this picture. For the GHZ state and the operators given in Eq. (2.40), quantum mechanics predicts

$$P_{A_1B_1C_1}(a,b,c) = P_{A_2B_2C_2}(a,b,c) = \begin{cases} \frac{2-\sqrt{2}}{16} \approx 0.03661 & \text{if } abc = +1, \\ \frac{2+\sqrt{2}}{16} \approx 0.2134 & \text{if } abc = -1, \end{cases} \tag{6.14}$$

$$\text{all remaining distributions} = \begin{cases} \frac{2+\sqrt{2}}{16} \approx 0.2134 & \text{if } abc = +1, \\ \frac{2-\sqrt{2}}{16} \approx 0.03661 & \text{if } abc = -1. \end{cases} \tag{6.15}$$

These distributions achieve the maximal quantum mechanical violation, meaning that the left-hand side of Eq. (6.1) takes the value $4\sqrt{2}$. Writing the distributions in exponential form,

$$P_{A_1B_1C_1}(a,b,c) = P_{A_2B_2C_2}(a,b,c) = \exp\big[-\operatorname{arcoth}(\sqrt{2})abc - \tfrac{7}{2}\ln(2)\big], \tag{6.16}$$

$$\text{all remaining distributions} = \exp\big[+\operatorname{arcoth}(\sqrt{2})abc - \tfrac{7}{2}\ln(2)\big], \tag{6.17}$$

we again find Hamiltonians with only three-party interactions. Here,

$$C_1 = C_2 = 0 \qquad \text{and} \qquad C_{\text{tot}} = C_3 = \frac{1}{2\sqrt{2}} \log\Big[\frac{2+\sqrt{2}}{2-\sqrt{2}}\Big] - \frac{1}{2} \approx 0.3991. \tag{6.18}$$

We have seen that neither the no-signalling conditions nor the $\overline{\mathcal{E}}_2$-constraint alone preclude maximal violation of the Svetlichny inequality. We do not yet know the effect of both sets of conditions combined. Numerical maximization over all models subject to these constraints supports the following conjecture, which is the first main "result" of this section:

Conjecture 6.1. *A nonsignalling probabilistic model whose probability distributions are in the compactified exponential family $\overline{\mathcal{E}}_2$ cannot violate the Svetlichny inequality.*

Below the conjecture will be proven for the class of models satisfying a certain symmetry condition. The general case is still open. The numerical results should not be taken on trust, because the maximization problem is somewhat unwieldy: If we parametrize the probability distributions of all nonsignalling models in the 26 independent parameters (see above), the requirement that the distributions are in $\overline{\mathcal{E}}_2$ constitutes eight nonlinear constraints, in addition to the constraints imposed by the nonnegativity of the probabilities. Conversely, if we parametrize the Hamiltonians, the nonsignalling constraints are rather cumbersome.

To simplify the problem we consider now only models which satisfy

$$P_{A_1B_1C_2} = P_{A_1B_2C_1} = P_{A_2B_1C_1} \quad \text{and} \quad P_{A_1B_2C_2} = P_{A_2B_1C_2} = P_{A_2B_2C_1}. \tag{6.19}$$

This symmetry is suggested by the distribution of the signs in the Svetlichny inequality itself. Also note that all examples of models discussed so far obey this condition. This may indicate the symmetry assumption is not very restrictive. We are left with the four independent distributions

$$P_1 = P_{A_1B_1C_1}, \tag{6.20}$$
$$P_2 = P_{A_1B_1C_2} = P_{A_1B_2C_1} = P_{A_2B_1C_1}, \tag{6.21}$$
$$P_3 = P_{A_2B_2C_2} = P_{A_2B_1C_2} = P_{A_2B_2C_1}, \tag{6.22}$$
$$P_4 = P_{A_1B_2C_2}. \tag{6.23}$$

The no-signalling conditions [Eq. (6.8) and permutations] reduce to

$$\sum_a P_1(a,b,c) = \sum_a P_2(a,b,c) = \sum_a P_3(a,b,c) = \sum_a P_4(a,b,c) \qquad \forall\, b,c \tag{6.24}$$

and permutations. In other words, the two-party marginals of the P_i are all equal.

The following lemma is the second main result of this section:

Lemma 6.2. *A nonsignalling probabilistic model which obeys the symmetry assumption given in Eq.* (6.19) *and whose probability distributions are in the compactified exponential family* $\overline{\mathcal{E}}_2$ *cannot violate the Svetlichny inequality.*

Two proofs of this lemma will be given: The first one is based on information geometry, the second one is more elementary.

Information-geometric proof of Lemma 6.2. According to the second definition of the information projection [see Eq. (2.126)], the projection of P onto $\overline{\mathcal{E}}_2$ is completely determined by the two-party marginals of P. We already know that the four probability distributions that constitute the model have the same two-party marginals. Thus their projections onto $\overline{\mathcal{E}}_2$ are equal. On the other hand, by assumption each distribution is in $\overline{\mathcal{E}}_2$ and hence equal to its own projection. This shows that the four distributions of the model are all equal. Consequently the eight expectation values in the Svetlichny inequality (6.1) are equal. It follows that the absolute value of the left-hand side of the inequality is ≤ 4, which is the local hidden variable bound. □

Alternative proof of Lemma 6.2. We write the probability distributions that constitute the model as vectors $\mathbf{P}^{(i)} \in \mathbb{R}^8$. The constraint that two distributions $\mathbf{P}^{(i)}$ and $\mathbf{P}^{(j)}$ have the same two-party marginals can be written in the form $\hat{R}\mathbf{P}^{(i)} = \hat{R}\mathbf{P}^{(j)}$ with a 12×8-matrix $\hat{R}$. This matrix turns out to have rank 7; its nullspace is spanned by the vector $\mathbf{n} = (1, -1, -1, 1, -1, 1, 1, -1)$. It follows that $\mathbf{P}^{(j)} = \mathbf{P}^{(i)} + q\mathbf{n}$ for some q. Writing $\mathbf{P}^{(i)} = (p_1, \ldots, p_8)$, the distributions $\mathbf{P}^{(i)}$ and $\mathbf{P}^{(j)}$ are in $\mathcal{E}_2$ if [cf. Eq. (6.5)]

$$p_1 p_4 p_6 p_7 = p_2 p_3 p_5 p_8 \tag{6.25}$$

and

$$(p_1 + q)(p_4 + q)(p_6 + q)(p_7 + q) = (p_2 - q)(p_3 - q)(p_5 - q)(p_8 - q). \tag{6.26}$$

This implies $q = 0$. If $\mathbf{P}^{(i)}$ does not have full support, at least one factor of the left-hand side and one factor on the right-hand side of Eq. (6.25), say p_1 and p_2, have to be zero. This implies $q = 0$, because $p_2 - q$ is a probability and must be nonnegative. We have shown that the four distributions of the model are all equal. Proceeding as in the first proof, we see that all eight expectation values in the Svetlichny inequality (6.1) are equal and that the absolute value of the left-hand side of the inequality is ≤ 4, which is the local hidden variable bound. □

6.2 Exponential families of quantum states

In this section the theory of exponential families of interaction spaces is generalized to the quantum domain, with quantum states taking the place of probability distributions. The results presented here are due to Zhou [132, 133, 134].

6.2.1 Exponential and Bloch representation

We consider n-qubit states throughout; the generalization to higher Hilbert space dimensions will be obvious.

Any n-qubit quantum state with full rank can be written in the *exponential representation*

$$\rho_{\exp}(\boldsymbol{\theta}) = \exp\Big(\sum_{i_1,\ldots,i_n} \theta_{i_1,\ldots,i_n} \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n} \Big) \tag{6.27}$$

with indices running from 0 to 3. It will be convenient to use a multi-index notation,

$$\rho_{\exp}(\boldsymbol{\theta}) = \exp\big(\sum_{\alpha} \theta_\alpha \sigma_\alpha\big), \tag{6.28}$$

where $\sigma_\alpha = \sigma_{\alpha_1} \otimes \cdots \otimes \sigma_{\alpha_n}$. Greek indices will always be multi-indices. The coefficient $\theta_{\mathbf{0}}$ of the identity $\sigma_{\mathbf{0}} = \mathbb{1}^{\otimes n}$ is not free, but rather determined by normalization. Explicitly, $\theta_{\mathbf{0}} = -\psi(\boldsymbol{\theta})$ where

$$\psi(\boldsymbol{\theta}) = \ln\big\{ \mathrm{Tr}\big[\exp\big(\sum_{\alpha \neq 0} \theta_\alpha \sigma_\alpha \big) \big] \big\}. \tag{6.29}$$

We call the number of factors in the Pauli operator $\sigma_\alpha = \sigma_{\alpha_1} \otimes \cdots \otimes \sigma_{\alpha_n}$ which are different from the identity its *weight* and denote it by $W(\alpha)$. In other words, the weight $W(\alpha)$ of a multi-index α is the number of nonzero elements α_i.

We often think of the exponent in the exponential representation as a Hamiltonian and of the state as a thermal or Gibbs state. As in the classical case this Hamiltonian does not necessarily correspond to an actual physical system. The function ψ is minus the free energy.

For any $1 \leq k \leq n$ we define the exponential family $\mathcal{Q}_k$ of thermal states of k-party Hamiltonians,

$$\mathcal{Q}_k = \{\rho \mid \rho = \exp\Big(\sum_{\substack{\alpha \text{ with} \\ W(\alpha)\leq k}} \theta_\alpha \sigma_\alpha\Big)\}. \tag{6.30}$$

This defines a hierarchy

$$\mathcal{Q}_1 \subset \mathcal{Q}_2 \subset \cdots \subset \mathcal{Q}_n, \tag{6.31}$$

where $\mathcal{Q}_n$ is the set of all states with full rank and $\mathcal{Q}_1$ the set of all product states with full rank. It we want to include states without full rank, we work with the compactified exponential families $\overline{\mathcal{Q}}_k$.

Alternatively, any state can be written in the *Bloch representation*[1]

$$\rho_{\text{aff}}(\boldsymbol{\eta}) = \frac{1}{2^n} \sum_\alpha \eta_\alpha \sigma_\alpha. \tag{6.32}$$

Here the normalization condition is $\eta_{\mathbf{0}} = 1$.

We calculate the relative entropy of two full-rank states $\rho = \rho_{\exp}(\boldsymbol{\theta}) = \rho_{\text{aff}}(\boldsymbol{\eta})$ and $\rho' = \rho_{\exp}(\boldsymbol{\theta}') = \rho_{\text{aff}}(\boldsymbol{\eta}')$,

$$\begin{aligned} \ln(2) D(\rho\|\rho') &= -\ln(2) S(\rho_{\text{aff}}(\boldsymbol{\eta})) - \operatorname{Tr}\Big\{ \frac{1}{2^n} \Big[\mathbb{1} + \sum_{\alpha\neq 0} \eta_\alpha \sigma_\alpha\Big] \Big[\sum_{\beta\neq 0} \theta_\beta \sigma_\beta - \psi(\boldsymbol{\theta}')\Big]\Big\} \\ &= \phi(\boldsymbol{\eta}) + \psi(\boldsymbol{\theta}') - \sum_{\alpha\neq 0} \eta_\alpha \theta'_\alpha, \end{aligned} \tag{6.33}$$

where the function

$$\phi(\boldsymbol{\eta}) = -\ln(2) S(\rho_{\text{aff}}(\boldsymbol{\eta})) \tag{6.34}$$

was introduced. With the scalar product

$$\boldsymbol{\eta} \cdot \boldsymbol{\theta}' = \sum_{\alpha\neq 0} \eta_\alpha \theta'_\alpha \tag{6.35}$$

this result takes the form

$$\ln(2) D(\rho\|\rho') = \phi(\boldsymbol{\eta}) + \psi(\boldsymbol{\theta}') - \boldsymbol{\eta} \cdot \boldsymbol{\theta}'. \tag{6.36}$$

In the special case $\rho = \rho'$,

$$\phi(\boldsymbol{\eta}) + \psi(\boldsymbol{\theta}) - \boldsymbol{\eta} \cdot \boldsymbol{\theta} = 0. \tag{6.37}$$

[1] In Ref. 134 the η_α are called *affine parameters*.

The last equation shows that the exponential and the Bloch representation are related by a Legendre transformation,

$$\eta_\alpha = \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta_\alpha} \qquad \text{and} \qquad \theta_\alpha = \frac{\partial \phi(\boldsymbol{\eta})}{\partial \eta_\alpha} \tag{6.38}$$

for all $\alpha \neq 0$. It is instructive to formulate the results obtained so far in the language of statistical mechanics: We define a thermodynamic ensemble by the requirement that the observables σ_α have expectation values η_α. Maximizing the entropy under these constraints, we find the thermal state of this ensemble to be $\rho = \rho_{\exp}(\boldsymbol{\theta})$, where the θ_α arise as Lagrange multipliers. In statistical mechanics it is well-known that the Lagrange multipliers θ_α are related to the expectation values η_α by a Legendre transformation.

For the pairwise relative entropies of three full-rank states ρ, ρ' and ρ'' we obtain

$$\begin{aligned} &D(\rho\|\rho'') - D(\rho\|\rho') - D(\rho'\|\rho'') \\ &\qquad = D(\rho\|\rho'') - D(\rho\|\rho') - D(\rho'\|\rho'') + D(\rho'\|\rho') \\ &\qquad = \frac{1}{\ln(2)}\Big[\phi(\boldsymbol{\eta}) + \psi(\boldsymbol{\theta}'') - \boldsymbol{\eta}\cdot\boldsymbol{\theta}'' - \phi(\boldsymbol{\eta}) - \psi(\boldsymbol{\theta}') + \boldsymbol{\eta}\cdot\boldsymbol{\theta}' \\ &\qquad\qquad - \phi(\boldsymbol{\eta}') - \psi(\boldsymbol{\theta}'') + \boldsymbol{\eta}'\cdot\boldsymbol{\theta}'' + \phi(\boldsymbol{\eta}') + \psi(\boldsymbol{\theta}') - \boldsymbol{\eta}'\cdot\boldsymbol{\theta}'\Big] \\ &\qquad = \frac{1}{\ln(2)}(\boldsymbol{\eta} - \boldsymbol{\eta}')\cdot(\boldsymbol{\theta}' - \boldsymbol{\theta}'') \end{aligned} \tag{6.39}$$

and thus

$$D(\rho\|\rho'') = D(\rho\|\rho') + D(\rho'\|\rho'') + \frac{1}{\ln(2)}(\boldsymbol{\eta} - \boldsymbol{\eta}')\cdot(\boldsymbol{\theta}' - \boldsymbol{\theta}''). \tag{6.40}$$

If the scalar product vanishes, we call this relation the *generalized Pythagoras theorem*.

6.2.2 Information projection

Proceeding as in the classical case, we give three definitions of the information projection. We then prove their equivalence. The proof given here, which uses methods from statistical mechanics, fails if the information projection does not have full rank. To be on the safe side, all definitions and results will be formulated only for the non-compactified exponential families. It is to be expected, though, that the results remain valid in the general case.

Definition 6.3. The *information projection* $\widetilde{\rho}_k$ of a quantum state ρ is the element of the exponential family $\mathcal{Q}_k$ which is closest to ρ with respect to the quantum relative entropy,

$$\widetilde{\rho}_k = \operatorname*{argmin}_{\rho' \in \mathcal{Q}_k} D(\rho\|\rho'). \tag{6.41}$$

For any state ρ we define the set $M_k(\rho)$ of states with the same k-party reduced density matrices,

$$M_k(\rho) = \big\{\rho' \;\big|\; \rho'_A = \rho_A \text{ for all } A \subseteq \{1,\ldots,n\} \text{ with } |A| = k\big\}, \tag{6.42}$$

where $\rho_A = \mathrm{Tr}_{\{i_1,\ldots,i_n\}\setminus A}(\rho)$. In the language of information geometry this is a linear family. Explicitly,

$$M_k(\rho_{\mathrm{aff}}(\boldsymbol{\eta})) = \big\{\rho_{\mathrm{aff}}(\boldsymbol{\eta}') \;\big|\; \eta'_\alpha = \eta_\alpha \text{ for all } \alpha \text{ with } W(\alpha) \leq k\big\}. \tag{6.43}$$

The following lemmas contain the second and the third definition of the information projection:

Lemma 6.4. *The information projection $\widetilde{\rho}_k$ of a quantum state ρ is the maximizer of the von Neumann entropy among all states with the same k-party reduced density matrices as ρ,*

$$\widetilde{\rho}_k = \operatorname*{argmax}_{\rho' \in M_k(\rho)} S(\rho'). \tag{6.44}$$

Lemma 6.5. *The information projection $\widetilde{\rho}_k$ of a quantum state ρ is the uniquely defined element of the exponential family $\mathcal{Q}_k$ with the same k-party reduced density matrices as ρ,*

$$\{\widetilde{\rho}_k\} = \mathcal{Q}_k \cap M_k(\rho). \tag{6.45}$$

We will now show that the definitions are equivalent and define a unique state.

Proof of Lemmas 6.4 and 6.5. Define $\widetilde{\rho}_k$ according to Eq. (6.44). We have to maximize the entropy of ρ' under the constraints $\mathrm{Tr}(\rho'\sigma_\alpha) = \mathrm{Tr}(\rho\sigma_\alpha)$ for all α with $W(\alpha) \leq k$. This maximization is analogous to the well-known derivation of the thermal state in statistical mechanics, but it will be repeated here for completeness. Introducing Lagrange multipliers $\widetilde{\theta}_\alpha$ for $0 \leq W(\alpha) \leq k$, the information projection is the solution to

$$\mathrm{Tr}\Big[\Big(-\ln(\widetilde{\rho}_k) - \mathbb{1} + \sum_{0\leq W(\alpha)\leq k} \widetilde{\theta}_\alpha \sigma_\alpha\Big)\delta\widetilde{\rho}_k\Big] = 0, \tag{6.46}$$

which is

$$\widetilde{\rho}_k = \exp(\sum_\alpha \widetilde{\theta}_\alpha \sigma_\alpha), \tag{6.47}$$

where we re-defined $\widetilde{\theta}_\mathbf{0} - 1 \to \widetilde{\theta}_\mathbf{0}$. This $\widetilde{\rho}_k$ obviously is in $\mathcal{Q}_k$. (If the system of equations determining the Lagrange multipliers does not have a solution, the information projection does not exist in the non-compactified exponential family.) We still have to show that the stationary point is indeed a maximum and that it is unique. Let $\rho' = \rho_{\mathrm{aff}}(\boldsymbol{\eta}')$ be any other state satisfying the constraints $\mathrm{Tr}(\rho'\sigma_\alpha) = \mathrm{Tr}(\rho\sigma_\alpha)$ for all α with $W(\alpha) \leq k$. Using Eq. (6.36),

$$\ln(2) D(\rho'\|\widetilde{\rho}_k) = -\ln(2) S(\rho') + \psi(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\eta}' \cdot \widetilde{\boldsymbol{\theta}}. \tag{6.48}$$

The terms in the scalar product $\boldsymbol{\eta}' \cdot \widetilde{\boldsymbol{\theta}} = \sum_\alpha \eta'_\alpha \widetilde{\theta}_\alpha$ with $W(\alpha) > k$ vanish because $\widetilde{\theta}_\alpha = 0$ for these α. The coefficients η'_α with $W(\alpha) \leq k$ are fixed by the constraints $\mathrm{Tr}(\rho'\sigma_\alpha) = \mathrm{Tr}(\rho\sigma_\alpha)$ and are thus equal to $\widetilde{\eta}_\alpha$. This shows

$$\begin{aligned} \ln(2) D(\rho'\|\widetilde{\rho}_k) &= -\ln(2) S(\rho') + \psi(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\eta}' \cdot \widetilde{\boldsymbol{\theta}} \\ &= -\ln(2) S(\rho') + \psi(\widetilde{\boldsymbol{\theta}}) - \widetilde{\boldsymbol{\eta}} \cdot \widetilde{\boldsymbol{\theta}} \\ &= -\ln(2) S(\rho') + \ln(2) S(\widetilde{\rho}_k), \end{aligned} \tag{6.49}$$

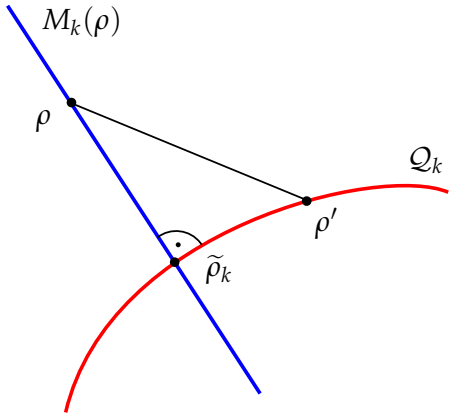

Figure 6.1: Illustration of the information projection onto a quantum exponential family. Shown are the linear family $M_k(\rho)$ of distributions with the same k-party reduced density matrices as ρ (blue line), the exponential family $\mathcal{Q}_k$ of thermal states of k-party Hamiltonians (red curve) and the information projection $\widetilde{\rho}_k$ of ρ onto $\mathcal{Q}_k$; and ρ' represents an arbitrary state in $\mathcal{Q}_k$. This figure is the quantum version of Fig. 2.4.

where Eq. (6.37) was used. From the positive definiteness of the relative entropy (see Section 2.3.3) we conclude that $\widetilde{\rho}_k$ is indeed the unique maximum.

We now apply Eq. (6.40) to the states $\rho = \rho_{\text{aff}}(\boldsymbol{\eta})$, $\widetilde{\rho}_k = \rho_{\text{aff}}(\widetilde{\boldsymbol{\eta}}) = \rho_{\text{exp}}(\widetilde{\boldsymbol{\theta}})$ and an arbitrary $\rho' = \rho_{\text{exp}}(\boldsymbol{\theta}')$ in $\mathcal{Q}_k$, obtaining

$$D(\rho\|\rho') = D(\rho\|\widetilde{\rho}_k) + D(\widetilde{\rho}_k\|\rho') + \frac{1}{\ln(2)}(\boldsymbol{\eta} - \widetilde{\boldsymbol{\eta}}) \cdot (\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}'). \tag{6.50}$$

The terms in the scalar product with $W(\alpha) \leq k$ vanish because $\eta_\alpha = \widetilde{\eta}_\alpha$ for these α, and the terms with $W(\alpha) > k$ vanish because of $\widetilde{\theta}_\alpha = \theta'_\alpha = 0$. The Pythagoras theorem

$$D(\rho\|\rho') = D(\rho\|\widetilde{\rho}_k) + D(\widetilde{\rho}_k\|\rho'), \tag{6.51}$$

which we have just shown, implies that $\widetilde{\rho}_k$, which was defined according to Eq. (6.44), is also the unique state minimizing the relative entropy $D(\rho\|\rho')$ among all $\rho' \in \mathcal{Q}_k$. This establishes the equivalence of the first and the second definition of the information projection. The state $\widetilde{\rho}_k$ defined in this way obviously satisfies Eq. (6.45). The uniqueness of the third definition follows again from the Pythagoras theorem. □

Figure 6.1 illustrates the geometric relations which we have established. It is completely analogous to Fig. 2.4 in the classical case.

Lemma 6.4 implies that any state which is completely determined by its k-party reduced density matrices is in the exponential family $\overline{\mathcal{Q}}_k$ (ignoring for the moment the issue of states without full rank). The question which states are completely determined by their reduced density matrices has already received attention. For the special case of pure states a number of interesting results have been obtained: Almost every pure n-party state with equidimensional subsystems is uniquely determined among all states

(whether pure or mixed) by its reduced k-party density matrices, where k scales like $n/2$. (This was shown in Ref. 60; see also Refs. 76,77 for earlier works). For a qubit system, only the generalized GHZ states $\cos(\theta/2)|0\rangle^{\otimes n} + \mathrm{e}^{\mathrm{i}\phi}\sin(\theta/2)|1\rangle^{\otimes n}$ and their LU-equivalents are undetermined by the $(n-1)$-party reduced density matrices [121,122].

Analogously to the classical case, we denote the distance from a quantum state ρ to its projection $\widetilde{\rho}_k$ in terms of the relative entropy by

$$D_k(\rho) = D(\rho\|\widetilde{\rho}_k), \qquad k = 1,\ldots,n-1 \tag{6.52}$$

and define the *degree of irreducible k-party interaction* as

$$C_k(\rho) = D_{k-1}(\rho) - D_k(\rho), \qquad k = 2,\ldots,n \tag{6.53}$$

(where $D_n \equiv 0$). By the generalized Pythagoras theorem, the last definition is equivalent to

$$C_k(\rho) = D(\widetilde{\rho}_k\|\widetilde{\rho}_{k-1}), \qquad k = 2,\ldots,n-1. \tag{6.54}$$

Writing $\rho = \rho_{\mathrm{aff}}(\boldsymbol{\eta})$ and $\widetilde{\rho}_k = \rho_{\mathrm{aff}}(\widetilde{\boldsymbol{\eta}}) = \rho_{\exp}(\widetilde{\boldsymbol{\theta}})$, one obtains with Eqs. (6.36) and (6.37)

$$\begin{aligned} \ln(2) D_k(\rho) &= \ln(2) D(\rho\|\widetilde{\rho}_k) \\ &= -\ln(2) S(\rho) + \psi(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\eta}\cdot\widetilde{\boldsymbol{\theta}} \\ &= -\ln(2) S(\rho) + \psi(\widetilde{\boldsymbol{\theta}}) - \widetilde{\boldsymbol{\eta}}\cdot\widetilde{\boldsymbol{\theta}} \\ &= -\ln(2) S(\rho) + \ln(2) S(\widetilde{\rho}_k). \end{aligned} \tag{6.55}$$

Here the fact was used that $\widetilde{\theta}_\alpha = 0$ for $W(\alpha) > k$ and $\widetilde{\eta}_\alpha = \eta_\alpha$ for $W(\alpha) \leq k$. This shows

$$D_k(\rho) = S(\widetilde{\rho}_k) - S(\rho), \qquad k = 1,\ldots,n-1 \tag{6.56}$$

and

$$C_k(\rho) = S(\widetilde{\rho}_{k-1}) - S(\widetilde{\rho}_k), \qquad k = 2,\ldots,n-1. \tag{6.57}$$

The multi-information or *degree of total interaction* has an expansion into a telescopic sum of entropy differences

$$C_{\mathrm{tot}}(\rho) = D_1(\rho) = \sum_{k=2}^{n} C_k(\rho). \tag{6.58}$$

This is an orthogonal decomposition in the sense of the generalized Pythagoras theorem.

The exponential family $\mathcal{E}_1$ is comprised of all product states (with full rank). The projection of a state ρ onto this family is given by the tensor product of the one-party reduced density matrices,

$$\widetilde{\rho}_1 = \rho_{\{1\}} \otimes \cdots \otimes \rho_{\{n\}} \qquad \text{where} \qquad \rho_{\{i\}} = \mathrm{Tr}_{\{1,\ldots,n\}\setminus\{i\}}\,\rho. \tag{6.59}$$

For the other projections there is no explicit formula.

6.2.3 Information projections of stabilizer states and generalized GHZ states

In Ref. 132 Zhou gave a method for the explicit calculation of the information projections for two classes of states, namely, stabilizer states and generalized GHZ states. In this subsection the method is described, examples are given, and relations to other works are discussed.

Let ρ_S be an n-qubit stabilizer state with stabilizer group S. This state does not need to be pure. The rank of ρ_S is related to the cardinality of the stabilizer group by $\mathrm{rank}(\rho_S) = 2^{n-m}$ where $|S| = 2^m$. The stabilizing operators can be classified according to the numbers of qubits on which they act nontrivially. Recall that we defined the weight $W(\alpha)$ of a Pauli operator $\sigma_\alpha = \sigma_{\alpha_1} \otimes \cdots \otimes \sigma_{\alpha_n}$ as the number of factors which are not equal to the identity. The set of stabilizing operators of weight less than or equal to k is in general not a group. For any $k < n$ we define $S^{(k)}$ as the smallest subgroup of S that contains all elements $\pm\sigma_\alpha$ of S with $W(\alpha) \leq k$. This group is again a stabilizer group; we will show that the corresponding stabilizer state is the information projection of ρ_S.

In general $S^{(k)}$ contains operators of weight larger than k, but by definition one can find a set $g^{(k)}$ of generators with weight less than or equal to k. The generating set of $S^{(k)}$ can be completed to a generating set of $S^{(k+1)}$, since $S^{(k)} \subseteq S^{(k+1)}$. We choose

$$g^{(1)} \subseteq g^{(2)} \subseteq \cdots \subseteq g^{(m-1)} \subseteq g, \tag{6.60}$$

where each $g^{(k)}$ is a generating set of $S^{(k)}$ containing only operators of weight less than or equal to k; and $g = \{g_1, \ldots, g_m\}$ is a generating set of S.

Since ρ_S does not have full rank we will regularize it. Recall that ρ_S can be written in terms of the generators as [see Eq. (2.72)]

$$\rho_S = \frac{1}{2^n} \prod_{i=1}^{m} (\mathbb{1} + g_i). \tag{6.61}$$

This state is obtained in the limit $\lambda \to \infty$ of the following parametrized family of full-rank states:

$$\begin{aligned} \rho_S(\lambda) &= \frac{1}{2^n \cosh^m(\lambda)} \exp\bigl(\lambda \sum_{i=1}^{m} g_i\bigr) \\ &= \frac{1}{2^n} \prod_{i=1}^{m} \bigl[\mathbb{1} + \tanh(\lambda) g_i\bigr]. \end{aligned} \tag{6.62}$$

Expanding the product in the last equation, we obtain

$$\rho_S(\lambda) = \frac{1}{2^n} \sum_{i=1}^{2^m} \tanh^{s_i}(\lambda) M_i, \tag{6.63}$$

where the M_i are the stabilizing operators of ρ_S and s_i is the number of factors when writing M_i as a product of generators.

Similarly the generator set $g^{(k)}$, whose elements will be denoted as

$$g^{(k)} = \bigl\{g_1^{(k)}, \ldots, g_{m_k}^{(k)}\bigr\}, \tag{6.64}$$

defines the regularized stabilizer state

$$\begin{aligned}\widetilde{\rho}_k(\lambda) &= \frac{1}{2^n \cosh^m(\lambda)} \exp\big(\lambda \sum_{i=1}^{m_k} g_i^{(k)}\big) \\ &= \frac{1}{2^n} \prod_{i=1}^{m_k} [\mathbb{1} + \tanh(\lambda) g_i^{(k)}] \\ &= \frac{1}{2^n} \sum_{i=1}^{2^{m_k}} \tanh^{s_i}(\lambda) M_i^{(k)}.\end{aligned} \tag{6.65}$$

By construction this state is in the exponential family $\mathcal{Q}_k$. The $M_i^{(k)}$ are the elements of the group $S^{(k)}$. By definition these are all M in the expansion in Eq. (6.63) which have weight less than or equal to k. This shows that $\widetilde{\rho}_k(\lambda)$ has the same k-party reduced density matrices as $\rho_S(\lambda)$. According to Lemma 6.5, the state $\widetilde{\rho}_k(\lambda)$ is the information projection of $\rho_S(\lambda)$. In the limit $\lambda \to \infty$,

$$\widetilde{\rho}_k(\lambda) \to \frac{1}{2^n} \prod_{i=1}^{m_k} (\mathbb{1} + g_i^{(k)}). \tag{6.66}$$

The von Neumann entropy of a stabilizer state is easily seen to be $S(\rho_S) = n - m$, where m is the number of generators. We summarize our findings in the following theorem:

Theorem 6.6 (Ref. 132). *The information projection $\widetilde{\rho}_k$ of the stabilizer state ρ_S with stabilizer group S is again a stabilizer state, given by the smallest subgroup $S^{(k)}$ of S which contains all stabilizing operators with weight less than or equal to k. The distances to the exponential families and the degrees of interaction are given by*

$$D_k(\rho_S) = m - m_k \tag{6.67}$$

and

$$C_k(\rho_S) = m_k - m_{k-1} \qquad \text{and} \qquad C_{tot}(\rho_S) = m - m_1, \tag{6.68}$$

respectively, where 2^{m_k} is the cardinality of $S^{(k)}$ and 2^m is the cardinality of S.

As our first example we consider the four-qubit GHZ state, whose stabilizer group was given in Eq. (2.96). This group does not contain operators of weight 1, therefore $S^{(1)} = \{\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\}$. The stabilizing operators of weight 2 are $\mathbb{1}\mathbb{1}ZZ$ and its permutations. The set formed by those six operators and the identity is not closed under multiplication. The smallest group containing this set also includes $ZZZZ$,

$$S^{(2)} = \{\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1},\ \mathbb{1}\mathbb{1}ZZ \text{ and permutations},\ ZZZZ\}. \tag{6.69}$$

As there are no stabilizing operators of weight 3, we have $S^{(3)} = S^{(2)}$. For the nested generating sets in Eq. (6.60) one can choose

$$g^{(1)} = \emptyset, \tag{6.70}$$

$$g^{(3)} = g^{(2)} = \{\mathbb{1}\mathbb{1}ZZ,\ \mathbb{1}Z\mathbb{1}Z,\ ZZ\mathbb{1}\mathbb{1}\}, \tag{6.71}$$

$$g = g^{(4)} = \{\mathbb{1}\mathbb{1}ZZ,\ \mathbb{1}Z\mathbb{1}Z,\ ZZ\mathbb{1}\mathbb{1}, XXXX\}. \tag{6.72}$$

We define $\emptyset$ rather than $\{\mathbb{1}^{\otimes n}\}$ as the generating set of the trivial group $\{\mathbb{1}^{\otimes n}\}$ for reasons of consistency, ensuring that the cardinality of a stabilizer group is always 2^m if m is the number of generators. The distances from the GHZ state to the exponential families are given by

$$D_1(\rho_{\mathrm{GHZ}_4}) = 4, \quad D_2(\rho_{\mathrm{GHZ}_4}) = 1, \quad D_3(\rho_{\mathrm{GHZ}_4}) = 1 \tag{6.73}$$

and its degrees of interaction by

$$C_2(\rho_{\mathrm{GHZ}_4}) = 3, \quad C_3(\rho_{\mathrm{GHZ}_4}) = 0, \quad C_4(\rho_{\mathrm{GHZ}_4}) = 1, \quad C_{\mathrm{tot}}(\rho_{\mathrm{GHZ}_4}) = 4. \tag{6.74}$$

As a caveat note the following: While the numbers m_k, which determine the degrees of interaction, can be found by counting the generating operators with weight k in Eq. (6.72), this does not work with an arbitrary generating set. For example, $g = \{\mathbb{1}\mathbb{1}ZZ, \mathbb{1}Z\mathbb{1}Z, ZZZZ, XXXX\}$ is a legitimate generating set for the GHZ state, but gives the wrong numbers m_k. The reason is that the operators of weight 2 contained in this set, namely $\{\mathbb{1}\mathbb{1}ZZ, \mathbb{1}Z\mathbb{1}Z\}$, do not generate $S^{(2)}$, since $ZZ\mathbb{1}\mathbb{1}$ is missing. In other words, the generating set is not of the form of Eq. (6.60).

As our second example we take the four-qubit linear cluster state with stabilizer group Eq. (2.99). Here, $S^{(1)} = \{\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\}$ and

$$S^{(2)} = \{\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1},\ \mathbb{1}\mathbb{1}ZZ,\ ZZ\mathbb{1}\mathbb{1},\ ZZZZ\}. \tag{6.75}$$

The group $S^{(3)}$ contains at least the four elements of $S^{(2)}$ and the eight stabilizing operators of weight 3, so the cardinality of this group is at least 12. But because the cardinality of a stabilizer group is a power of 2, this group must have cardinality 16 and is equal to the complete stabilizer group of the cluster state. The distances and the interaction measures take values

$$D_1(\rho_{\mathrm{C}_4}) = 4, \quad D_2(\rho_{\mathrm{C}_4}) = 2, \quad D_3(\rho_{\mathrm{C}_4}) = 0 \tag{6.76}$$

and

$$C_2(\rho_{\mathrm{C}_4}) = 2, \quad C_3(\rho_{\mathrm{C}_4}) = 2, \quad C_4(\rho_{\mathrm{C}_4}) = 0, \quad C_{\mathrm{tot}}(\rho_{\mathrm{C}_4}) = 4, \tag{6.77}$$

respectively.

A related question which has been addressed in the literature is under which conditions stabilizer states can occur as ground states of k-party Hamiltonians. In particular, it would be advantageous if states which are known to be universal resources for measurement-based quantum computation, such as the two-dimensional cluster state, occurred as nondegenerate ground states of physically reasonable Hamiltonians. In this case, they could be prepared by cooling, obviating the need to manipulate the qubits individually. In Ref. 116 Van den Nest et al. have shown the following: The minimal k for a pure, fully entangled stabilizer state ρ_S to be the (possibly degenerate) ground state of a nontrivial[2] k-party Hamiltonian is given by the minimal weight of its stabilizing operators (excluding the identity),

$$k_{\min} = \min_{\pm\sigma_\alpha \in S \setminus \{\mathbb{1}\}} W(\alpha). \tag{6.78}$$

[2]In this context a Hamiltonian is called *nontrivial* if it is not a multiple of the identity.

If one asks for a nondegenerate ground state, the threshold is given by the minimal k such that the stabilizing operators of weight at most k generate the whole stabilizer group. In our language this is the minimal k such that the state is contained in the compactified exponential family,

$$k_{\min} = \min\{k \mid S^{(k)} = S\} = \min\{k \mid \rho_S \in \overline{Q}_k\}. \tag{6.79}$$

In particular, for more than two qubits this $k_{\min}$ is always at least 3. Furthermore, it was shown in that reference that a stabilizer state can only be close (in the sense of the fidelity) to the ground state of a Hamiltonian whose k is lower than the threshold in Eq. (6.79) if the energy gap of the Hamiltonian (relative to the total energy scale) is small. (See also Ref. 48 for related results.) The topic of states that cannot be written as general eigenstates (not necessarily ground states) of few-party Hamiltonians has also received attention [33, 47].

We resume the description of Zhou's method [132], discussing the explicit calculation of the information projections of the generalized GHZ state of n qubits ($n \geq 3$),

$$|\mathrm{GHZ}_n(\theta,\phi)\rangle = \cos\left(\tfrac{\theta}{2}\right)|0\rangle^{\otimes n} + \mathrm{e}^{\mathrm{i}\phi}\sin\left(\tfrac{\theta}{2}\right)|1\rangle^{\otimes n}. \tag{6.80}$$

This state can be written in a form reminiscent of a stabilizer state,

$$\rho_{\mathrm{GHZ}_n}(\theta,\phi) = \frac{1}{2^n}\left[\mathbb{1} + \Sigma_{\theta,\phi}\right]\prod_{i=2}^{n}\left[\mathbb{1} + \sigma_z^{(1)}\sigma_z^{(i)}\right] \tag{6.81}$$

where $\Sigma_{\theta,\phi} = \mathbf{e}_{\theta,\phi}\cdot\mathbf{\Sigma}$ with

$$\mathbf{e}_{\theta,\phi} = \begin{pmatrix}\sin(\theta)\cos(\phi)\\ \sin(\theta)\sin(\phi)\\ \cos(\theta)\end{pmatrix} \qquad \text{and} \qquad \mathbf{\Sigma} = \begin{pmatrix}\sigma_x^{(1)}\sigma_x^{(2)}\cdots\sigma_x^{(n)}\\ \sigma_y^{(1)}\sigma_x^{(2)}\cdots\sigma_x^{(n)}\\ \sigma_z^{(1)}\end{pmatrix}. \tag{6.82}$$

This can be seen as follows: First note that

$$\Sigma_{\theta,\phi}^2 = \mathbb{1} \qquad \text{and} \qquad \left[\Sigma_{\theta,\phi}, \sigma_z^{(1)}\sigma_z^{(i)}\right] = 0. \tag{6.83}$$

The operators $\sigma_z^{(1)}\sigma_z^{(i)}$ can be regarded as generators of a stabilizer group, stabilizing a two-dimensional subspace. This subspace is spanned by $|0\rangle^{\otimes n}$ and $|1\rangle^{\otimes n}$. The operator $\Sigma_{\theta,\phi}$ acts on this subspace as

$$\Sigma_{\theta,\phi}\left(\alpha|0\rangle^{\otimes n} + \beta|1\rangle^{\otimes n}\right) = \alpha'|0\rangle^{\otimes n} + \beta'|1\rangle^{\otimes n} \tag{6.84}$$

where

$$\begin{pmatrix}\alpha'\\ \beta'\end{pmatrix} = \begin{pmatrix}\cos(\theta) & \mathrm{e}^{-\mathrm{i}\phi}\sin(\theta)\\ \mathrm{e}^{\mathrm{i}\phi}\sin(\theta) & -\cos(\theta)\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}. \tag{6.85}$$

The matrix in the last equation has eigenvalues ± 1; the eigenvector corresponding to $+1$ is $(\cos(\theta/2), \mathrm{e}^{\mathrm{i}\phi}\sin(\theta/2))$.

The generalized GHZ state is obtained in the limit

$$\rho_{\mathrm{GHZ}_n}(\theta,\phi) = \lim_{\mu\to\infty} \rho_{\mathrm{GHZ}_n}(\mu,\mu,\theta,\phi) \tag{6.86}$$

of the parametrized family of full-rank states

$$\begin{aligned}\rho_{\mathrm{GHZ}_n}(\lambda,\mu;\theta,\phi) &= \frac{1}{2^n}\big[\mathbb{1} + \tanh(\lambda)\Sigma_{\theta,\phi}\big]\prod_{i=2}^{n}\big[\mathbb{1} + \tanh(\mu)\sigma_z^{(1)}\sigma_z^{(i)}\big] \\ &= \frac{1}{2^n\cosh(\lambda)\cosh^{n-1}(\mu)}\exp\Big[\lambda\Sigma_{\theta,\phi} + \mu\sum_{i=2}^{n}\sigma_z^{(1)}\sigma_z^{(i)}\Big].\end{aligned} \tag{6.87}$$

Now observe that multiplication of the x- or the y-component of $\mathbf{\Sigma}$ [which was defined in Eq. (6.82)] with the product $\prod_{i=2}^{n}[\mathbb{1} + \tanh(\mu)\sigma_z{}^{(1)}\sigma_z{}^{(i)}]$ gives only operators of weight n. This shows that only the z-component of $\mathbf{\Sigma}$ contributes to the reduced one-party density matrices of $\rho_{\mathrm{GHZ}_n}(\lambda,\mu;\theta,\phi)$. Therefore this state has the same information projections as $\rho_{\mathrm{GHZ}_n}(\lambda',\mu;0,0)$ if

$$\tanh(\lambda') = \tanh(\lambda)\cos(\theta). \tag{6.88}$$

The latter state is in $\mathcal{Q}_2$. Let $\lambda'(\mu,\theta) = \operatorname{artanh}(\tanh(\mu)\cos(\theta))$. Then

$$\begin{aligned}\lim_{\mu\to\infty}\rho_{\mathrm{GHZ}_n}\big(\lambda'(\mu,\theta),\mu,0,0\big) &= \frac{1}{2^n}\big[\mathbb{1} + \cos(\theta)\sigma_z^{(1)}\big]\prod_{i=2}^{n}\big[\mathbb{1} + \sigma_z^{(1)}\sigma_z^{(i)}\big] \\ &= \cos^2\!\big(\tfrac{\theta}{2}\big)|0\rangle\langle 0|^{\otimes n} + \sin^2\!\big(\tfrac{\theta}{2}\big)|1\rangle\langle 1|^{\otimes n}.\end{aligned} \tag{6.89}$$

This state is the information projection of the generalized GHZ state for $k \geq 2$. The projection for $k = 1$ is as always given by the tensor product of the reduced density matrices, which is

$$\widetilde{\rho}_1 = \Big(\cos^2\!\big(\tfrac{\theta}{2}\big)|0\rangle\langle 0| + \sin^2\!\big(\tfrac{\theta}{2}\big)|1\rangle\langle 1|\Big)^{\otimes n}. \tag{6.90}$$

A direct calculation gives the interaction measures (see below).

If one is not afraid of working with states without full rank, one can obtain these results more easily by arguing that the generalized GHZ state has the same reduced density matrices, and hence the same information projections, as the state in Eq. (6.89).

We summarize the results in the following theorem:

Theorem 6.7 (Ref. 132). *The generalized n-qubit GHZ state*

$$|GHZ_n(\theta,\phi)\rangle = \cos\!\big(\tfrac{\theta}{2}\big)|0\rangle^{\otimes n} + \mathrm{e}^{\mathrm{i}\phi}\sin\!\big(\tfrac{\theta}{2}\big)|1\rangle^{\otimes n} \tag{6.91}$$

has information projections

$$\widetilde{\rho}_1 = \Big(\cos^2\!\big(\tfrac{\theta}{2}\big)|0\rangle\langle 0| + \sin^2\!\big(\tfrac{\theta}{2}\big)|1\rangle\langle 1|\Big)^{\otimes n} \tag{6.92}$$

and

$$\widetilde{\rho}_k = \cos^2\!\big(\tfrac{\theta}{2}\big)|0\rangle\langle 0|^{\otimes n} + \sin^2\!\big(\tfrac{\theta}{2}\big)|1\rangle\langle 1|^{\otimes n} \quad \textit{for} \quad 2 \leq k \leq n-1. \tag{6.93}$$

It has distances to the exponential families given by

$$D_k(\rho_{GHZ_n}) = \begin{cases} nS_2\big(\cos^2(\theta)\big) & \text{for } k = 1, \\ (n-1)S_2\big(\cos^2(\theta)\big) & \text{for } 2 \leq k \leq n-1, \end{cases} \tag{6.94}$$

degrees of irreducible k-party interactions

$$C_k(\rho_{GHZ_n}) = \begin{cases} S_2\big(\cos^2(\theta)\big) & \text{for } k = 2, \\ 0 & \text{for } 3 \leq k \leq n-1, \\ (n-1)S_2\big(\cos^2(\theta)\big) & \text{for } k = n \end{cases} \tag{6.95}$$

and degree of total interaction

$$C_{tot}(\rho_{GHZ_n}) = nS_2\big(\cos^2(\theta)\big), \tag{6.96}$$

where $S_2(p) = -p\log(p) - (1-p)\log(1-p)$ is the binary entropy function.

6.3 Witness operators for exponential families

It would be interesting to show that an experimental state is not in a certain exponential family, in other words, that is contains irreducible interactions of higher order. This is similar to the task of entanglement detection (cf. Section 2.1.3). Exploiting this similarity, we adopt the method of witness operators, which is widely used in entanglement detection. The aim of this section is to prove statements of the following type: *If the fidelity of a state ρ_{exp} with a target state $|T\rangle$, which is not in the quantum exponential family $\overline{\mathcal{Q}}_k$, is larger than certain bound F_0, then ρ_{exp} is not in $\overline{\mathcal{Q}}_k$ either.* In this case

$$W = F_0\mathbb{1} - |T\rangle\langle T| \tag{6.97}$$

is a witness operator for irreducible $(k+1)$-party interactions. Because of the linearity of the witness criterion, a state detected by the witness is not in the convex hull of $\overline{\mathcal{Q}}_k$ either. The existence of a witness for $\overline{\mathcal{Q}}_k$ thus shows that the convex hull of this set is not the complete space.

The fidelity bounds that will be derived here are much too close to 1 to be useful. Therefore the present section should be understood as a proof-of-principle, showing that such bounds can be found at all.

As target state the five-qubit ring cluster state $|\mathrm{R}_5\rangle$ will be used, which is the graph state defined by the graph in Fig. 2.2 (b). Its state vector and stabilizer group are given in Eqs. (2.100) and (2.101). For the purpose of this section it is only important that the stabilizer group contains no operators of weight 1 of 2. For the distances and the interaction measures one finds with the method of the previous section

$$D_1(\rho_{\mathrm{R}_5}) = D_3(\rho_{\mathrm{R}_5}) = 5, \qquad D_3(\rho_{\mathrm{R}_5}) = D_4(\rho_{\mathrm{R}_5}) = 0 \tag{6.98}$$

and

$$C_2(\rho_{\mathrm{R}_5}) = C_4(\rho_{\mathrm{R}_5}) = C_5(\rho_{\mathrm{R}_5}) = 0, \qquad C_{\mathrm{tot}}(\rho_{\mathrm{R}_5}) = C_3(\rho_{\mathrm{R}_5}) = 5. \tag{6.99}$$

Figure 6.4 below shows these quantities for the ring cluster state with white noise.

We expand the experimental state into Pauli operators,

$$\rho_{\mathrm{exp}} = \frac{1}{32}\sum_{\alpha}\eta_\alpha\sigma_\alpha. \tag{6.100}$$

If this state has a high fidelity with the ring cluster state, the coefficients η_α corresponding to the stabilizing operators of the ring cluster state will be large, and the coefficients corresponding to other Pauli operators will be small. In particular all Pauli operators of weight 1 and 2 will have small coefficients. This observation will be used to show that for sufficiently high fidelity of ρ_{exp} the operator ρ' defined by

$$\rho' = \frac{1}{32}\Big(\mathbb{1} + \sum_{W(\alpha)=1,2}\eta_\alpha\sigma_\alpha\Big) \tag{6.101}$$

(with the same η_α for $W(\alpha) = 1, 2$ as ρ_{exp}) is a valid quantum state and has a higher von Neumann entropy than ρ_{exp}. In this case ρ_{exp} cannot be equal to its own projection onto $\overline{\mathcal{Q}_2}$, because there is another state, namely ρ', with the same two-party reduced density matrices, but a higher entropy. This shows that ρ_{exp} is not in $\overline{\mathcal{Q}_2}$. (In general ρ' is not the information projection of ρ_{exp}, but this is not the point.)

Let S be the stabilizer group of the ring cluster state and $S^* = S \setminus \{\mathbb{1}\}$. Suppose we measure a fidelity

$$F = \langle \mathrm{R}_5|\rho_{\mathrm{exp}}|\mathrm{R}_5\rangle = \frac{1}{32}\Big(1 + \sum_{M\in S^*}\langle M\rangle\Big) \geq 1-\varepsilon. \tag{6.102}$$

With the inequality

$$\Big|\frac{1}{N}\sum_{i=1}^{N} x_i\Big| \leq \frac{1}{N}\sum_{i=1}^{N}|x_i| \leq \Big(\frac{1}{N}\sum_{i=1}^{N} x_i^2\Big)^{1/2}, \tag{6.103}$$

which holds for all real numbers x_i, we obtain

$$\sum_{M\in S^*}\langle M\rangle^2 \geq 31\big(1-\tfrac{32}{31}\varepsilon\big)^2. \tag{6.104}$$

From $\mathrm{Tr}(\rho_{\mathrm{exp}}^2) \leq 1$ it follows that $\sum_{\alpha\neq 0}\eta_\alpha^2 \leq 31$ and thus

$$\sum_{W(\alpha)=1,2}\eta_\alpha^2 \leq \sum_{\alpha\neq 0}\eta_\alpha^2 - \sum_{M\in S^*}\langle M\rangle^2 \leq 31 - 31\big(1-\tfrac{32}{31}\varepsilon\big)^2 = 64\varepsilon\big(1-\tfrac{16}{31}\varepsilon\big). \tag{6.105}$$

Here it is essential that the stabilizer group does not contain any operators of weight 1 or 2.

We shall need the following lemma:

Lemma 6.8. *The one- and two-qubit Pauli operators of a five-qubit system can be divided into 15 sets of 7 operators each such that all operators in the same set anticommute pairwise.*

Proof. The elements of the set

$$\{\sigma_x^{(j)},\ \sigma_x^{(i)}\sigma_y^{(j)},\ \sigma_y^{(i)}\sigma_y^{(j)},\ \sigma_z^{(i)}\sigma_y^{(j)},\ \sigma_z^{(j)}\sigma_x^{(k)},\ \sigma_z^{(j)}\sigma_y^{(k)},\ \sigma_z^{(j)}\sigma_z^{(k)}\} \tag{6.106}$$

anticommute pairwise. Consider this set together with the two sets that can be obtained from it by the cyclic replacement of Pauli operators $\sigma_x \to \sigma_y \to \sigma_z \to \sigma_x$. Put together these three sets are comprised of the one-qubit Pauli operators on qubit j and the the two-qubit Pauli operators on the pairs of qubits (i,j) and (j,k). Now choose $(i,j,k) = (n, n+1, n+3)$ for $n = 1,\ldots,5$ (qubit labels to be understood modulo 5, of course). □

This lemma allows us to apply a result on dichotomic anticommuting observables which we derived in the context of uncertainty relations, namely the meta-uncertainty relation Lemma 5.6. With the above lemma and the meta-uncertainty relation we can show:

Lemma 6.9. *If the coefficients η_α for $W(\alpha) = 1,2$ are bounded by $\sum_{W(\alpha)=1,2} \eta_\alpha^2 \leq 1/15$, then $\rho' = \frac{1}{32}(\mathbb{1} + \sum_{W(\alpha)=1,2} \eta_\alpha \sigma_\alpha)$ is a valid quantum state.*

Proof. With Lemma 6.8 we can write the state ρ' as

$$\rho' = \sum_{i=1}^{15} \mu_i \rho_i \qquad \text{with} \qquad \rho_i = \frac{1}{32}\Big(\mathbb{1} + \frac{1}{\mu_i}\sum_{j=1}^{7} \eta_{ij}\sigma_{\alpha_{ij}}\Big) \tag{6.107}$$

such that the $\sigma_{\alpha_{ij}}$ with the same i anticommute pairwise. We can choose the μ_i such that

$$\sum_j \eta_{ij}^2 \leq \mu_i^2 \quad \text{and} \quad \sum_i \mu_i = 1. \tag{6.108}$$

[How? Let $\mu_i = (\sum_j \eta_{ij}^2)^{1/2}$. Then $\sum_i \mu_i^2 \leq 1/15$. Using again Eq. (6.103), we obtain $\sum_i \mu_i \leq 1$. Then increase some of the μ_i until $\sum_i \mu_i = 1$.] The meta-uncertainty relation Lemma 5.6 shows that the ρ_i are positive semidefinite. Then ρ' is a convex combination of valid states. □

The last lemma together with Eq. (6.105) show that $64\varepsilon(1 - 16\varepsilon/31) \leq 1/15$ is a sufficient condition for the positive semidefiniteness of ρ'. This translates to

$$\varepsilon \leq \frac{31}{32} - \frac{1}{8}\sqrt{\frac{899}{15}} \approx 0.001042. \tag{6.109}$$

It remains to show that $S(\rho') > S(\rho_{\text{exp}})$. First an upper bound on $S(\rho_{\text{exp}})$ will be derived. Let p_i, $i = 1,\ldots,32$ be the probabilities for the measurement of ρ_{exp} in the graph state basis of the ring cluster state. This gives the bound

$$S(\rho_{\text{exp}}) \leq S\big((p_1,\ldots,p_{32})\big) \tag{6.110}$$

[see Eq. (2.58)]. By assumption $p_1 = F \geq 1 - \varepsilon$, where p_1 corresponds to the ring cluster state. The probability distribution with the highest entropy which is consistent with this constraint is $(1-\varepsilon, \frac{\varepsilon}{31}, \ldots, \frac{\varepsilon}{31})$, and so

$$S(\rho_{\text{exp}}) \leq S\big((1-\varepsilon, \tfrac{\varepsilon}{31}, \ldots, \tfrac{\varepsilon}{31})\big) = -(1-\varepsilon)\log(1-\varepsilon) - \varepsilon \log(\tfrac{\varepsilon}{31}). \tag{6.111}$$

This gives $S(\rho_{\text{exp}}) \lesssim 0.01699$ for the ε from Eq. (6.109).

We will now derive a lower bound on $S(\rho')$. From Eq. (6.105) we obtain, using once again Eq. (6.103),

$$\sum_{W(\alpha)=1,2} |\eta_\alpha| \leq \sqrt{105 \cdot 64\varepsilon(1 - \tfrac{16}{31}\varepsilon)}. \tag{6.112}$$

Since for any normalized state $|\psi\rangle$

$$\langle\psi|\rho'|\psi\rangle = \frac{1}{32}\Big(1 + \sum_\alpha \eta_\alpha \langle\psi|\sigma_\alpha|\psi\rangle\Big) \leq \frac{1}{32}\Big(1 + \sum_\alpha |\eta_\alpha|\Big), \tag{6.113}$$

inequality (6.112) gives an upper bound on the eigenvalues p_i of ρ', namely

$$p_i \leq \frac{1}{32}\Big[1 + \sqrt{105 \cdot 64\varepsilon(1 - \tfrac{16}{31}\varepsilon)}\,\Big]. \tag{6.114}$$

This shows that

$$S(\rho') \geq S\big((p, 1-p, 0, \ldots, 0)\big) = -p\log(p) - (1-p)\log(1-p) \tag{6.115}$$

where

$$p = \frac{1}{32}\Big[1 + \sqrt{105 \cdot 64\varepsilon(1 - \tfrac{16}{31}\varepsilon)}\,\Big]. \tag{6.116}$$

This gives $S(\rho') \gtrsim 0.5117$ for the value from Eq. (6.109). In conclusion, $S(\rho') > S(\rho_{\text{exp}})$ for all values of ε for which we have shown that ρ' is positive semidefinite.

Let us discuss the implications of our result. As was briefly mentioned above, any state detected by the witness $W = F_0 \mathbb{1} - |\mathrm{R}_5\rangle\langle\mathrm{R}_5|$ (where F_0 is the threshold fidelity) does not lie in the convex hull of $\overline{\mathcal{Q}}_k$. In this context it is worth noting that in the classical case even the convex hull of $\overline{\mathcal{E}}_1$, which consists of all convex combinations of product distributions, is the complete space. By constructing a witness for $\overline{\mathcal{Q}}_2$ we have thus shown the quite nontrivial result that quantum exponential families are different in this respect, and that lower bounds on the distance of a state to the convex hull of such a family can be found. For experimental applications it only remains to improve the bound in Eq. (6.109).

6.4 Iterative computation of the quantum information projection

In this section an algorithm for the computation of the quantum information projection is developed. It is based on the iterative scaling procedure for the classical case,

which was described in Section 2.6.3. However, it is not a straightforward generalization of that algorithm. Rather, it involves a certain approximation in the iteration step to circumvent the difficulties associated with the matrix exponential.

We recall the iteration step of the classical scaling algorithm: For a subset $A \subseteq \{1, \ldots, n\}$ of the parties with $|A| = k$, the approximation Q to the projection $\widetilde{P}_k$ of P is updated in such a way that after the update Q has the same A-marginal as P:

$$Q \to Q' \quad \text{such that} \quad Q'_A = P_A. \tag{6.117}$$

The algorithm for the quantum case is based on the idea of adapting this update rule to the quantum case as follows: Let ρ be the state whose projection onto $\mathcal{Q}_k$ we want to calculate, and let $\sigma = \mathrm{e}^H / \operatorname{Tr}(\mathrm{e}^H)$ with a k-party Hamiltonian H be the approximation to $\widetilde{\rho}_k$ that we have obtained so far. For an ℓ-party observable A with $\ell \leq k$ we add εA to the Hamiltonian with ε chosen such that afterwards A has the correct expectation value:

$$\sigma = \frac{\mathrm{e}^H}{\operatorname{Tr}(\mathrm{e}^H)} \to \sigma' = \frac{\mathrm{e}^{H+\varepsilon A}}{\operatorname{Tr}(\mathrm{e}^{H+\varepsilon A})} \quad \text{such that} \quad \operatorname{Tr}(A\sigma') = \operatorname{Tr}(A\rho). \tag{6.118}$$

The difficulty lies in finding ε. We linearize the equation which determines it,

$$\operatorname{Tr}(A\sigma') = \operatorname{Tr}(A\sigma) + \varepsilon \partial_\varepsilon\big|_{\varepsilon=0} \operatorname{Tr}(A\sigma') + \mathcal{O}(\varepsilon^2) \tag{6.119}$$

with the derivative

$$\partial_\varepsilon \operatorname{Tr}(A\sigma') = \operatorname{Tr}\Big[A \frac{\partial_\varepsilon \mathrm{e}^{H+\varepsilon A}}{\operatorname{Tr}(\mathrm{e}^{H+\varepsilon A})}\Big] - \operatorname{Tr}\Big[A \frac{\mathrm{e}^{H+\varepsilon A}}{\operatorname{Tr}(\mathrm{e}^{H+\varepsilon A})}\Big] \frac{\operatorname{Tr}(\partial_\varepsilon \mathrm{e}^{H+\varepsilon A})}{\operatorname{Tr}(\mathrm{e}^{H+\varepsilon A})}. \tag{6.120}$$

When evaluating the derivative of the exponential one has to deal with noncommuting H and A. With the identity [see for example Eq. (4.1) in Ref. 130]

$$\partial_t \mathrm{e}^{M(t)} = \int_0^1 \mathrm{d}s\, \mathrm{e}^{sM(t)} M'(t) \mathrm{e}^{-sM(t)} \mathrm{e}^{M(t)} \tag{6.121}$$

one obtains for $\varepsilon = 0$

$$\begin{aligned}
\partial_\varepsilon\big|_{\varepsilon=0} \operatorname{Tr}(A\sigma') &= \operatorname{Tr}\Big[A \frac{1}{\operatorname{Tr}(\mathrm{e}^H)} \int_0^1 \mathrm{d}s\, \mathrm{e}^{sH} A e^{-sH} e^H\Big] \\
&\quad - \operatorname{Tr}\Big[A \frac{\mathrm{e}^H}{\operatorname{Tr}(\mathrm{e}^H)}\Big] \frac{1}{\operatorname{Tr}(\mathrm{e}^H)} \operatorname{Tr}\Big[\int_0^1 \mathrm{d}s\, \mathrm{e}^{sH} A e^{-sH} e^H\Big] \\
&= \frac{1}{\operatorname{Tr}(\mathrm{e}^H)} \int_0^1 \mathrm{d}s \operatorname{Tr}\big(A \mathrm{e}^{sH} A \mathrm{e}^{-sH} \mathrm{e}^H\big) - \Big\{\operatorname{Tr}\Big[A \frac{\mathrm{e}^H}{\operatorname{Tr}(\mathrm{e}^H)}\Big]\Big\}^2 \\
&= \frac{1}{\operatorname{Tr}(\mathrm{e}^H)} \int_0^1 \mathrm{d}s \operatorname{Tr}\big(A \mathrm{e}^{sH} A \mathrm{e}^{-sH} \mathrm{e}^H\big) - \big[\operatorname{Tr}(A\sigma)\big]^2.
\end{aligned} \tag{6.122}$$

We crudely approximate the integral by evaluating the integrand only at the two end points $s = 0$ and $s = 1$ of the integration interval,

$$\frac{1}{\operatorname{Tr}(\mathrm{e}^H)} \int_0^1 \mathrm{d}s \operatorname{Tr}\big(A \mathrm{e}^{sH} A \mathrm{e}^{-sH} \mathrm{e}^H\big) \approx \frac{1}{\operatorname{Tr}(\mathrm{e}^H)} \frac{1}{2} \Big[\operatorname{Tr}(A \mathrm{e}^H A) + \operatorname{Tr}(A^2 e^H)\Big] = \operatorname{Tr}(A^2 \sigma). \tag{6.123}$$

which leads to

$$\mathrm{Tr}(A\sigma') \approx \mathrm{Tr}(A\sigma) + \varepsilon\big\{\mathrm{Tr}(A^2\sigma) - [\mathrm{Tr}(A\sigma)]^2\big\}. \tag{6.124}$$

Like the above expansion in ε, this approximation is only justified a posteriori by the performance of the algorithm. The solution for ε is

$$\varepsilon \approx \frac{\mathrm{Tr}(A\rho) - \mathrm{Tr}(A\sigma)}{\mathrm{Tr}(A^2\sigma) - [\mathrm{Tr}(A\sigma)]^2} = \frac{\langle A\rangle_\rho - \langle A\rangle_\sigma}{\Delta^2_\sigma(A)}. \tag{6.125}$$

To compute the information projection onto $\mathcal{Q}_k$, one chooses an orthogonal basis V_k in the space of k-party observables (omitting the identity) and updates σ for each $A \in V_k$ in turn. For an n-qubit system one can choose the Pauli operators

$$V_k = \big\{\sigma_\alpha \;\big|\; 1 \le W(\alpha) \le k\big\}. \tag{6.126}$$

The complete algorithm is as follows:

Algorithm 6.10 (Iterative computation of the quantum information projection). Problem: *Given an n-qubit state ρ, compute its information projection $\widetilde{\rho}_k$ onto the exponential family $\overline{\mathcal{Q}}_k$.*

1. *For each element A of an orthonormal basis V_k of the space of k-party observables compute the expectation value $\langle A\rangle_\rho$.*
2. *Initialize $\sigma = \mathbb{1}/2^n$ as the completely mixed state.*
3. *Looping through all observables $A \in V_k$, update σ according to the rule*

$$\sigma = \frac{\mathrm{e}^H}{\mathrm{Tr}(\mathrm{e}^H)} \to \sigma' = \frac{\mathrm{e}^{H+\varepsilon A}}{\mathrm{Tr}(\mathrm{e}^{H+\varepsilon A})} \quad \textit{where} \quad \varepsilon = \frac{\langle A\rangle_\rho - \langle A\rangle_\sigma}{\Delta^2_\sigma(A)}. \tag{6.127}$$

4. *Repeat the last step.*

When implementing the algorithm, it turns out to be useful to introduce an additional parameter ω which controls the size of the steps in the space of Hamiltonians,

$$\sigma = \frac{\mathrm{e}^H}{\mathrm{Tr}(\mathrm{e}^H)} \to \sigma' = \frac{\mathrm{e}^{H+\omega\varepsilon A}}{\mathrm{Tr}(\mathrm{e}^{H+\omega\varepsilon A})} \tag{6.128}$$

with ε as above. Choosing $\omega < 1$ is arguably at odds with the idea of the algorithm, but sometimes improves the convergence rate. (Figure 6.5 has been obtained in this way.)

Note that we did not prove that the algorithm converges. One could also think of improving it by using a better approximation to the integral. However, the numerical results shown below demonstrate that the algorithm as described here works remarkably well.

Having obtained the projections $\widetilde{\rho}_k$ of ρ, one can compute the distances $D_k(\rho) = D(\rho\|\widetilde{\rho}_k)$ and the interaction measures $C_k(\rho)$. The latter can be calculated in three different ways, namely

$$C_k(\rho) = D_{k-1}(\rho) - D_k(\rho), \tag{6.129}$$
$$C_k(\rho) = D(\widetilde{\rho}_k\|\widetilde{\rho}_{k-1}), \tag{6.130}$$
$$C_k(\rho) = S(\widetilde{\rho}_{k-1}) - S(\widetilde{\rho}_k). \tag{6.131}$$

If $\widetilde{\rho}_k$ is not the correct information projection, these three expressions will in general give different values for $C_k(\rho)$. This is useful as a consistency check for the numerical value of $\widetilde{\rho}_k$.

In order to test the algorithm, the information projections have been computed for a number of four- and five-qubit states, each with additional white noise,

$$\rho(p) = p\frac{\mathbb{1}}{2^n} + (1-p)\rho. \tag{6.132}$$

These states are the four-qubit GHZ state, the four-qubit Smolin state

$$\rho_{\mathrm{Smo}} = \frac{1}{16}\big(\mathbb{1} + XXXX + YYYY + ZZZZ\big), \tag{6.133}$$

the five-qubit W state

$$|\mathrm{W}_5\rangle = \frac{1}{\sqrt{5}}\big(|10000\rangle + |01000\rangle + |00100\rangle + |00010\rangle + |00001\rangle\big) \tag{6.134}$$

and the five-qubit ring cluster state [Eq. (2.100)]. Figures 6.2–6.5 show the distances D_k and the interaction measures C_k and C_{tot} as functions of the noise level p. The first three of these figures reproduce results which were obtained before by Zhou [133] with a different method.[3]

In Ref. 133 Zhou seems to compute the information projections as follows: The exponential family $\mathcal{Q}_k$ is parametrized by the coefficients of the Hamiltonian in an operator basis. The information projection $\widetilde{\rho}_k$ is determined among all states in $\mathcal{Q}_k$ by the requirement that it has the same k-party reduced density matrices as ρ. This constitutes a system of nonlinear equations for the parameters of the Hamiltonian, which is then solved numerically. The numerical method which is employed is not described. However, this method apparently requires to be provided with a carefully chosen initial guess for the roots of the equations. In fact, in that reference the parametrized families of the form Eq. (6.132) are introduced explicitly for the purpose of using the result for a higher noise level p as initial guess for a lower p. Starting from a state close to the completely mixed state, the information projections of states with progressively lower noise levels are calculated.

[3]Note that in this reference the base of the logarithm is handled inconsistently: In the equations the natural logarithm is used, but the numerical results in the figures use the binary logarithm. Also, there is a mistake in the definition of the Smolin state, but for the calculation the correct state was apparently used.

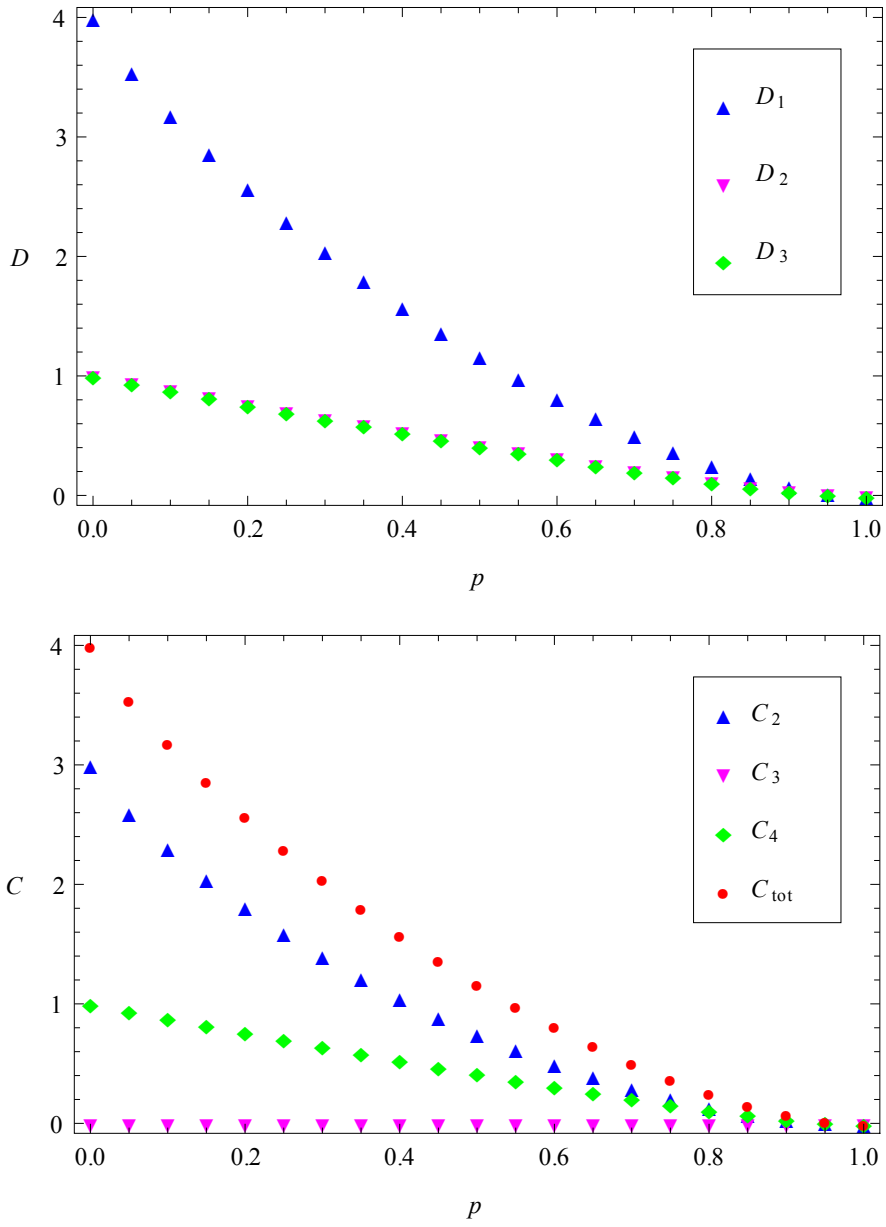

Figure 6.2: Distances to the exponential families (upper figure) and interaction measures (lower figure) as functions of the level of white noise for the four-qubit GHZ state. This reproduces results which were obtained in Ref. 133 with a different method.

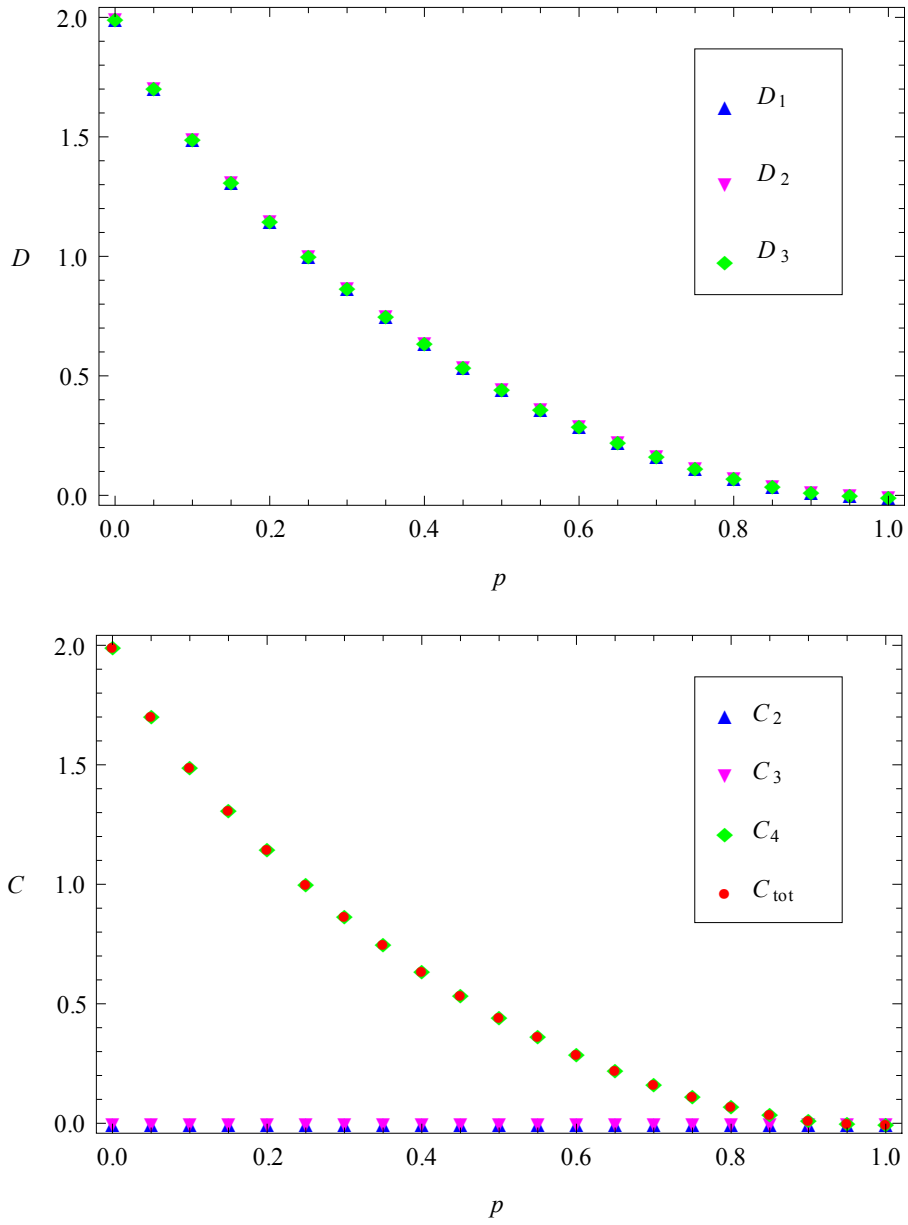

Figure 6.3: Distances to the exponential families (upper figure) and interaction measures (lower figure) as functions of the level of white noise for the four-qubit Smolin state [Eq. (6.133)]. This reproduces results which were obtained in Ref. 133 with a different method.

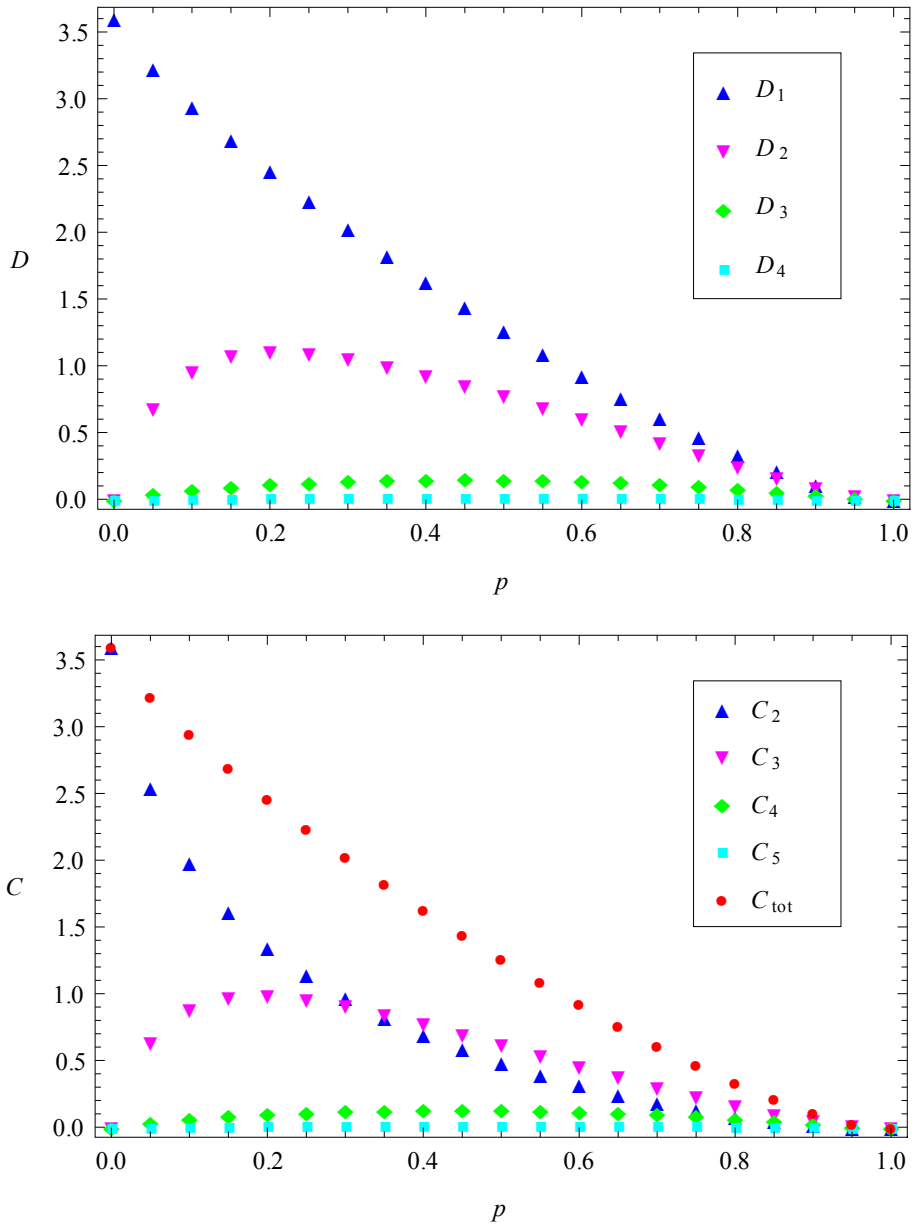

Figure 6.4: Distances to the exponential families (upper figure) and interaction measures (lower figure) as functions of the level of white noise for the five-qubit W state [Eq. (6.134)]. This reproduces results which were obtained in Ref. 133 with a different method.

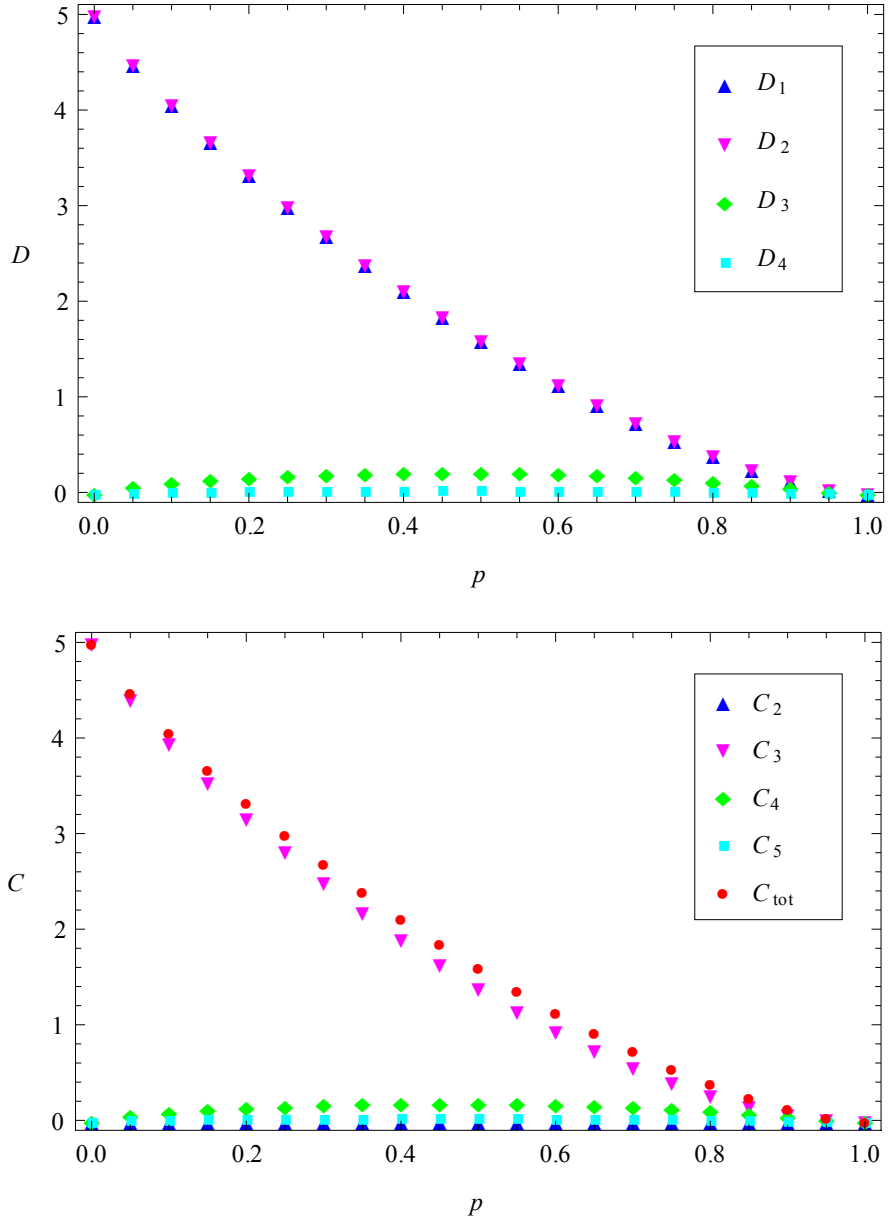

Figure 6.5: Distances to the exponential families (upper figure) and interaction measures (lower figure) as functions of the level of white noise for the five-qubit ring cluster state [Eq. (2.100)].

Lacking detailed information on Zhou's algorithm, it is difficult to compare the performances of the algorithms objectively. However, it seems fair to say that the algorithm developed here makes better use of the structure of the problem.

While the data shown in the figures were produced for the purpose of testing the algorithm, understanding the results might also help to understand the geometry of the exponential families in the space of density matrices. Generally speaking, both the distances D_k and the interaction measures C_{tot} and C_k decrease with increasing noise level p, as one might expect. Interestingly, there are exceptions to this rule, where some of the quantities increase with p within a certain p-interval. This behaviour is most pronounced in D_2 and C_3 for the five-qubit W state (see Fig. 6.4). Since the depolarizing channel, whose effect it is to add white noise to a state, can be implemented locally, this shows again the known fact that the quantities can increase under local operations. This interpretation was already given in Ref. 133; an explanation of when this phenomenon occurs is still lacking, though.

There is an interesting connection between the quantum information projection and the *maximum likelihood–maximum entropy (MLME)* state reconstruction scheme which has been proposed in Ref. 107 for estimating a state from data obtained by an incomplete tomography: Let Π_i with $\sum_i \Pi_i = \mathbb{1}$ be the projection operators describing the measurements, and let $F = \{f_i\}$ be the vector of relative frequencies of the measurement outcomes. Then the likelihood is

$$L(\rho|F) = \prod_i \left[\operatorname{Tr}(\Pi_i \rho)\right]^{f_i}. \tag{6.135}$$

If the tomography is incomplete, the maximum-likelihood estimate

$$\rho_{\text{ML}} = \operatorname*{argmax}_\rho L(\rho|F) \tag{6.136}$$

will in general not be uniquely defined. The authors of Ref. 107 propose to estimate the state by the maximizer of the von Neumann entropy among all states maximizing the likelihood. To this end, they introduce the function

$$I(\rho; \lambda, F) = \lambda S(\rho) + \log[L(\rho|F)], \tag{6.137}$$

where λ is a positive parameter. Then

$$\rho_{\text{MLME}} = \lim_{\lambda \to 0} \operatorname*{argmax}_\rho I(\rho; \lambda, F) \tag{6.138}$$

is the desired estimate. For the calculation of the maximum for fixed λ they use an iterative procedure based on the likelihood maximization algorithm of Refs. 78, 85, 120.

To see the connection to the information projection, suppose that the measurements Π_i correspond to a complete set of k-party observables. Consider now the case that the f_i in the likelihood are the exact values of the measurement probabilities for a valid state, $f_i = \operatorname{Tr}(\Pi_i \rho)$, without any statistical error. Then the log-likelihood is given by

$$\log[L(\rho'|F)] = -D(F\|\{p_i\}) - S(F) \qquad \text{with} \qquad p_i = \operatorname{Tr}(\Pi_i \rho'). \tag{6.139}$$

As the relative entropy is positive definite (see Section 2.3.1), the log-likelihood is maximized by any state ρ' which gives probabilities $p_i = f_i$. Since we assumed the Π_i to correspond to a complete set of k-party observables, this is the case precisely if ρ' has the same k-party reduced density matrices as the state ρ, which defined the f_i. The MLME estimate, which was defined as the maximizer of the Shannon entropy among all states maximizing the likelihood, is thus nothing but the information projection. The results in Figs. 6.2–6.5 can indeed be obtained with the MLME algorithm of Ref. 107 as well, though the MLME algorithm seems to needs more time to achieve a similar accuracy.

Another interesting connection is provided by the above-mentioned likelihood maximization algorithm, which is remarkably similar to classical iterative scaling.

The function $I(\rho; p, F)$ in Eq. (6.137) is concave in the state ρ. This means that the MLME estimate is found by maximizing a concave function over the convex set of all states. Therefore it is to be expected that, in addition to the algorithm of Ref. 107 and the one introduced in this thesis, other efficient numerical methods can be applied to this problem as well.

6.5 Outlook

The results reported in this chapter raise a number of new questions. Recall that the generalization of Lemma 6.2 in Section 6.1 remains a conjecture. Moreover, a proof would only be a first step towards understanding possible connections between information geometry and multipartite nonlocality in the sense of Svetlichny. Concerning the theory of exponential families of quantum states (Section 6.2), it has already been mentioned that the proofs fail if the information projection does not exist as a full-rank state. Presumably the methods used in Ch. 3 of Ref. 26 for the classical case can be adapted to prove rigorously that the quantum information projection always exists as a state in the compactified exponential family with the expected properties. In this context, it would be interesting to see how the results of Ref. 116 on stabilizer states as ground states of k-party Hamiltonians (see above) fit into the general scheme. In Section 6.3 it was demonstrated that witness operators can be used to detect higher-order interactions. The next step would be to improve the bound on the fidelity until it reaches an experimentally accessible value. It might also be possible to adapt the information-projection algorithm of Section 6.4 for the computation of the maximal overlap of an exponential family with a target state. Finally, at the end of Section 6.4 it has already been pointed out that there are connections to the maximum likelihood–maximum entropy scheme that warrant further investigation.

7 Conclusion

The main subject of this thesis have been different characterizations of multipartite quantum correlations; another subject have been uncertainty relations.

In Chapter 3 the effect of finite measurement statistics on the detection of entanglement was studied with focus on witness operators and Bell inequalities. For either detection method, the statistical significance was defined as the ratio of the violation to the statistical error. Within a naive model, where the statistical error of an observable is estimated by its standard deviation, it was shown that for any fixed target state that is detected by a given entanglement witness the significance can be made arbitrarily large by adding a positive operator to the witness. This is the opposite of witness optimization in the conventional sense, where a positive operator is subtracted from the witness in order to maximize the set of detected states. An error model for experiments with polarization-entangled photons was described, and the underlying assumptions were discussed. In this scenario the significances of the four-qubit Mermin and Ardehali inequality were compared, showing that a Bell inequality with a lower violation can have a higher significance. The dependence of this effect on the fidelity of the state was studied. This analysis motivated an experiment implementing these two Bell inequalities, in which bit-flip noise was introduced on purpose. The experimental results confirmed the prediction that the Mermin inequality has a higher significance than the Ardehali inequality for fidelities above a critical value.

Chapter 4 was concerned with the experimental discrimination of different classes of multipartite entangled states. More precisely, the question was for the best set of observables to show that an experiment which aimed at preparing a certain target state did not result in a state from some class of undesired states. In order to formulate the optimization problem precisely, two measures were defined for the discrimination strength of an observable. The first of these measures is based on the difference of the expectation values of two states and can be interpreted as a noise tolerance. The second measure is given by the relative entropy of the probability distributions for the measurement results; its interpretation is based on the theory of statistical hypothesis testing. These measures were applied to finding optimal families of observables for several examples. Here, the sets of undesired states were given as LU equivalence classes, though this was done only to facilitate the calculations; the discrimination measures themselves are equally applicable to different scenarios, such as SLOCC classes. Results of a four-photon experiment were used to demonstrate the suitability of the measures for assessing experimental data.

The subject of Chapter 5 were entropic uncertainty relations. The question for pairs of measurement bases for which the well-known Maassen-Uffink relation is tight led in a natural way to a generalization of mutual unbiasedness. It was shown that pairs of stabilizer bases have precisely this property. The deeper reason for this fact lies in

the close relation between the geometric properties of pairs of stabilizer bases and the group-theoretic properties of the corresponding pairs of stabilizer groups. For the more special case of graph state bases, some explicit results were obtained. The second part of the chapter was concerned with the many-observable setting. For dichotomic, pairwise anticommuting observables the tuples of expectation values that correspond to valid quantum states could be characterized completely (this was known before), and from this characterization entropic uncertainty relations were derived for a large class of entropy functions. The variance as a characterization of uncertainty also fitted into this scheme, in fact, it was seen to give the strongest possible uncertainty relation for these observables.

In Chapter 6 the theory of exponential families of interaction spaces was applied to different questions concerning the characterization of quantum correlations. In an attempt to understand the properties of this classification scheme better, it was first used in its variant for classical probability distributions by applying it to the measurement probabilities of a three-party Bell experiment. After that, the corresponding theory for quantum states was outlined, presenting previously known results in a coherent fashion. Particular attention was given to the different equivalent characterizations of the information projection. It was emphasized that the calculation of the projection is completely equivalent to the computation of the equilibrium state of a suitably defined thermodynamic ensemble. Also mentioned was the relation to the question (frequently studied in the literature) if graph states can be approximated by ground states of local-interaction Hamiltonians. In analogy to the task of entanglement detection, witness operators to detect higher-order interactions were constructed. An efficient algorithm was developed for the computation of the information projection. This algorithm was employed to compute the distances to the exponential families, which have an interpretation as interaction measures, for certain four-and five-qubit states. The equivalence of the underlying maximization problem to a certain maximum likelihood–maximum entropy state estimation procedure was pointed out.

References

[1] A. Acín, D. Bruß, M. Lewenstein and A. Sanpera, *Classification of mixed three-qubit states*, Phys. Rev. Lett. **87**, 040401 (2001).

[2] S.-i. Amari, *Information geometry on hierarchy of probability distributions*, IEEE Trans. Inf. Theory **47**, 1701–1711 (2001).

[3] S.-i. Amari and H. Nagaoka, *Methods of Information Geometry* (American Mathematical Society, Providence, 2007), translated from the Japanese by D. Harada.

[4] M. Ardehali, *Bell inequalities with a magnitude of violation that grows exponentially with the number of particles*, Phys. Rev. A **46**, 5375–5378 (1992).

[5] A. Aspect, J. Dalibard and G. Roger, *Experimental test of Bell's inequalities using time-varying analyzers*, Phys. Rev. Lett. **49**, 1804–1807 (1982).

[6] P. Badziąg, Č. Brukner, W. Laskowski, T. Paterek and M. Żukowski, *Experimentally friendly geometrical criteria for entanglement*, Phys. Rev. Lett. **100**, 140403 (2008).

[7] L. E. Ballentine, *The statistical interpretation of quantum mechanics*, Rev. Mod. Phys. **42**, 358–381 (1970).

[8] J.-D. Bancal, C. Branciard, N. Gisin and S. Pironio, *Quantifying multipartite nonlocality*, Phys. Rev. Lett. **103**, 090503 (2009).

[9] J.-D. Bancal, N. Brunner, N. Gisin and Y.-C. Liang, *Detecting genuine multipartite quantum nonlocality: A simple approach and generalization to arbitrary dimensions*, Phys. Rev. Lett. **106**, 020405 (2011).

[10] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury and F. Vatan, *A new proof for the existence of mutually unbiased bases*, Algorithmica **34**, 512–528 (2002).

[11] J. T. Barreiro, P. Schindler, O. Gühne, T. Monz, M. Chwalla, C. F. Roos, M. Hennrich and R. Blatt, *Experimental multiparticle entanglement dynamics induced by decoherence*, Nature Physics **6**, 943–946 (2010).

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

[13] J. S. Bell, *On the Einstein-Podolsky-Rosen paradox*, Physics **1**, 195–200 (1964); reprinted in: J. S. Bell, *Speakable and unspeakable in quantum mechanics* (Cambridge University Press, Cambridge, 1987), pp. 14–21.

[14] M. Berta, M. Christandl, R. Colbeck, J. M. Renes and R. Renner, *The uncertainty principle in the presence of quantum memory*, Nature Physics **6**, 659–662 (2010).

[15] I. Bialynicki-Birula and Ł. Rudnicki, *Entropic uncertainty relations in quantum physics*, in: K. D. Sen (ed.), *Statistical Complexity* (Springer, Dordrecht, 2011), pp. 1–34.

[16] S. L. Braunstein, C. M. Caves, R. Jozsa, N. Linden, S. Popescu and R. Schack, *Separability of very noisy mixed states and implications for NMR quantum computing*, Phys. Rev. Lett. **83**, 1054–1057 (1999).

[17] P. Busch, T. Heinonen and P. Lahti, *Heisenberg's uncertainty principle*, Physics Reports **452**, 155–176 (2007).

[18] A. Cabello, A. J. López-Tarrida, P. Moreno and J. R. Portillo, *Entanglement in eight-qubit graph states*, Phys. Lett. A **373**, 2219–2225 (2009); erratum: Phys. Lett. A **374**, 3991 (2010).

[19] A. Chefles, *Quantum state discrimination*, Contemp. Phys. **41**, 401–424 (2000).

[20] Y.-A. Chen, A.-N. Zhang, Z. Zhao, X.-Q. Zhou and J.-W. Pan, *Experimental quantum error rejection for quantum communication*, Phys. Rev. Lett. **96**, 220504 (2006).

[21] B. S. Cirel'son, *Quantum generalizations of Bell's inequality*, Lett. Math. Phys. **4**, 93–100 (1980).

[22] J. F. Clauser and M. A. Horne, *Experimental consequences of objective local theories*, Phys. Rev. D **10**, 526–535 (1974).

[23] J. F. Clauser, M. A. Horne, A. Shimony and R. A. Holt, *Proposed experiment to test local hidden-variable theories*, Phys. Rev. Lett. **23**, 880–884 (1969); erratum: Phys. Rev. Lett. **24**, 549 (1970).

[24] D. Collins, N. Gisin, S. Popescu, D. Roberts and V. Scarani, *Bell-type inequalities to detect true n-body nonseparability*, Phys. Rev. Lett. **88**, 170405 (2002).

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, 2006), second edn.

[26] I. Csiszár and P. C. Shields, *Information theory and statistics: A tutorial*, Found. and Trends in Communications and Inf. Theory **1**, 417–528 (2004).

[27] W. van Dam, R. D. Gill and P. D. Grünwald, *The statistical strength of nonlocality proofs*, IEEE Trans. Inf. Theory **51**, 2812–2835 (2005); preprint containing additional material: arXiv:quant-ph/0307125.

[28] I. B. Damgård, S. Fehr, R. Renner, L. Salvail and C. Schaffner, *A tight high-order entropic quantum uncertainty relation with applications*, in: A. Menezes (ed.), *Advances in Cryptology – CRYPTO 2007*, vol. 4622 of *Lecture Notes in Computer Science* (Springer, Berlin, 2007), pp. 360–378.

[29] J. Dehaene and B. De Moor, *Clifford group, stabilizer states, and linear and quadratic operations over GF(2)*, Phys. Rev. A **68**, 042318 (2003).

[30] D. Deutsch, *Uncertainty in quantum measurements*, Phys. Rev. Lett. **50**, 631–633 (1983).

[31] W. Dür, G. Vidal and J. I. Cirac, *Three qubits can be entangled in two inequivalent ways*, Phys. Rev. A **62**, 062314 (2000).

[32] A. Einstein, B. Podolsky and N. Rosen, *Can quantum-mechanical description of physical reality be considered complete?*, Phys. Rev. **47**, 777–780 (1935).

[33] P. Facchi, G. Florio, S. Pascazio and F. V. Pepe, *Greenberger-Horne-Zeilinger states and few-body Hamiltonians*, Phys. Rev. Lett. **107**, 260502 (2011).

[34] A. Fine, *Hidden variables, joint probability, and the Bell inequalities*, Phys. Rev. Lett. **48**, 291–295 (1982); comment by A. Garg and N. D. Mermin: Phys. Rev. Lett. **49**, 242 (1982); rejoinder: Phys. Rev. Lett. **49**, 243 (1982).

[35] T. Galla and O. Gühne, *Complexity measures, emergence, and multiparticle correlations* (2011), arXiv:1107.1180.

[36] S. Gharibian, *Strong NP-hardness of the quantum separability problem*, Quant. Inf. Comp. **10**, 343–360 (2010).

[37] G. Ghirardi, L. Marinatto and R. Romano, *An optimal entropic uncertainty relation in a two-dimensional Hilbert space*, Phys. Lett. A **317**, 32–36 (2003).

[38] O. Gittsovich, P. Hyllus and O. Gühne, *Multiparticle covariance matrices and the impossibility of detecting graph-state entanglement with two-particle correlations*, Phys. Rev. A **82**, 032306 (2010).

[39] D. Gottesman, *Class of quantum error-correcting codes saturating the quantum Hamming bound*, Phys. Rev. A **54**, 1862–1868 (1996).

[40] D. Gottesman, *Stabilizer Codes and Quantum Error Correction*, Ph. D. thesis, California Institute of Technology (1997), arXiv:quant-ph/9705052.

[41] M. Grassl, A. Klappenecker and M. Rötteler, *Graphs, quadratic forms, and quantum codes*, in: *Proc. 2002 IEEE Int. Symp. on Information Theory (ISIT 2002)*, p. 45, arXiv:quant-ph/0703112.

[42] O. Gühne, *Characterizing entanglement via uncertainty relations*, Phys. Rev. Lett. **92**, 117903 (2004).

[43] O. Gühne and M. Lewenstein, *Entropic uncertainty relations and entanglement*, Phys. Rev. A **70**, 022316 (2004).

[44] O. Gühne and G. Tóth, *Entanglement detection*, Physics Reports **474**, 1–75 (2009).

[45] L. Gurvits, *Classical deterministic complexity of Edmonds' problem and quantum entanglement*, in: *Proc. 35th Ann. ACM Symp. on Theory of Computing (STOC 2003)* (ACM, New York, 2003), pp. 10–19.

[46] P. Hall, *On representatives of subsets*, J. London Math. Soc. **10**, 26–30 (1935).

[47] H. L. Haselgrove, M. A. Nielsen and T. J. Osborne, *Quantum states far from the energy eigenstates of any local Hamiltonian*, Phys. Rev. Lett. **91**, 210401 (2003).

[48] H. L. Haselgrove, M. A. Nielsen and T. J. Osborne, *Entanglement, correlations, and the energy gap in many-body quantum systems*, Phys. Rev. A **69**, 032303 (2004).

[49] J. Havrda and F. Charvát, *Quantification method of classification processes. Concept of stuctural a-entropy*, Kybernetika **3**, 30–35 (1967).

[50] M. Hein, W. Dür, J. Eisert, R. Raussendorf, M. Van den Nest and H. J. Briegel, *Entanglement in graph states and its applications*, in: G. Casati, D. L. Shepelyansky, P. Zoller and G. Benenti (eds.), *Quantum Computers, Algorithms and Chaos*, no. 162 in *Proceedings of the International School of Physics Enrico Fermi* (IOS Press, Amsterdam, 2006), p. 115, arXiv:quant-ph/0602096.

[51] M. Hein, J. Eisert and H. J. Briegel, *Multiparty entanglement in graph states*, Phys. Rev. A **69**, 062311 (2004).

[52] W. Heisenberg, *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, Z. Phys. **43**, 172–198 (1927);
for an English translation see: J. A. Wheeler and W. H. Zurek (eds.), *Quantum Theory and Measurement* (Princeton University Press, Princeton, 1983), pp. 62–84.

[53] H. F. Hofmann and S. Takeuchi, *Violation of local uncertainty relations as a signature of entanglement*, Phys. Rev. A **68**, 032103 (2003).

[54] M. Horodecki, P. Horodecki and R. Horodecki, *Separability of mixed states: necessary and sufficient conditions*, Phys. Lett. A **223**, 1–8 (1996).

[55] R. Horodecki, P. Horodecki, M. Horodecki and K. Horodecki, *Quantum entanglement*, Rev. Mod. Phys. **81**, 865–942 (2009).

[56] P. Hyllus, O. Gühne and A. Smerzi, *Not all pure entangled states are useful for sub-shot-noise interferometry*, Phys. Rev. A **82**, 012337 (2010).

[57] L. M. Ioannou, *Computational complexity of the quantum separability problem*, Quant. Inf. Comp. **7**, 335–370 (2007).

[58] D. F. V. James, P. G. Kwiat, W. J. Munro and A. G. White, *Measurement of qubits*, Phys. Rev. A **64**, 052312 (2001).

[59] Z. Ji, J. Chen, Z. Wei and M. Ying, *The LU-LC conjecture is false*, Quant. Inf. Comp. **10**, 97–108 (2010).

[60] N. S. Jones and N. Linden, *Parts of quantum states*, Phys. Rev. A **71**, 012324 (2005).

[61] B. Jungnitsch, *Criteria for genuine multiparticle quantum correlations*, Ph. D. thesis, Ludwig-Franzens-Universität Innsbruck (2012), in preparation.

[62] T. Kahle, E. Olbrich, J. Jost and N. Ay, *Complexity measures from interaction structures*, Phys. Rev. E **79**, 026201 (2009).

[63] T. Kahle, W. Wenzel and N. Ay, *Hierarchical models, marginal polytopes, and linear codes*, Kybernetika **45**, 189–207 (2009).

[64] E. H. Kennard, *Zur Quantenmechanik einfacher Bewegungstypen*, Z. Phys. **44**, 326–352 (1927).

[65] N. Kiesel, C. Schmid, U. Weber, G. Tóth, O. Gühne, R. Ursin and H. Weinfurter, *Experimental analysis of a four-qubit photon cluster state*, Phys. Rev. Lett. **95**, 210502 (2005).

[66] M. Koashi, *Unconditional security of quantum key distribution and the uncertainty principle*, J. Phys.: Conf. Ser. **36**, 98–102 (2006).

[67] M. Koashi, *Simple security proof of quantum key distribution based on complementarity*, New J. Phys. **11**, 045018 (2009).

[68] B. Kraus, *Local unitary equivalence of multipartite pure states*, Phys. Rev. Lett. **104**, 020504 (2010).

[69] K. Kraus, *Complementary observables and uncertainty relations*, Phys. Rev. D **35**, 3070–3075 (1987).

[70] M. Krishna and K. R. Parthasarathy, *An entropic uncertainty principle for quantum measurements*, Sankhyā: Indian J. Statistics Ser. A **64**, 842–851 (2002).

[71] P. Kurzyński, T. Paterek, R. Ramanathan, W. Laskowski and D. Kaszlikowski, *Correlation complementarity yields Bell monogamy relations*, Phys. Rev. Lett. **106**, 180402 (2011).

[72] P. G. Kwiat, K. Mattle, H. Weinfurter, A. Zeilinger, A. V. Sergienko and Y. Shih, *New high-intensity source of polarization-entangled photon pairs*, Phys. Rev. Lett. **75**, 4337–4341 (1995).

[73] J. Lavoie, R. Kaltenbaek and K. J. Resch, *Experimental violation of Svetlichny's inequality*, New J. Phys. **11**, 073051 (2009).

[74] M. Lewenstein, B. Kraus, J. I. Cirac and P. Horodecki, *Optimization of entanglement witnesses*, Phys. Rev. A **62**, 052310 (2000).

[75] C.-F. Li, J.-S. Xu, X.-Y. Xu, K. Li and G.-C. Guo, *Experimental investigation of the entanglement-assisted entropic uncertainty principle*, Nature Physics **7**, 752–756 (2011).

[76] N. Linden, S. Popescu and W. K. Wootters, *Almost every pure state of three qubits is completely determined by its two-particle reduced density matrices*, Phys. Rev. Lett. **89**, 207901 (2002).

[77] N. Linden and W. K. Wootters, *The parts determine the whole in a generic pure quantum state*, Phys. Rev. Lett. **89**, 277906 (2002).

[78] A. I. Lvovsky, *Iterative maximum-likelihood reconstruction in quantum homodyne tomography*, J. Opt. B: Quantum Semiclass. Opt. **6**, S556 (2004).

[79] H. Maassen, *A discrete entropic uncertainty relation*, in: L. Accardi and W. von Waldenfels (eds.), *Quantum Probability and Applications V*, vol. 1442 of *Lecture Notes in Mathematics* (Springer, Berlin, 1990), pp. 263–266.

[80] H. Maassen and J. B. M. Uffink, *Generalized entropic uncertainty relations*, Phys. Rev. Lett. **60**, 1103–1106 (1988).

[81] P. Mandayam, S. Wehner and N. Balachandran, *A transform of complementary aspects with applications to entropic uncertainty relations*, J. Math. Phys. **51**, 082201 (2010).

[82] D. Markham, A. Miyake and S. Virmani, *Entanglement and local information access for graph states*, New J. Phys. **9**, 194 (2007).

[83] N. D. Mermin, *Extreme quantum entanglement in a superposition of macroscopically distinct states*, Phys. Rev. Lett. **65**, 1838–1840 (1990).

[84] P. Mitchell, S. Popescu and D. Roberts, *Conditions for the confirmation of three-particle nonlocality*, Phys. Rev. A **70**, 060101 (2004).

[85] G. Molina-Terriza, A. Vaziri, J. Řeháček, Z. Hradil and A. Zeilinger, *Triggered qutrits for quantum communication protocols*, Phys. Rev. Lett. **92**, 167903 (2004).

[86] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[87] M. H. Partovi, *Majorization formulation of uncertainty in quantum mechanics*, Phys. Rev. A **84**, 052117 (2011).

[88] A. Peres, *Separability criterion for density matrices*, Phys. Rev. Lett. **77**, 1413–1415 (1996).

[89] A. Peres, *Bayesian analysis of Bell inequalities*, Fortschr. Phys. **48**, 531–535 (2000).

[90] S. Popescu and D. Rohrlich, *Quantum nonlocality as an axiom*, Found. Phys. **24**, 379–385 (1994).

[91] R. Prevedel, D. R. Hamel, R. Colbeck, K. Fisher and K. J. Resch, *Experimental investigation of the uncertainty principle in the presence of quantum memory and its application to witnessing entanglement*, Nature Physics **7**, 757–761 (2011).

[92] A. Rényi, *On measures of entropy and information*, in: J. Neyman (ed.), *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability, Vol. 1: Contributions to the Theory of Statistics* (University of California Press, Berkeley, 1960), pp. 547–561.

[93] A. Rényi, *Probability Theory* (North-Holland, Amsterdam, 1970).

[94] H. P. Robertson, *The uncertainty principle*, Phys. Rev. **34**, 163–164 (1929).

[95] H. P. Robertson, *An indeterminacy relation for several observables and its classical interpretation*, Phys. Rev. **46**, 794–801 (1934).

[96] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe and D. J. Wineland, *Experimental violation of a Bell's inequality with efficient detection*, Nature **409**, 791–794 (2001).

[97] V. Scarani, A. Acín, E. Schenck and M. Aspelmeyer, *Nonlocality of cluster states of qubits*, Phys. Rev. A **71**, 042325 (2005).

[98] D. Schlingemann, *Stabilizer codes can be realized as graph codes*, Quant. Inf. Comp. **2**, 307–323 (2002).

[99] C. Schmid, N. Kiesel, W. Laskowski, W. Wieczorek, M. Żukowski and H. Weinfurter, *Discriminating multipartite entangled states*, Phys. Rev. Lett. **100**, 200407 (2008).

[100] E. Schrödinger, *Die gegenwärtige Situation in der Quantenmechanik*, Die Naturwissenschaften **23**, 807–812, 823–828, 844–849 (1935);
for an English translation see: J. D. Trimmer, *The present situation in quantum mechanics: A translation of Schrödinger's "cat paradox" paper*, Proc. Am. Phil. Soc. **124**, 323–338 (1980).

[101] E. Schrödinger, *Discussion of probability relations between separated systems*, Math. Proc. Cambridge Phil. Soc. **31**, 555–563 (1935).

[102] E. Schrödinger, *Probability relations between separated systems*, Math. Proc. Cambridge Phil. Soc. **32**, 446–452 (1936).

[103] M. Seevinck and G. Svetlichny, *Bell-type inequalities for partial separability in N-particle systems and quantum mechanical violations*, Phys. Rev. Lett. **89**, 060401 (2002).

[104] L. Steiner and T. Kahle, *Computing information projections iteratively with CIPI*, github.com/tom111/cipi (accessed 27 Oct. 2011).

[105] G. Svetlichny, *Distinguishing three-body from two-body nonseparability by a Bell-type inequality*, Phys. Rev. D **35**, 3066–3069 (1987).

[106] Y. Tanaka, D. Markham and M. Murao, *Local encoding of classical information onto quantum states*, J. Mod. Opt. **54**, 2259–2273 (2007).

[107] Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček and Z. Hradil, *Quantum-state reconstruction by maximizing likelihood and entropy*, Phys. Rev. Lett. **107**, 020404 (2011).

[108] B. M. Terhal, *Bell inequalities and the separability criterion*, Phys. Lett. A **271**, 319–326 (2000).

[109] G. Tóth and O. Gühne, *Entanglement detection in the stabilizer formalism*, Phys. Rev. A **72**, 022340 (2005).

[110] D. A. Trifonov, *Generalizations of Heisenberg uncertainty relation*, Eur. Phys. J. B **29**, 349–353 (2002).

[111] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, J. Stat. Phys. **52**, 479–487 (1988).

[112] M. Van den Nest, *Local equivalence of stabilizer states and codes*, Ph. D. thesis, Katholieke Universiteit Leuven (2005), ftp.esat.kuleuven.be/pub/SISTA/mvandenn/reports/05-99.ps

[113] M. Van den Nest, *Classical simulation of quantum computation, the Gottesman-Knill theorem, and slightly beyond*, Quant. Inf. Comp. **10**, 258–271 (2010).

[114] M. Van den Nest, J. Dehaene and B. De Moor, *Graphical description of the action of local Clifford transformations on graph states*, Phys. Rev. A **69**, 022316 (2004).

[115] M. Van den Nest, J. Dehaene and B. De Moor, *Local equivalence of stabilizer states*, in: *Proc. 16th Int. Symp. on Mathematical Theory of Networks and Systems (MTNS)* (Leuven, 2004), ftp.esat.kuleuven.ac.be/pub/SISTA/mvandenn/reports/04-210.ps

[116] M. Van den Nest, K. Luttmer, W. Dür and H. J. Briegel, *Graph states as ground states of many-body spin-1/2 Hamiltonians*, Phys. Rev. A **77**, 012301 (2008).

[117] M. Van den Nest, A. Miyake, W. Dür and H. J. Briegel, *Universal resources for measurement-based quantum computation*, Phys. Rev. Lett. **97**, 150504 (2006).

[118] V. Vedral, *The role of relative entropy in quantum information theory*, Rev. Mod. Phys. **74**, 197–234 (2002).

[119] J. I. de Vicente and J. Sánchez-Ruiz, *Improved bounds on entropic uncertainty relations*, Phys. Rev. A **77**, 042110 (2008).

[120] J. Řeháček, Z. Hradil, E. Knill and A. I. Lvovsky, *Diluted maximum-likelihood algorithm for quantum tomography*, Phys. Rev. A **75**, 042108 (2007).

[121] S. N. Walck and D. W. Lyons, *Only n-qubit Greenberger-Horne-Zeilinger states are undetermined by their reduced density matrices*, Phys. Rev. Lett. **100**, 050501 (2008).

[122] S. N. Walck and D. W. Lyons, *Only n-qubit Greenberger-Horne-Zeilinger states contain n-partite information*, Phys. Rev. A **79**, 032326 (2009).

[123] S. Wehner and A. Winter, *Higher entropic uncertainty relations for anti-commuting observables*, J. Math. Phys. **49**, 062105 (2008).

[124] S. Wehner and A. Winter, *Entropic uncertainty relations – a survey*, New J. Phys. **12**, 025009 (2010).

[125] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter and A. Zeilinger, *Violation of Bell's inequality under strict Einstein locality conditions*, Phys. Rev. Lett. **81**, 5039–5043 (1998).

[126] R. F. Werner, *Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model*, Phys. Rev. A **40**, 4277–4281 (1989).

[127] R. F. Werner, *The uncertainty relation for joint measurement of position and momentum*, Quant. Inf. Comp. **4**, 546–562 (2004).

[128] R. F. Werner and M. M. Wolf, *Bell inequalities and entanglement* Quant. Inf. Comp. **1**, 1–25 (2001).

[129] W. Wieczorek, C. Schmid, N. Kiesel, R. Pohlner, O. Gühne and H. Weinfurter, *Experimental observation of an entire family of four-photon entangled states*, Phys. Rev. Lett. **101**, 010503 (2008).

[130] R. M. Wilcox, *Exponential operators and parameter differentiation in quantum physics*, J. Math. Phys. **8**, 962–982 (1967).

[131] W. K. Wootters and B. D. Fields, *Optimal state-determination by mutually unbiased measurements*, Ann. Phys. **191**, 363–381 (1989).

[132] D. L. Zhou, *Irreducible multiparty correlations in quantum states without maximal rank*, Phys. Rev. Lett. **101**, 180505 (2008).

[133] D. L. Zhou, *An efficient numerical algorithm on irreducible multiparty correlations* (2009), arXiv:0909.3700.

[134] D. L. Zhou, *Irreducible multiparty correlations can be created by local operations*, Phys. Rev. A **80**, 022113 (2009).

List of publications

Chapters 3, 4 and 5 are based on Publications B, C and D, respectively. Publication A is unrelated to this thesis.

[A] S. Niekamp, T. Wirth and H. Frahm, *The XXZ model with anti-periodic twisted boundary conditions*, J. Phys. A: Math. Theor. **42**, 195008 (2009), arXiv:0902.1079.

[B] B. Jungnitsch, S. Niekamp, M. Kleinmann, O. Gühne, H. Lu, W.-B. Gao, Y.-A. Chen, Z.-B. Chen and J.-W. Pan, *Increasing the statistical significance of entanglement detection in experiments*, Phys. Rev. Lett. **104**, 210401 (2010), arXiv:0912.0645.

[C] S. Niekamp, M. Kleinmann and O. Gühne, *Discrimination strategies for inequivalent classes of multipartite entangled states*, Phys. Rev. A **82**, 022322 (2010), arXiv:1006.1313.

[D] S. Niekamp, M. Kleinmann and O. Gühne, *Entropic uncertainty relations and the stabilizer formalism*, J. Math. Phys. **53**, 012202 (2012), arXiv:1103.2316.

Acknowledgements

Many people have helped me in one way or another to finish this work, and I would like to thank all of them.

First of all I thank Prof. Dr. Otfried Gühne for accepting me as his Ph. D. student, for his constant support and his unshakeable optimism.

I am particularly indebted to Dr. Matthias Kleinmann for the collaboration on several projects, his interest in my work and his willingness to answer all my questions.

I thank Bastian Jungnitsch for working with me on our first project and also Prof. Dr. Jian-Wei Pan's group, who carried out the experiment. Similarly, I thank Dr. Tobias Galla for the collaboration on our most recent project and the invitations to Manchester.

Although this thesis was completed in Siegen, most of the work was done in Innsbruck. I am grateful to Prof. Dr. Hans J. Briegel and his extended group for the enjoyable time. In particular, I thank Otfried Gühne's group in Innsbruck and Siegen: in addition to those already mentioned, Dr. Oleg Gittsovich, Dr. Tobias Moroder, Dr. Mazhar Ali, Martin Hofmann, Marcel Bergmann and Jochen Szangolies.

I thank Prof. Dr. Dagmar Bruß for kindly agreeing to referee this thesis.

Finally, I thank my parents for supporting me during all those years as a student.