

Characterizing classical and quantum systems from marginal correlations

DISSERTATION
zur Erlangung des Grades eines Doktors
der Naturwissenschaften

vorgelegt von
MSc. Nikolai Miklin

eingereicht bei der Naturwissenschaftlich-Technischen Fakultät
der Universität Siegen
Siegen, 03. Juli, 2017

Gutachter:

- Prof. Dr. Otfried Gühne
- Prof. Dr. David Gross

Datum der mündlichen Prüfung: 18.08.2017

Prüfer:

- Prof. Dr. Otfried Gühne (Vorsitz der Prüfungskommission)
- Prof. Dr. David Gross
- Prof. Dr. Thomas Mannel
- Prof. Dr. Markus Risse

Abstract

This thesis studies different problems in quantum information theory and the foundations of quantum mechanics. These include the quantum marginal problem, the problem of causal inference in quantum mechanics, and the problem of indefinite causal order in quantum processes.

We start by considering an instance of the quantum marginal problem in which our goal is to detect genuine multiparticle entanglement from the marginal information, i.e. correlations in the subsystems. In simple words, genuine multiparticle entanglement means that all particles are entangled with each other. Moreover, we consider an exotic case where the marginals themselves are separable, i.e. do not manifest entanglement if considered separately. Our results show that this phenomenon, which we call emergence of multiparticle entanglement, occurs frequently and for an arbitrary number of particles. In particular, we present a systematic method to look for such states and present various examples of systems up to 6 qubits (two-level systems). Interestingly, already for four qubits there exist a pure state with this properties which suggests that this phenomenon can be observed in the experiment.

In the subsequent part of the thesis we define and study a particular class of genuine entangled states, called hypergraph states, in systems of qudits (d -level systems). This class of states is a generalization of graph states, which are used in measurement-based quantum computing and error-correcting codes. Hypergraph states can be obtained by applying certain sequence of entangling gates, associated with hyperedges, on systems of qudits, associated with vertices. In this thesis we provide a detailed analysis of equivalence of tripartite hypergraph states in dimension 3 and 4 under local operations.

Then we pass on to the problem of explaining correlations observed in experiment by classical causal models. A particular example of a causal model is a local hidden variable model of Bell's test. Cause-effect relations, or causal links, in causal models are given by the underlying causal structures, which often can be represented by graphs. Given a causal structure one can derive constraints for correlations to be compatible with this structure, which in the case of Bell's theorem are the famous Bell inequalities. Alternatively, given experimental data, the task would be to determine the underlined causal model, which is a problem of causal inference. In some experiments of causal inference the correlations among all variables cannot be accessed or are not collected. In this case one faces a type of the marginal problem where one has to judge about possible underlined causal structures from marginal data. Clearly, the success of the causal inference in this case depends strongly on the configuration

of accessible marginals, which is known as the marginal scenario. In this thesis we provide a general theory connecting marginal scenario and possible causal structures. We derive a necessary condition on causal structures to be distinguishable from a given marginal scenario. Among others, this result can help us to find new interesting scenarios for nonlocality tests.

Finally, we discuss the problem of indefinite causal order in quantum mechanics. Causal order puts restrictions on causal relations in two systems of random variables generated by two different events. In particular, it restricts these causal links to be directed in the same way, from one event to the other. An example of causal order is a space-time manifold. Recently, it has been realized that physical theories do not necessarily have to comply with the idea of a definite causal order. For example, one can imagine a theory where the causal order is a dynamical element of this theory and can be in a sort of "quantum superposition". In this thesis we derive inequalities, similar to those of Bell, but for testing indefiniteness of causal order. In particular, inequalities are derived for information-theoretic quantities allowing for testing information flow in processes with indefinite causal order.

Zusammenfassung

Diese Arbeit untersucht verschiedene Probleme in der Quanteninformationstheorie und die Grundlagen der Quantenmechanik. Dies beinhaltet das Quantenmarginalienproblem, das Problem der Kausalfolgerung in der Quantenmechanik und das Problem der unbestimmten Kausalordnung in Quantenprozessen.

Wir beginnen mit einem Beispiel aus dem Bereich des Quantenmarginalienproblems. Hierbei ist es unser Ziel, eine echte Mehrteilchen-Verschränkung aus der Marginalieninformation, d.h. aus Korrelationen in den Teilsystemen, zu erkennen. Echte Mehrteilchen-Verschränkung bedeutet, dass alle Teilchen miteinander verschränkt sind. Darüber hinaus betrachten wir einen exotischen Fall, bei dem die Marginale selbst separabel sind, d.h. keine Verschränkung beinhalten. Unsere Ergebnisse zeigen, dass dieses Phänomen, das wir als Auftreten der Mehrteilchen-Verschränkung bezeichnen, häufig und für eine beliebige Anzahl von Teilchen auftritt. Insbesondere stellen wir eine systematische Methode vor, um solche Zustände zu suchen. Hierzu geben wir verschiedene Beispiele von Systemen, die aus bis zu sechs Qubits (zweistufige Systeme) bestehen. Interessanterweise gibt es bereits für vier Qubits einen reinen Zustand mit diesen Eigenschaften, was darauf hindeutet, dass dieses Phänomen im Experiment beobachtet werden kann.

Im darauffolgenden Teil der Arbeit definieren und analysieren wir eine bestimmte Klasse von echt verschränkten Zuständen, sogenannte Hypergraph-Zustände, in Systemen von Qudits (d-Level-Systeme). Diese Klasse von Zuständen ist eine Verallgemeinerung von Graphenzuständen, die in messbasierten Quantencomputern und fehlerkorrigierenden Codes verwendet werden. Hypergraph-Zustände können durch Anwenden einer bestimmten Abfolge von Verschränkungsgattern, die mit Hyperkanten assoziiert sind, auf Systeme von Qudits, die mit Knoten assoziiert sind, erzeugt werden. In dieser Arbeit geben wir eine detaillierte Analyse der Äquivalenz von dreiteiligen Hypergraph-Zuständen unter lokalen Operationen in Dimension drei und vier.

Danach befassen wir uns mit dem Problem der Erklärung von Korrelationen, die im Experiment durch klassische Kausalmodelle beobachtet wurden. Ein besonderes Beispiel für ein Kausalmodell ist ein lokales verstecktes Variablenmodell im Bell Test. Ursache-Wirkungs-Beziehungen, oder Kausal-Verbindungen, in Kausal-Modellen werden durch die zugrunde liegenden Kausalstrukturen gegeben, die oft durch Graphen dargestellt werden können. Für eine gegebene kausale Struktur kann man Einschränkungen für Korrelationen ableiten, um mit dieser Struktur kompatibel zu sein, was im Fall von Bell's Theorem die berühmten Bell-Ungleichungen sind. Alternativ, bei gegebenen experimentellen Daten, wäre die Aufgabe, das zugrundeliegende Kausalmodell

zu bestimmen, was ein Problem der kausalen Inferenz ist. In einigen Experimenten der kausalen Inferenz kann auf die Korrelationen zwischen allen Variablen nicht zugegriffen werden oder sie werden nicht gesammelt. In diesem Fall steht man einer Art des Marginalienproblems gegenüber, wo man über mögliche zugrundeliegende Kausalstrukturen aus Marginaliendaten folgern muss. Offensichtlich hängt der Erfolg der kausalen Inferenz in diesem Fall stark von der Konfiguration der zugänglichen Marginalien ab, die als das Marginalienszenario bekannt ist. In dieser Arbeit geben wir eine allgemeine Theorie, die das marginale Szenario und mögliche Kausalstrukturen verbindet. Wir erhalten eine notwendige Bedingung für kausale Strukturen, um diese von einem gegebenen Marginalienszenario zu unterscheiden. Dieses Ergebnis kann uns dabei helfen, neue interessante Szenarien für Nichtlokalitätsprüfungen zu finden.

Schließlich diskutieren wir das Problem der unbestimmten Kausalordnung in der Quantenmechanik. Ein Beispiel für Kausalordnung ist ein Raum-Zeit-Mannigfaltigkeit. Neure Erkenntnisse zeigen, dass physikalische Theorien nicht unbedingt mit der Idee einer bestimmten Kausalordnung übereinstimmen müssen. Zum Beispiel kann man sich eine Theorie vorstellen, in der die Kausalordnung ein dynamisches Element dieser Theorie ist und in einer Art Quantenüberlagerung sein kann. In dieser Arbeit leiten wir Ungleichheiten, ähnlich denen von Bell, aber für die Prüfung der Unbestimmtheit von Kausalordnung her. Insbesondere werden Ungleichungen für informationstheoretische Größen abgeleitet, die es ermöglichen, den Informationsfluss in Prozessen mit unbestimmter Kausalordnung zu testen.

Contents

Symbols	ix
1 Introduction	1
2 Preliminaries	4
2.1 Entanglement	4
2.1.1 Bipartite entanglement	4
2.1.2 Multipartite entanglement	9
2.1.3 SLOCC classification	11
2.1.4 Graph and hypergraph states	12
2.2 Nonlocality	15
2.2.1 Bell's theorem	15
2.2.2 Bell inequalities	17
2.2.3 Correlation polytopes	18
2.3 Causal models	19
2.3.1 Directed and undirected graphs	19
2.3.2 Cause-effect relations	20
2.3.3 Bayesian networks	22
2.3.4 Markov random fields	23
2.3.5 Hidden variables	24
2.3.6 Causal models in Bell's test	25
2.3.7 Indefinite causal order	25
2.4 Marginal problem	28
2.5 Entropic inequalities	29
2.5.1 Entropy cone	30
2.5.2 Probability structures	31
2.5.3 The entropic characterization of Bayesian networks	32
2.5.4 The entropic characterization of counterfactuals	33
2.6 Semidefinite Programming	35
3 Multipartite entanglement as an emergent phenomenon	38
3.1 Emergence of entanglement	38
3.2 Statement of the problem	40
3.3 Construction method	41
3.4 Results	43
3.4.1 Three qubits	43
3.4.2 Four and five qubits	44

3.4.3	Generalization to more particles	45
3.4.4	Uniqueness of the global state	46
3.5	Extensions of the problem	46
3.5.1	Separability of the higher-order marginals	46
3.5.2	Proving entanglement from a subset of marginals	47
3.5.3	Higher-dimensional systems	49
3.5.4	No localizable entanglement in the marginals	50
3.6	Conclusions	51
4	Qudit hypergraph states	53
4.1	Introduction	53
4.2	Definition of qudit hypergraph states	54
4.3	SLOCC and LU classes of hypergraphs	55
4.3.1	Elementary hypergraphs	56
4.3.2	Tools for SLOCC classification	57
4.3.3	Tools for LU classification	60
4.4	SLOCC classification of tripartite hypergraph states in dimensions 3 and 4.	60
4.4.1	Classification of $3 \otimes 3 \otimes 3$	61
4.4.2	Classification of $4 \otimes 4 \otimes 4$	62
4.5	Conclusions	67
5	Indistinguishability of causal relations from limited marginals	68
5.0.1	Properties of graphs and hypergraphs	68
5.1	Adhesivity and independence constraints associated with a marginal scenario	71
5.1.1	Adhesivity of probabilities	72
5.1.2	Marginal scenarios admitting a global extension	73
5.1.3	Maximal set of independence conditions associated with a marginal scenario	74
5.2	Optimal characterization of the marginal scenario for probabilities and entropies	75
5.2.1	Triangulation	76
5.2.2	Probabilities	76
5.2.3	Entropies	77
5.2.4	Outer approximations of the entropy cone	78
5.3	Indistinguishability of causal structures	79
5.4	Examples and computational results	82
5.4.1	Inclusions in Obs. 5.4	82
5.4.2	Three cases in Theorem 5.5	84
5.5	Conclusions	87
6	The entropic approach to causal correlations	89
6.1	Introduction	90
6.2	Bipartite entropic causal inequalities	90
6.2.1	Characterization based on causal Bayesian networks	91
6.2.2	Characterization based on counterfactual variables	100
6.3	Multipartite entropic causal inequalities	104
6.3.1	Causal Bayesian network method	105

6.3.2 Counterfactual variable method	107
6.4 Information bounds in causal games	108
6.5 Discussion	110
Conclusions	112
A An Appendix to Chapter 6	114
Bibliography	119
List of publications	132

Symbols

\mathcal{H}	Hilbert space
\otimes	tensor product
$ \psi\rangle$	vector of a pure quantum state
ρ	density matrix of a quantum state
$\mathbb{1}$	identity operator
tr	trace operator
$(\)^T$	transpose operator
$(\)^{T_A}$	partial transpose operator (with respect to system A)
A, B, Λ, \dots	random variables (uppercase letters)
a, b, λ, \dots	outcomes of random variables (lowercase letters)
$P(a)$	probability of $A = a$
$P(A)$	probability distribution of A
$\langle A \rangle$	expectation value of A
$X \perp Y$	independence of variables X and Y
$X \perp Y Z$	conditional independence
$\Gamma_{[n]}^*$	entropy cone for the set of variables $[n]$
$\Gamma_{[n]}$	Shannon cone
$\Pi_{\mathcal{M}}(\Gamma)$	projection of cone Γ on \mathcal{M}
$\text{conv}(\Gamma_1, \Gamma_2)$	convex hull of cones Γ_1 and Γ_2
$A \prec B$	causal order in which A is in the causal past of B
$\text{gcd}(n, m)$	greatest common divisor of n and m

Chapter 1

Introduction

A central property of complex quantum systems is the possibility of being entangled, meaning in the simplest case that the wave function of the system does not factorize. This phenomenon was first described by Erwin Schrödinger in 1935 [1] and it has been under extensive investigation ever since. In the early days of quantum mechanics entanglement created a lot of confusion among scientists, including Einstein, Podolsky, and Rosen (EPR), who criticized quantum theory in their famous argument of 1935 [2]. In 1964 John Bell showed in his famous theorem [3] that the assumptions of EPR argument are inconsistent with predictions of quantum mechanics and thus should be rejected. In 1971 the Bell's theorem was tested experimentally by Franson and in 1981 by Alan Aspect [4], however, it was not until 2015 when this argument was indisputably proven experimentally [5-7].

One reason why quantum mechanics is so counter-intuitive is because it contradicts our usual understanding of causality. For example, if one of the entangled particles is being measured, it changes the quantum state of the whole system which might seem to affect the outcomes of measurements on the other particles. This phenomenon was described by Einstein as "spooky action at a distance", because it looked as if the particle on one side of the experiment can signal to the other particle faster than the speed of light. Indeed, measurements on a complex quantum system cannot be explained using classical ideas of cause and effect, which was systematically investigated by Wood and Spekkens [8] on the example of Bell's theorem [3].

Apart from contradicting classical models, entanglement is the central resource for quantum information tasks. It can be used for lossless transfer of quantum information over large distances via protocols of quantum teleportation [9] or remote state preparation [10]. Entanglement allows for establishing of a secret key among trusted parties where security is provided by a fundamental property of entanglement, called

monogamy [11]. Moreover, entanglement provides speed up in computations [12–15] and gives advantages in numerous communication tasks, including dense coding [16], random access coding [17], etc.

The resource, used in most of these applications, is the pair of entangled particles. This is due to the fact that bipartite entanglement is well understood and is easier to achieve experimentally. Entanglement of many-body systems, on the other hand, can reveal many surprising phenomena, the use of which is yet to be discovered. One of such phenomena is studied in one part of this thesis and can be regarded as *emergence* of multiparticle entanglement. The idea is that in multiparticle systems entanglement can not be present in the subsystems, or marginals, but at the same time correlations in these marginals indicate that in the global state all particles are entangled with each other.

The so-called graph states [18] are one of the most famous examples of multipartite entangled states with fruitful practical applications [15]. Recently, their direct generalization, hypergraph states [19], were shown [20] to display an exponential violation of local realism, similar to graph states, with an additional property of robustness against a particle loss. In this thesis we consider a generalization of hypergraph states, defined for two-level systems, to d -level systems.

Analogously to entanglement, most of the studied scenarios displaying the discrepancy of quantum theory from classical models are of the bipartite nature. One example is the Bell's theorem. However, studying more complicated scenarios and models can lead to a better understanding of nonlocality, the property of quantum mechanics predicted by the theorem of Bell [3]. In another part of this thesis we apply the machinery of causal structures, the graphical language to represent dependencies among random variables, to study such complex scenarios. We derive a necessary condition on causal structures to be distinguishable from observations in a given set of subsystems. As an application, this result indicates which causal structures can be used to test nonlocality under such restrictions.

An important technique, used in the aforementioned analysis of causal structures, is the entropic approach [21, 22]. This method, which sometimes is called entropy vector method [23], was initially introduced in Ref. [24] to calculate bounds on transmission rates in complex communication networks. In this method an important notion is non-Shannon-type inequalities, which when taken into consideration, help to derive tighter information-theoretic bounds. Together with the result on indistinguishability of causal structures in this thesis we provide an improvement to the method of entropy vectors, which leads to the derivation of non-Shannon-type inequalities.

In the last part of the thesis we consider processes for which the definite casual order may not exist. Recently, such processes were defined in the formalism of so-called process matrices [25]. We derive inequalities on entropy region of processes compatible with a definite causal structure. These inequalities are similar to Bell inequalities that bound the correlations compatible with the theories assuming local realism.

This thesis is based on five publications Refs. [26–30] also listed in the [end of the thesis](#). Three of this papers are journal publications and two are submitted to the publication. The main content of the thesis is divided into six chapters. In the current chapter we have given a general introduction to the fields of studies and in the subsequent Chapter 2 a more detailed description of the preliminary results is given. Chapter 3 describes the results of Refs. [26, 29] on which I have collaborated with Marius Paraschiv, Tobias Moroder and Otfried Gühne. Chapter 4 gives a partial report of Ref. [28], the joint work with Frank E.S. Steinhoff, Christina Ritz, and Otfried Gühne. Chapter 5 is dedicated to the results of Ref. [27], a project in collaboration with Costantino Budroni and Rafael Chaves. The last Chapter 6 summarizes the work done in collaboration with Alastair Abbott, Cyril Branciard, Rafael Chaves, and Costantino Budroni [30].

Chapter 2

Preliminaries

This chapter introduces notions and definitions, which will be used later in this thesis. We will start with a formal definition of entanglement. After, we will discuss the argument of Bell and its formulation in terms of classical causal models. Finally, we give a broad description of causal models, including the framework of indefinite causal order.

Basic notions of Hilbert spaces, quantum states, quantum transformations, and quantum measurements are used in this thesis without definitions. For these definitions please see Ref. [31]. We will use terms *qubit* and *qudit* for two- and d -level quantum systems respectively. A qubit is the quantum analogy of a bit in classical information theory.

2.1 Entanglement

2.1.1 Bipartite entanglement

Pure states

We begin with the notion of a pure entangled state of two systems, controlled by two experimenters Alice and Bob. For each system we associate a Hilbert space \mathcal{H}_A and \mathcal{H}_B of dimensions d_A and d_B respectively, then a pure state of the combined system is represented by a normalized vector $|\psi\rangle$ of the product space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ of dimension $d = d_A d_B$. A state $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_{i=1}^{d_A} \sum_{j=1}^{d_B} c_{ij} |a_i\rangle \otimes |b_j\rangle, \quad (2.1)$$

where $|a_i\rangle$ and $|b_j\rangle$ are orthonormal basis vectors of spaces \mathcal{H}_A and \mathcal{H}_B respectively and c_{ij} are some complex coefficients, chosen such that the state $|\psi\rangle$ is normalized. A state which can be written as Eq. (2.1) with only one element of the sum is called a *product state*. In this thesis a shorthand notation $|a_i b_j\rangle$ is used for all product states $|a_i\rangle \otimes |b_j\rangle$. If a pure state $|\psi\rangle$ is not product, it is called *entangled*.

Mixed states

In general, the quantum state of a system is not pure and can be described by a positive linear Hermitian operator of a unit trace. We will call this operator a *density operator*, or a *density matrix*¹ and denote as ρ . We say that a matrix (or operator) is positive definite if all its eigenvalues are positive. For a pure state $|\psi\rangle$ density matrix is defined as an outer product of vector $|\psi\rangle$ with its Hermitian conjugate, i.e. $\rho = |\psi\rangle\langle\psi|$. In general, any density matrix can be written as a convex mixture of pure states, i.e. as

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \quad (2.2)$$

where p_i are positive coefficients, and $\sum_i p_i = 1$. Coefficients p_i can be considered as probabilities with which the quantum system is in one of the states $|\psi_i\rangle$.

A quantum system can be in a mixed state if it is part of a larger system, which itself is in a pure state. For example, consider the state from Eq. (2.1). The state of the system controlled by Alice can be written as

$$\rho_A = \text{tr}_B(|\psi\rangle\langle\psi|) = \sum_{i,k=1}^{d_A} \sum_{j=1}^{d_B} c_{ij} c_{kj}^* |a_i\rangle\langle a_k|, \quad (2.3)$$

where tr_B denotes a partial trace over the Bob's system. If Alice and Bob perform measurements on the state $|\psi\rangle$ and do not communicate, the expectation value of Alice measurement of some observable \mathcal{A} will be described as $\text{tr}(\mathcal{A}\rho_A)$. Everywhere throughout the thesis when speaking about expectation values we will always mean the asymptotic limit of the statistics of measurements of the same observable on the copies of the same state. States of the subsystems, like ρ_A , are often called *reduced states* and the state of the whole system – *global state*. The problem of studying relations between global state and reduced states is known as *marginal problem* and is one of the central parts of this thesis.

Let us continue with definition of entanglement for bipartite mixed states.

¹More precisely, density matrix is a matrix of moments of the corresponding density operator taken with respect to some orthonormal basis. However, in this thesis both terms are used interchangeably.

Definition 2.1. A bipartite state ρ_{AB} is called *separable* if there exist a decomposition

$$\rho_{AB} = \sum_i p_i \rho_A^i \otimes \rho_B^i, \quad (2.4)$$

with $p_i \geq 0$, $\forall i$ and where ρ_A^i and ρ_B^i are mixed states of Alice's and Bob's systems. If the above decomposition cannot be found, the state is called *entangled*.

Separable states can be introduced in an equivalent way as states, which can be prepared from product states by means of local operations and classical communication (LOCC). LOCC are the class of operations which two parties can perform by acting locally on quantum systems in their laboratories with possible coordination by classical communication. First, note that taking into account the definition (2.2) every separable state can be written as

$$\rho_{AB} = \sum_i \tilde{p}_i |a_i\rangle\langle a_i|_A \otimes |b_i\rangle\langle b_i|_B, \quad (2.5)$$

for some new distribution \tilde{p}_i and $|a_i\rangle$, $|b_i\rangle$ which are now pure states of Alice's and Bob's systems. It is then easy to see that state of the form (2.5) can be prepared by means of LOCC. Let Alice have access to a random variable I with distribution \tilde{p}_i from (2.5) and let Alice's and Bob's systems to be in the initial state $|00\rangle$. After obtaining an outcome i of the variable I , Alice sends it to Bob and then depending on i Alice and Bob can apply local unitary transformations on their systems bringing the state to $|a_i b_i\rangle$. The resulting state is exactly the one, presented in Eq. (2.5).

From the same reasoning it follows that no entanglement can be created via LOCC. Additionally, as we mentioned in the introductory chapter, entanglement is important for performing various tasks like quantum teleportation, quantum key distribution, quantum computation, etc. [32]. These facts give rise to an alternative definition of entanglement as a resource for quantum information tasks. As any resource, entanglement needs to be quantified which can be done by so-called entanglement measures (see Ref. [33]).

The entanglement of pure bipartite states is fully characterized by their *Schmidt decomposition*.

Definition 2.2. The Schmidt decomposition of a state $|\psi\rangle$ is a sum

$$|\psi\rangle = \sum_{i=1}^r \lambda_i |\alpha_i\rangle \otimes |\beta_i\rangle, \quad (2.6)$$

with positive coefficients λ_i and vectors $|\alpha_i\rangle \in \mathcal{H}_A$ and $|\beta_i\rangle \in \mathcal{H}_B$, such that $\sum_{i=1}^r \lambda_i^2 = 1$, and $\langle \alpha_i | \alpha_j \rangle = \langle \beta_i | \beta_j \rangle = \delta_{ij}$. The proof that every pure state can be written in a form of Eq. (2.6) can be found in Ref. [32]. Schmidt coefficients λ_i are usually taken to be real and positive. This can be always done by keeping the phase in one of the vectors $|\alpha_i\rangle$ or $|\beta_i\rangle$. Squares of λ_i equal to the common eigenvalues of both reduced states ρ_A and ρ_B and are usually calculated in this way. It is clear that if $|\psi\rangle$ is entangled, then ρ_A and ρ_B are mixed and vice-versa. Based on that fact, one can introduce a measure of entanglement of pure states as, for instance, entropy of the reduce state.

The sum limit r in Eq. (2.6) is called the *Schmidt rank* and it characterizes the type of entanglement observed in the state. We will discuss the importance of this types, or classes, of entanglement later in Section 2.1.3.

For the case of two qubits the Schmidt decomposition (2.6) can contain either 1 or 2 terms, which corresponds to product states and entangled states respectively. Ones of most famous two-qubit entangled states are the so-called Bell states defined as

$$\begin{aligned} |\Phi^+\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), & |\Phi^-\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\ |\Psi^+\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), & |\Psi^-\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \end{aligned} \quad (2.7)$$

PPT criterion

Although the Schmidt decomposition (2.6) completely characterizes entanglement of pure bipartite states, achieving the same in mixed states is not as straightforward. However, some other criteria can be used to detect entanglement in the states. Here we will mention one of the most famous one, called PPT criterion (or Peres-Horodecki criterion). For the review on entanglement criteria see Ref. [33].

Definition 2.3. A bipartite state ρ shared between two parties A and B is said to have a positive partial transpose (PPT) if

$$\rho^{TA} \geq 0, \quad (2.8)$$

that is, ρ^{TA} has no negative eigenvalues. ρ^{TA} denotes the partial transpose of the state ρ with respect to party A and can be written as follows

$$\rho^{TA} = \sum_{i,j=1}^{d_A} \sum_{k,l=1}^{d_B} \rho_{ij,kl} |j\rangle \langle i| \otimes |k\rangle \langle l|, \quad (2.9)$$

for the following decomposition of the state ρ

$$\rho = \sum_{i,j=1}^{d_A} \sum_{k,l=1}^{d_B} \rho_{ij,kl} |i\rangle\langle j| \otimes |k\rangle\langle l|. \quad (2.10)$$

For simplicity we will call states which have positive partial transposition *PPT states* and, conversely, states for which ρ^{T_A} has negative eigenvalues we will call *NPT states*. From (2.4) it is easy to see that separable states are PPT, since

$$\rho_{AB}^{T_A} = \sum_i p_i (\rho_A^i)^T \otimes \rho_B^i = \sum_i p_i (\rho_A^i)^* \otimes \rho_B^i, \quad (2.11)$$

where $(\rho_A^i)^*$ is a complex conjugate of ρ_A^i . Clearly, (2.11) is a valid separable state, i.e. is positive semidefinite. PPT criterion was first introduced in Ref. [34] as a necessary condition for separability and later in Ref. [35] was proven to be sufficient for $2 \otimes 2$ and $2 \otimes 3$ systems.

This criterion gives rise to entanglement measure called negativity, which can be found as follows:

$$\mathcal{N}(\rho_{AB}) = \sum_i \frac{|\lambda_i| - \lambda_i}{2}, \quad (2.12)$$

where the sum is taken over all the eigenvalues of the partial transposition $\rho_{AB}^{T_A}$ of the state. It is easy to see that (2.12) is nothing but the absolute of the sum of negative eigenvalues of the partial transposition.

Apart from the fact that PPT criterion is very easy to test, it gives rise to the very important tool of decomposable entanglement witnesses.

Entanglement witness

It can be easily seen from Eq. (2.4) that the set of separable states is convex. It ensures that in the Hilbert space of bipartite states for any entangled state there exists a hyperplane separating this state from the set of separable states. In entanglement theory such hyperplanes are called *entanglement witnesses* [36].

Definition 2.4. An entanglement witness is an observable \mathcal{W} that is non-negative on all separable states and has a negative expectation value on at least one entangled state.

If \mathcal{W} is an entanglement witness and $\text{tr}(\mathcal{W}\rho) < 0$, we say that the entanglement of the state ρ is detected by \mathcal{W} .

For the two-party case a witness \mathcal{W} is called *decomposable* if it can be written in terms of two positive semidefinite operators P and Q , ($P \geq 0$ and $Q \geq 0$) as

$$\mathcal{W} = P + Q^{T_A}, \quad (2.13)$$

where T_A denotes partial transposition with respect to party A . One can easily see² that \mathcal{W} from Eq. (2.13) is an entanglement witness, since for any separable (and thus PPT) state ρ , $\text{tr}(\mathcal{W}\rho) = \text{tr}(P\rho) + \text{tr}(Q^{T_A}\rho) = \text{tr}(P\rho) + \text{tr}(Q\rho^{T_A}) \geq 0$ and for any NPT state σ with an eigenstate $|\psi\rangle$ of σ^{T_A} , corresponding to the negative eigenvalue of σ^{T_A} , the witness (2.13) with $P = 0, Q = |\psi\rangle\langle\psi|$ detects σ .

In general, finding an entanglement witness is a hard task, on the other hand, decomposable witnesses can be effectively found with the help of semidefinite programming (SDP)³. Decomposable entanglement witnesses can be generalized to efficiently detect genuine multipartite entanglement [37] in systems of many particles.

2.1.2 Multipartite entanglement

It is clear that in multipartite quantum systems some subsystems can be entangled among each other and some not. However, perhaps the most interesting situation is when all particles are entangled with each other, which is often called genuine multipartite entanglement. To give a strict definition, we will consider three-particle case, however, the definition can be extended to an arbitrary number of particle. Detailed discussions can be found in Refs. [33, 37, 38]. First, a state ρ_{ABC} is said to be fully separable if it can be written as

$$\rho = \sum_i p_i \rho_A^i \otimes \rho_B^i \otimes \rho_C^i, \quad (2.14)$$

where the p_i form a probability distribution. A less strict condition is separability of a state ρ_{ABC} with respect to some bipartition, e.g. $A|BC$, which in this case means that the state can be written as

$$\rho = \sum_i p_i \rho_A^i \otimes \rho_{BC}^i, \quad (2.15)$$

where, again, $p_i \geq 0$ and $\sum_i p_i = 1$. If the global state of the system can be written as

$$\rho^{\text{bs}} = p_1 \rho_{A|BC}^{\text{sep}} + p_2 \rho_{B|AC}^{\text{sep}} + p_3 \rho_{C|AB}^{\text{sep}}, \quad \text{with } p_1 + p_2 + p_3 = 1, p_{1,2,3} \geq 0, \quad (2.16)$$

²We have used here the fact that for two Hermitian operators Q and q from $\mathcal{H}_A \otimes \mathcal{H}_B$ $\text{tr}(Q^{T_A}q) = \text{tr}(Qq^{T_A})$. Latter can be imitatively seen after writing down decomposition of both operators into the sum of direct products in \mathcal{H}_A and \mathcal{H}_B .

³We will review SDP later in this chapter.

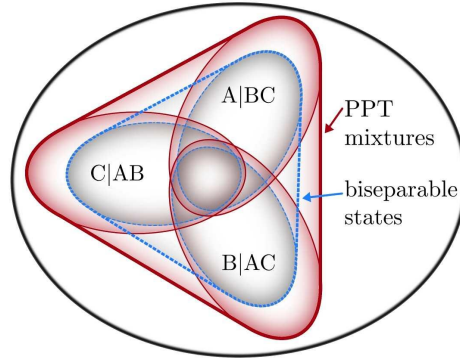


FIGURE 2.1: Illustration of biseparable states and PPT mixtures for a three-particle system, see text for further details. The figure is taken from Ref. [37].

it is called biseparable. If a state is not biseparable, i.e. it cannot be written in the form of Eq. (2.16), it is *genuinely multipartite entangled*.

As in the case of bipartite systems, the biseparability of states is not easy to check. However, as in the case of PPT criterion, we can consider a larger sets of states, which can be written as a mixture of PPT states with respect to each bipartition. States, which can be written in the following way

$$\rho^{\text{pmix}} = \tilde{p}_1 \rho_{A|BC}^{\text{ppt}} + \tilde{p}_2 \rho_{B|AC}^{\text{ppt}} + \tilde{p}_3 \rho_{C|AB}^{\text{ppt}}, \quad (2.17)$$

where again $\tilde{p}_1 + \tilde{p}_2 + \tilde{p}_3 = 1$, $\tilde{p}_{1,2,3} \geq 0$, are called PPT mixture states [37].

Looking at Fig. 2.1, the convex hull of all states separable with respect to a fixed bipartition is the set of biseparable states. In a similar way, the convex hull of states which are PPT with respect to a bipartition is the set of PPT mixtures. It is clear that every biseparable state is also a PPT mixture. Thus, if we can prove that a state is not a PPT mixture, then it is genuinely multipartite entangled.

Now, having a suitable criterion for entanglement, we can write down an entanglement witness that can detect a state which is not a PPT mixture. A decomposable witness for multipartite case, which can be written as

$$\mathcal{W} = P_M + Q_M^{T_M}, \quad (2.18)$$

for any bipartition $M|\bar{M}$ of the system is called *fully decomposable*. The connection to the notion of PPT mixtures is the following:

Observation 2.5. If ρ is not a PPT mixture, then there exists a fully decomposable witness \mathcal{W} that detects it.

The proof can be found in Ref. [37]. Similar to decomposable witnesses, fully decomposable witnesses can be efficiently found with the help of SDP.

2.1.3 SLOCC classification

Remarkably, not all genuine multipartite entangled states are entangled in the same way [39]. This distinction comes from the same ideas of entanglement being a resource in setups where operations are restricted to be local.

Let us first refine the definition of local operations which we gave in Section 2.1.1. Two pure n -partite states $|\phi\rangle$ and $|\psi\rangle$ are equivalent under local unitaries if they are related as follows

$$|\phi\rangle = \bigotimes_{i=1}^n U_i |\psi\rangle, \quad (2.19)$$

where the U_i are unitary matrices, acting on the i -th particle. The question whether two multiqubit states are LU equivalent or not can be decided by bringing the states into a normal form under LU transformations [40].

More generally, the states are equivalent under stochastic local operations and classical communication (SLOCC) iff there exist invertible local operators (ILOs) A_i such that

$$|\phi\rangle = \bigotimes_{i=1}^n A_i |\psi\rangle. \quad (2.20)$$

Physically, this means that $|\phi\rangle$ can be reached starting from $|\psi\rangle$ by local operations and classical communication with a non-zero probability.

LOCC equivalence of two states, which we have already mentioned before, is a particular case of SLOCC equivalence where it is possible to perform SLOCC transformation with probability 1.

Based on pairwise SLOCC-equivalence states can be divided into classes, which we will call *SLOCC classes*. SLOCC equivalence of bipartite states is governed by their Schmidt rank, introduced in Section 2.1.1: two states are equivalent under SLOCC if their Schmidt ranks coincide. In case of two qubits this results in just one SLOCC class of entangled states.

Interestingly, for three qubits there exist two SLOCC classes of genuine multipartite entangled states, so-called GHZ-class and W-class. GHZ class is the class of states

which are equivalent under SLOCC to the following state

$$|GHZ\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle), \quad (2.21)$$

which is often called GHZ state after the authors of Ref. [41]. Another class of states are equivalent to the state, called W state [39], which can be written as

$$|W\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle). \quad (2.22)$$

W class is of the measure zero in the space of all states, which means that number of free parameters in characterization of a state from this class is less than the number of parameters in the density matrix of general three-qubit state.

The situation becomes much more complicated as number of particles grows. For instance, for four qubits one can distinguish 9 families of entangled states [42]. Family differs from SLOCC class by the fact that, in general, one cannot assign a single state like (2.21) or (2.22) to which all states from the family are SLOCC equivalent. Instead, each family corresponds to a certain form of the states defined by some set of free parameters.

Although general criteria for SLOCC (or LOCC) equivalence of multipartite states do not exist, it is possible to find general necessary conditions for the equivalence. For instance, SLOCC transformations can clearly not change the rank of a reduced state. Moreover, in special cases (e.g. for special classes of states) sufficient conditions for SLOCC-equivalence can be found. In this thesis we will introduce several tools to determine SLOCC equivalence of multiparticle states. These tools are based on the inductive SLOCC classification, introduced in Ref. [43].

2.1.4 Graph and hypergraph states

Some of the most notable examples of multipartite entangled states include GHZ states, which are defined similar to (2.21), Dicke states [44], cluster states [45], graph and hypergraph states [18, 19].

Here, we will define qudit graph states to introduce ideas, which will be used later in this thesis to define qudit hypergraph states (see Chapter 4). For the definition of qubit hypergraph states please see Ref. [19]. The main idea of graph and hypergraph states is the bijection between mathematical objects of graphs and hypergraphs and a certain class of quantum states. Such connection opens a possibility to find graphical rules

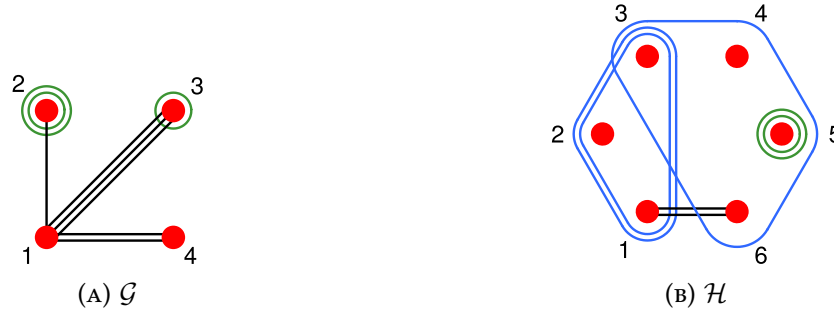


FIGURE 2.2: **(A)** Example of a multi-graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} = \{1, 2, 3, 4\}$ and $\mathcal{E} = \{\{1, 2\}, \{1, 3\}, \{1, 3\}, \{1, 3\}, \{1, 4\}, \{1, 4\}, \{2\}, \{2\}, \{3\}\}$; **(B)** Example of a multi-hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{E})$, with $\mathcal{N} = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{E} = \{\{1, 2, 3\}, \{1, 2, 3\}, \{1, 6\}, \{1, 6\}, \{5\}, \{5\}, \{3, 4, 5, 6\}\}$.

describing operations on the states from this class. Examples include LU equivalence, rules for measurements in Pauli X or Z basis [46–49].

Let us first start with a definition of graphs and hypergraphs. A graph is a pair $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where \mathcal{N} is the set of nodes (or vertices) and \mathcal{E} is a set comprised of 2-element subsets of \mathcal{N} called edges. Likewise, a hypergraph is a pair $\mathcal{H} = (\mathcal{N}, \mathcal{E})$, where \mathcal{N} are the vertices and \mathcal{E} is a set comprised of subsets of \mathcal{N} with arbitrary number of elements. In simple words, a hyperedge is an edge that can connect more than two vertices. We say that an edge $e \in \mathcal{E}$ has cardinality m if it connects m vertices. A multi-(hyper)graph is a set where the (hyper)edges are allowed to appear repeated. An example of a multi-graph can be found in Fig. 2.2a, while one of a multi-hypergraph can be found in Fig. 2.2b. We will say that e has multiplicity n if it appears n times in \mathcal{E} .

The Pauli group and its normalizer

Prior to the definition of qudit graph states, we need to review the groups of operators known as Pauli and Clifford groups. In a d -dimensional system with computational basis $\{|q\rangle\}_{q=0}^{d-1}$, let us consider the unitary operators given by

$$Z = \sum_{q=0}^{d-1} \omega^q |q\rangle\langle q|, \quad X = \sum_{q=0}^{d-1} |q \oplus 1\rangle\langle q| \quad (2.23)$$

with the properties $X^d = Z^d = I$ and $X^m Z^n = \omega^{-mn} Z^n X^m$, where $\omega = e^{2\pi i/d}$ is the d -th root of unity and \oplus denotes addition modulo d . The group generated by these operators is known as the Pauli group and the operators $X^\alpha Z^\beta$, for $\alpha, \beta \in \mathbb{Z}_d$ are referred as Pauli operators. For $d = 2$, these operators reduce to the well-known Pauli

matrices for qubits. In general, these operators enable a phase-space picture for finite-dimensional systems, via the relations $Z = e^{\frac{2\pi i}{d}Q}$, $X = e^{-\frac{2\pi i}{d}P}$, where $Q = \sum_{q=0}^{d-1} q|q\rangle\langle q|$ and $P = \sum_{q=0}^{d-1} q|p_q\rangle\langle p_q|$ are discrete versions of the position and momentum operators; here $|p_q\rangle = F|q\rangle$ and $F = d^{-1/2} \sum_{q',q=0}^{d-1} \omega^{q'q}|q'\rangle\langle q|$ is the discrete Fourier transform. Thus, X performs displacements in the computational (position) basis, while Z performs displacements in its Fourier transformed (momentum) basis.

Another set of important operators are the so-called Clifford or symplectic operators, defined as

$$S(\zeta, 0, 0) = \sum_{q=0}^{d-1} |\zeta q\rangle\langle q|; \quad (2.24)$$

$$S(1, \zeta, 0) = \sum_{q=0}^{d-1} \omega^{\zeta q^2} |q\rangle\langle q|; \quad (2.25)$$

$$S(1, 0, \zeta) = \sum_{q=0}^{d-1} \omega^{-\zeta q^2} |p_q\rangle\langle p_q|. \quad (2.26)$$

These operators are unitary whenever the values of ζ and d are coprime and generate the normalizer of the Pauli group, which is usually referred as the Clifford group. The interested reader can check a more broad formulation in terms of a discrete phase-space in the Ref. [50].

Qudit graph states

Now we can briefly review the theory of the so-called qudit graph states, which is well established in the literature [51–55]. The mathematical object used is a multigraph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$; we call $m_e \in \mathbb{Z}_d$ the multiplicity of the edge e , i.e., the number of times the edge appears. Given a multigraph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, we associate a quantum state $|G\rangle$ in a d -dimensional system in the following way:

- To each vertex $i \in \mathcal{N}$ we associate a local state $|+\rangle = |p_0\rangle = d^{-1/2} \sum_{q=0}^{d-1} |q\rangle$.
- For each edge $e = \{i, j\}$ and multiplicity m_e we apply the unitary

$$Z_e^{m_e} = \sum_{q=0}^{d-1} |q_i\rangle\langle q_i| \otimes (Z_j^{m_e})^q \quad (2.27)$$

on the state $|+\rangle^{\mathcal{N}} = \bigotimes_{i \in \mathcal{N}} |+\rangle_i$. Thus, the graph state is defined as

$$|G\rangle = \prod_{e \in E} Z_e^{m_e} |+\rangle^{\mathcal{N}}. \quad (2.28)$$

We allow among the edges $e \in \mathcal{E}$ the presence of “loops”, i.e., an edge that contains only a single vertex. A loop of multiplicity m on vertex k means here that a local gate $(Z_k)^m$ is applied to the graph state. An example of a qudit graph state, which corresponds to the graph in Fig. 2.2a, is $|G\rangle = Z_{12}Z_{13}^3Z_{14}^2Z_2^2Z_3|+\rangle^{\mathcal{N}}$.

An equivalent way of defining a qudit graph state is via the stabilizer formalism [51–55]. Given a multi-graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, define for each vertex $i \in \mathcal{N}$ the operator $K_i = X_i \prod_{e \in \mathcal{E}} Z_{e \setminus \{i\}}$. The set K_i generates an abelian group known as the stabilizer. The unique +1 common eigenstate of these operators is precisely the state $|G\rangle$ associated to the multi-graph \mathcal{G} . Moreover, the set of common eigenstates of these operators form a basis of the global state space, the so-called graph state basis. It is important to note that the fully-connected graph states are LU-equivalent to the GHZ state (2.21) of the same dimension (see Ref. [55]). Alternatively, it means that the GHZ states are also stabilizer states.

The local action of the generalized Pauli group on a graph state is easy to picture and clearly preserves the graph state structure. As already said, the action of Z_l^m corresponds to a loop of multiplicity m on the qudit l , while the action of X_l^m corresponds to loops of multiplicity m on the qudits in the neighbourhood of the qudit l .

2.2 Nonlocality

2.2.1 Bell’s theorem

Interestingly, to prove that state is entangled one does not always need to know its full density matrix. In some cases it is enough for parties to measure expectation values of only few observables. One of such examples is violation of so-called *Bell inequalities*. When a quantum state violates a Bell inequality it is said to be *Bell-nonlocal*. However, as we will see in this section there is more to Bell inequalities and Bell’s theorem than just entanglement detection. Let us start with formulation of Bell’s theorem [3].

Theorem 2.6. [Bell 1964]

No local hidden variable (LHV) theory can reproduce all the predictions of quantum mechanics.

Local hidden variable (LHV) theories (or models) are an attempt to describe correlations observed in nature, which includes correlations from quantum experiments, while keeping the assumptions of *local realism*. Local realism is a common term for two assumptions underlined below:

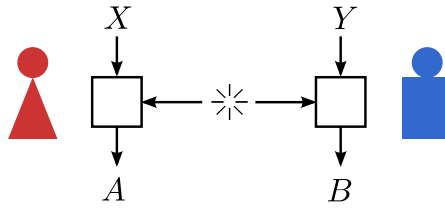


FIGURE 2.3: Schematic setup of Bell's experiment. X, Y – choices of Alice's and Bob's measurements. A, B – their measurement outcomes.

- (i) Signaling, i.e. sending information, is possible with at most the speed of light. It means, in particular, that if two events of Alice's measurement and Bob's measurement are space-like separated the probability of outcomes of Alice's measurement cannot depend on measurement settings of Bob. The same holds for probabilities of Bob's outcomes and Alice's choices of operations.
- (ii) The outcomes of all measurements are predefined. In other words, the randomness, which we observe in the experiment, is due to lack of knowledge and therefore there must exist a complete theory, generalizing quantum mechanics, in which outcomes of measurements are definite.

An additional assumption, which is not listed above, is the freedom of choice. It means that each party Alice and Bob can pick their measurement settings x and y at random. Assumptions of local realism were introduced by Einstein, Podolsky and Rosen in Ref. [2] and were shown to be inconsistent with predictions of quantum mechanics by Bell [3].

Let us now look closer at the experimental setup of Bell's experiment [3]. In this setup a pair of correlated quantum systems (e.g. spin 1/2 particles) is distributed between two parties, Alice and Bob as shown on Fig. 2.3. At random, the parties choose settings for their measurements (e.g. measure the spin in z or in x direction). Let us denote these choices of settings as random variables X and Y . We will denote as A and B the results of the measurements. After several runs of the experiment, the parties meet or give results to a referee and the expectation values $\langle A_x B_y \rangle$ for each setting x, y are analyzed. Here A_x is a new random variable, which can be obtained from A when conditioned on a particular context $X = x$. The same notation is used for B and y .

From the assumptions of local realism it follows that there exist a random variable Λ , distributed according to some distribution $P(\Lambda)$, and response functions $P(a|x, \lambda)$, $P(b|y, \lambda)$ for Alice and Bob such that probability of outcomes a, b of their measurements, specified by x and y , is given by

$$P(ab|xy) = \sum_{\lambda} P(a|x, \lambda)P(b|y, \lambda)P(\lambda). \quad (2.29)$$

In Eq. (2.29) the hidden variable Λ is taken to be discrete, however, in general it can be a continuous parameter or even a set of parameters. However, this restriction on generality does not play any role as Λ can be removed from consideration, leaving only constraints on observed correlations $P(ab | xy)$, known as Bell inequalities.

2.2.2 Bell inequalities

The first Bell inequalities were obtained by Bell in Ref. [3] and Clauser, Horne, Shimony, and Holt in Ref. [56]. Up to permutation of indexes the latter can be written as

$$-2 \leq \langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \leq 2 \quad (2.30)$$

and is often called *CHSH inequality*. Here $A_x, B_y \in \{-1, 1\}$, $X, Y \in \{1, 2\}$ and the expectation values can be found as

$$\langle A_x B_y \rangle = \sum_{a,b \in \{-1,1\}} ab P(ab | xy). \quad (2.31)$$

As realized by Pitowsky [57], the set of probability distributions compatible with the LHV model form a convex set of correlations, the so called correlation polytope. This polytope is characterized by finitely many extremal points, which correspond to deterministic response functions $P(a | x, \lambda)$ and $P(b | y, \lambda)$ in (2.29). Equivalently, this polytope is characterized by finitely many linear inequalities, the non-trivial ones being exactly the Bell inequalities.

Following Pitowsky [57], it is easy to prove the LHV bound in Eq. (2.30). From convexity, it follows that any linear function over the correlation polytope is maximized by one of the extremal points. It means that all we need to do is to go through all the deterministic assignments $A_x, B_y = \pm 1$, and take the maximum. In particular, from $\langle A_1 B_1 \rangle = \langle A_1 B_2 \rangle = \langle A_2 B_1 \rangle = 1$ we conclude that $A_1 = B_1$, $A_1 = B_2$, and $A_2 = B_1$, which leaves the only option $A_2 = B_2$ and thus $\langle A_2 B_2 \rangle = 1$.

The statement of the Bell's theorem is manifested by the violation of Bell inequalities by measurements on entangled states. In particular, the expression in (2.30) can take the value of $2\sqrt{2}$ for optimal measurements on one of the Bell states (2.7). Interestingly, this is the upper bound on (2.30) allowed by quantum theory independently on the dimension of the quantum systems [58].

It is important to note that nonlocality, i.e. violation of Bell inequalities, implies entanglement of the state. This can be easily seen from the fact that any local measurements

on separable states (2.4) gives the form (2.29) for conditional probabilities. However, as first pointed out by Werner [59] not all entangled states lead to the violation of Bell inequalities.

2.2.3 Correlation polytopes

We have already mentioned the results of Pitowsky [57], who proved that the set of correlations, compatible with LHV models, is a polytope. However, different methods to characterize such correlation polytopes are available. Here, we will describe a particular method obtained via Fine's theorem [60] where the use of the concept of marginal scenario becomes apparent.

Fine's theorem shows that the existence of a LHV model of the form (2.29) is equivalent to the existence of a joint probability distribution $P(A_1, \dots, A_m, B_1, \dots, B_m)$ that marginalizes to the observed distribution $P(A_x B_y)$ with $X, Y = 1, \dots, m$. The existence of a well-defined joint distribution over all variables implies that such distribution must respect some constraints, namely, positivity and normalization. That is, $P(A_1, \dots, A_m, B_1, \dots, B_m)$ must lie inside a simplex polytope [61]. From this geometric perspective, the correlation polytope is nothing other than the projection of the simplex polytope –characterizing the joint distribution– to a subspace of it that is given by the marginal scenario⁴ in question, that is, a projection to the subspace spanned by the observable components $P(A_x B_y)$ with $X, Y = 1, \dots, m$. Such a projection can be obtained by eliminating, from the corresponding system of linear inequalities describing the simplex, all terms that correspond to non-observables probabilities. This can be achieved, for example, via a standard algorithm known as Fourier-Motzkin elimination [62]. Once the redundant inequalities are removed, the remaining set gives the facets of the correlations polytope, that is, tight Bell inequalities.

For the simplest case of two measurement settings and two possible measurement outcomes, one needs to consider the 16-dimensional vector

$$\vec{p} = (1, P(a_1), P(a_2), P(b_1), P(b_2), P(a_1 a_2), P(b_1 b_2), P(a_1 b_1), P(a_1 b_2), P(a_2 b_1), P(a_2 b_2), P(a_1 a_2 b_1), P(a_1 a_2 b_2), P(a_1 b_1 b_2), P(a_2 b_1 b_2), P(a_1 a_2 b_1 b_2)), \quad (2.32)$$

where we used the notations $P(a_1 a_2 b_1) = P(A_1 = -1, A_2 = -1, B_1 = -1)$ (we consider here as before $A_x, B_y = \pm 1$). The simplex inequalities in this case are the ones bounding each term in this vector to be in the region $[0, 1]$ and, additionally, ones which are implied by the positivity of probabilities, which are not represented in

⁴For a formal definition of marginal scenario see Sec. 2.4.

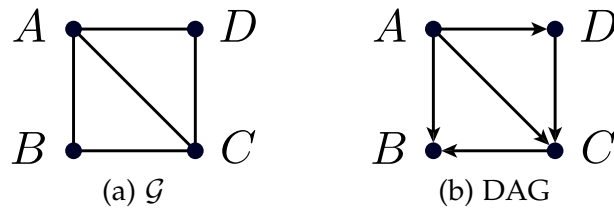


FIGURE 2.4: Examples of graphs. (a) A graph \mathcal{G} containing a loop. (b) A similar DAG with no loops.

Eq. (2.32). For example, the condition $P(a_1) \geq P(a_1b_1)$ is implied by the positivity of $P(A_1 = 0, B_1 = 1)$. To obtain CHSH inequality one needs to project the polytope of simplex inequalities on the subspace of $\{1, P(a_1), P(a_2), P(b_1), P(b_2), P(a_1b_1), P(a_1b_2), P(a_2b_1), P(a_2b_2)\}$ and note that the correlators $\langle A_x B_y \rangle$ are just linear functions of the elements of probabilities in this subspace. For instance, $\langle A_1 B_1 \rangle = 1 + 4P(a_1b_1) - 2P(a_1) - 2P(b_1)$.

Unfortunately, such a nice linear convex picture does not hold for more complicated scenarios (e.g. bilocality) [63–66] that now require computationally expensive and highly intractable methods from algebraic geometry [67] in order to deal with the non-linear constraints arising from such models. As it will be explained later, this is one of the reasons why the entropic approach has become a more viable option in dealing with such problems. The entropic approach plays an important role in this thesis and will be discussed in detail in Sec. 2.5.

2.3 Causal models

We have already pointed out in the introduction that the correlations, observed in a quantum experiments, challenge our understanding of cause and effect relations among random variables. Finding a suitable LHV model is an attempt to explain the Bell experiment from classical ideas of causality. Here, the hidden variable Λ is a cause for correlations observed in A and B . In this section we will discuss causal models beyond simple LHV model and give a background for the results of Chapters 5 and 6. We will start by reviewing some notions of the graph theory.

2.3.1 Directed and undirected graphs

We have already considered graphs and hypergraphs in Section 2.1.4. What we did not consider yet is a notion of directed graph, i.e., a graph which has directed edges corresponding to ordered pairs $(i, j) \in \mathcal{E}$, denoted by an arrow from i to j . An important subclass of directed graphs are *directed acyclic graphs (DAGs)*, which means that

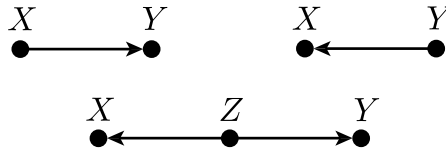


FIGURE 2.5: Three cases, considered in Reichenbach's Principle. The dependence among variables X, Y, Z is represented by DAGs.

there is no sequence of edges $(i_1, i_2), (i_2, i_3), \dots, (i_n, i_1)$ which form a closed path. For a formal definition of paths and cycles see Sec. 5.0.1. See Fig. 2.4 for examples of a undirected graph and a DAG.

DAGs is a widely used tool to represent cause-effect relations in complex systems of random variables [68]. In this way the acyclicity of DAGs prevent the existence of self-causation which results in so-called grandfather paradox.

2.3.2 Cause-effect relations

One of the central topics of this thesis is investigation of cause-effect relations in systems of random variables. These relations can be represented graphically, using the language, introduced in the previous section. As before, we will use upper-case letters (e.g., X) to denote random variables, and corresponding lower-case letters (e.g., x) the specific value they take.

The idea of looking for causal explanation of observed correlations was first formalized by Hans Reichenbach. In his work [69] he pointed out that for two random variables X and Y which are correlated, meaning that $P(xy) \neq P(x)P(y)$, one of the following must hold:

- (a) X is a cause of Y , meaning that distribution of Y can depend on the choice of X , but not the other way around.
- (b) Y is a cause of X .
- (c) There exist a random variable Z , which is not influenced by any of two X or Y , but which is a common cause for both.

All three cases can be represented pictorially as on Fig. 2.5 by directed acyclic graphs (DAG), which were introduced in Section 5.0.1. Ideas of Reichenbach were later developed into mathematical theory of causal structures [68]. As we will see later, the LHV model in Bell test is similar to the common cause model in Reichenbach's principle. In this way quantum nonlocality contradicts the principle and thus it challenges the classical ideas of cause and effect in general.

For a set of random variables X_1, \dots, X_n two common examples of causal relations, or dependencies, are given below:

- *Deterministic dependence*: The variable X_i is said to be a deterministic function of X_j if their joint probability distribution satisfies $P(x_i, x_j) = \delta_{x_i, F(x_j)} P(x_j)$, where the deterministic dependence is given by $x_i = F(x_j)$. Such types of constraints are usually present in network coding [70]. While not strictly necessary, deterministic constraints also play an important role in the derivation of Bell inequalities [60].
- *Conditional or unconditional independence*: The variable X_i is said to be independent of the variable X_j if the joint probability distribution satisfies $P(x_i, x_j) = P(x_i)P(x_j)$. Similarly, variable X_i is said to be conditionally independent (CI) of X_j given X_k , if $P(x_i, x_j, x_k) = P(x_i|x_k)P(x_j|x_k)P(x_k)$; that is, X_k screens off the correlations between the two other variables. We will denote the two situations as $(X_i \perp X_j)$ and $(X_i \perp X_j | X_k)$, respectively. This notation is straightforwardly extendible to many random variables, e.g. $(X_i \perp X_j \perp X_k)$ means that $P(x_i, x_j, x_k) = P(x_i)P(x_j)P(x_k)$.

Definition 2.7. We will call a *causal structure* a set of deterministic or conditional dependencies for a given set of variables X_1, \dots, X_n . Causal structure can be specified by a set of ordered pairs (X_i, X_j) for all i and j denoting that X_i is a potential cause of X_j .

Note that in the above definition we left the possibility for bidirectional dependence, i.e. if both (X_i, X_j) and (X_j, X_i) are contained in the causal structure. It means that X_i and X_j are dependent, but the direction of the dependence is not known, e.g. it can be a subject to some hidden parameter.

Definition 2.8. A *causal model* is a causal structure together with specified functional dependencies (e.g. $x_i = F(x_j)$) or specified conditional probability distributions (e.g. $P(x_j|x_k), P(x_k)$).

For the case of a common cause in Reichenbach's principle, $\{(Z, X), (Z, Y)\}$ is a causal structure, while specifying that $x = z$ and $y = \bar{z}$ is a causal model (in this case deterministic).

When studying causal models two central problems are considered:

- (i) Given experimental data find causal models, which would explain the observed correlations. This task is often called *causal discovery* or causal inference.

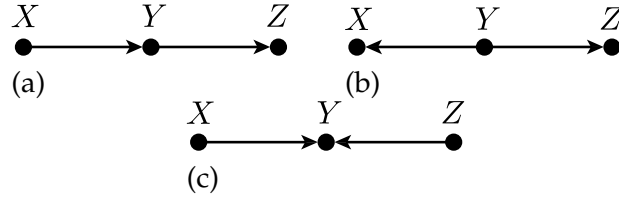


FIGURE 2.6: **Examples of different Bayesian networks.** (a) A DAG representing a Markov chain $X \rightarrow Y \rightarrow Z$ implying the CI $(X \perp Z | Y)$. (b) A DAG where the variable Y is a common parent of X and Z , once more implying the CI $(X \perp Z | Y)$. (c) A DAG where the variables X and Z have a common child Y . In this case $(X \perp Z)$, but $(X \not\perp Z | Y)$.

- (ii) Given a causal structure find restrictions on correlations which this structure imposes.

Below we will consider two types of causal models, Bayesian networks and Markov random fields [71].

2.3.3 Bayesian networks

A Bayesian network (BN) is a probabilistic model for which conditional dependencies can be represented via a directed acyclic hypergraph (DAG). More precisely, the probability distribution factorizes as

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{Pa}_i), \quad (2.33)$$

where Pa_i denotes the *parents* of the node i , i.e., the nodes with arrows pointing at i . The above factorization of the probability distribution gives rise to the *local Markov property*

$$(X_i \perp \text{Nd}_i | \text{Pa}_i), \quad (2.34)$$

namely that X_i is independent of its *nondescendants* Nd_i , i.e., nodes reachable from X_i via a directed path, given its parents.

More generally, one has the set of conditional independence relations

$$\mathcal{I}(\mathcal{G}) = \{(X_A \perp X_B | X_C) \mid \text{dsep}_{\mathcal{G}}(A : B | C)\}, \quad (2.35)$$

where $\text{dsep}_{\mathcal{G}}(A : B | C)$ refers to the d -separation properties of nodes in A and B with respect to nodes in C , namely that every path from $a \in A$ to $b \in B$, or vice versa, is *blocked* by a node in C . The path is said to be blocked if it contains one of the following: $x \rightarrow c \rightarrow y$, or $x \leftarrow c \leftarrow y$, or $x \rightarrow z \leftarrow y$, for x, y, z, c in the path, $c \in C$, $z \notin C$ and no descendant of z is in C (cf. Ref. [68]).

Bayesian networks are of particular relevance to formalize causal relations. Within this context, such causal models have been called causal Bayesian networks [68], as opposed to traditional Bayesian networks that formalize conditional independence relations without having necessarily a causal interpretation. To exemplify, consider the three DAGs shown in Fig. 2.6. DAGs (a) and (b) clearly imply a different set of causal relations between variables X , Y and Z : in both cases the correlations between X and Z are mediated via Y but in (a) X is a parent of Y while in (b) the reverse is true. In spite of their clear causal differences, both causal models imply the same set of CIs, namely that $(X \perp Z | Y)$. That is, every observable probability distribution $p(x, y, z)$ compatible with (a) is also compatible with (b), thus both models are indistinguishable from observations alone ⁵. The DAG (c) in Fig. 2.6 can nonetheless be distinguished from (a) and (b), since it implies that $(X \not\perp Z | Y)$ and the only CI is given by the independence constraint $(X \perp Z)$.

2.3.4 Markov random fields

Similarly to Bayesian networks, Markov random fields (MRF) correspond to probabilistic models for which conditional dependencies can be represented by a graph \mathcal{G} . In this case, the graph is undirected and it may contain cycles. More precisely, independence relations are given by the *global Markov property*

$$(X_A \perp X_B | X_C), \quad (2.36)$$

if every path from a node in A to a node in B passes through a node in C , i.e., if C is a separator for A and B in \mathcal{G} , a fact denoted as $\text{sep}_{\mathcal{G}}(A : B | C)$. We will denote the corresponding set of independence relations as

$$\mathcal{I}(\mathcal{G}) = \{(X_A \perp X_B | X_C) \mid \text{sep}_{\mathcal{G}}(A : B | C)\}. \quad (2.37)$$

As opposed to Bayesian networks (cf. Eq. (2.33)), MRFs do not admit a unique factorization of the probability distribution. However, for the special case of a triangulated graph, denoting with C_1, \dots, C_k the set of maximal cliques with the running intersection property and $S_i := C_i \cap (C_1 \cup \dots \cup C_{i-1})$, as in Eq. (5.1), one can write

$$P(x_1, \dots, x_n) = \prod_{i=1}^k \frac{P(x_{C_i})}{P(x_{S_i})}. \quad (2.38)$$

⁵In such cases, in order to distinguish different causal structures, one has to rely on another crucial concept of the mathematical theory of causality, that of an intervention [68]

Both MRFs and Bayesian networks can be specified by a pair $(\mathcal{G}, \mathcal{P})$, where \mathcal{G} specifies the graph (directed or undirected) and \mathcal{P} specifies the probabilities in Eq. (2.38) or Eq. (2.33).

2.3.5 Hidden variables

An important element of causal models is the possibility to consider *latent*, or *hidden*, variables. We will use the word “hidden” as it is more common in the field of quantum foundations. Hidden variables are random variables, which are for one or the other reason cannot be observed, or missed from consideration. Examples of hidden variable is a common cause Z in the Reichenbach principle and local hidden variable Λ in Bell test.

Below we introduce two central requirements to causal models. These conditions are the most important when considering hidden variables.

Minimality: Let us consider two models $(\mathcal{G}_1, \mathcal{P}_1)$ and $(\mathcal{G}_2, \mathcal{P}_2)$ with which we want to explain observed correlations P (or to model P). We say that model with causal structure \mathcal{G}_1 can simulate another model with structure \mathcal{G}_2 if for every $(\mathcal{G}_2, \mathcal{P}'_2)$ producing P' there is a model $(\mathcal{G}_1, \mathcal{P}'_1)$ producing P' . If now both $(\mathcal{G}_1, \mathcal{P}_1)$ and $(\mathcal{G}_2, \mathcal{P}_2)$ yield P and, additionally, \mathcal{G}_1 can simulate \mathcal{G}_2 but not the other way around, we say that $(\mathcal{G}_2, \mathcal{P}_2)$ is preferable to $(\mathcal{G}_1, \mathcal{P}_1)$. The model which is preferable to any other model producing P is called *minimal*.

Faithfulness: Let us denote the set of independence constraints, which is observed in P , as \mathcal{I} . A model $(\mathcal{G}, \mathcal{P})$ is called *faithful* if for any \mathcal{P}' , yielding distribution P' , the set of independence constraints, observed in P' is exactly \mathcal{I} .

The idea of minimality is introduced to make a model more easily disprovable, since, by the definition, the corresponding causal structure can produce less possible probability distributions than the other models. In particular, it prevents from inflating the causal structure with redundant hidden variables.

Faithfulness requires all of the independence constraints to be associated only with the causal structure, i.e. the graph. The models which are not faithful are said to be *fine-tuned*. In Ref. [8] authors showed that correlations observed on Bell experiment can only be explained by fine-tuned classical causal models. The results of this work go beyond the scope of this introductory part of the thesis, however, we will briefly discuss some of the models of Bell test below.

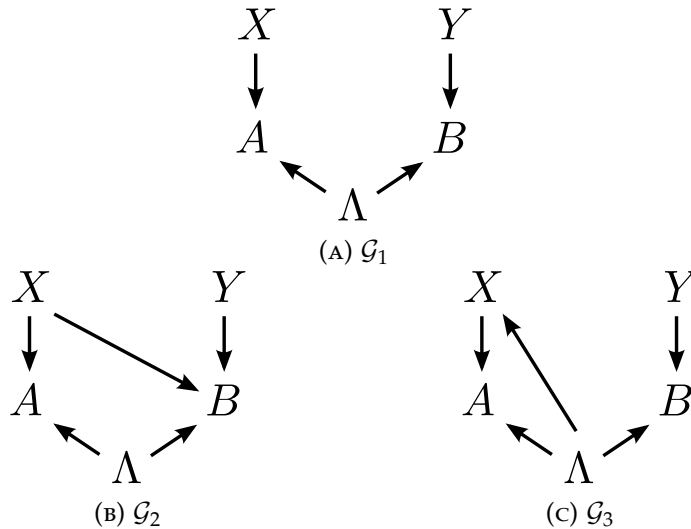


FIGURE 2.7: Causal models for Bell test. (A) LHV model, (B) signaling model, (C) superdeterministic model.

2.3.6 Causal models in Bell's test

We can now look back at the Bell test and notice that it is an example of causal discovery problem. Given experimental data, i.e. correlations $\langle A_x B_y \rangle$, one is interested to find a causal model, which would explain this correlations.

Let us consider tree models, which correspond to graphs shown on Fig. 2.7. The first causal structure \mathcal{G}_1 (Fig. 2.7a) corresponds to LHV model and it is easy to see that for any model with \mathcal{G}_1 factorization of probability distribution (2.33) corresponds to conditional probabilities of the type (2.29). Models with \mathcal{G}_1 do not violate any assumptions of Bell theorem, however, they do not generate distribution which display nonlocality. On the other hand, models, corresponding to \mathcal{G}_2 and \mathcal{G}_3 (Fig. 2.7b and Fig. 2.7c), can reproduce correlations violating Bell inequalities. However, it can be seen [8] that in order to satisfy the conditions of Bell experiment this models must be fine-tuned.

2.3.7 Indefinite causal order

When describing most physical phenomena, it seems natural to assume that physical events take place in a well-defined causal structure. For instance, earlier events can influence later ones but not the opposite, or, if two events are distant enough (typically, space-like separated) from each other, any correlation between them can only be due to some common cause in their past.

As already mentioned, quantum correlations obtained by measurements on distant entangled parties are incompatible with Reichenbach’s principle [72, 73] or, more generally, with classical theories of causality, forcing us to generalize the notion of causal models [22, 74–78]. In a scenario where different experimenters interact only once with a given system that is exchanged between them, one could expect that no simultaneous causal influences between each other should be possible but rather only one-way influences. However, it has been realized that physical theories do not necessarily have to comply with the idea of a definite causal order [25, 79]. One can also imagine theories where the causal order itself is in a sort of “quantum superposition” [25, 80], which can be verified using so-called causal witnesses [81, 82].

As for entanglement witnesses [35, 38], the use of causal witnesses assumes that we have a precise description of the measurement apparatus, that is, they are relevant in a device-dependent framework. Nevertheless, by allowing physical theories that are locally equivalent to quantum mechanics but relax the assumption of a fixed global causal structure, it is possible to verify causal indefiniteness also in a device-independent manner. With the aim of providing a general framework to such scenarios, the process matrix formalism [25] has been introduced and shown to allow for the violation of so-called causal inequalities [25, 83–87], which are device-independent constraints that play a similar role to that of Bell inequalities [3]. However, whether violations of causal inequalities can be experimentally observed is still an important open question.

Causal correlations

Causal correlations are most easily introduced in the bipartite case, where we consider two parties, Alice (A) and Bob (B), who together conduct a joint experiment while each having control over a separate closed laboratory. During each round of the experiment, Alice and Bob each receive, operate on, and send out a single physical system, which is the only means by which they may communicate. In addition, they each receive some (external) classical inputs X and Y , for Alice and Bob respectively, and produce some classical outputs A and B , respectively. As before we use upper-case letters (e.g., X) to denote random variables, and corresponding lower-case letters (e.g., x) to denote the specific values they take. Their probability distributions will generically be denoted by P ; we will also use the shorthand notations $P(x)$ for $P(X = x)$, $P(x, y)$ for $P(X = x, Y = y)$, $P(a|x)$ for $P(A = a|X = x)$, etc.

The joint conditional probability distributions $P(ab|xy)$ that can be produced in such an experiment depend on the causal relation between Alice and Bob. If Bob cannot

signal to Alice their correlations should obey $P(a|xy) = P(a|xy')$ for all x, y, y', a , where $P(a|xy) = \sum_b P(ab|xy)$. We denote this situation by $A \prec B$, and write $P = P^{A \prec B}$. Note that this does not necessarily imply that Alice is in the causal past of Bob since the events could be space-like separated, but merely that the correlation is compatible with such a causal order. Similarly, if the correlation is compatible with Bob being in the causal past of Alice we write $B \prec A$ and we have $P^{B \prec A}(b|xy) = P^{B \prec A}(b|x'y)$ for all x, x', y, b . The correlations that satisfy both these conditions (and are thus consistent both with $A \prec B$ and $B \prec A$) are precisely the nonsignaling correlations [88].

More generally, we are interested in the correlations achievable under the assumption of a definite causal order in each round of the experiment, even if the causal relation between Alice and Bob may be different (e.g., chosen randomly) for each individual round. We, thus, say that a correlation $P(ab|xy)$ is *causal* if it can be written as

$$P(ab|xy) = q_0 P^{A \prec B}(ab|xy) + q_1 P^{B \prec A}(ab|xy), \quad (2.39)$$

with $q_0, q_1 \in [0, 1]$ and $q_0 + q_1 = 1$, where $P^{A \prec B}(ab|xy)$ and $P^{B \prec A}(ab|xy)$ satisfy the respective (one-way) no-signaling conditions defined above [25].

It was shown in Ref. [86] that the set of bipartite causal correlations forms a convex polytope, whose vertices are simply the deterministic causal correlations (i.e., causal correlations for which the outputs A, B are deterministic functions of the inputs X, Y). The facets of this polytope thus specify *causal inequalities*, analogous to Bell inequalities for local correlations, that any causal correlation must satisfy [25]. The situation with binary input and output variables was characterized completely in [86], where it was shown that there are only two nonequivalent causal inequalities (up to symmetries). The simplest of these is perhaps the “guess your neighbor’s input” (GYNI) inequality, which has a simple interpretation as a game (up to a relabeling of the inputs and outputs) in which the inputs X, Y are chosen uniformly at random and the goal is for each party to output the other party’s input. One such form of this inequality can be written [86]

$$\frac{1}{4} \sum_{x,y,a,b} \delta_{a,y} \delta_{b,x} P(ab|xy) \leq \frac{1}{2}, \quad (2.40)$$

where δ is the Kronecker delta function.

The notion of causal correlations can be generalized to more parties, although one has to take into account the fact that, in a given round of the experiment, the causal order of some parties may depend on the inputs and outputs of previous parties [85, 87]. In this thesis we will primarily focus on applying the entropic approach to bipartite causal correlations and derive entropic causal inequalities. After, we will describe the generalization to the multipartite case. These results are described in Chapter 6.

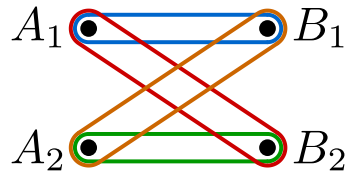


FIGURE 2.8: Hypergraph of the marginal scenario associated with a Bell-CHSH experiment. The observed probabilities correspond to the marginal for $\{A_x, B_y\}$, for $x, y = 1, 2$.

2.4 Marginal problem

In many relevant situations, one may have only partial information of the distribution of the variables. This may be due to practical limitations in collecting data, e.g., hidden variables which cannot be measured, or fundamental limitation, e.g., the impossibility of performing a joint measurement of incompatible quantum observables. This is common in Bell experiments, where one has access only to a limited set of joint probability distributions. Also in purely classical contexts the role of partial information can hardly be overemphasized [68, 89]. For instance, in the so called instrumentality tests modeling randomized experiments [68, 90], the effects from a drug in the recovery of patients is allowed to depend on some unobserved factors that are not under experimental control (social or economical background, etc). Whenever one needs to decide about properties of the system (e.g. global probability distribution) from the data observed in its subsystems we say that one faces the *marginal problem*.

For this reason, we introduce the notion of a marginal scenario. Given a set of random variables $X = \{X_1, \dots, X_n\}$, a *marginal scenario* is a collection of subsets $\mathcal{M} = \{M_1, \dots, M_{|\mathcal{M}|}\}$, $M_i \subset X$ of them representing variables that can be jointly measured, i.e., for each M_i , we have access to a probability distribution $P_{M_i}(x_{M_i})$. Moreover, if a set of variables are jointly measurable, we require that the same holds for any subset, i.e., $M \in \mathcal{M}$ and $M' \subset M$ imply $M' \in \mathcal{M}$. Equivalently, one can take only the maximal subsets $S \in \mathcal{M}$.

A marginal scenario can be naturally considered as an hypergraph, with \mathcal{M} the set of hyperedges and $\cup_i M_i$ the set of nodes. We will adopt the maximal subsets convention above and assume that $S \in \mathcal{M}$ are only maximal subsets, i.e., the hypergraph is reduced. We will call such a hypergraph the *marginal scenario hypergraph*, or simply marginal scenario when it is clear we are referring to the hypergraph.

It is again instructive to consider a simple example to fix the above notions. A standard example is given by a Bell experiment [3], in particular by the CHSH scenario [56]. The observed probabilities will then amount to the marginals for $\{A_x, B_y\}$. The corresponding marginal scenario hypergraph is depicted in Fig. 2.8.

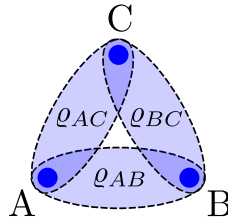


FIGURE 2.9: An example of the marginal problem for three-particle states. Marginals, denoted by blue filled ellipses are known and used to infer properties (e.g. entanglement) of the global state.

The idea of marginal scenario can be extended to quantum states. In this case, we allow arbitrary many measurements (tomography) on parts of the shared quantum systems, which means that we can recover the density operators of the subsystems. These density matrices are most often called reduced states, however, in this thesis we will refer to them simply as marginals. To demonstrate, let us consider an example of three-particle quantum system where we have access at most to the two-particle correlations as shown on Fig. 2.9. It means that we have access to ρ_{AB} , ρ_{AC} , and ρ_{BC} and from that information we might wish to recover the full density matrix ρ_{ABC} , or to prove that the global state is necessarily entangled. The problem of characterizing the global quantum states, or their properties, from marginals is called *quantum marginal problem* [91] and it is another central part of this thesis.

2.5 Entropic inequalities

In this section we will review an approach to study nonlocality scenarios, and causal structures in general, from entropic perspective. The idea of this approach is to derive entropic inequalities for a given marginal scenario from axioms of entropy, which are known as Shannon inequalities (or von Neumann inequalities), and a set of independence constraints associated with the causal structure. This set of inequalities forms a polyhedral convex cone in the space of joint entropies similar to correlation polytopes in the space of probabilities (see Sec. 2.2.3). As in the case of probabilities the entropic inequalities for the marginal scenario can be obtained via variable elimination (e.g. Fourier-Motzkin method).

The first reference to the approach discussed in this section is [24], although later it was used to study nonlocality scenarios in Ref. [21] and translated to quantum domain in Ref. [22]. Since then it was proven to be a powerful tool in deriving entropic inequalities for various phenomena in quantum information and foundations of quantum mechanics [27, 30, 92–96]. We will start the detailed description of the entropic approach with the formal definition of entropic cones.

2.5.1 Entropy cone

Given a collection of n discrete random variables X_1, \dots, X_n with an associated joint probability distribution $P(x_1, \dots, x_n)$, and denoting with X_S the random vector $(X_i)_{i \in S}$, for any subset $S \subset [n] := \{1, \dots, n\}$, the Shannon entropy $H : 2^{[n]} \rightarrow \mathbb{R}$ is defined as

$$H(S) := H(X_S) = - \sum_{x_S} P(x_S) \log_2 P(x_S). \quad (2.41)$$

The above entropies can be arranged in a vector

$$h = (H(\emptyset), H(X_1), \dots, H(X_1, X_2), \dots, H(X_1, \dots, X_n)) \in \mathbb{R}^{2^n} \quad (2.42)$$

The region

$$\Gamma_{[n]}^* := \overline{\{h \in \mathbb{R}^{2^n} \mid h = (H(S))_{S \subset [n]} \text{ for some entropy } H\}}, \quad (2.43)$$

where the overline denotes the closure in \mathbb{R}^{2^n} , is known to be a convex cone, also called the *entropy cone* and it has been studied extensively in information theory [70]. A tight and explicit description, however, has not yet been found for $n > 3$, but only some outer approximations of $\Gamma_{[n]}^*$ via polyhedral cones, i.e., cones described by a finite system of linear inequalities $Ax \geq b$, where A is an $m \times n$ real matrix and b an m -dimensional real vector. The most famous outer approximation to the entropic cone is the so-called *Shannon cone* $\Gamma_{[n]}$, defined by

$$h([n] \setminus \{i\}) \leq h([n]), \quad (2.44a)$$

$$h(S) + h(S \cup \{i, j\}) \leq h(S \cup \{i\}) + h(S \cup \{j\}), \quad (2.44b)$$

$$h(\emptyset) = 0, \quad (2.44c)$$

for all $i, j \in [n]$, $i \neq j$, and $S \subset [n] \setminus \{i, j\}$. That is, the Shannon cone associated with n variables is described by $2^{n-2} \binom{n}{2} + n$ inequalities plus one equality constraint (normalization).

The above is the minimal set of inequalities that implies monotonicity of entropy, i.e., $H(A|B) \geq 0$, and the submodularity (or strong subadditivity), i.e., $I(A : B|C) := H(A, C) + H(B, C) - H(A, B, C) - H(C) \geq 0$, for any disjoint subsets $A, B, C \subset [n]$ (cf. Ref. [70]).

Inequalities in Eq. (2.44) are known as *Shannon-type inequalities* in information theory or *polymatroidal axioms* in combinatorial optimization [70]. Given a finite set N and

real-valued function $f : 2^N \rightarrow \mathbb{R}$, the pair (N, f) is called a polymatroid if f satisfies Eqs. (2.44) above for $[n] = N$ and $S, \{i, j\} \subset N$.

In general we are interested in characterizing the entropy cone $\Gamma_{\mathcal{M}}^*$ associated with a marginal scenario \mathcal{M} , thus obtaining constraints implied by the global entropy cone on the marginal subspace of interest. Geometrically, this corresponds to the projection of the original entropy cone onto the subspace corresponding to the variables in \mathcal{M} . Since, in practice, we work with the Shannon cone $\Gamma_{[n]}$ – possibly constrained by some further linear constraints specifying a linear subspace L_C , as described previously – which is characterized by a finite system of inequalities, this projection corresponds to a simple variable elimination of all the terms not contained in \mathcal{M} [62, 92, 97]. After removing redundant inequalities, the remaining inequalities are facets (i.e., the boundaries) of the Shannon cone, or more generally polyhedron⁶, in the observable marginal subspace. Formally, the marginal Shannon polyhedron $\Gamma_{\mathcal{M}}$ is defined as

$$\Gamma_{\mathcal{M}} = \Pi_{\mathcal{M}} \left(\Gamma_{[n]} \cap L_C \right), \quad (2.45)$$

where $\Pi_{\mathcal{M}}$ denotes the projection onto the coordinates associated with the marginal scenario \mathcal{M} .

2.5.2 Probability structures

The characterization of entropy cones and marginal problems outlined above can be easily extended to the case where we no longer assume that there is a well-defined global probability distribution over all the variables in the set $[n]$. Instead, we may assume that only certain subsets of variables have such a joint distribution, and that only the marginals of certain subsets of these subsets are empirically accessible. This type of restriction may be imposed by assumptions about the underlying physical theory being described, as will be clear in the example we discuss in Sec. 2.5.4.

We will denote the collection of subsets of $[n]$ for which we assume joint probability distributions exist by $\mathcal{S} = \{S_1, \dots, S_{|\mathcal{S}|}\}$, with each $S_i \subset [n]$, such that $\cup_i S_i = [n]$; we call \mathcal{S} the *probability structure*. As for the marginal scenario, we will represent \mathcal{S} by just its maximal subsets in a slight abuse of notation; the complete representation of \mathcal{S} , that explicitly includes all (not necessarily maximal) subsets for which a joint probability distribution exists, will be denoted $\mathcal{S}^c = \{T \mid T \subset S_i, S_i \in \mathcal{S}\}$. In such a situation the entropies $H(T)$ cannot be defined for all subsets $T \subset [n]$, but only for the subsets in \mathcal{S}^c . The entropy vectors we shall consider will thus be defined here as

⁶Although, polyhedron is often means the polytope in dimension 3 in this thesis we use the terms polyhedron and polytope interchangeably.

$\mathbf{h} = (H(T))_{T \in \mathcal{S}^c} \in \mathbb{R}^{|\mathcal{S}^c|}$. Again, no explicit characterization is known for the set of valid entropy vectors; we will instead rely on its outer approximation characterized via the Shannon constraints, now restricted to each subset $S_i \in \mathcal{S}$. Namely, the Shannon cone of interest is now

$$\Gamma^{\mathcal{S}} = \bigcap_{S_i \in \mathcal{S}} \Gamma_{S_i}, \quad (2.46)$$

where $\Gamma_{S_i} \subset \mathbb{R}^{|\mathcal{S}^c|}$ is the cone defined by the Shannon inequalities on the variables in S_i , which, in particular, leave the other variables in $[n] \setminus S_i$ unconstrained. In the extremal case where we do assume a global joint probability distribution for all variables we have $\mathcal{S} = [n]$, $\mathcal{S}^c = 2^{[n]}$, and we recover $\Gamma^{\mathcal{S}} = \Gamma_{[n]}$.

One can similarly consider marginal scenarios under a given probability structure \mathcal{S} , with the constraint that marginals must arise from existing probability distributions, i.e., for all $M_j \in \mathcal{M}$ there must exist an $S_i \in \mathcal{S}$ such that $M_j \subset S_i$. One can also add linear constraints to the entropy vectors under consideration, as before, represented by some subset L_C . We can thus define the marginal Shannon polyhedron associated with \mathcal{S} , \mathcal{M} , and L_C as

$$\Gamma_{\mathcal{M}}^{\mathcal{S}} = \Pi_{\mathcal{M}} \left(\Gamma^{\mathcal{S}} \cap L_C \right). \quad (2.47)$$

The choice of probability structure can generally be considered on a case-by-case basis depending on the scenario being modeled. Unless otherwise stated we will take $\mathcal{S} = [n]$ but, as we will discuss, this will not always be the most pertinent choice.

2.5.3 The entropic characterization of Bayesian networks

Let us consider now an application of the entropic approach to study causal structures. In this section we will consider Bayesian networks, although for Markov Random Fields the approach is similar. As discussed in Section 2.3.3 causal structures of Bayesian networks can be represented by DAGs which are associated with the set of independence constraints, obtainable by the d -separation criterion [68]. Entropically, these CIs correspond to simple linear relations:

$$\begin{aligned} X_i \perp X_j &\rightarrow H(X_i X_j) = H(X_i) + H(X_j), \\ X_i \perp X_k \mid X_k &\rightarrow H(X_i X_j \mid X_k) = H(X_i \mid X_k) + H(X_j \mid X_k). \end{aligned} \quad (2.48)$$

As a result, the set of entropy vectors compatible with a given DAG is the intersection of the entropy cone $\Gamma_{[n]}^*$ with the linear subspace L_C defined by the set of linear constraints that characterize the CIs associated with the DAG [94, 95]. In practice, we

again rely on the outer approximation given by the intersection of the Shannon cone $\Gamma_{[n]}$ with L_C .

If all the variables in a DAG are observable, in order to check the compatibility of a given entropy vector with the DAG it suffices to check whether all the entropic CIs are satisfied. However, we are often interested in DAGs containing latent, non-observable, variables (see Sec. 2.3.5). Splitting the n variables making up the DAG into j observable variables O_1, \dots, O_j and $n - j$ latent variables $\Lambda_1, \dots, \Lambda_{n-j}$ we thus need to compute the marginal Shannon cone $\Pi_{\mathcal{M}}(\Gamma_{[n]} \cap L_C)$ where $\mathcal{M} = \{\{O_1, \dots, O_j\}\}$.

2.5.4 The entropic characterization of counterfactuals

Application of the above method to the Bell scenario (see Fig. 2.7a) fails to provide nontrivial constraints, a result that can be extended to a larger class of “line-like” Bayesian networks [23]. However, it has been known for some time that entropic Bell inequalities can be derived using different methods [98]. Interestingly, these inequalities can even be turned into necessary and sufficient conditions for a given probability distribution to satisfy Bell’s local causality assumption [93].

The method that allows such inequalities to be derived is motivated by the realization that the entropic approach can be applied to any marginal scenario for a relevant set of random variables [92], and not only those arising from causal Bayesian networks. In particular, when we are interested in constraints on conditional distributions of the form $P(ab|xy)$, where we have distinct sets of input and output variables, we may consider the output variables conditioned on certain relevant input variables (e.g. A_{xy} and B_{xy} , where the notation A_{xy} denotes the random variable $A|(X = x, Y = y)$).⁷ The choice of relevant input variables to condition on, as well as the appropriate probability structure, will depend on the physical situation being considered. In general, a global probability distribution may not exist on such “counterfactual” variables even if one does on the unconditioned variables.

Let us illustrate how this method may be applied by considering again its application to the Bell scenario. Instead of considering all the input and output variables as in the DAG approach (e.g. X, Y, A, B), one can consider copies of the output variables conditioned on the corresponding party’s input, i.e., A_x, B_y , where A_x denotes the random variable $A|(X = x)$. Indeed, due to the DAG constraints (no-signaling), the output variables can only depend on the corresponding local input. Furthermore, from Fine’s

⁷We focus on the bipartite case for concreteness, but the method readily generalizes to multipartite scenarios.

Theorem [60] we know that Bell’s local causality assumption is equivalent to the existence of a well defined (although empirically inaccessible) joint probability distribution $P(a_1, \dots, a_{|\mathcal{X}|}, b_1, \dots, b_{|\mathcal{Y}|})$ (where $\mathcal{X} = \{1, \dots, |\mathcal{X}|\}$ and $\mathcal{Y} = \{1, \dots, |\mathcal{Y}|\}$ denote the alphabets of Alice and Bob’s inputs) on these variables⁸ that marginalizes to the observable one given by $P(ab|xy) = P(a_x b_y)$. Hence, the appropriate probability structure for local correlations in the Bell scenario is $\mathcal{S} = [n]$ with $[n] = \{A_1 \dots, A_{|\mathcal{X}|}, B_1, \dots, B_{|\mathcal{Y}|}\}$, and we consider the Shannon cone $\Gamma_{[n]} = \Gamma^{\mathcal{S}}$ that contains all $2^{|\mathcal{X}|+|\mathcal{Y}|}$ -dimensional entropy vectors $(H(\emptyset), H(A_1), \dots, H(B_1), \dots, H(A_1 \dots A_{|\mathcal{X}|} B_1 \dots B_{|\mathcal{Y}|}))$. The marginal scenario in this case is simply $\mathcal{M} = \{\{A_x, B_y\}\}_{x,y}$ and local correlations are then characterized by the cone $\Pi_{\mathcal{M}}(\Gamma_{[n]})$.

In contrast to the characterization based directly on the DAG variables, this approach leads to nontrivial entropic inequalities (i.e., not obtainable from the elemental inequalities in Eqs. (2.44)) in the Bell scenario. For example, for two measurement settings per party, which we take as $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, one obtains the Braunstein-Caves inequality [98] together with its symmetries obtained by relabeling the inputs, namely,

$$I(A_0 : B_0) + I(A_0 : B_1) + I(A_1 : B_0) - I(A_1 : B_1) - H(A_0) - H(B_0) \leq 0, \quad (2.49)$$

where $I(A_x : B_y) := H(A_x) + H(B_y) - H(A_x B_y)$ is the mutual information between the variables A_x and B_y . This inequality can be understood as the entropic counterpart of the paradigmatic CHSH inequality [56].

In general (i.e., beyond the simplest Bell scenario), both methods based on the variables in a causal Bayesian network and on counterfactual variables can lead to nontrivial constraints [64, 75, 94–96, 99]. To conclude this section, let us nonetheless highlight an important difference between the two methods: while the former is valid for arbitrary input alphabets, the latter fixes the number of inputs to which the inequalities apply.

Although the choice of probability structure above corresponds, via Fine’s theorem, to the assumption of a local hidden variable theory, one can also consider other possibilities. For instance, taking $\mathcal{S} = \mathcal{M}$ amounts to assuming a nonsignaling theory [88]. In this case, the entropy cone is characterized only by the Shannon inequalities and one can obtain a characterization of the extremal rays of the cone, corresponding to the entropic analogue of Popescu-Rohrlich (PR) boxes [96].

The approach of entropic inequalities appears in several parts of this thesis. In Chapter 5 we will provide a modification to this approach. In particular, we will show

⁸In particular, by invoking Fine’s Theorem we do not need to explicitly include the hidden variable Λ in this method, contrary to the DAG method outlined previously.

that depending on the hypergraph of marginal scenario additional independence constraints may apply. Such additional linear constraints lead to a tighter description of the marginal polyhedron $\Gamma_{\mathcal{M}}$, i.e. to non-Shannon-type inequalities.

Moreover, in Chapter 6 we will apply the entropic approach to study correlations with indefinite causal order (see Sec. 2.3.7). We will provide analysis of both characterizations, Bayesian networks and counterfactuals, of processes compatible with a definite causal order. We will show that both characterizations lead to nontrivial constraints, i.e. entropic causal inequalities.

2.6 Semidefinite Programming

Finally, we will give a brief introduction to the tool of semidefinite programming, used in many parts of this theses. A semidefinite programming SDP [61] optimization problem can be formulated as

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & F(x) = F_0 + \sum_{i=1}^m x_i F_i \\ & F(x) \geq 0. \end{aligned} \tag{2.50}$$

The vector $c \in \mathbb{R}^m$ and the $m + 1$ symmetric matrices $F_0, \dots, F_m \in \mathbb{R}^{n \times n}$ represent the problem data, while the $F(x) \geq 0$ constraint means that $F(x)$ is a positive semidefinite matrix, $z^T F(x) z \geq 0$, or, alternately, all eigenvalues of $F(x)$ are non-negative.

To a given SDP one can associate the so-called dual semidefinite program, which is of the form

$$\begin{aligned} \max_Z \quad & -\text{tr}(F_0 Z) \\ \text{subject to} \quad & \text{tr}(F_i Z) = c_i, \\ & Z \geq 0, \end{aligned} \tag{2.51}$$

for all $i = 1, \dots, m$. The weak duality theorem state that if both primal and dual SDPs are feasible, then

$$\min(c^T x) \geq \max(-\text{tr}(F_0 Z)). \tag{2.52}$$

This, in particular means that if the dual problem is infeasible, the primal one is unbounded. In this way the dual SDP provides a certificate for the bound of the primal program, namely the matrix Z from 2.51, which give the lower bound on the

minimum of the primal problem. This turns the numerical results into analytical proofs. An example of certificates for deciding whether the state is PPT are exactly the decomposable witnesses (2.13). In this case SDP takes the form

$$\begin{aligned} & \max_W \quad -\operatorname{tr}(\rho W) \\ & \text{subject to } W = P + Q^{T_A}, \\ & \quad P \geq 0, Q \geq 0, \end{aligned} \tag{2.53}$$

where ρ is an entangled state for which we wish to find a witness. Other examples of the application of SDPs in problems in foundations of quantum mechanics are in deriving upper bounds on violation of Bell inequalities by quantum states [58, 100], characterization of the phenomenon called steering [1, 101].

SDPs can be solve efficiently with freely accessible tools [102–104] available for MATLAB and Python.

In Section 2.4 we have introduced the quantum marginal problem where the task is to deduce properties of the global quantum system from information contained in the subsystems, or marginals [91]. Usually, the properties that one wants to decide upon are existence and uniqueness of the global state. Let us consider an example where we have access to a set of two-body marginal states ρ_{AB} , ρ_{AC} , and ρ_{BC} . As it can be clear from Section 2.6 one can use SDP to decide whether there exist a state ρ_{ABC} for which the given marginal states ρ_{AB} , ρ_{AC} , and ρ_{BC} are its reduced states. This SDP can be formulated as follows

$$\begin{aligned} & \underset{\rho}{\text{minimize:}} \quad 0, \\ & \text{subject to:} \quad \operatorname{tr}_A(\rho) = \rho_{BC}, \operatorname{tr}_B(\rho) = \rho_{AC}, \operatorname{tr}_C(\rho) = \rho_{AB}, \\ & \quad \rho \geq 0, \end{aligned} \tag{2.54}$$

where as before tr_A denotes the partial trace of the subsystem A and so on. If a solution ρ_{ABC} to SDP (2.54) is found, one might want to determine whether this solution is unique. Unfortunately, this question cannot be answered with certainty using SDPs, however, one can modify the above problem (2.54) to get an evidence whether this is true or not. Indeed, if instead of 0 in the objective function in (2.54) one puts the minus trace distance to the known solution ρ_{ABC} ($\operatorname{tr}(\rho - \rho_{ABC})$), and as a result gets another state, then, clearly, the global state for marginals ρ_{AB} , ρ_{AC} , and ρ_{BC} is not unique. On the other hand, if one arrives at the same state ρ_{ABC} , it provides a strong evidence that the state ρ_{ABC} is unique.

In this thesis we will use SDPs, similar to (2.54) and (2.53), to find states for which it is possible to detect multipartite entanglement in these states from their separable two-body marginals (see Chapter 3).

Chapter 3

Multiparticle entanglement as an emergent phenomenon

This chapter describes the results of Refs. [26, 29]. The corresponding sections from the introduction chapter 2 are Sec. 2.1, 2.4, and 2.6.

In this chapter we address a question when and under which circumstances the global entanglement of a multiparticle quantum system can be inferred from its local properties. Such questions are of great relevance for the theory of quantum correlations as well as for experimental implementations. In particular, we present a method to systematically find quantum states, for which the two- or three-body marginals do not contain any entanglement; nevertheless, the knowledge of these reduced states is sufficient to prove genuine multiparticle entanglement of the global state. With this we show that the emergence of global entanglement from separable local quantum states occurs frequently and for an arbitrary number of particles. After, we discuss various extensions of the phenomenon. First, we consider a problem of detecting genuine multiparticle entanglement from a subset of two-body marginals with an emphasis on configurations where only nearest-neighbor correlations are known. Finally, we present examples where global entanglement can be proven from marginals, even if entanglement cannot be localized in the marginals with measurements on the other parties.

3.1 Emergence of entanglement

The relations between global properties of a system and the properties of its parts are central for many debates in science. One example, is the quantum marginal problem,

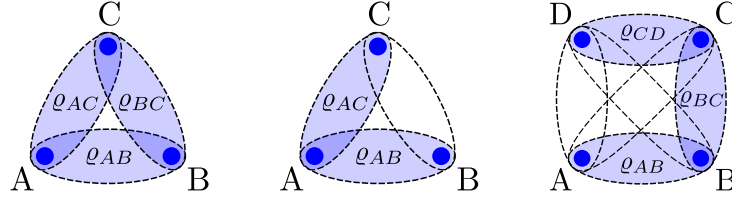


FIGURE 3.1: Possible applications of method provided in this chapter. Left: The method finds systematically three-particle states with no entanglement in the two-qubit marginals ρ_{AB} , ρ_{BC} , and ρ_{AC} , where nevertheless global genuine multipartite entanglement can be proven from knowing the separable marginals only. Middle, right: Possible variations of the problem considered in this work. All two-body reduced states are required to be separable (shown by dashed ellipses), but only some (denoted by blue filled ellipses) are known and sufficient to prove that all particles are entangled.

which, although, has been extensively studied for decades, remains a key problem in quantum mechanics [91]. Another interesting concept in these discussions is the notion of *emergence*, meaning that at a global scale properties may be present, which are not present in the parts of the system. An important requirement to the emerging property is that it can be anticipated already from the collective data about the parts, which do not exhibit the property themselves. In this chapter we will discuss a problem of detecting entanglement from reduced states, which are separable, i.e. do not display entanglement. This would mean that the global entanglement appear as an emergent phenomenon.

Emergence of entanglement has first been observed in spin models and in the context of spin squeezing, where the task of proving entanglement from two-body marginals arises naturally [105, 106]. The same observation has been made with the violation of Bell inequalities: here, the marginals may be compatible with a local hidden variable model, but such a model can be excluded for the global state by considering the marginals only [107, 108].

In all these works, however, only the simplest form of multipartite entanglement was considered. Namely, it was asked whether or not the global state is fully separable (see Eq. (2.14) for the definition). Clearly, proving genuine multipartite entanglement is a more demanding task than showing only non-separability and thus it is much more difficult to infer this type of entanglement from the marginals. In fact, there has been only one recent example of a three-qubit state, where the genuine multipartite entanglement can be concluded from separable marginals [109]. This state is given by the following density matrix

$$\begin{aligned} \rho &= \frac{2}{3}|\xi\rangle\langle\xi| + \frac{1}{3}|111\rangle\langle 111|, \\ \text{where } |\xi\rangle &= \frac{1}{2}|010\rangle + \frac{1}{2}|100\rangle + \frac{1}{\sqrt{2}}|001\rangle. \end{aligned} \quad (3.1)$$

However, this example and the underlying techniques, which were used to find it, are tailored to a specific situation and are not straightforward to generalize in any direction.

In this chapter, we present a systematic method to find multiparticle quantum states where genuine multiparticle entanglement can be proven from separable marginals (see Fig. 3.1). Implementing this method numerically we find various examples for up to five particles. Based on these findings we provide a scheme for constructing of states with the desired properties for *any* number of particles, proving the universality of the described phenomenon. Furthermore, we consider possible extensions of the problem. Situations where only a subset of the separable two-body marginals (for example, nearest-neighbor marginals) is sufficient to prove entanglement, can easily be identified, as well as the case, where additionally the separability of higher-order marginals is given.

Finally, a direct extension of our method delivers states where no entanglement can be generated between two arbitrary particles, even if measurements are made on the other particles. Still, global entanglement can be proven from the marginals. This shows that localizable entanglement is not a precondition for genuine multiparticle entanglement.

3.2 Statement of the problem

Let us define the problem, which we want to solve, in a precise manner. For three particles, we want to find a state ρ , such that:

- (i) All two-body marginals (reduced states) are separable,
- (ii) The state ρ itself is genuine multiparticle entangled and the entanglement can be proved from the two-body marginals.

A three-qubit GHZ state (2.21), which we mentioned in Section 2.1.2, is a genuine multipartite entangled state and it satisfies the condition (i). However, we can see that the reduced states of (2.21) are all the same $\rho_{AB} = \rho_{AC} = \rho_{BC} = (|00\rangle\langle 00| + |11\rangle\langle 11|)/2$ and also compatible with the global separable state $\rho_{sep} = (|000\rangle\langle 000| + |111\rangle\langle 111|)/2$. This means that GHZ state does not satisfy the condition (ii) since given only ρ_{AB} , ρ_{AC} , and ρ_{BC} in this case we cannot guarantee that the global state is the GHZ state (2.21). From the above reasoning it is clear that the two conditions (i) and (ii) are equivalent to the requirement that the separable known marginals must be only compatible with

global genuine multipartite entanglement. It is important to note that in our problem uniqueness of the global state is not required and the only condition is that all global states, compatible with the marginals, must be genuine multiparticle entangled.

One can extend the conditions in various ways. Concerning the first condition, one can require for four or more particles in addition to the two-body marginals that also the three-body marginals are fully separable. The second condition can be modified such that only some of the two-body marginals are known, while all of them are still separable (see Fig. 3.1). Clearly, both modifications make it more difficult to find the desired states and it is not clear, whether states with these properties exist at all. It is one of the main results of this work that for all reasonable modifications of the problem the corresponding states can be found.

3.3 Construction method

Our method for constructing states with the desired properties is formulated as an iteration process consisting of two semidefinite programs (SDPs), (see Sec. 2.6 for review on SDPs). The algorithm is based on the approach to multiparticle entanglement with PPT mixtures [37] (see as well Sec. 2.1.2). The question whether a state is a PPT mixture can directly be decided via SDPs, most importantly, it can also be decided when only partial information on the state is available (see also below).

Now we describe the steps of the searching algorithm. We formulate it first for qubits, the extension to higher-dimensional systems is explained afterwards.

Step 1. In the first part we apply the criterion for PPT mixtures to a random initial state ϱ_0 using only its marginals. For that, we have to solve the following SDP

$$\begin{aligned}
 & \underset{\mathcal{W}, P_M, Q_M}{\text{minimize:}} && \text{tr}(\mathcal{W}\varrho_0), && (3.2) \\
 & \text{subject to:} && \text{tr}(\mathcal{W}) = 1, \text{ where} \\
 & && \mathcal{W} = \sum_{ij} w_{ij}^{\alpha,\beta} \sigma_i^\alpha \otimes \sigma_j^\beta \otimes \mathbb{1}^{\otimes(N-2)} + \text{perm.} \\
 & && \text{and for all bipartitions } M|\overline{M}: \\
 & && \mathcal{W} = P_M + Q_M^{T_M}, \quad Q_M \geq 0, \quad P_M \geq 0.
 \end{aligned}$$

This optimization program means that for the state ϱ_0 we construct the optimal \mathcal{W} decomposable witness for all bipartitions $M|\overline{M}$. If the expectation value $\text{tr}(\mathcal{W}\varrho_0)$ for such a witness is negative, the state is genuine multipartite entangled (see Sec. 2.1.2). The first constraint on the witness' trace maximizes the white-noise tolerance of the

entanglement detection. The second linear constraint requires the witness to contain only two-body correlations, since we want to certify entanglement of our states from their two-body marginals only. Permutations are taken for all pairs α, β of qubits. This constraint is additional to the original PPT mixture problem and it is one of the algorithm's main features.

Step 2. In the second part of the iteration we determine the state that gives the most negative value for the given witness from the first step. In line with our approach, we require separability of the reduced two-party marginals. For the case of qubits this part can be formulated as a simple SDP, since a two-qubit state is separable, if and only if it is PPT. So we consider:

$$\begin{aligned}
 & \underset{\varrho}{\text{minimize:}} && \text{tr}(\mathcal{W}\varrho) && (3.3) \\
 & \text{subject to:} && \varrho \geq 0, \text{tr}(\varrho) = 1, \\
 & && \text{and for all } \alpha, \beta \quad \varrho_{\alpha\beta}^{T_\alpha} \geq 0,
 \end{aligned}$$

where α, β denote pairs of qubits and $\varrho_{\alpha\beta}$ is the two-qubit marginal.

Combining both steps 1 and 2 and putting the output ϱ of the second part as an input into the first part one can run an iteration process. If, at some point of the iteration the second step gives a negative value, we found already a state that has separable two-body marginals, where the entanglement can be proven from the marginals only. In practice, if a pure random state is taken as a seed of the algorithm, the iteration process typically gives a state with the desired properties after the second or the third iteration. Running the iteration further maximizes the entanglement and noise robustness of the state, while keeping the desired properties. Besides, the output states for different inputs ϱ_0 mainly differ only up to some local unitary transformation.

The algorithm can be extended to modifications of the problem in several ways. First, if higher-dimensional systems are considered, one has to use the appropriate generalizations of the Pauli matrices in the construction of the witness in Eq. (3.2). Also, in higher dimensions the PPT criterion is not necessary for separability, so for the state resulting in the second step separability of the marginals is not guaranteed. Nevertheless, if a state is found, the separability can later be checked with existing effective algorithms for proving separability of quantum states [110, 111]. Second, if only some of the marginals are known [as in Fig. (3.1)] one just has to modify the definition of the witness in Eq. (3.2) and take only correlations from the known marginals. Finally, in the case of more than three parties, if also the full separability of the three-body marginals is required, one can modify the second step by changing the conditions that

now the three-body marginals are PPT for any bipartition, and later verify the full separability of them with existing approaches [110, 111].

3.4 Results

3.4.1 Three qubits

Although one example (3.1) of a state with the desired properties for three qubits was already reported in Ref. [109], we start with this case, as this allows to explain our methods.

Taking the most robust algorithm's final state, applying local unitary operations and searching for an analytical expression, we find the following state which has separable two-body marginals which suffice to prove genuine multiparticle entanglement:

$$\begin{aligned} \varrho_N^{(3)} &= \frac{2}{3}|\xi\rangle\langle\xi| + \frac{1}{3}|\bar{W}_3\rangle\langle\bar{W}_3|, \\ |\xi\rangle &= \sqrt{\frac{1}{3}}|W_3^*\rangle + \sqrt{\frac{2}{3}}|111\rangle, \end{aligned} \quad (3.4)$$

where

$$\begin{aligned} |\bar{W}_3\rangle &= \frac{1}{\sqrt{3}}(|011\rangle + |101\rangle + |110\rangle) \\ \text{and} & \\ |W_3^*\rangle &= \frac{1}{\sqrt{3}}(e^{i\frac{\pi}{3}}|001\rangle + e^{-i\frac{\pi}{3}}|010\rangle - |100\rangle). \end{aligned} \quad (3.5)$$

The latter in Eq. (3.5) is just a W -state (2.22) with equally distributed phases. Note that besides these phases the state (3.4) itself would be permutationally invariant. However, the asymmetry is necessary, since it can be shown that for symmetric states the studied phenomenon cannot exist [106]¹. We add that one can also see the set of marginals ϱ_{AB} , ϱ_{BC} and ϱ_{AC} as the output of the algorithm, we will discuss below the extent to which the marginals determine the state completely.

To compare our result with the result from Ref. [109], we note that the phenomenon described here is not fragile, as can be seen by the white noise tolerance. For that we consider the following mixture of the target state and white noise

$$\varrho(p) = (1-p)\varrho_N^{(3)} + p\mathbb{1}/8, \quad (3.6)$$

¹Nevertheless, we found with our algorithm permutationally invariant states with the desired properties.

and ask for which values of p the state has still the desired properties. Clearly, the marginals remain separable if white noise is added, so one only has to check whether the three-qubit entanglement can be proven from the marginals. One finds that the state in Eq. (3.4) remains this property with possible 13.7% of white-noise added, which means that this phenomenon is quite robust and it is realistic that this phenomenon can be observed experimentally. In addition, the white-noise tolerance of the state itself is 28.6%, but then the entanglement cannot be proven from the marginals only. For comparison, the state (3.1) keeps its properties only up to 5.2% of possible white-noise. Note as well that such estimates are impossible with the methods from Ref. [109], this it is another advantage of our algorithm.

3.4.2 Four and five qubits

For four qubits we find many analytical examples of states that have the desired properties. Remarkably, there are now also pure states with separable marginals, from which entanglement can be proved. One of the simplest solutions for four qubits is a Dicke-type [44] state without one term and with one π -phase:

$$|N^{(4)}\rangle = \frac{1}{\sqrt{5}}(|0011\rangle + |0101\rangle + |0110\rangle + |1001\rangle - |1010\rangle). \quad (3.7)$$

After local unitaries this state may also be expressed as a cluster state [45] with an extra term $|N^{(4)}\rangle = 2/\sqrt{5}|CL_4\rangle + 1/\sqrt{5}|0110\rangle$. This state has 21.2% white-noise tolerance (here and later we mean by white-noise tolerance the tolerance of the studied properties). A further property is that the state $|N^{(4)}\rangle$ is uniquely determined by its two-body marginals, this fact has far-reaching consequences as we will see below.

There are other four-qubit states with the desired properties. For example, the most noise-robust four-qubit state that we found contains a Dicke part and a GHZ part with the Dicke part having asymmetric amplitudes. It is given by:

$$\begin{aligned} |\Psi^{(4)}\rangle &= \frac{1}{\sqrt{2}}|\eta\rangle + \frac{1}{\sqrt{2}}|GHZ_4\rangle, \\ |\eta\rangle &= \frac{1}{2\sqrt{2}}(-i|0011\rangle + |0101\rangle + |0110\rangle - \sqrt{3}|1001\rangle + i|1010\rangle + |1100\rangle). \end{aligned} \quad (3.8)$$

This state has white-noise tolerance of 22.4% and it is the closest one to the numerical solution.

With our method we were able to go as far as to six-qubit case. The following state is a genuine multiparticle entangled five-qubit state with separable marginals, which

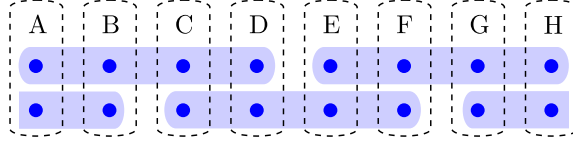


FIGURE 3.2: Illustration of the construction of the desired state for eight parties. Dashed lines represent different parties, each party possesses two qubits. Blue rounded rectangles depict the entangled four-qubit state $|N^{(4)}\rangle$ from Eq. (3.7).

are compatible with this state only.

$$\begin{aligned}
 |\Psi^{(5)}\rangle = & \frac{1}{\sqrt{6}}|00\rangle|000\rangle + \frac{1}{\sqrt{8}}|11\rangle(\sqrt{\frac{2}{3}}|000\rangle - |001\rangle - |010\rangle - |100\rangle) \\
 & - \frac{1}{\sqrt{48}}|01\rangle(\sqrt{2}|001\rangle + \sqrt{2}e^{i\frac{2\pi}{3}}|010\rangle + e^{i\frac{\pi}{3}}|011\rangle + \sqrt{2}e^{-i\frac{2\pi}{3}}|100\rangle + e^{-i\frac{\pi}{3}}|101\rangle - |110\rangle) \\
 & - \frac{1}{\sqrt{48}}|10\rangle(\sqrt{2}|001\rangle + \sqrt{2}e^{-i\frac{2\pi}{3}}|010\rangle + e^{i\frac{2\pi}{3}}|011\rangle + \sqrt{2}e^{i\frac{2\pi}{3}}|100\rangle + e^{-i\frac{2\pi}{3}}|101\rangle + |110\rangle).
 \end{aligned} \tag{3.9}$$

This state corresponds to 17.3% white-noise tolerance.

3.4.3 Generalization to more particles

Using the state in Eq. (3.7) one can construct examples of states for an arbitrary number of particles. We can formulate:

Observation 3.1. For any number of particles greater than four one can find a pure state with separable two-body marginals, where genuine multiparticle entanglement can be proven from these marginals only.

We illustrate the construction for eight parties, where each of the parties possesses a four-level system, represented by two qubits, so the total system consists of 16 qubits. We can distribute four copies of the state $|N^{(4)}\rangle$ in Eq. (3.7) as depicted in Fig. (3.2). Now every two-party marginal is separable since it is a direct product of two two-qubit separable states. Furthermore, knowledge of the two-party marginals implies knowledge on the states $|N^{(4)}\rangle$ and their distribution among the 16 qubits, since $|N^{(4)}\rangle$ is uniquely determined by its marginals. So the state of eight parties is uniquely determined by its two-body marginals as well. The global state is pure and does not factorize for any bipartition of the eight parties, so it is genuine multiparticle entangled and the entanglement can be proven from the separable marginals.

A similar reasoning is valid for any number of parties. For five parties one takes two copies of the state $|N^{(4)}\rangle$ and distributes the first copy to the parties A, B, C, D and the second copy to the parties B, C, D, E , for other numbers of qubits the construction is

analogous. Note that the purity of the state $|N^{(4)}\rangle$ is essential for the argument, so the construction would not work with other states, e.g. the state from Eq. (3.1). Also, we needed that the state is uniquely determined by its two-qubit reduced states. Below we discuss this property for the other states we found using the method.

3.4.4 Uniqueness of the global state

The property of uniqueness of the global state was noted before for the state from Eq. (3.1). In fact, this is the case for both three and four-qubit states from the Eqs. (3.4, 3.7). However, this property does not necessarily follow from the constraints and a counterexample for four qubits is the following state:

$$|\Phi^{(4)}\rangle = \frac{1}{\sqrt{2}}|\widetilde{D}_2^4\rangle + \frac{1}{\sqrt{2}}|GHZ_4\rangle \quad (3.10)$$

with

$$|\widetilde{D}_2^4\rangle = \frac{1}{\sqrt{6}}\left(|0011\rangle + |0101\rangle + |0110\rangle + e^{i\varphi}|1001\rangle + e^{i\varphi}|1010\rangle + e^{-i\varphi}|1100\rangle\right), \quad (3.11)$$

where $\varphi = \arccos(-1/3)$ and $|GHZ_4\rangle = (|0000\rangle + |1111\rangle)/\sqrt{2}$ is a four-qubit GHZ state (2.21). In this case, the two-qubit marginals are also compatible with the state of the same form as Eq. (3.10), but with the opposite phases $\varphi \rightarrow -\varphi$, meaning that there is a set of states, which are convex combinations of $|\Phi^{(4)}(\varphi)\rangle$ and $|\Phi^{(4)}(-\varphi)\rangle$ and which are compatible with the same reduced two-qubit states. This demonstrates that in our problem the set of two-body reduced states need not be compatible with only one global state, the only condition is that set of global states, compatible with these marginals, must be enclosed into the set of genuine multiparticle entangled states.

3.5 Extensions of the problem

In this section we will consider several problem's extensions, which we mentioned in the beginning of this chapter.

3.5.1 Separability of the higher-order marginals

First, one may ask whether there is any four-qubit state with separable two *and* three-body marginals, where the genuine multiparticle entanglement can be proven from the two-qubit reduced states. To find the desired state we have used our iterative

method, in which in the SDP of Step 2 (3.3) we additionally require that the three-body marginals are also PPT. Still in the SDP of Step 1 (3.2) we try to prove entanglement from the two-body marginals. Our method results in the following state

$$\begin{aligned}
 \rho_N^{(4)} &= \frac{1}{2}|\zeta_1\rangle\langle\zeta_1| + \frac{1}{2}|\zeta_2\rangle\langle\zeta_2|, \\
 |\zeta_1\rangle &= \sqrt{\frac{4}{5}}|GHZ_4\rangle + \sqrt{\frac{1}{5}}|\Psi^+\rangle_{AB} \otimes |\Psi^+\rangle_{CD}, \\
 |\zeta_2\rangle &= \sqrt{\frac{2}{5}}(|0011\rangle + |1100\rangle) + \sqrt{\frac{1}{5}}|\Psi^-\rangle_{AB} \otimes |\Psi^-\rangle_{CD},
 \end{aligned} \tag{3.12}$$

where $|\Psi^\pm\rangle$ are the Bell states (2.7) with the subscript denoting the parties which share these states. The state (3.12) has the described properties, even if up to 21.8% of white-noise are added. However, while the PPT property of the two-body marginals implies separability this is not the case for the three-body marginals: Even if a three-qubit state is PPT for all bipartitions, this does not mean that it is fully separable. Therefore, we checked whether the three-qubit marginals are fully separable with the separability testing algorithm from Ref. [110] and found that if more than 13.5 % of white noise is added, the marginals are indeed fully separable. This proves that the state $\rho(p) = (1-p)\rho_N^{(4)} + p\mathbb{1}/16$ for $p \in [0.135, 0.218]$ has the desired properties.

3.5.2 Proving entanglement from a subset of marginals

Second extension, which we have considered, is the possibility to detect entanglement from only a part of two-body correlations, while we still require all two-body marginals to be separable. Clearly, we will only consider the situation where every particle is at least in one of the measured marginals, otherwise, the state where this particle is in a product with the rest of the particles will always be consistent with the data.

In order to look for such states with our iterative method, we modify the SDP (3.2) of Step 1 and keep the SDP of Step 2 unchanged. In (3.2) we put a further restriction on \mathcal{W} such that it contains terms corresponding to sets $\{\alpha, \beta\} \in \mathcal{M}$, for some configuration \mathcal{M}^2 .

For the three-qubit case there is only one possible configuration (up to permutations of parties), which is shown on Fig. 3.1 (middle), where $\mathcal{M} = \{\{AB\}, \{AC\}\}$, which means that only marginals ρ_{AB} and ρ_{AC} are known. Below is an example of a state

²Configuration \mathcal{M} is very similar to the notion of the marginal scenario from Sec. 2.4, however, we will keep it distinct since we use the notion of marginal scenario in the context of probability distributions.

with this properties.

$$\begin{aligned} \varrho_{\mathcal{M}}^{(3)} &= \frac{1}{2}|\xi_1\rangle\langle\xi_1| + \frac{1}{2}|\xi_2\rangle\langle\xi_2|, \\ |\xi_1\rangle &= \sqrt{\frac{1}{10}}(\sqrt{5}|000\rangle + \sqrt{4}e^{-i\frac{3}{4}\pi}|011\rangle + e^{-i\frac{3}{4}\pi}|101\rangle), \\ |\xi_2\rangle &= \sqrt{\frac{1}{10}}(\sqrt{3}(|001\rangle + e^{i\frac{2}{3}\pi}|010\rangle + e^{-i\frac{1}{3}\pi}|100\rangle) + |111\rangle). \end{aligned} \quad (3.13)$$

This effect has about 5% of white-noise tolerance.

For the four-qubit case we first note that the genuine multiparticle entanglement of the state from the Eq. (3.10) can be detected by knowing only ϱ_{AB} , ϱ_{BC} , and ϱ_{CD} with possible 3% of white noise added. This configuration of marginals is shown on Fig. 3.1 (right). However, the state (3.10) was not tailored for this configuration and the fact that genuine multiparticle entanglement of this state can be detected from only three out of six marginals manifests the robustness of the described phenomenon for four-party case.

One can go further and consider various configurations \mathcal{M} . It is, in particular, interesting to look for \mathcal{M} corresponding to acyclic graphs. Physically it means that we have access only to the nearest-neighbor correlations which often happens in practice. In Table 3.1 we show various configurations of marginals for qubit systems up to 6 particles. For examples of states and more details, please, see Refs. [29, 112].

Number of qubits	Considered configurations
4 qubits	
5 qubits	
6 qubits	

TABLE 3.1: Considered configurations \mathcal{M} . Interestingly, the optimal state for the first configuration of 4 qubits, which was found by the algorithm, is mixed. However, as we already noted, the state from Eq. (3.10) has the same properties and is pure. Thus, for every considered configuration there is pure state with the desired properties. More details can be found in Ref. [29].

Concerning the problem of detecting genuine multipartite entanglement from separable nearest-neighbor marginals we can reason, similar to Observation 3.1, that it is always possible to construct a state displaying this phenomenon for an arbitrary number of parties and an arbitrary configuration graph. For an example see Fig. 3.3.

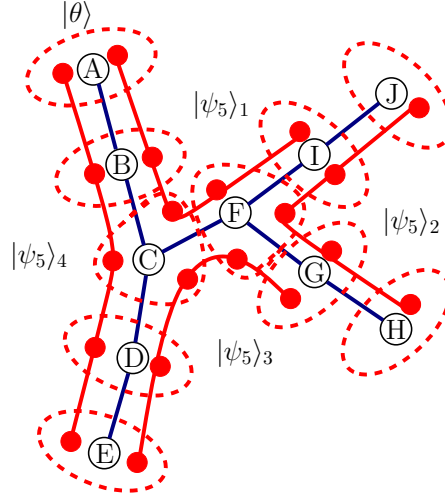


FIGURE 3.3: Demonstration of the construction of a state $|\theta\rangle$ with desired properties for some configuration for parties $A \dots H$ by distributing copies of 5 qubit state $|\psi_5\rangle_i$ of line configuration to the parties.

3.5.3 Higher-dimensional systems

As a further extension, we found also examples for three qutrits (three-dimensional systems). When considering higher dimensions, it is worth noting that for every example of states, which we present in this chapter, there exist trivial extensions to higher dimensions. For instance, if we take a three-qubit state $\varrho_N^{(3)}$ from Eq. (3.4), then, as noted in Ref. [109], a following family of states

$$\varrho_d^{(3)} = p_1 \varrho_N^{(3)} + \sum_{m=2}^d p_m |mmm\rangle\langle mmm|, \quad \text{for } (p_1 > 0, p_m \geq 0), \quad (3.14)$$

satisfy all the conditions of the desired states. However, this extension is rather trivial and does not give the best possible effect for qudit states. Using our algorithm for the three-qutrit case, we have found the following state.

$$\begin{aligned} \varrho_3^{(3)} &= \frac{1}{2} |\eta_1\rangle\langle\eta_1| + \frac{1}{2} |\eta_2\rangle\langle\eta_2|, & (3.15) \\ |\eta_1\rangle &= \frac{1}{\sqrt{12}} (|000\rangle - |222\rangle) - \frac{i\sqrt{5}}{6} (|012\rangle + |021\rangle - |102\rangle + |120\rangle + |201\rangle + |210\rangle), \\ |\eta_2\rangle &= \frac{1}{\sqrt{6}} |111\rangle - \frac{\sqrt{5}}{6} (|012\rangle - |021\rangle + |102\rangle + |120\rangle + |201\rangle - |210\rangle). \end{aligned}$$

for which the two-body marginals are PPT, but genuine multiparticle entanglement can be proven from them. The white-noise tolerance of the discussed properties of the state is 29.5%. Separability of its reduced states can be proven using the algorithm [110] if 5.3% of white-noise is added. So, the three-qutrit state $\sigma = (1 - p)\varrho + p\mathbb{1}/27$

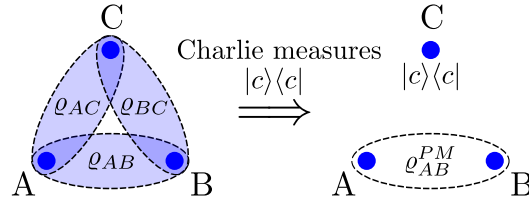


FIGURE 3.4: Illustration of the emergence of multipartite entanglement in the case of vanishing localizable entanglement. The post-measurement state $\varrho_{AB}^{PM} \sim \langle c| \varrho_{ABC} |c\rangle$ is separable after arbitrary measurements of Charlie, but knowing the reduced states ϱ_{AB} , ϱ_{AC} , and ϱ_{BC} is still enough to prove genuine tripartite entanglement of the original state ϱ_{ABC} .

has for $p \in [0.053, 0.275]$ the property that genuine multipartite entanglement can be proven from separable marginals.

Judging by the simulations we made, it seems that it is possible to find a state for any imaginable configuration of measurable correlations as soon as they contain correlations between every bipartition of the system.

3.5.4 No localizable entanglement in the marginals

Finally, we demonstrate that an extension of our method can be used to find a three-qubit state, where the marginals do not contain any entanglement, even after arbitrary measurements on the third particle. Nevertheless, global entanglement can be proved from the marginals only.

To start, one can ask whether the two-particle marginals are also separable, if the third party has made a measurement, described by a projector $|c\rangle\langle c|$ (see Fig. 3.4). For projective measurements, the post-measurement state is described by $\varrho_{AB}^{PM} \sim \langle c| \varrho_{ABC} |c\rangle$ and one can see that if this state is separable for all $|c\rangle$ then the marginal is also separable, if Charlie makes a generalized measurement or an operation³. So the question arises whether there exist genuine tripartite entangled states with the property that whatever projective measurement one party performs, the remaining two parties share a separable state. One can think about it as a state with no creatable entanglement in its subsystems, or a state with vanishing localizable entanglement [113]. In addition, we want the entanglement to be provable from the two-body marginals q_{ij} , but the knowledge of the conditional states q_{ij}^{PM} should not be required for the entanglement proof.

³The most general marginal after an operation on C is given by $\varrho_{AB}^{PM} = \text{Tr}_C[\mathbb{I}_A \otimes \mathbb{I}_B \otimes \Lambda_C(\varrho_{ABC})]$, where $\Lambda_C(\rho) = \sum_k A_k \rho A_k^\dagger$ with $\sum_k A_k^\dagger A_k \leq \mathbb{1}$ is a trace-non-increasing operation. Using the polar decomposition $A_k = U[s_1|c\rangle\langle c| + s_2|c^\perp\rangle\langle c^\perp|]$ with $s_i \geq 0$ we have for a single term that $\text{Tr}_C[A_k(\varrho_{ABC})A_k^\dagger] = s_1\langle c|\varrho_{ABC}|c\rangle + s_2\langle c^\perp|\varrho_{ABC}|c^\perp\rangle$ is separable as a convex combination of separable states, so also ϱ_{AB}^{PM} is separable.

To look for such a state one should add an infinite number of additional constraints to the semidefinite program (3.3) of the Step 2 in the algorithms, corresponding to all possible local projective measurements. This is, of course, infeasible. However, it is sufficient to add a *finite* number of constraints corresponding to some measurements (preferably equally distributed over the Bloch sphere) and require a strict positivity (e.g., $[\varrho_{AB}^{PM}]^{T_A} \geq \epsilon \mathbb{1}$ for some $\epsilon > 0$) of the partial transposition of the post-measurement states. If a state is finally found, one can check by direct numerical evaluation whether the post-measurement states are separable even for arbitrary measurements.

Applying this in practice, one directly finds states with the desired properties, an example is given below.

$$\varrho_L^{(3)} = \frac{1}{3}|\chi_1\rangle\langle\chi_1| + \frac{1}{3}|\chi_2\rangle\langle\chi_2| + \frac{1}{6}|\chi_3\rangle\langle\chi_3| + \frac{1}{6}|\chi_4\rangle\langle\chi_4|, \quad (3.16)$$

where the eigenvectors are

$$\begin{aligned} |\chi_1\rangle &= \sqrt{\frac{5}{21}}(|001\rangle + e^{-i\frac{\pi}{6}}|010\rangle + e^{-i\frac{3}{4}\pi}|100\rangle) + \sqrt{\frac{2}{21}}(e^{i\frac{\pi}{5}}|011\rangle + e^{i\pi}|101\rangle + e^{i\frac{\pi}{9}}|110\rangle), \\ |\chi_2\rangle &= \frac{1}{3}e^{i\frac{4}{5}\pi}|000\rangle + \sqrt{\frac{3}{7}}|111\rangle + \sqrt{\frac{1}{42}}(e^{i\frac{5}{6}\pi}|001\rangle + e^{-i\frac{2}{3}\pi}|010\rangle + e^{-i\frac{3}{5}\pi}|100\rangle) \\ &+ \sqrt{\frac{7}{54}}(e^{-i\frac{3}{5}\pi}|011\rangle + e^{-i\frac{5}{9}\pi}|101\rangle + |110\rangle), \\ |\chi_3\rangle &= \frac{\sqrt{18}}{5}|000\rangle + \frac{1}{5}e^{i\frac{\pi}{5}}|111\rangle + \frac{\sqrt{2}}{5}(e^{i\pi}|001\rangle + e^{-i\frac{\pi}{2}}|010\rangle + e^{-i\frac{2}{5}\pi}|100\rangle), \\ |\chi_4\rangle &= \frac{1}{\sqrt{3}}(|001\rangle + e^{-i\frac{5}{6}\pi}|010\rangle + |100\rangle). \end{aligned} \quad (3.17)$$

This state keeps its properties with possible level of white noise of 2%. As discussed above, to find this state we added a finite number of constraints, corresponding to various projective measurements, and then required strict positivity of the post-measurement states. In our implementation we defined ~ 1000 constraints and required the eigenvalues of the partial transposition of the post-measurement states to be larger than $\sim 10^{-4}$. This example proves that genuine multiparticle entanglement can emerge, even if the localizable entanglement vanishes.

3.6 Conclusions

In conclusion, we have provided a method to study systematically the emergence of genuine multiparticle entanglement from separable marginals. Our findings show the rich structure of multiparticle entanglement, where essentially all possible entanglement properties of the marginals can be combined. We believe that the entanglement

properties of the resulting quantum states deserve further study, moreover, it would be interesting to observe the here described effects experimentally.

Chapter 4

Qudit hypergraph states

This chapter describes some partial results of Ref. [28]. The corresponding preliminary sections are 2.1,(in particular 2.1.4 and 2.1.3) and 2.6.

In Ref. [28] we generalize the class of hypergraph states to multipartite systems of qudits. For simple hypergraphs, the different equivalence classes under local operations are shown to be governed by a greatest common divisor hierarchy. Moreover, the special cases of three qutrits and three ququarts are analysed in detail.

4.1 Introduction

We have mentioned in the introductory chapter that the class of graph states [18] contains as a special case the whole class of cluster states, which are the key ingredients in paradigms of quantum computing, e.g., the one-way quantum computer [45]. Apart from that, graph states are used in quantum error correcting codes [114], in the construction of Bell-like theorems [115], entanglement witnesses [116], models of topological quantum computing [117], and others.

Recently, there has been an interest in the generalization of graph states to a broader class of states known as hypergraph states [19]. In a hypergraph an edge can connect more than two vertices, so hypergraph states are associated with many-body interactions beyond the usual two-body ones. Interestingly, the mathematical description of hypergraph states is still very simple and elegant and in Ref. [47] a full classification of the local unitary equivalence classes of hypergraphs states up to four qubits was obtained. Also, in Refs. [20, 47] Bell and Kochen-Specker inequalities have been derived and it has been shown that some hypergraph states violate local realism in a way that is exponentially increasing with the number of qubits. Finally, recent studies

in condensed matter theory showed that this class of states occur naturally in physical systems associated with topological phases [118]. Originally, hypergraph states were defined as members of an even broader class of states known as locally maximally entangleable (LME) states [119], which are associated to applications such as quantum fingerprinting protocols [120]. Hypergraph states are then known as π -LME states and display the main important features of the general class of LME states.

Prior to Ref. [28], hypergraph states were defined only in the multi-qubit setting, while graph states can be defined in systems with arbitrary dimensions (see Sec. 2.1.4). In higher dimensions *graph* states have many interesting properties not present in the two-dimensional setting. For example, there are considerable differences between systems where the underlying local dimensions are a prime or non-prime [50]. Another difference is the construction of Bell-like arguments for higher-dimensional systems [121].

In the present work, we extend the definition of hypergraph states to multipartite systems of arbitrary dimensions (qudits) and analyse their entanglement properties. Especially, we focus on the equivalence relations under local unitary (LU) operations or under stochastic local operations assisted by classical communication (SLOCC). Note that the whole class of qudit graph states is a special case of our formulation.

4.2 Definition of qudit hypergraph states

We introduce now the class of hypergraph states in a system with underlying finite dimension d . Given a multi-hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ (see Fig. 2.2b for an example of a multi-hypergraph), we associate a quantum state $|H\rangle$ in a d -dimensional system in the following way:

- To each vertex $i \in \mathcal{N}$ we associate a local state $|+\rangle = d^{-1/2} \sum_{q=0}^{d-1} |q\rangle$.
- For each hyperedge $e \in \mathcal{E}$ with multiplicity m_e we apply the controlled-unitary $Z_e^{m_e}$ on the state $|+\rangle^{\mathcal{N}} = \otimes_{i \in \mathcal{N}} |+\rangle_i$. Thus, the hypergraph state is defined as

$$|H\rangle = \prod_e Z_e^{m_e} |+\rangle^{\mathcal{N}}. \quad (4.1)$$

Here, m_e is a multiplicity of the hyperedge e which defines how many times the controlled Z_e is applied to $|+\rangle^{\mathcal{N}}$. The controlled unitary operation Z_e is analogous to the controlled operation, defined in Eq. (2.27) for graphs (see Sec. 2.1.4), however, in

this case the control system is not one qudit, but many. Explicitly, the controlled Z gate on n particles is given by

$$\begin{aligned} Z_e &= \sum_{q_1=0}^d \dots \sum_{q_{n-1}=0}^d |q_1 \dots q_{n-1}\rangle \langle q_1 \dots q_{n-1}| Z^{q_1 \dots q_{n-1}} \\ &= \sum_{q_1=0}^d \dots \sum_{q_n=0}^d |q_1 \dots q_n\rangle \langle q_1 \dots q_n| \omega^{q_1 \dots q_n} \end{aligned} \quad (4.2)$$

where $\omega = e^{2\pi i/d}$ as for the case of qudit graph states. For the definition of the Z gate, please, see Eq. (2.23) from Sec. 2.1.4. It is more naturally to define controlled operations over many qudits in a recursive way, for details see Ref. [28].

Note that in our definition we allow among the hyperedges $e \in \mathcal{E}$ the presence of “loops”, i.e., an edge that contains only a single vertex. Also empty edges are allowed, they correspond to a global sign. A loop of multiplicity m on vertex k means here that a local gate $(Z_k)^m$ is applied to the hypergraph state.

Equivalently, one can define a hypergraph state as the unique $+1$ eigenstate of a maximal set of commuting stabilizer operators $K_i = X_i \prod_{e \in \mathcal{E}} Z_e^\dagger$; In contrast to the case of qudit graph states, the stabilizers of hypergraph states are generally nonlocal, for more details, please, see Ref. [28].

An important special class of hypergraph states are the so-called n -elementary hypergraph states, which are those constituted of a single hyperedge e between n qudits, i.e., the state has the simple form $|H\rangle = Z_e^{m_e} |+\rangle^{\mathcal{N}}$. In Ref. [28] we provide full SLOCC classification of n -elementary hypergraph states, which we will briefly mention in the next section.

4.3 SLOCC and LU classes of hypergraphs

We have already discussed the importance of SLOCC classification of entangled states in Section 2.1.3. In the beginning of this section we will only state the general result of Ref. [28] on SLOCC-equivalence in n -elementary hypergraph states and later we will provide in detail a complete classification of tripartite hypergraph states in dimension 3 and 4.

4.3.1 Elementary hypergraphs

Before going to the classification of hypergraph states, we will need to mention the following

Proposition 4.1. *Let $k, k' \in \mathbb{Z}_d$ be such that $\gcd(d, k) = \gcd(d, k') = g$. Then there exists a Clifford operator S such that $S(Z_e)^k S^\dagger = (Z_e)^{k'}$.*

For the definition of Clifford operations see Eqs. (2.24-2.26) from Sec. 2.1.4. For the proof of the above proposition see Ref. [122] or Ref. [28].

Using Proposition 4.1 we can prove the following

Theorem 4.2. *For a d -dimensional n -partite system, two n -elementary hypergraph states with hyperedge multiplicities k and k' are equivalent under LU, and hence also under SLOCC if $\gcd(d, k) = \gcd(d, k')$. For the case that $\gcd(d, k) \neq \gcd(d, k')$ the states are inequivalent under SLOCC.*

Here $\gcd(d, k)$ denotes the greatest common divisor of d and k . For the complete proof of Theorem 4.2, please, see Ref. [28]. From the above theorem it follows that the number of different elementary hypergraph SLOCC classes is the number of different values (modulo d) of $\gcd(d, k)$, which is obviously the number of divisors of d . It is remarkable that in this case SLOCC equivalence is the same as equivalence under LU operations. For d prime, all values $k \in \mathbb{Z}_d$ are obviously coprime with d and hence the following implication is straightforward:

Corollary 4.3. *For d prime, all n -elementary hypergraph states are equivalent under SLOCC.*

Finally, we note that also some other hypergraphs states are LU equivalent to elementary hypergraph states. In particular, we state the following

Observation 4.4. *Action of the local gate X_i^\dagger on i -th qudit of an n -elementary hypergraph state creates a $(n - 1)$ -hyperedge on the neighbourhood of i with equal multiplicity m_e of the n -hyperedge e . Acting k times with this local gate, i.e., application of $(X_i^\dagger)^k$, results in inducing in the neighbourhood of the qudit i a $(n - 1)$ -hyperedge of multiplicity km_e .*

For the proof see Re. [28], for a demonstration see Fig. 4.1 (case (a)).

4.3.2 Tools for SLOCC classification

In this section we will explain some more refined criteria for proving or disproving SLOCC equivalence in multipartite entangled states. As already mentioned in the introductory chapter (Sec. 2.1.3), the rank of the reduced states is a simple way of identifying inequivalences.

To find a finer distinction we employ a method based on Ref. [43] that uses a $1|23\dots n$ split of the system to identify types of inequivalent bases of the $(2, 3, \dots, n)$ -subspace, which results in a lower bound on the number of actual SLOCC-classes. As we want to infer for a given state its SLOCC-class, there remains the following problem to be solved: identifying the basis which has minimal entanglement in its basis vectors. Accordingly, we refer to this tool as minimally entangled basis (MEB) criterion. A major disadvantage of this method is that with growing number n of subsystems, the entanglement structure within the bases becomes more complex, as it rises recursively from the total number of SLOCC-classes of the $(n - 1)$ -partite systems.

The MEB of an n -partite quantum state is defined as follows:

Definition 4.5. Consider a state

$$|\psi_{12\dots n}\rangle = \sum_{a_1, a_2, \dots, a_n=0}^{d-1, d-1, \dots, d-1} c_{a_1, a_2, \dots, a_n} |a_1, a_2, \dots, a_n\rangle$$

in a $d^{\times n}$ system. According to Ref. [43], we define the $d \times (d^{n-1})$ coefficient-matrix $C_{1|2\dots n}$ in the canonical basis $\{e_i\}$ as follows:

$$C_{1|2\dots n} = \sum_{a_1, a_2, \dots, a_n=0}^{d-1, d-1, \dots, d-1} c_{a_1, a_2, \dots, a_n} e_1 (e_2^T \otimes \dots \otimes e_n^T)$$

This matrix holds all information about the total state. From the singular value decomposition (SVD) of this matrix, $C_{1|2\dots n} = U_1 D V_{2\dots n}^\dagger$, we can identify a basis $\{v_k\}$ of the right subspace $(2, \dots, n)$, where individual basis vectors $|v_k\rangle$ might be entangled.

Within this framework, we define a minimally entangled basis (MEB) $\{v_k\}_{MEB}$ of $|\psi_{12\dots n}\rangle$ as the one within which the number of full product vectors is maximal under the condition that it spans the same subspace as $\{v_k\}$.

Below we will give two lemmata without the proof. The proof can be found in Ref. [28].

Lemma 4.6. *Two n -partite quantum states $|\phi\rangle, |\psi\rangle$ of the same subsystem-dimensionality and equal reduced ranks are SLOCC-inequivalent, if their MEBs have a different number of product vectors.*

In the above lemma we consider states $|\phi\rangle, |\psi\rangle$ that have equal reduced ranks, because otherwise these states are automatically SLOCC-inequivalent and there is no point in calculating their MEBs.

Notice that inequivalent MEBs can exclude SLOCC equivalence, but an equivalence of MEBs does not, in general, guarantee SLOCC-equivalence. An exception is the case where the right subspace is spanned by a complete product basis. The reason is that in this case they are SLOCC equivalent to a generalized GHZ state:

Lemma 4.7. *Two genuine multipartite entangled states $|\phi\rangle, |\psi\rangle$ of n particles of the same subsystem-dimensionality and equal reduced single-particle ranks are SLOCC-equivalent, if their MEBs are complete product bases.*

Based on the lemmata presented in this subsection we wrote computer programs which we regard as tools which we use later for classification of tripartite hypergraphs of dimension 3 and 4.

Tool # 1

The first program checks whether there exist a state ρ in the subspace spanned by a given set of pure states $|v_i\rangle$ for $i = 1, \dots, K$, $K \leq d$ that has a positive partial transpose (PPT) with respect to any bipartition (see Sec. 2.1.1 for the definition of PPT). This problem can be formulated as following semidefinite program (SDP)¹

$$\begin{aligned} \min_{\lambda} \quad & 0 & (4.3) \\ \text{subject to} \quad & \rho = \sum_{ij} \lambda_{ij} |v_i\rangle \langle v_j|, \\ & \rho \geq 0, \\ & \forall \text{ bipartitions } M | \overline{M}, \rho^{T_M} \geq 0, \\ & \lambda^\dagger = \lambda, \text{tr}(\lambda) = 1, \end{aligned}$$

where last condition means λ is a hermitian $K \times K$ matrix with trace 1, and ρ^{T_M} denotes partial transpose of matrix ρ with respect to the subsystem M .

This tool can be used to prove that there is no product vector in the right subspace $(2, \dots, n)$ of an n -partite state $|\phi\rangle$, where K is the number of basis vectors in the right subspace. If the above SDP is infeasible it implies that there is no separable state in the subspace $(2, \dots, n)$, which in turns implies that there is no product vector. If for

¹For the basic theory of SDPs see Sec. 2.6.

some other n -partite state $|\psi\rangle$ there is a product vector in the right subspace $(2, \dots, n)$ the two states $|\phi\rangle$ and $|\psi\rangle$ are SLOCC-inequivalent according to Lemma 4.6.

Tool # 2

The second program is a slight modification of the first one and it checks whether there exist a PPT state of rank K in the subspace spanned by K linearly independent vectors $|v_i\rangle$, $i = 1, \dots, K$, $K \leq d$. If the optimal value ϵ of the following semidefinite program

$$\begin{aligned}
 \min_{\lambda, \epsilon} \quad & \epsilon & (4.4) \\
 \text{subject to} \quad & \varrho = \sum_{ij} \lambda_{ij} |v_i\rangle\langle v_j|, \\
 & \varrho \geq 0, \\
 & \forall \text{ bipartitions } M|\overline{M}, \varrho^{T_M} \geq 0, \\
 & \varrho \geq \epsilon \left(\sum_i |v_i\rangle\langle v_i| \right), \\
 & \lambda^\dagger = \lambda, \text{tr}(\lambda) = 1,
 \end{aligned}$$

is greater than 0, and if the found PPT state ϱ can be proven to be (fully) separable, then by the range criterion (see Ref. [123]) it means that in the subspace spanned by $|v_i\rangle$ there are K product states which span the same subspace. This program can be used to prove SLOCC-equivalence of states $|\phi\rangle$ and $|\psi\rangle$ according to Lemma 4.7, if for both states the above conditions are satisfied for their right subspace of at least one bipartition.

Tool # 3

Finally, it is convenient to perform a numerical optimization in order to find product states $|v_i^p\rangle$, $i = 1, \dots, K'$, $K' \leq K$ in the subspace spanned by the given set of vectors $|v_i\rangle$ for $i = 1, \dots, K$, $K \leq d$. This can be done by, let us say, maximizing the purity of the reduced states (that is, $1 - \text{tr}(\varrho_M^2)$, where $\varrho_M = \text{tr}_M(\varrho)$ is the reduced state of the subsystem M) for each bipartition and minimizing the scalar product $|\langle v_i^p | v_j^p \rangle|^2$ between each pair of product vectors for each unique pair $\{i, j\}$, $i, j \in \{1, \dots, K'\}$. Minimizing the scalar products makes the program to look for linearly independent vectors which in the best case are orthogonal.

As we will see in the next section, for most of the tripartite hypergraph states of dimension 3 and 4 numerical optimization, described above, gives explicit form of

product states in the right subspace if they exist. Moreover, knowing the exact form of product states for the case where a full product basis exists for both states $|\phi\rangle$ and $|\psi\rangle$ allows us to find an explicit SLOCC transformation between these states.

4.3.3 Tools for LU classification

Let us now discuss tools how LU equivalence can be characterized. In principle, this question can be decided using the methods of Ref. [40], but for the examples in the next section some other methods turn out to be useful.

If LU equivalence should be proven, an obvious possibility is to find directly the corresponding LU transformation. For proving non-equivalence, one can use entanglement measures such as geometric measure [124], since such measures are invariant under LU transformations. Another possibility is the white-noise tolerance of witnessing genuine multipartite entanglement [37] (see as well Sec. 2.1.2). The latter method works as follows: For an entangled state which is detected by a witness one can assign an upper limit of white noise which can be added to the state, such that the state can still be detected by that witness. Clearly, if two states are equivalent under LU, they have the same level of white-noise tolerance of entanglement detection. In the previous Chapter 3 we have already used the white-noise tolerance to determine the robustness of the phenomenon of detecting genuine multipartite entanglement from separable marginals.

Using the tools described above we present a classification in terms of SLOCC- and LU-equivalence of tripartite hypergraph states for dimensions 3 and 4 in the next section.

4.4 SLOCC classification of tripartite hypergraph states in dimensions 3 and 4.

We now consider the special cases of a tripartite system with prime dimension 3 and a tripartite system with the smallest non-prime dimension 4 as examples. We will be interested only in states corresponding to hypergraphs with no isolated vertices, i.e. in genuine multipartite entangled hypergraph states².

²Technically, the genuine multipartite entanglement of these states follows from witnessing it with the method from Ref. [37].

4.4.1 Classification of $3 \otimes 3 \otimes 3$

From Proposition 4.1 it follows that in dimension 3 a (hyper)edge of multiplicity 2 can be converted to a (hyper)edge of multiplicity 1 via local symplectic operations. It means that in dimension 3 we need to consider only edges and hyperedges of multiplicity 1.

In the case of a tripartite system of qutrits (3-level systems), there is only one SLOCC equivalence class of hypergraph states and two LU equivalence classes. The representatives of these two classes are the graph state and the 3-elementary hypergraph state respectively.

Equivalence of states from the first LU class (graph states) is governed by the operation called local complementation. For the details see Ref. [55]³. LU equivalence among the second LU class follows from the Observation 4.4. For instance, by acting with X^\dagger on qudit 2 of the first state in the class (see Table 4.1) we create an edge $\{1\ 3\}$, thus obtaining the second state from this LU-class. By applying further a local gate X^\dagger on the first qudit, we create an edge $\{2\ 3\}$ and an additional local loop (a hyperedge of one vertex) around qudit 3. This is a consequence of already existing edge $\{1\ 3\}$, however, clearly local loops can be removed by applying Z^\dagger . The obtained state is equivalent to the second state of this LU class up to permutation of parties ($2 \leftrightarrow 3$).

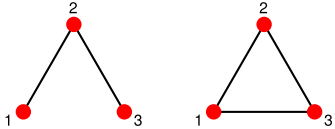
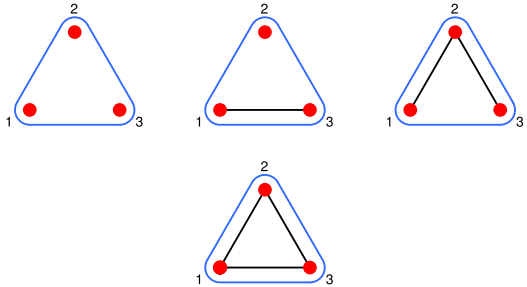
Class	Schmidt ranks	Representatives	Geom. measure/ w-noise tolerance
1	1 23 3 2 13 3 3 12 3		<p>~ 0.66 62.5%</p>
			<p>~ 0.53 $\sim 76.0\%$</p>

TABLE 4.1: Table of SLOCC and LU classes of 3-qutrit hypergraph multipartite entangled states.

³In Ref. [55] it is shown as well that both states are LU-equivalent to the GHZ state, which for d -level systems is given by $|GHZ\rangle = \sum_{i=0}^{d-1} |iii\rangle / \sqrt{d}$.

We can directly see that states from the two different LU classes are LU-inequivalent due to different values of geometric measure of entanglement and white noise tolerance (see Table (4.1)).

Finally, in order to show the SLOCC equivalence, local invertible operations connecting these two LU subclasses can be achieved by applying A_1 to one of the qutrits of the graph state and $A_{2,3}$ to the other two, where

$$A_1 = \frac{1}{4\sqrt{3}} \begin{pmatrix} -2\sqrt{3} - 2i & 4i & 4i \\ -4\sqrt{3} + 4i & \sqrt{3} + i & -5\sqrt{3} + 7i \\ -6\sqrt{3} - 2i & -\sqrt{3} - 5i & -\sqrt{3} + 7i \end{pmatrix} \quad (4.5)$$

and

$$A_{2,3} = \frac{1}{3} \begin{pmatrix} e^{i2\pi/3} & 1 & 1 \\ \sqrt{3}e^{i\pi/6} & \sqrt{3}e^{i\pi/6} & \sqrt{3}e^{i\pi/6} \\ e^{i2\pi/3} & e^{i2\pi/3} & \frac{5-\sqrt{3}i}{2} \end{pmatrix}. \quad (4.6)$$

This local operations were found with the help of the tool# 3 (numeric optimization program described in the previous section), which gives in this case full product basis for right subspaces of all states from Table 4.1.

4.4.2 Classification of $4 \otimes 4 \otimes 4$

In dimension 4 there are two types of (hyper)edges that need to be considered: ones of multiplicity 1 and ones of multiplicity 2. The hyperedges of multiplicity 3 are equivalent to the ones of multiplicity 1. As in the previous case this is the consequence of Proposition 4.1 (see also Fig. 4.1, case (b)).

In the case of a tripartite system of ququarts (4-level systems), there are five SLOCC and six LU equivalence classes of hypergraph states. All possible states with respect to permutations and equivalence of edge multiplicities are shown in the Tables 4.2, 4.3 and the interconversion between representatives within the same class are explained in detail in what follows.

Class 1

Class 1 contains hypergraph states with at least two edges of multiplicity 1 and with either no hyperedges, or with hyperedge of multiplicity 2. All these states belong to the same LU-equivalence class.

Class	Schmidt ranks	Representatives	Geom. measure/ w-noise tolerance
$\mathbf{1}$	$1 23 \quad 4$ $2 13 \quad 4$ $3 12 \quad 4$		0.75 $\sim 84.2\%$
$\mathbf{1}'$	$1 23 \quad 4$ $2 13 \quad 4$ $3 12 \quad 4$		~ 0.58 $\sim 87.1\%$

TABLE 4.2: Table of SLOCC classes $\mathbf{1}$ and $\mathbf{1}'$ of 3-ququart hypergraph multipartite entangled states.

LU-equivalence among first three state of class $\mathbf{1}$ (see Table 4.2) is governed by standard local complementation operations [55], which can be used to create a new edge of multiplicity 1 in the neighborhood of qudit 2, while applying these operations once more generates an edge of multiplicity 2 in the neighborhood of qudit 2. The same local complementation is responsible for LU equivalence among three last states this class.

To prove LU equivalence between these two subgroups of states (with no hyperedge and with 2-hyperedge) we find with the Tool# 3 explicit form of their MEBs, which appear to consist of product vectors. It can be shown then that local transformation between these states is unitary. Here we present such local unitary for transformation

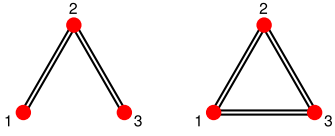
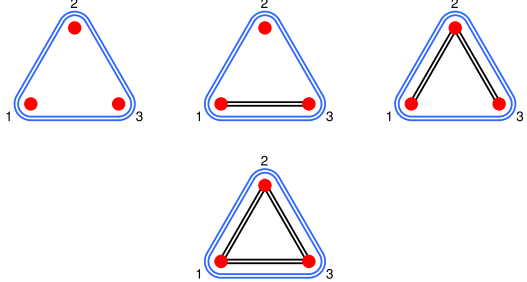
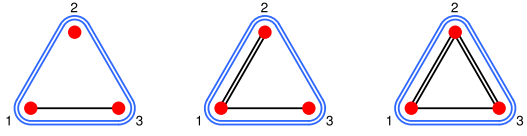
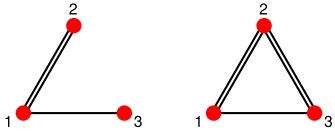
Class	Schmidt ranks	Representatives	Geom. measure/ w-noise tolerance
2	$1 23$ 2 $2 13$ 2 $3 12$ 2		0.50 $\sim 91.4\%$
			~ 0.32 $\sim 88.7\%$
3	$1 23$ 4 $2 13$ 2 $3 12$ 4		0.75 $\sim 86.1\%$
4	$1 23$ 4 $2 13$ 2 $3 12$ 4		0.75 $\sim 88.8\%$

TABLE 4.3: Table of SLOCC classes 2 – 4 and LU classes of 3-ququart hypergraph multipartite entangled states.

from the Figure 4.1 case (e).

$$U_{1,2,3} = \frac{1}{2} \begin{pmatrix} 1+i & 0 & 1-i & 0 \\ 0 & 0 & 0 & -2 \\ 1-i & 0 & 1+i & 0 \\ 0 & -2 & 0 & 0 \end{pmatrix}. \quad (4.7)$$

Class 1'

Class 1' contains all hypergraph states which have 3-hyperedge of multiplicity 1. LU equivalence of the states within this class is governed by the unitary $(X^\dagger)^m$, which, when allied to some qudit, generates edges of multiplicity m on the neighbourhood of the qudit (see Observation 4.4).

Class 2

Class 2 consists of two LU equivalence classes. The representative of the first LU-equivalence class are the graph states composed of two and three edges of multiplicity 2, though the representatives of the second LU class are the hypergraph state with a 3-hyperedge of multiplicity 2 with possible edges of multiplicity 2.

We can perform some form of "local complementation" between two states from the first LU class by applying the following unitaries in the basis $\{|p_0\rangle, |p_1\rangle, |p_2\rangle, |p_3\rangle\}$:

$$U_{1,3} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & i & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ -i & 0 & -1 & 0 \\ 0 & 0 & 0 & \sqrt{2} \end{pmatrix}; \quad U_2 = \begin{pmatrix} i & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Applying the local $(X^\dagger)^m$ unitary to some qudit of the states from the second LU class generates edges of multiplicity $2m$ (i.e., 0 or 2) on the neighbourhood of that qudit.

Using the Tool #3 one can find the local operation corresponding to SLOCC equivalence between these LU classes. For the representatives shown on the Fig. 4.1 case (d) the corresponding LO is

$$A_{1,2,3} = \frac{1}{2} \begin{pmatrix} -i(1 + \sqrt[3]{4}) & 0 & (1 - \sqrt[3]{4}) & 0 \\ 0 & 2 & 0 & 0 \\ i & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}. \quad (4.8)$$

One can easily check that $A_{1,2,3}$ is invertible, but not unitary. To show that there is no local unitary transformation possible, one can look at the entanglement measures for these LU classes (see Table 4.3).

Class 3

The representatives of class 3 are the elementary hypergraph states with a 3-hyperedge of multiplicity 2, one edge of multiplicity 1 and possible edges of multiplicity 2. These three states are in the same LU class and the local transformation between them is (X^\dagger) applied on one of the qudits.

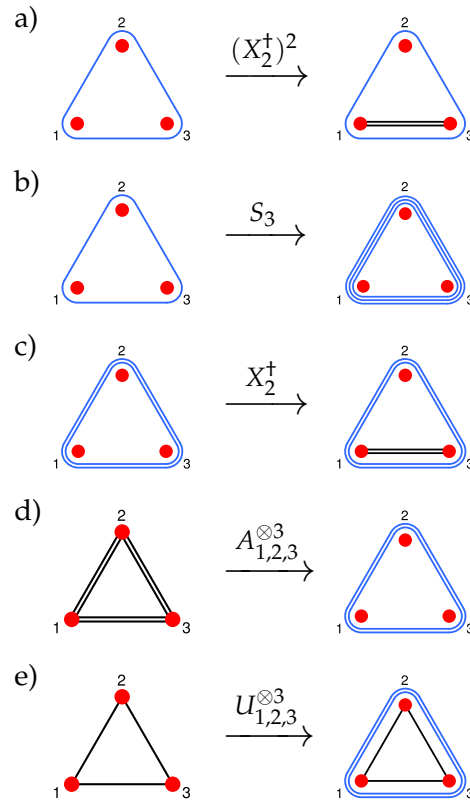


FIGURE 4.1: SLOCC equivalence among representatives of the same class.

Class 4

The representatives of class 4 are graph states composed of one or two edges of multiplicity 2 and one edge of multiplicity 1. Applying the local unitaries $U_1 = (S(1, 1, 0))^4$, $U_2 = S(1, 0, 1)$, $U_3 = S(1, 1, 0)$ to the first state creates an edge of multiplicity 2 between qudits 2 and 3.

SLOCC-inequivalence of Classes 1-4,1'

To prove the SLOCC-inequivalence of states of most of the classes it is sufficient to look at their Schmidt ranks for each bipartition (see Tables 4.2, 4.3). Exceptions are pairs of classes 1, 1' and 3, 4.

To prove that there is no SLOCC transformation between states from classes 3 and 4 let us consider the vectors from the right subspace for bipartition 2|13 for two representatives from each class. From the Schmidt decomposition of the state from class 3 one finds directly that there is at least one product vectors in the right subspace of parties 13, i.e. MEB contains at least one product vector. For the state from class 4 we can prove that in the corresponding subspace there are no product vectors in the MEB

using the Tool #1. Thus from Lemma 4.6 it follows that these states belong to different SLOCC classes.

Unfortunately, we were not able to prove SLOCC-inequivalence of states from classes 1 and 1' using the tools presented above. In fact, using the Tool #3 we found that the states from class 1 have a full product basis in their right subspace for each bipartition and the Tool #2 showed that for the states from class 1' there are states with PPT and full rank in their right subspace. However, the optimal value ϵ of the SDP of the Tool #2 for the states in class 1' was in order of 10^{-5} . Besides, the direct numerical search for SLOCC transformation bringing a states in class 1 to some state in class 1' returned states of fidelity of almost 1, though the numerical search for SLOCC transformation in the opposite direction, from a state in 1' to some state in 1, succeed in returning states of fidelity of only 0.875. This difference in fidelities of local transformations in different directions is typical for the three-qubit states of GHZ and W classes, which suggests that classes 1 and 1' are inequivalent.

4.5 Conclusions

In this work we generalized the class of hypergraph states to systems of arbitrary finite dimensions. For the special class of elementary hypergraph states we obtained the full SLOCC classification in terms of the greatest common divisor, which also governs other properties such as the ranks of reduced states. For tripartite systems of local dimensions 3 and 4, we obtained all SLOCC and LU classes by developing new theoretical and numerical methods based on the original concept of MEBs.

Some open questions are worth to mention. In the multiqubit case, hypergraph states are a special case of LME states; it would be interesting to generalize the class of LME states to arbitrary dimensions and see if a similar relation holds. Nonlocal properties of qudit hypergraph states were not a part of this work and deserve a separate consideration. Finally, possible applications of these states as a resource for quantum computing should be investigated.

Chapter 5

Indistinguishability of causal relations from limited marginals

In this chapter we will describe the results of Ref. [27]. The preliminaries for this chapter are introduced in Sections 2.3, 2.4, and 2.5 of Chapter 2.

In this work we have investigated the possibility of distinguishing among different causal structures starting from a limited set of marginals. The main tool used is the notion of adhesivity, that is, the extension of probability or entropies defined only on subsets of variables, which provides additional independence constraints among them. Our results provide a criterion for recognizing which causal structures are indistinguishable when only limited marginal information is accessible. Furthermore, the existence of such extensions greatly simplifies the characterization of a marginal scenario, a result that facilitates the derivation of Bell inequalities both in the probabilistic and entropic frameworks, and the identification of marginal scenarios where classical, quantum, and postquantum probabilities coincide.

5.0.1 Properties of graphs and hypergraphs

We will start by giving some notions from graph theory, such as tree graph, acyclic hypergraph, etc. which we will need in this chapter. As already defined in Section 2.1.4, a hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ is defined by a finite set of nodes $\mathcal{N} = \{1, \dots, n\}$ and a set of (hyper)edges corresponding to subsets of \mathcal{N} , i.e., $\mathcal{E} \subset 2^{\mathcal{N}}$. A graph \mathcal{G} is a special case of an hypergraph where edges have cardinality 2, i.e., $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, with $|E| = 2$ for all $E \in \mathcal{E}$. In this chapter we will only consider hypergraphs with multiplicity of edges 1 (i.e. no double edges). See Fig. 5.1 for examples of graphs, directed graphs, hypergraphs, and additional notions discussed below. As it will become clear later,

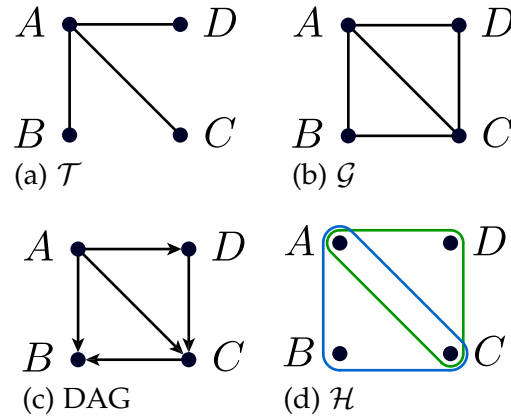


FIGURE 5.1: **Examples of graphs and hypergraphs.** (a) A tree graph \mathcal{T} where $\{B, A, D\}$ is a path. (b) A graph \mathcal{G} where $\{A, C, B, A\}$ is a loop and $\{A, B, C\}$ and $\{A, C, D\}$ are cliques. B and D are separated by $\{A, C\}$. (c) A directed graph where $\{A, C, B, A\}$ is not a directed path, because direction of (A, B) is not respected. This graph does not contain loops i.e. it is a directed acyclic graph (DAG). B and D are as well separated by $\{A, C\}$, but $\{C\}$ is a minimal separator. (d) A hypergraph \mathcal{H} where nodes B and D are separated by $\{A, C\}$.

we will be interested only in hypergraphs without isolated nodes, thus we will assume that $\mathcal{N} = \cup_{E \in \mathcal{E}} E$, when not stated otherwise, and we will sometimes denote the hypergraph simply by the set of edges \mathcal{E} .

Paths, cycles, and acyclicity are fundamental notions in graph theory. A *path* is a sequence of distinct nodes v_0, \dots, v_n (except possibly the first and last) connected by edges (v_k, v_{k+1}) $k = 0, \dots, n$, and a *closed path* or a *loop* is a path with first and last node coinciding, i.e., $v_0 = v_n$. For directed graphs, the definition is analogous with (v_k, v_{k+1}) representing a directed edge. *Acyclic graphs*, also called *tree graphs*, are graphs not containing loops. A graph is *connected* if, for every pair of nodes, there is a path connecting them.

A *clique* is a set of nodes v_0, \dots, v_n pairwise connected by an edge, i.e. $(v_i, v_j) \in \mathcal{E}$ for all $i, j = 1, \dots, n, i \neq j$. Given a graph \mathcal{G} , we can construct a hypergraph from it, called the *clique hypergraph* $\mathcal{H}_{\mathcal{G}}^{\text{cl}}$, with the same nodes and hyperedges in $\mathcal{H}_{\mathcal{G}}^{\text{cl}}$ corresponding to cliques in \mathcal{G} . Similarly, a hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ can be transformed into a graph by constructing the *2-section* $[\mathcal{H}]_2$: we connect by edges in \mathcal{G} all nodes that are connected by at least one hyperedge in \mathcal{H} . Notice that given a hypergraph \mathcal{H} , the clique graph of its 2-section will have, in general, extra hyperedges with respect to \mathcal{H} (cf. Fig. 5.3).

A hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ is a *partial hypergraph* of $\mathcal{H}' = (\mathcal{N}, \mathcal{E}')$ if for any $E \in \mathcal{E}$ there exist $E' \in \mathcal{E}'$ such that $E \subset E'$. Equivalently, we will say that \mathcal{H}' *extends*, or is an *extension* of, \mathcal{H} (cf. Fig. 5.3).

Given two disjoint subsets of nodes A, B they are said to be *separated* by a subset C if for each pair $a \in A, b \in B$, all the paths from a to b pass through C , i.e., if we remove

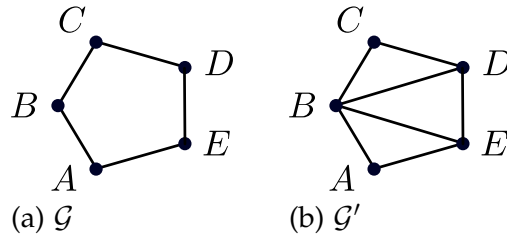


FIGURE 5.2: Example of a graph (a) and one of its possible triangulations (b).

C , A and B are no longer connected. In addition, C is called a minimal separator if $C \setminus \{v_i\}$ is no longer a separator for any $v_i \in C$.

An important notion is also that of *triangulated*, or *chordal* graphs, namely, graphs for which every cycle v_0, \dots, v_n of length $n \geq 4$, contains a *chord*, i.e., an edge connecting (v_i, v_{i+2}) . Given any graph, \mathcal{G} , additional edges can be added such that the obtained graph, \mathcal{G}' , is triangulated, and we will refer to \mathcal{G}' as the *triangulation* of \mathcal{G} , see, e.g., Fig. 5.2.

For hypergraphs, the generalization of the notions of acyclicity and tree is not straightforward and several definitions have been proposed (cf. Ref. [125]). For reasons that will be clear in later in this chapter, here we will focus on the notion of α -acyclicity, developed in the framework of database theory, which we will simply call acyclicity. There are several equivalent characterizations of this property (cf. Refs. [71, 125]), but we will focus on three of them: a characterization via the so-called *Graham algorithm*, one via the *running intersection property* of hyperedges, and the characterization as a clique hypergraph of a chordal graph.

The Graham algorithm is defined as follows. Given a hypergraph described by hyperedges $\mathcal{E} = \{E_1, \dots, E_n\}$, apply the following operations whenever they are possible

- a) Delete a node i if it appears in exactly one hyperedge.
- b) Delete a hyperedge E if $E \subset E'$ for some E' .

Acyclic hypergraphs are those for which the Graham algorithm returns the empty set. Given a hypergraph \mathcal{H} , its *reduced hypergraph* is the hypergraph obtained by applying only operation b) of the Graham algorithm.

A hypergraph has the running intersection property if there exists an ordering of the edges, E_1, \dots, E_n such that

$$S_i := E_i \cap (E_1 \cup \dots \cup E_{i-1}) \subset E_j, \text{ with } j < i. \quad (5.1)$$

In addition, for a connected and reduced hypergraph, the set $\{S_i\}$ corresponds to the *set of minimal separators* of the graph, i.e., S_i separates $R_i := E_i \setminus S_i$ from $(E_1 \cup \dots \cup E_{i-1}) \setminus S_i$. It can be proven that the running intersection property is equivalent to the empty set output for the Graham algorithm, so it can be used as an alternative definition of an acyclic hypergraph (see, e.g., [71]).

The third equivalent property is defined in terms of graphs: a hypergraph is acyclic iff its hyperedges correspond to the set of cliques of a triangulated graph (see e.g., Ref. [71]).

In order to clarify the above notions, it is instructive to apply them to the simple example depicted in Fig. 5.3. For instance, we can apply the Graham algorithm to the hypergraph in Fig. 5.3 (a). By applying operation (a), we remove the nodes A, D, F and we are left with the edges $\{\{B, C\}, \{B, E\}, \{C, E\}\}$. At this point the algorithm stops, because we cannot remove any edge via operation (b), or any other node with operation (a). The hypergraph \mathcal{H} is thus not acyclic. We can apply the same procedure to \mathcal{H}' : by removing the nodes A, D, F we are left with the edges $\{\{B, C, E\}, \{B, C\}, \{B, E\}, \{C, E\}\}$. We can then continue and remove the edges $\{B, C\}, \{B, E\}, \{C, E\}$ via operation (b), and finally the nodes B, E, C connected by a single edge $\{B, C, E\}$. The hypergraph \mathcal{H}' is thus acyclic. Equivalently, one can see that the graph \mathcal{G} , obtained as the 2-section $[\mathcal{H}]_2 = [\mathcal{H}']_2$, has as cliques exactly the hyperedges of \mathcal{H}' , but not those of \mathcal{H} . Finally, for the running intersection property of \mathcal{H}' , we can choose the ordering $E_1 = \{B, C, E\}$, and for E_2, E_3, E_4 any ordering of the remaining edges. It is clear that any intersection of edges is contained in E_1 . One can also straightforwardly check that \mathcal{H} does not have the running intersection property. Similarly, one can easily check the property of separators of the sets $S_i = E_i \cap (E_1 \cup \dots \cup E_{i-1})$, $i = 2, 3, 4$, for the hypergraph \mathcal{H}' .

As we will see in the next sections the property of α -acyclicity in hypergraphs representing marginal scenario allows for the construction of the global probability distributions from marginal distributions.

5.1 Adhesivity and independence constraints associated with a marginal scenario

The main result of this section is that when one has only partial information about (i.e., only some marginals of) a probability distribution, such marginals are always consistent with a global distribution where additional independence constraints are

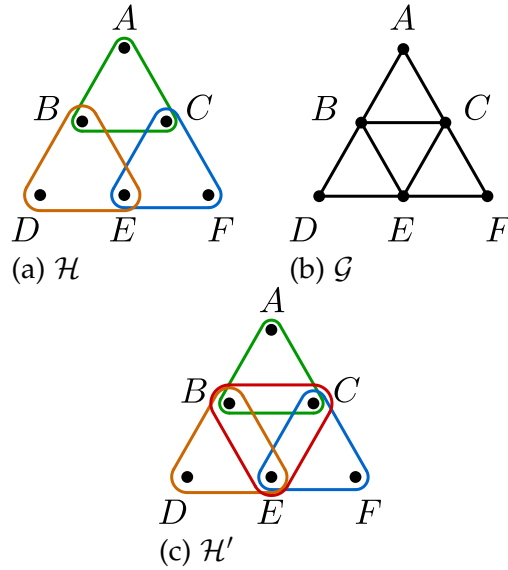


FIGURE 5.3: Example of graphs and hypergraphs. (a) A cyclic hypergraph \mathcal{H} . (b) A 2-section graph $\mathcal{G} = [\mathcal{H}]_2$ of the hypergraph. (c) A clique hypergraph \mathcal{H}' of \mathcal{G} . Notice that $\mathcal{H}' \neq \mathcal{H}$. In fact, \mathcal{H}' is an extension of \mathcal{H} .

imposed. We start by introducing the notion of adhesivity and restating in our language a theorem by Vorob'ev [126] (Th. 5.1), then we connect this result with the notion of marginal scenario to prove which independence constraints are always compatible with a set of marginals (Th. 5.2).

On the one hand, such additional constraints simplify the characterization of the entropy cone and correlation polytope associated with the marginal scenario. On the other hand, this result allows us to identify which causal structures can be distinguished when we have access only to some restricted set of marginals.

The notion of adhesivity, albeit in different terms, was first introduced for probability distributions [126, 127] and subsequently extended to entropies [128]. In the framework of Bell and noncontextuality inequalities, similar ideas have been investigated by several authors [94, 97, 129–131], but never in full generality.

5.1.1 Adhesivity of probabilities

Adhesivity can be explained in simple terms as follows. Given two sets of variables $X_I = (X_i)_{i \in I}$ and $X_J = (X_j)_{j \in J}$ and two probability distributions $p(x_I)$ and $p'(x_J)$ such that p and p' coincide on the variables $X_{I \cap J}$, we can define a probability distribution on $I \cup J$ as

$$P(x_{I \cup J}) = \begin{cases} 0 & \text{if } p(x_{I \cap J}) = 0, \\ \frac{p(x_I)p'(x_J)}{p(x_{I \cap J})} & \text{otherwise.} \end{cases} \quad (5.2)$$

One can easily check that this is a valid probability distribution on the set of variables in $I \cup J$.

The construction in Eq. (5.2) implies that every two marginals of a probability distribution are always consistent with a probability distribution conditionally independent of their intersection, i.e., $(X_{I \setminus J} \perp\!\!\!\perp X_{J \setminus I} \mid X_{I \cap J})$, since $p(x_I)p'(x_J)/p(x_{I \cap J}) = p(x_I|x_{I \cap J})p'(x_J|x_{I \cap J})p(x_{I \cap J})$. We call such an extension of p and p' an *adhesive extension*.

Similarly, two polymatroids (N, h) and (M, g) coinciding on $N \cap M$ are said to *adhere* or to have an *adhesive extension* if there exists a polymatroid $(N \cup M, f)$ extending h, g , i.e., $f(I) = h(I)$ for $I \subset N$, $f(J) = g(J)$ for $J \subset M$, which is also *modular* on N and M , that is, $f(N \cup M) = f(N) + f(M) - f(N \cap M)$ or, equivalently, such that N and M are conditionally independent on the intersection $N \cap M$. As a consequence of the construction in Eq. (5.2) for probabilities, restrictions of entropies have an adhesive extension, whereas general polymatroids do not (cf. Ref. [128]).

This observation is at the basis of the derivation of several *non-Shannon* information inequalities, i.e., information inequalities that do not follow from Eqs. (2.44) (cf. Ref. [128]). Starting from the first non-Shannon inequality derived by Zhang and Yeung [132], infinitely many inequalities have been derived by Matúš [133], and several others authors investigated the problem [134–136].

5.1.2 Marginal scenarios admitting a global extension

From the adhesivity property of probability distributions, one can extend probabilities defined on a marginal scenario to a joint probability distribution over all variables, which satisfies extra conditional independence constraints that depends on the marginal scenario.

The theorem below was first stated without proof by Vorob'ev in Ref. [126], and subsequently explicitly proven in Ref. [127], but also independently derived by other authors [137–139]. The original proof, however, used a quite different terminology. It is helpful to restate it in the language of hypergraphs, and to present a sketch of it, in order to understand the role of the adhesivity property.

Theorem 5.1. [Vorob'ev] *A set of probabilities associated with an acyclic marginal scenario hypergraph \mathcal{M} admits a global extension to a single probability distribution. Moreover, the extension can be chosen as a MRF described by the 2-section graph $[\mathcal{M}]_2$.*

Sketch of the proof.— Let \mathcal{M} be the marginal scenario hypergraph, by definition, it is a reduced hypergraph. If \mathcal{M} is acyclic, we can find an ordering M_1, \dots, M_n of its hyperedges respecting the running intersection property. The construction of a global probability distribution can then be obtained by induction on n , the number of hyperedges. For $n = 1$, $P(M_1)$ is a valid probability distribution (to simplify the notation, we will use $P(M_i)$ as a shorthand for $P(x_{M_i})$, etc). We then apply the inductive hypothesis. Let us assume that for $n - 1$ $P(M_1 \cup \dots \cup M_{n-1})$ is a valid probability distribution extending the marginals $P(M_i)$ for $1 \leq i \leq n - 1$. We want to extend it to $P(M_1 \cup \dots \cup M_{n-1} \cup M_n)$. By the running intersection property, $M_n \cap (M_1 \cup \dots \cup M_{n-1}) =: S_n \subset M_j$ for $j < n$. Denoting by P_{M_i} the marginal probability distribution on M_i , we define $R_n := M_n \setminus S_n$ and

$$P(R_n | S_n) := \frac{P_{M_n}(M_n)}{P_{M_j}(S_n)}, \quad (5.3)$$

defining $0/0$ to be zero as in Eq. (5.2), and for the joint distribution

$$\begin{aligned} P(M_1 \cup \dots \cup M_{n-1} \cup M_n) &:= P(R_n | S_n) \\ &P(M_1 \cup \dots \cup M_{n-1} \setminus S_n | S_n) P(S_n). \end{aligned} \quad (5.4)$$

By the adhesivity property, this is a valid probability distribution, and its marginals coincide with $P(M_i)$ for $1 \leq i \leq n$, so it is an extension of the marginal scenario. In addition, it is modular over the intersection, i.e.

$$(R_n \perp (M_1 \cup \dots \cup M_{n-1}) \setminus S_n | S_n) \quad (5.5)$$

Since \mathcal{M} is connected and reduced, the set of minimal separators precisely corresponds to the set S_n above. The modularities of the constructed distribution are thus precisely those implied by the MRF defined by $[\mathcal{M}]_2$. \square

In the next section, we will see the application of this result to the general marginal scenario, i.e., not necessarily acyclic.

5.1.3 Maximal set of independence conditions associated with a marginal scenario

We will now see the implications of Vorob'ev's theorem on general marginal scenarios. More precisely, we will discuss which independence conditions, arising as MRF conditions, are consistent with a given marginal scenario and how to compute maximal sets of such conditions.

The main result is the following.

Theorem 5.2. *Given a joint probability distribution P on n variables X_1, \dots, X_n , and a marginal scenario \mathcal{M} , the marginals P_{M_i} for $M_i \in \mathcal{M}$ are consistent with a probability distribution arising from a MRF associated with the 2-section graph $[\mathcal{T}]_2$, where \mathcal{T} is an acyclic hypergraph extending \mathcal{M} .*

Proof.– An acyclic hypergraph \mathcal{T} extending \mathcal{M} can always be found. It is the clique hypergraph of a triangulation of the graph $[\mathcal{M}]_2$. The marginals in \mathcal{M} are consistent with the marginals in \mathcal{T} extracted from the same probability distribution P . Since \mathcal{T} is an acyclic hypergraph, we can apply the construction in Th. 5.1 to obtain a MRF with independence relations described by the 2-section graph $[\mathcal{T}]_2$. \square

For any given marginal scenario \mathcal{M} , an acyclic hypergraph extending it corresponds to the clique hypergraph of the triangulation of the 2-section $[\mathcal{M}]_2$, hence the maximum set of independence constraints will correspond to the triangulation with the minimum number of edges, also called the *minimum triangulation*. The problem of computing the minimum triangulation is known to be NP-hard [140]. However, it is much easier to calculate a *minimal triangulation*, namely, a triangulation such that by removing any edge the obtained graph is no longer a chordal graph. Notice that such a minimal triangulation is not necessarily the one with the smallest number of edges among all possible triangulations. Several algorithms have been developed to compute a minimal triangulation, which run in $O(n + m)$ steps, n being the number of nodes and m the number of edges of a graph [140].

In the following, we will adopt the above terminology also for hypergraphs, namely, we will speak about the minimum acyclic hypergraph extending \mathcal{M} , in the sense of the minimum triangulation, and a minimal hypergraph extending \mathcal{M} , in the sense of a minimal triangulation.

5.2 Optimal characterization of the marginal scenario for probabilities and entropies

As a consequence of the above results, the characterization of a given marginal scenario \mathcal{M} , in terms of inequalities for the probability vector or entropy vector, can be computed from those associated with a minimal acyclic hypergraph extending \mathcal{M} . This approach offers advantages both for the probabilistic and the entropic approach. Here, we will give a brief summary of the two results, but later we mostly discuss the entropic approach.

A similar approach, albeit with a different terminology, and using a less general version of Th. 5.2, has been already used in relation with Bell and noncontextuality inequalities. For instance, the decomposition of the CHSH scenario in Fig. 5.4 was used in the proof of the necessity and sufficiency of Bell inequalities for the existence of a LHV model by Fine [60]. Special cases of Th. 5.2 have been discussed in Refs. [129–131], and their application to more general scenarios has been discussed for probabilities [97, 141] and for entropies [94].

5.2.1 Triangulation

The first step is to compute a minimal acyclic hypergraph \mathcal{T} extending the marginal scenario \mathcal{M} . It can be done as follows:

- a1) Compute the 2-section graph $[\mathcal{M}]_2$.
- a2) Compute its minimal triangulation.
- a3) Take as \mathcal{T} the corresponding clique hypergraph.

In the following, we discuss the above procedure with a simple example. Consider the hypergraph with edges $\{\{A_x, B_y\}\}_{x,y=1,2}$ associated with a Bell experiment and discussed in Sec. 2.4. One starts with the marginal scenario hypergraph \mathcal{M} of Fig. 5.4 (a) and computes its 2-section $[\mathcal{M}]_2$ [cf. Fig. 5.4 (b)]. In this simple case $[\mathcal{M}]_2$ can be triangulated by adding an extra edge connecting A_1 and A_2 [cf. Fig. 5.4 (c)], or equivalently, connecting B_1 and B_2 . Finally, one takes as \mathcal{T} the clique hypergraph of the triangulation [cf. Fig. 5.4 (d)]. The corresponding MRF independence condition consistent with the marginals in \mathcal{M} is $(B_1 \perp B_2 | A_1, A_2)$.

5.2.2 Probabilities

Once \mathcal{T} has been obtained, the probabilistic inequalities describing the marginals consistent with the given scenario \mathcal{M} (describing a correlation polytope, see Sec. 2.2.3) can be computed as follows:

- b1) Write down the simplex inequalities associated with each maximal clique C_i , i.e., the inequalities corresponding to a classical probability on $|C_i|$ variables (cf. Ref. [142] and Sec. 2.2.3 for further details on the simplex polytope).
- b2) Project such inequalities onto the initial marginal scenario \mathcal{M} (for instance, applying the Fourier-Motzkin elimination, see Sects. 2.5 and 2.2.3).

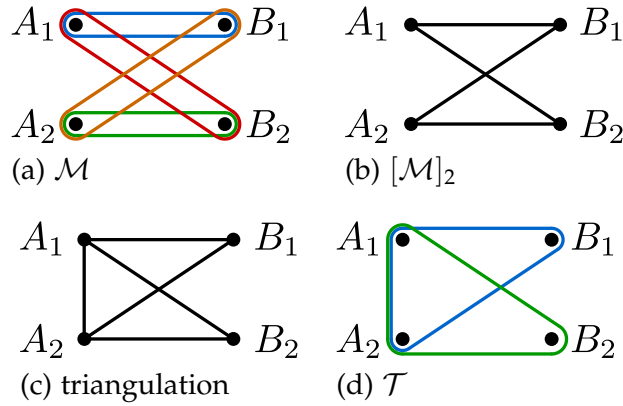


FIGURE 5.4: Procedure to compute the minimal hypergraph \mathcal{T} extending the marginal scenario \mathcal{M} for the CHSH scenario. (a) Initial marginal scenario hypergraph corresponding to the CHSH case. (b) 2-section $[\mathcal{M}]_2$ of the original hypergraph. (c) Triangulation of the 2-section graph. (d) Clique hypergraph \mathcal{T} of the triangulation. \mathcal{T} extends \mathcal{M} .

5.2.3 Entropies

A similar approach can be applied for deriving entropic inequalities, but it only gives an outer approximation if an exact characterization of the entropy cone is not known [cf. Eq. (5.7) below].

An alternative approach can be summarized as follows:

- c1) Compute the independence constraints $\mathcal{I}(\mathcal{T})$ associated with the MRF graph $[\mathcal{T}]_2$,
- c2) Consider the Shannon cone on n variables Γ , with the reduced set of polymatroid axioms associated with $\mathcal{I}(\mathcal{T})$.
- c3) Use the linear constraints associated with $\mathcal{I}(\mathcal{T})$, i.e., the vanishing of some conditional mutual information terms, for a partial projection of the full cone onto the marginal scenario. Then, complete the projection with the usual Fourier-Motzkin algorithm.

It is clear that the above approach can be adapted to any type of linear constraints, including those arising from some assumed causal structure, i.e., a BN or MRF, and those arising as deterministic dependence conditions, corresponding to the vanishing of some conditional entropy, discussed in Sec. 2.3.

In the next section, we will see in details what approaches are possible for characterizing entropic marginals.

5.2.4 Outer approximations of the entropy cone

The adhesivity property for restrictions of an entropic polymatroid can be used to obtain outer approximations of the entropy cone as follows.

Theorem 5.3. *Let \mathcal{M} be the marginal scenario hypergraph and \mathcal{T} an acyclic hypergraph extending it. Let us denote with $\Gamma_{\mathcal{T}}^*$ the entropy cone intersected with the linear constraints defined by $\mathcal{I}(\mathcal{T})$. Then, we have that*

$$\Pi_{\mathcal{M}}(\Gamma^*) = \Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}}^*) \quad (5.6)$$

where $\Pi_{\mathcal{M}}$ denotes the projection onto the coordinates associated with the marginal scenario \mathcal{M} .

Proof.— The result, basically, follows from the fact that the marginals in \mathcal{M} are consistent with the linear constraints $\mathcal{I}(\mathcal{T})$. Given an entropic polymatroid, it is sufficient to take the associated probability distribution and apply Th. 5.2. The obtained distribution will be consistent with the MRF $[\mathcal{T}]_2$, hence, the associated marginal entropies will be identical to the original ones and inside $\Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}}^*)$ by construction. The same construction can be combined with the limits necessary (see, e.g., Refs. [70, 143]) to obtain the closure of entropic polymatroid Γ^* . \square

Inspired by the notion of adhesivity, one can also consider the outer approximation

$$\Gamma_{\mathcal{T}}^* \subset \bigcap_k \Gamma_{C_k}^* \quad (5.7)$$

where C_k are the hyperedges of \mathcal{T} and each $\Gamma_{C_k}^*$ is the entropic cone associated with the variables in C_k embedded in the space of all variables, i.e., the remaining variables are constrained. However, except for the case of the entropic polymatroid arising as a restriction of a single polymatroid and few other cases (cf. Ref. [128]), it is not clear whether entropic polymatroids are adhesive; hence, the inclusion in Eq. (5.7) may be strict.

Now if we take an outer approximation Γ (e.g., the Shannon cone) of the entropic cone Γ^* , and intersect it with the linear subspaces defined by $\mathcal{I}(\mathcal{T})$, we obtain the cone $\Gamma_{\mathcal{T}}$ which satisfies

$$\Pi_{\mathcal{M}}(\Gamma^*) = \Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}}^*) \subset \Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}}) \subset \Pi_{\mathcal{M}}\left(\bigcap_k \Gamma_{C_k}\right). \quad (5.8)$$

In general, given a marginal scenario \mathcal{M} there exist several minimal acyclic hypergraphs $\{\mathcal{T}_i\}$ extending it. The inclusion in Eq. (5.8) is valid for each $\Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}_i})$, consequently, also for the intersection

$$\Pi_{\mathcal{M}}(\Gamma^*) \subset \bigcap_i \Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}_i}). \quad (5.9)$$

As a conclusion, for each marginal scenario \mathcal{M} , we have three different outer approximations for $\Pi_{\mathcal{M}}(\Gamma^*)$, namely,

- (i) the intersection of the projections of the full Shannon cones $\Gamma_{[n]}$ with modularity conditions $\{\mathcal{I}(\mathcal{T}_i)\}$, namely, $\bigcap_i \Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}_i})$; $\{\mathcal{T}_i\}$ is the set of minimal acyclic hypergraphs extending \mathcal{M} ;
- (ii) the projection of the full Shannon cone, namely, $\Pi_{\mathcal{M}}(\Gamma)$;
- (iii) the projection of the intersection of Shannon cones associated with $\{C_k^{(i)}\}_k$ is the set of hyperedges of \mathcal{T}_i , i.e., $\bigcap_i \Pi_{\mathcal{M}}\left(\bigcap_k \Gamma_{C_k^{(i)}}\right)$, where each $\Gamma_{C_k^{(i)}}$ is seen as a cone in the space of all variables, and the variables not appearing in $C_k^{(i)}$ are unconstrained [cf. Eq. 5.7].

We can then summarize the relations among the above cones as follows

Observation 5.4. The above approximations satisfy the inclusion relations

$$\Pi_{\mathcal{M}}(\Gamma^*) \subset \bigcap_i \Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}_i}) \subset \Pi_{\mathcal{M}}(\Gamma) \subset \bigcap_i \Pi_{\mathcal{M}}\left(\bigcap_k \Gamma_{C_k^{(i)}}\right). \quad (5.10)$$

In general, the inclusion relations from Observation 5.4 are proper. In particular, it means that the outer approximation $\bigcap_{i \in \{\mathcal{T}_i\}} \Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}_i})$ is tighter than the projection of the Shannon cone—the most widely used method in the literature; see for instance [70, 128, 132, 133]—and thus may contain nonconstrained non-Shannon-type inequalities. We will provide some examples of this in Sec. 5.4.

5.3 Indistinguishability of causal structures

In this section, we will investigate the role of the above results for the case of probability distributions and entropies, where some underlying causal structure is assumed, i.e., some additional conditional independence constraints are present.

The general goal is to characterize the region of probability distributions or entropies compatible with a given causal structure. Via such a characterization, for instance via Bell inequalities, one can check whether some observed data are consistent or inconsistent with the assumed causal structure, thus being of fundamental importance in both quantum information and any other field where causal discovery may play a relevant role. Furthermore, notice that unless one is able to intervene in the physical system under investigation [68], one can never unambiguously prove what is the underlying causal structure. Rather, based on observations alone, one can only prove the compatibility or incompatibility of a given presumed set of causal relations. As expected, the less constraints a given causal structure implies on the distributions compatible with it, the more correlations such models can explain and the smaller is the possibility of falsifying it.

The ideas to be discussed next apply not only to the case of classical causal structures, but also to the quantum [22, 64, 74, 77, 144, 145] and even post-quantum [75, 96, 146] generalizations as well. In the following, we will focus on the classical case, that is, all nodes in the associated Bayesian networks or MRFs represent random variables for which a global joint probability distribution can always be assumed to exist. The case of quantum and post-quantum theories will be briefly discussed at the end of this section and presented in full details elsewhere.

Let us consider the causal structure defined by a graph \mathcal{G} , which may be either a DAG corresponding to a Bayesian network, or a graph corresponding to a MRF. We will denote the set of independence conditions associated with \mathcal{G} as $\mathcal{I}(\mathcal{G})$, as in Eqs.(2.35,2.37). Let us now assume to have a fixed marginal scenario, with \mathcal{M} the associated hypergraph. Let $\{\mathcal{T}_i\}_i$ be the set of acyclic hypergraph extending \mathcal{M} as in Th. 5.1. We will denote by $\mathcal{I}(\mathcal{T}_i)$, the set of independence conditions of the corresponding MRF defined by its 2-section $[\mathcal{T}_i]_2$. We have the following result

Theorem 5.5. *Given \mathcal{M}, \mathcal{G} and $\{\mathcal{T}_i\}_i$, we have three possible cases:*

- (i) $\exists i$ such that $\mathcal{I}(\mathcal{G}) \subset \mathcal{I}(\mathcal{T}_i)$.
- (ii) $\forall i \mathcal{I}(\mathcal{T}_i) \subset \mathcal{I}(\mathcal{G})$.
- (iii) $\forall i \mathcal{I}(\mathcal{G}) \not\subset \mathcal{I}(\mathcal{T}_i)$, and $\exists j$ such that $\mathcal{I}(\mathcal{T}_j) \not\subset \mathcal{I}(\mathcal{G})$.

Then:

In case (i), it is impossible to falsify the causal structure described by \mathcal{G} . This follows since for any probability distribution P , its marginals in \mathcal{M} are always consistent with the causal structure described by \mathcal{G} . Approach (c1)–(c3) of Sec. 5.2 can be used to characterize the marginals associated with \mathcal{M} and \mathcal{G} .

In case (ii), marginals associated with \mathcal{M} can still generate correlations that are incompatible with the causal structure associated with $\mathcal{I}(\mathcal{G})$. Approach (c1)–(c3) can be used, but the obtained constraints are redundant with respect to $\mathcal{I}(\mathcal{G})$. It is, therefore, more convenient to apply approach (c1)–(c3) directly with the conditional independence relations $\mathcal{I}(\mathcal{G})$.

In case (iii), the marginal correlations associated with \mathcal{M} can again be incompatible with the causal structure associated with $\mathcal{I}(\mathcal{G})$. However, the marginal scenario implies constraints that cannot be combined with those of the causal structure \mathcal{G} . Hence, the approach (c1)–(c3) cannot be used with the constraints $\mathcal{I}(\mathcal{T})$.

Proof.— In case (i), the independence constraints of the causal structure are just a subset of the independence constraints consistent with the marginal scenario. As a consequence, given the marginal probabilities $\{P_M\}_{M \in \mathcal{M}}$, one can repeat the construction of Th. 5.1 and obtain a valid joint probability distribution P that is consistent with the causal structure defined by \mathcal{G} .

The above result implies that the approach of Sec. 5.2, both the constructions (b1)–(b3) for probabilities and (c1)–(c3) for entropies, applies also to the case of a causal structure with $\mathcal{I}(\mathcal{G}) \subset \mathcal{I}(\mathcal{T}_i)$.

In case (ii), the marginal scenario implies fewer constraints than the causal structure \mathcal{G} , hence, it is clear that the marginals associated with \mathcal{M} are still able to detect inconsistencies with the causal structure associated with $\mathcal{I}(\mathcal{G})$.

From the point of view of the characterization, one can still use the approach (c1)–(c3) of Sec. 5.2. However, it is more convenient to use in (c3) the linear constraints implied by $\mathcal{I}(\mathcal{G})$, since they also include those associated with $\mathcal{I}(\mathcal{T}_i)$ for all i .

In case (iii), it is again clear that the marginals associated with \mathcal{M} are still able to detect inconsistencies with the causal structure associated with $\mathcal{I}(\mathcal{G})$. However, the approach (c1)–(c3) of Sec. 5.2 cannot be used as it generates constraints inconsistent with the causal structure. The situation is clarified by the following inclusion relations among entropy cones in Eqs. (5.11)–(5.13). Let us denote by $L_{\mathcal{G}}$ the subspace of entropy vectors in \mathbb{R}^{2^n} where the linear constraints imposed by $\mathcal{I}(\mathcal{G})$ are satisfied, and similarly for $L_{\mathcal{T}_i}$. The entropy cone associated with a given causal structure, either \mathcal{G} or \mathcal{T}_i , will be $\Gamma^* \cap L_{\mathcal{G}, \mathcal{T}_i}$. We can write down the relation between the associated entropy cones as

$$\Pi_{\mathcal{M}}(\Gamma^*) = \Pi_{\mathcal{M}}(\Gamma^* \cap L_{\mathcal{T}_i}), \quad (5.11)$$

$$\Pi_{\mathcal{M}}(\Gamma^* \cap L_{\mathcal{G}} \cap L_{\mathcal{T}_i}) \subset \Pi_{\mathcal{M}}(\Gamma^* \cap L_{\mathcal{G}}), \quad (5.12)$$

$$\begin{aligned} \Pi_{\mathcal{M}}(\Gamma^* \cap L_{\mathcal{G}}) &\subset \Pi_{\mathcal{M}}(\Gamma^*) \cap \Pi_{\mathcal{M}}(L_{\mathcal{G}}) \\ &= \Pi_{\mathcal{M}}(\Gamma^* \cap L_{\mathcal{T}_i}) \cap \Pi_{\mathcal{M}}(L_{\mathcal{G}}) \end{aligned} \quad (5.13)$$

It is then clear that by imposing both the $\mathcal{I}(\mathcal{G})$ and $\mathcal{I}(\mathcal{T}_i)$ conditions one obtains an inner approximation of the entropy cone associated with the causal structure. Vice versa, imposing the causal structure conditions after the projection gives an outer approximation. \square

To clarify this last part, in particular the impossibility of combining the independence relations $\mathcal{I}(\mathcal{G})$ arising from the causal structure, with the $\{\mathcal{I}(\mathcal{T}_i)\}_i$ arising from the marginal scenario, it is helpful to look at some specific examples. We will discuss them in details in Sec. 5.4.2.

A natural question is, then, how to extend the above results to the case of quantum and post-quantum causal structures. For instance, in the postquantum case if \mathcal{M} is an acyclic hypergraph, then by Th. 5.1 the observed marginals are always consistent with a classical probability distribution. The same holds, in particular, in the quantum case. However, the fact that different rules for causal inference arise in the quantum and postquantum cases (cf. [22, 75, 77, 96, 144, 146]) together with the different characterization of the associated entropy regions (cf. [22, 96]) makes the above investigation more complex and worth a separate discussion elsewhere.

5.4 Examples and computational results

In order to clarify the results and methods presented in the previous sections, we discuss examples of marginal scenarios and causal structures, together with some computational results.

5.4.1 Inclusions in Obs. 5.4

In the following, we will discuss the possible cases presented in Obs. 5.4. In particular, we will see examples of strict and non-strict inclusion for the outer approximations of the entropy cone.

$$\text{Case: } \bigcap_i \Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}_i}) \subsetneq \Pi_{\mathcal{M}}(\Gamma) \subsetneq \bigcap_i \Pi_{\mathcal{M}}\left(\bigcap_k \Gamma_{C_k^{(i)}}\right)$$

As already noted by Matúš [128], the proper inclusion $\bigcap_i \Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}_i}) \subsetneq \Pi_{\mathcal{M}}(\Gamma)$ means existence of non-Shannon-type inequalities in $\Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}_i})$. On the other hand, the strict inclusion $\Pi_{\mathcal{M}}(\Gamma) \subsetneq \bigcap_i \Pi_{\mathcal{M}}\left(\bigcap_k \Gamma_{C_k^{(i)}}\right)$ is related to the non-adhesivity of general polymatroids.

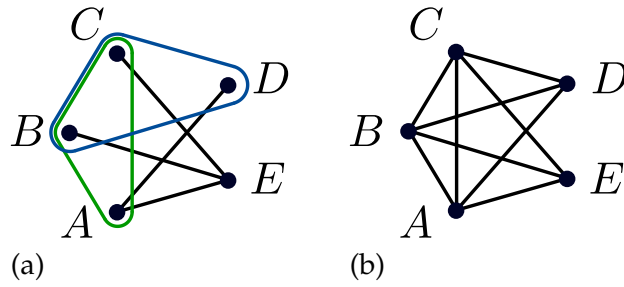


FIGURE 5.5: Marginal scenario hypergraph (a) and (b) its 2-section

To construct an example, is necessary to take at least four variables [128], and our example will consist of five variables. Let us consider the following marginal scenario $\mathcal{M} = \{ABC, BCD, AE, BE, CE, AD\}$, shown on the Fig. 5.5 (a). One can easily see that the 2-section of the hypergraph \mathcal{M} is a triangulated graph shown in Fig. 5.5 (b) and thus there is only one corresponding clique hypergraph $\mathcal{T} = \{ABCD, ABCE\}$. The independence constraint $I(D : E|ABC) = 0$ arising from the adhesivity property, gives rise, after projection on the set of entropies given by $\{B, C, D, AD, AE, BD, BE, CD, CE, ABC, BCD\}$, to the following 3 non-redundant non-Shannon-type inequalities

$$\begin{aligned} &H(E|C) + H(C|D) + H(E|B) + H(B, D) + H(A, D) \\ &-s_1H(B, C, D) - s_2H(A, E) + s_3H(A, B, C) \geq 0, \end{aligned} \quad (5.14)$$

where the coefficient triplet $(s_1, s_2, s_3) \in \{(1, 2, 1), (2, 1, 1), (2, 2, 2)\}$.

Another interesting aspect of this example is the reduction in the computational time required to compute the projection on a usual desktop computer. More precisely, adding the linear constraint $I(D : E|ABC) = 0$ reduced the time of our computation for the projection from approximately 320 to only 27 seconds.

Case: $\bigcap_i \Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}_i}) = \Pi_{\mathcal{M}}(\Gamma) = \bigcap_i \Pi_{\mathcal{M}}\left(\bigcap_k \Gamma_{C_k^{(i)}}\right)$

Consider the marginal scenario $\mathcal{M} = \{A_i B_j\}, \forall i, j \in \{1, 2, 3\}$, shown in Fig. 5.6 (a), corresponding to a bipartite Bell scenario with three measurement settings per party. The clique hypergraph of one of the triangulations of \mathcal{M} is shown on Fig. 5.6 (b).

If we consider an intersection of cones $\Gamma_{C_k^{(1)}}$ for cliques $C_k^{(1)} = \{A_1, A_2, A_3, B_k\}, k = 1, 2, 3$, which are the edges of the hypergraph shown on Fig. 5.6 (b), we find that its projection on the marginal scenario \mathcal{M} differs from the projection of the full cone such that 108 out of 217 rays of the projection $\Pi_{\mathcal{M}}\left(\bigcap_k \Gamma_{C_k^{(1)}}\right)$ are outside of $\Pi_{\mathcal{M}}(\Gamma)$. This

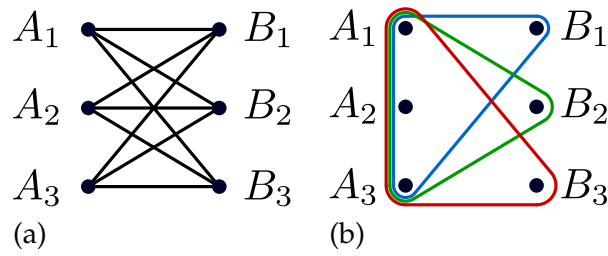


FIGURE 5.6: Marginal scenario (a) and a clique hypergraph (b)

can be checked via linear programming (cf. Ref. [96] Sec. II of the Supplemental Material) by simply checking whether such rays are compatible with the basic Shannon inequalities characterizing Γ .

However, if we now consider cliques of the second possible triangulation of \mathcal{M} , which are $C_k^{(2)} = \{A_k, B_1, B_2, B_3\}$, $k = 1, 2, 3$ and compute the intersection $\bigcap_{i=1,2} \Pi_{\mathcal{M}} \left(\bigcap_k \Gamma_{C_k^{(i)}} \right)$ we find, again via linear programming, that all its extremal rays are inside, not only $\Pi_{\mathcal{M}}(\Gamma)$, but also $\Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}_i})$, for all i . Hence we have that all the outer approximations coincide.

The reasons why such an equivalence is interesting is that the calculation of the intersection $\bigcap_{i=1,2} \Pi_{\mathcal{M}} \left(\bigcap_k \Gamma_{C_k^{(i)}} \right)$ with a standard Fourier-Motzkin algorithm on a standard desktop takes few minutes, however, a direct computation of $\Pi_{\mathcal{M}}(\Gamma)$ or $\Pi_{\mathcal{M}}(\Gamma_{\mathcal{T}_i})$ seems to be out of computational reach (at least on a usual desktop computer).

5.4.2 Three cases in Theorem 5.5

In this subsection, we will discuss in detail and provide examples for the different cases presented in Theorem 5.5. Let us consider four random variables A, B, C, D .

Case (i): $\exists i$ such that $\mathcal{I}(\mathcal{G}) \subset \mathcal{I}(\mathcal{T}_i)$

Let us consider the marginal scenario given by $\mathcal{M}_1 = \{AB, BD, BC\}$ and shown in Fig. 5.7 (b). One can easily see that the clique hypergraph \mathcal{T} of the corresponding triangulation of the 2-section graph is unique and coincides with \mathcal{M}_1 . The independence constraints implied by \mathcal{T} are given by

$$\begin{aligned}
 (A \perp C | B) \\
 (C \perp D | B) \\
 (D \perp A | B)
 \end{aligned} \tag{5.15}$$

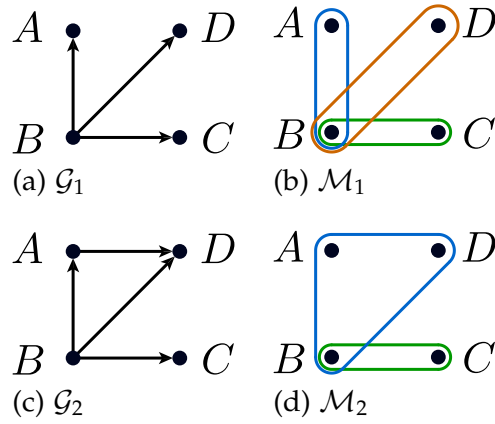


FIGURE 5.7: Example of causal structures (a),(c), and marginal scenarios (b),(d).

Consider now the two causal structures \mathcal{G}_1 and \mathcal{G}_2 shown in Fig. 5.7 (a),(c). The corresponding independence constraints are $\{(A \perp C | B); (C \perp D | B); (D \perp A | B)\}$ for \mathcal{G}_1 and $\{(A \perp C | B); (C \perp D | B)\}$ for \mathcal{G}_2 . In both cases $\mathcal{I}(\mathcal{G}_{1,2}) \subset \mathcal{I}(\mathcal{T})$ which means that marginal scenario \mathcal{M}_1 from Fig. 5.7 (b) is insufficient to distinguish causal structures \mathcal{G}_1 and \mathcal{G}_2 . In turn, a marginal scenario, which would be enough to distinguish between these two causal structures is given by $\mathcal{M}_2 = \{ABD, BC\}$ and shown in Fig. 5.7 (d).

Case (ii): $\forall i \mathcal{I}(\mathcal{T}_i) \subset \mathcal{I}(\mathcal{G})$

Consider again the causal graph \mathcal{G}_1 . As we already noted the independence constraints associated with this graph are

$$\begin{aligned}
 (A \perp C | B) \\
 (C \perp D | B) \\
 (D \perp A | B)
 \end{aligned} \tag{5.16}$$

If we are now interested in the marginal scenario $\mathcal{M}_2 = \{ABD, BC\}$, then one can see that in that case there is again only one possible triangulation of the 2-section graph of \mathcal{M}_2 and consequently only one corresponding clique hypergraph. The set of independence constraints, consistent with \mathcal{M}_2 is $\{(A \perp C | B); (C \perp D | B)\}$, which is a subset of constraints from Eq. (5.16). In other words, constraints coming from marginal scenario \mathcal{M}_2 are redundant to those coming from the causal structure.

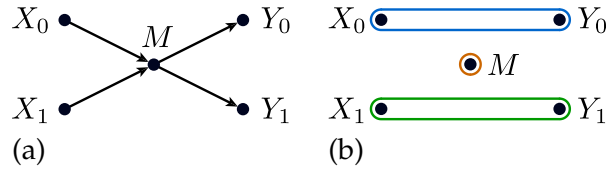


FIGURE 5.8: Causal structure (a) and marginal scenario (b) of classical case of information causality. X_0, X_1 are random inputs for Alice, Y_0, Y_1 are guesses for Bob, and M is a message which Alice sends to Bob.

Case (iii): $\forall i \mathcal{I}(\mathcal{G}) \not\subset \mathcal{I}(\mathcal{T}_i)$, and $\exists j$ such that $\mathcal{I}(\mathcal{T}_j) \not\subset \mathcal{I}(\mathcal{G})$

The third one is, arguably, the most interesting case: it shows that the independence constraints arising from the marginal scenario may be “inconsistent” with those associated with the causal structure. An example where this problem arises is the classical case of the information causality scenario [147]: Alice receives two independent inputs X_0, X_1 , she creates a message M depending on those inputs that is sent to Bob who provide guesses Y_0, Y_1 , respectively of X_0, X_1 , on the basis of the message M .

The corresponding causal structure is shown in Fig. 5.8 (a) and the marginal scenario \mathcal{M} in Fig. 5.8 (b). Once again the clique hypergraph \mathcal{T} coincides with \mathcal{M} , hence, it is unique.

We need to show that

$$\mathcal{I}(\mathcal{G}) \not\subset \mathcal{I}(\mathcal{T}), \text{ and } \mathcal{I}(\mathcal{T}) \not\subset \mathcal{I}(\mathcal{G}). \quad (5.17)$$

For showing $\mathcal{I}(\mathcal{G}) \not\subset \mathcal{I}(\mathcal{T})$, we consider the conditional independence between inputs X_0, X_1 and guesses Y_0, Y_1 . That is,

$$\{X_i \perp Y_j \mid M\}_{i,j=0,1} \not\subset \mathcal{I}(\mathcal{T}). \quad (5.18)$$

An example in the other direction is an independence of message M from the rest of the variables, which is implied by \mathcal{T} , and is not consistent with conditional independences $\mathcal{I}(\mathcal{G})$.

The projection $\Pi_{\mathcal{M}}(\Gamma \cap L_{\mathcal{G}})$ gives rise to the following inequalities

$$I(X_0 : Y_0) \geq 0, \quad I(X_1 : Y_1) \geq 0, \quad (5.19a)$$

$$H(Y_0|X_0) \geq 0, \quad H(X_0|Y_0) \geq 0, \quad (5.19b)$$

$$H(Y_1|X_1) \geq 0, \quad H(X_1|Y_1) \geq 0, \quad (5.19c)$$

$$I(X_0 : Y_0) + I(X_1 : Y_1) \leq H(M), \quad (5.19d)$$

where inequalities (5.19a), (5.19b), and (5.19c) are simply polymatroid axioms for the marginals $\{X_0Y_0, X_1Y_1\}$ and one obtains these 6 inequalities, if one computes $\Pi_{\mathcal{M}}(\Gamma \cap L_{\mathcal{T}})$. The last inequality Eq. (5.19d) is the information causality inequality and is not implied by $\mathcal{I}(\mathcal{T})$.

As a result of the relation from Eq. (5.17), one cannot combine $\mathcal{I}(\mathcal{T})$ and $\mathcal{I}(\mathcal{G})$. Due to the relation from Eq. (5.12) and the fact that the Shannon cone is an outer approximation for the case of more than 3 variables, the projection of the Shannon cone with combined constraints $\mathcal{I}(\mathcal{T})$ and $\mathcal{I}(\mathcal{G})$ in this case provides neither an outer nor an inner approximation of $\Pi_{\mathcal{M}}(\Gamma^* \cap L_{\mathcal{G}})$.

5.5 Conclusions

Deciding global features of a system of interest with limited information, the so-called marginal problem, is a task often encountered in many fundamental and practical problems. In turn, causal discovery, the inference of causal relations underlying the correlations between observed variables, is yet another basic goal in the most diverse fields. In this work, we use the notion of adhesivity to investigate marginal problems within causal inference. In particular, we show which causal relations are always compatible with some given marginal information. As a consequence, we are able to identify which causal structures, describing either a Bayesian network or a Markov random field, can be distinguished when only limited marginals are available. In addition, our results provide a method for a faster characterization (in terms of Bell inequalities) of the marginal scenarios associated with a given causal model. This holds true for the both the probabilistic and entropic approaches for Bell inequalities. In particular, in the entropic case our construction allows for a more accurate characterization of allowed regions for entropic marginals, as shown with explicit computational results.

An immediate and interesting open question is the possible generalization of these results to the case where the causal relations between the variables are mediated via quantum or postquantum (non-signalling) resources. Quantum generalizations of the notion of a causal structure have attracted growing attention [22, 74, 75, 77, 144, 145] and we believe that our results could constitute a viable option for the characterization of such quantum structures. Partial results, such as the fact that classical and postquantum correlations coincide for the case of acyclic marginal scenario hypergraphs (cf. Sec. 5.3), show that a similar approach can be extended also to the quantum and postquantum cases. In particular, this investigation could lead to new insights on

which causal structures can demonstrate some sort of non-locality [75, 146]. Finally, another possibility is to try to combine the notion of adhesivity and the algebraic geometry tools [67] required to characterize the set of compatible probabilities associated with complex causal structures [66, 148, 149].

Chapter 6

The entropic approach to causal correlations

This chapter describes the most recent results from Ref. [30]. The preliminary sections required for this chapter are Sec. 2.3 (in particular Sec. 2.3.7), 2.5, and partially Sec. 2.2.

The existence of a global causal order between events places constraints on the correlations that parties may share. Such “causal correlations” have been the focus of recent attention, driven by the realization that some extensions of quantum mechanics may violate so-called causal inequalities (see Sec. 2.3.7). In Ref. [30] we study causal correlations from an entropic perspective, and we show how to use this framework to derive entropic causal inequalities. We consider two different ways to derive such inequalities. Firstly, we consider a method based on the causal Bayesian networks describing the causal relations between the parties. In contrast to the Bell-nonlocality scenario, where this method has previously been shown to be ineffective, we show that it leads to several interesting entropic causal inequalities. Secondly, we consider an alternative method based on counterfactual variables that has previously been used to derive entropic Bell inequalities. We compare the inequalities obtained via these two methods and discuss their violation by noncausal correlations. As an application of our approach, we derive bounds on the quantity of information – which is more naturally expressed in the entropic framework – that parties can communicate when operating in a definite causal order.

6.1 Introduction

Our goal in this work is to introduce a new framework for the derivation of causal inequalities and the study of their potential violations: the entropic approach to causal correlations. The idea of using entropies to understand sets of correlations has its origin in the context of Bell inequalities [21, 92, 94, 98] but since then has also found various other applications in quantum contextuality [93, 131, 150], device-independent applications [151, 152], causal inference [75, 95, 146] and in the characterization of nonsignaling correlations [96]. As for these previous applications, the interest in characterizing the entropies compatible with causal correlations stems not only from practical and technical issues, but also from a more fundamental reason. To begin with, causal inequalities expressed in terms of probabilities are constructed for a fixed number of inputs and outputs, and their systematic derivation becomes harder as this number increases [86, 87]. In contrast, we will derive entropic causal inequalities that are valid for arbitrary finite alphabets either for the input and output variables, or just for the output variables. Furthermore, entropic inequalities can be easily combined with extra assumptions, such as conditional independence relations or information theoretic constraints (e.g., bounds on the amount of communication), which would be hard to treat in the probabilistic framework [66, 95, 96]. More fundamentally, given that entropies are a core concept in classical and quantum information theory, it is of clear relevance to have a framework that focuses on these quantities rather than on probabilities, and it may help connect causal inequalities with principles such as information causality [147].

6.2 Bipartite entropic causal inequalities

With the entropic approach to characterizing sets of correlations outlined, we can now proceed to apply this approach to causal correlations, so as to derive *entropic causal inequalities*. We consider in this section the bipartite case. We first show how the method based on causal Bayesian networks can be adapted to characterize causal correlations, before considering also the method based on counterfactual variables.

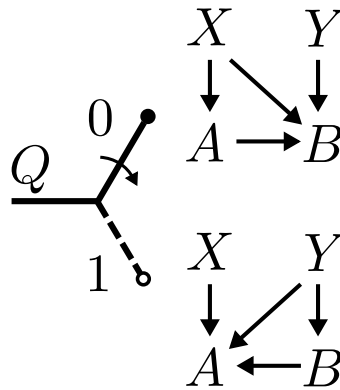


FIGURE 6.1: DAGs for bipartite causal correlations. The latent “switch” variable Q determines which DAG, corresponding to the fixed causal order $A \prec B$ (for $Q = 0$, top) or $B \prec A$ (for $Q = 1$, bottom), is “activated”.

6.2.1 Characterization based on causal Bayesian networks

Conditional DAGs for bipartite causal correlations

The ability to apply the entropic approach to DAGs, as outlined in Sec. 2.5.3, is a powerful tool for characterizing the correlations obtainable within arbitrary causal networks. However, the notion of causal correlations defined in Eq. (2.39) is somewhat more general and cannot be directly expressed within the framework of causal Bayesian networks. In order to see why this is the case, let us first note that the random variables of interest are X, Y, A, B , representing the inputs X, Y and outputs A, B for Alice and Bob. Note that since we consider signaling scenarios here, unlike in the Bell scenario, we do not need to include any latent variable Λ in our description to account for shared randomness, since this can be established via local randomness and communication.

If Alice and Bob share a correlation compatible with a fixed causal order (i.e. either $A \prec B$ or $B \prec A$, then the functional dependences between X, Y, A, B can indeed be expressed as a DAG (specifically, the two DAGs containing these variables in Fig. 6.1). However, a causal correlation may in general not be compatible with any fixed causal order, but may require a mixture thereof. This has some similarities with the situation in the Svetlichny definition of genuine multipartite nonlocality [153, 154] where a convex mixture of different DAGs has to be considered.

To tackle this problem it is necessary to find a way to take into account the constraints arising separately from each of the two fixed causal orders, and then to combine them to obtain those satisfied by causal correlations. In order to do this, we exploit the fact that any mixture of fixed-order causal correlations can be seen as arising from a latent variable that determines the causal order for each individual experiment [85]. We thus

introduce a new random variable Q which we call a “switch”, and which determines univocally the appropriate causal Bayesian network for each trial. The resulting causal model is shown in Fig. 6.1, where the DAG with $A \prec B$ is used for $Q = 0$, or the one with $B \prec A$ for $Q = 1$. By identifying q_0, q_1 in Eq. (2.39) as $q_0 = P(Q = 0)$, and $q_1 = P(Q = 1)$, one can readily see that this description is equivalent to the definition of a causal correlation in Eq. (2.39).

Both DAGs imply the independence of the inputs, $X \perp Y$. The DAG for $Q = 0$ (i.e., for $A \prec B$) also implies the CI $A \perp Y \mid X$ (i.e. that there is no signaling from B to A), while the DAG for $Q = 1$ implies $B \perp X \mid Y$ instead. In addition, the switch variable Q should be independent of Alice and Bob’s inputs X and Y , so that we have $XY \perp Q$, which, together with $X \perp Y$, implies that $X \perp Y \perp Q$.

Shannon polyhedron of causal correlations

In order to use the “conditional” causal Bayesian network in Fig. 6.1 to characterize the set of entropy vectors obtainable from causal correlations, we first note that we can directly use the techniques of Sec. 2.5.3 to characterize the Shannon cones for each of the two DAGs appearing in the figure conditioned on Q (i.e., for fixed-orders correlations with $A \prec B$ or $B \prec A$). Denoting these cones $\Gamma_{A \prec B}$ and $\Gamma_{B \prec A}$, we have

$$\Gamma_{A \prec B} = \Gamma_S \cap L_C^{A \prec B} \quad (6.1)$$

and

$$\Gamma_{B \prec A} = \Gamma_S \cap L_C^{B \prec A}, \quad (6.2)$$

where Γ_S is the Shannon cone for the four variables in $S = \{X, Y, A, B\}$, the probability structure is simply $\mathcal{S} = \{S\}$, and $L_C^{A \prec B}$ denotes linear subspace defined by the CI constraints for the case $A \prec B$, namely, the equations $H(XY) = H(X) + H(Y)$ and $H(YA|X) = H(Y|X) + H(A|X)$, and similarly for $B \prec A$. These cones are characterized by the systems of inequalities $I_0 \mathbf{h} \leq \mathbf{0}$ and $I_1 \mathbf{h} \leq \mathbf{0}$, where $\mathbf{h} = (H(T))_{T \subset S}$.

Recall that in the probabilistic case the polytope of causal correlations is simply the convex hull of the polytopes of correlations for $A \prec B$ and $B \prec A$ [86], and with the new variable Q the definition Eq. (2.39) can be rewritten as

$$P(ab|xy) = P(Q = 0)P^{A \prec B}(ab|xy, Q = 0) + P(Q = 1)P^{B \prec A}(ab|xy, Q = 1). \quad (6.3)$$

In contrast, the convex hull of the cones $\Gamma_{A \prec B}$ and $\Gamma_{B \prec A}$ does not contain all entropy vectors of causal correlations due to the concavity of the Shannon entropy. Indeed,

in Appendix A we provide an explicit example of a causal correlation whose entropy vector is not contained in the convex hull $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$.

To see more precisely why this is the case, and how to give a correct entropic characterization of causal correlations, observe that, when taking a convex mixture of two causal correlations with different causal orders, the “conditional entropy vectors” $\mathbf{h}_0 = (H(T|Q = 0))_{T \subset S}$ and $\mathbf{h}_1 = (H(T|Q = 1))_{T \subset S}$ must be contained in $\Gamma_{A \prec B}$ and $\Gamma_{B \prec A}$, respectively, and thus satisfy $I_0 \mathbf{h}_0 \leq \mathbf{0}$ and $I_1 \mathbf{h}_1 \leq \mathbf{0}$. For any causal correlation, the convex mixture

$$\mathbf{h}_{\text{conv}} = P(Q = 0)\mathbf{h}_0 + P(Q = 1)\mathbf{h}_1 \quad (6.4)$$

is thus contained in $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$. Observe now that, in contrast to the convex sum Eq. (6.3) defining causal correlations, \mathbf{h}_{conv} thus defined is equal to $(H(T|Q))_{T \subset S}$, rather than just $(H(T))_{T \subset S}$, and hence the convex hull of the fixed-order entropy cones characterizes the conditional entropies (conditioned on the switch variable Q) obtainable with causal correlations, rather than the entropy vectors of causal correlations directly.

With the appropriate transformation, the system of inequalities $I\mathbf{h} \leq \mathbf{0}$ characterizing¹ $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$ can be transformed into inequalities satisfied by the standard (i.e., non-conditional) entropy vector $\tilde{\mathbf{h}} = (H(T))_{T \subset \tilde{S}}$ for the variables now in $\tilde{S} = S \cup \{Q\}$ (and the probability structure is consequently extended to $\tilde{\mathcal{S}} = \{\tilde{S}\}$). Specifically, each row \mathbf{I} of the matrix I (defining each individual inequality $I\mathbf{h} \leq 0$) must undergo the linear transformation $\mathcal{T}_Q : \mathbb{R}^{2^{|S|}} \rightarrow \mathbb{R}^{2^{|S|+1}}$ mapping $\mathbf{I} \mapsto \tilde{\mathbf{I}} := \mathcal{T}_Q(\mathbf{I})$ with the components of $\tilde{\mathbf{I}}$ given by²

$$(\tilde{\mathbf{I}})_{T \cup \{Q\}} = (\mathbf{I})_T, \quad (\tilde{\mathbf{I}})_{\{Q\}} = - \sum_{T \neq \emptyset} (\mathbf{I})_T, \quad \text{and} \quad (\tilde{\mathbf{I}})_T = 0 \quad (6.5)$$

for all nonempty subsets $T \subset S$. We will denote by $\text{conv}_Q(\Gamma_{A \prec B}, \Gamma_{B \prec A})$ the cone of vectors $\tilde{\mathbf{h}}$ satisfying the resulting inequalities $\tilde{\mathbf{I}}\tilde{\mathbf{h}} \leq \mathbf{0}$.

To complete the characterization of entropy vectors for causal correlations, we recall that, in addition to the fact that any distribution on \tilde{S} must give an entropy vector in the Shannon cone $\Gamma_{\tilde{S}}$, the conditional DAG in Fig. 6.1 gives us the CI constraints $X \perp Y \perp Q$. Moreover, since Q is a binary variable (as there are only two orders to switch between) we have $H(Q) \leq 1$. A consequence of this final inequality constraint is that the set of entropy vectors under consideration will be characterized by an inhomogeneous system of inequalities of the form $\tilde{\mathbf{I}}\tilde{\mathbf{h}} \leq \tilde{\boldsymbol{\beta}}$ for some $\tilde{\boldsymbol{\beta}} \in \mathbb{R}^{2^{|S|+1}}$ and is thus

¹In practice these can be obtained by taking the union of the extremal rays of the two cones $\Gamma_{A \prec B}$ and $\Gamma_{B \prec A}$ and solving the facet enumeration problem to obtain the inequality representation of $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$ using standard software for convex polyhedra such as PANDA [155].

²Note that $(\tilde{\mathbf{I}})_{\emptyset}$ multiplies $H(\emptyset) = 0$ in the scalar product $\tilde{\mathbf{I}}\tilde{\mathbf{h}}$, so its value is irrelevant.

no longer a cone but a polyhedron. The polyhedron characterizing entropy vectors associated with the conditional DAG (when still including Q) is thus given by

$$\tilde{\Gamma}^{\text{causal}} = \Gamma_{\mathcal{S}} \cap \text{conv}_Q(\Gamma_{A \prec B}, \Gamma_{B \prec A}) \cap L_C(\{(X \perp Y \perp Q), H(Q) \leq 1\}), \quad (6.6)$$

where $L_C(\cdot)$ denotes the subspace or polyhedron defined by the corresponding linear constraints.

Finally, following the general approach presented in Sec. 2.5, it remains just to eliminate the terms containing the (unobservable) switch variable Q in order to obtain the inequalities characterizing bipartite causal correlations. This is done by projecting $\tilde{\Gamma}^{\text{causal}}$ onto the marginal scenario $\mathcal{M} = \{S\} = \{\{X, Y, A, B\}\}$. We thus finally obtain the polyhedron

$$\Gamma^{\text{causal}} = \Pi_{\mathcal{M}}(\tilde{\Gamma}^{\text{causal}}), \quad (6.7)$$

which we shall refer to as the *causal Shannon polyhedron* or simply the *causal polyhedron* and is again characterized by an inhomogeneous system of inequalities $I'h \leq \beta$ for some $\beta \in \mathbb{R}^{2^{|S|}}$.

We emphasize that the construction given above is in fact not at all restricted to the description of causal correlations, and can be used to characterize arbitrary convex mixtures of different Bayesian networks. Furthermore, as we will see in Sec. 6.3, this method can be generalized to convex combinations of more distributions, in our case corresponding to more than two causal orders in multipartite scenarios (and even correlations with “dynamical causal order” [85, 87, 156]).

Entropic causal inequalities and their violation

The constructive description of the causal polyhedron Γ^{causal} from Eqs. (6.6) and (6.7) also makes it clear how we can characterize it, in practice, as a system of linear inequalities. A description of $\tilde{\Gamma}^{\text{causal}}$ in terms of its facets is straightforwardly obtained by taking the union of the inequalities describing the individual cones, linear subspaces, and polyhedra appearing in Eq. (6.6). The inequalities characterizing Γ^{causal} can then be found by eliminating the terms not contained in the marginal scenario $\mathcal{M} = \{S\}$, either by Fourier-Motzkin elimination [62] or by finding its extremal rays and projecting out the unwanted coordinates.

The corresponding system of inequalities is thus satisfied by any bipartite causal correlation. However, many of these inequalities are either elemental inequalities (see Eq. (2.44)) or can be obtained from these by using the independence constraint $X \perp Y$,

and thus represent trivial constraints. After computing the polyhedron in Eq. (6.7) and eliminating all trivial inequalities, i.e., those satisfied by any distribution $P(xyab)$ with $X \perp Y$, we find 35 novel entropic causal inequalities. Several of these inequalities are equivalent under the exchange of parties (i.e., exchanging $(X, A) \leftrightarrow (Y, B)$), and under this symmetry there are in fact 20 equivalence classes of entropic causal inequalities, the full list of which is given in Appendix A. Of these, 10 have bounds of 0 (i.e., are of the form $I \cdot h \leq 0$), while the remaining 10 have nonzero bounds (resulting from a nontrivial dependence on $H(Q)$ before this variable was eliminated, see Appendix A). Simple interpretations of the entropic causal inequalities seem to be less forthcoming than for the bipartite causal inequalities in terms of probabilities [86] (for binary inputs and outputs – recall that the entropic inequalities given here are, in contrast, valid for any number of possible inputs and outputs). One of the simpler examples, which is symmetric under the exchange of parties, is

$$I(X : YA) + I(Y : XB) - H(AB) \leq 0. \quad (6.8)$$

Note that the fact that we find nontrivial inequalities is in stark contrast to the situation for Bell-type inequalities (and line-like causal Bayesian networks), where the DAG-based entropic method only leads to trivial inequalities obtainable from the elemental inequalities and no-signaling conditions [23].

While these entropic inequalities are obeyed by any bipartite causal correlation, we note that *a priori* they need not be tight. Indeed, recall that the Shannon cone is only an outer approximation to the true entropy cone, and the method we applied to bound convex combinations of fixed-order correlations may introduce extra slack. It is thus interesting to study the tightness and violation of these inequalities more carefully.

Although one generally would not expect every point on the boundary of Γ^{causal} to be obtainable by a causal correlation, it is nonetheless desirable to be able to saturate each inequality by some causal probability distribution for appropriate distributions for X and Y . By looking at deterministic causal distributions with binary inputs and outputs, which can easily be enumerated, we readily verified that all 10 families of inequalities that are bounded by 0 (given in Eq. (A.2)) can indeed be saturated when taking uniformly distributed inputs. However, we were unable to find causal distributions, either by mixing binary ones or by considering more outputs, that saturate the remaining inequalities, and their tightness remains an open question.

To understand the violation by noncausal distributions of the entropic inequalities, we consider the extremal rays of the constrained Shannon cone

$$\Gamma_S \cap \text{LC}(\{X \perp Y\}) \quad (6.9)$$

which violate the inequalities.³ A crucial question is whether or not these extremal rays actually correspond to valid probability distributions (i.e., whether they support entropy vectors), and if not, whether the inequalities can nonetheless be violated.

In order to look at this, it is instructive to first restrict our attention to distributions satisfying $H(X) \leq 1$, $H(Y) \leq 1$, $H(A) \leq 1$ and $H(B) \leq 1$. These constraints are satisfied by all distributions with binary inputs and outputs, and this therefore also allows us to compare the violation of the entropic causal inequalities to the violation of standard causal inequalities that are understood well in this scenario [86]. Imposing these constraints on the cone in Eq. (6.9), one obtains a polytope with extremal points corresponding to the extremal rays of the cone scaled to satisfy these constraints (together with the null vertex $\mathbf{0}$). Under these constraints we found that the 10 inequalities in Eq. (A.2) and the two Eq. (A.3) could be violated, although the latter are weaker than, and implied by, the former and are thus redundant. The remaining 8 inequalities in Eqs. (A.4) and (A.5) cannot be violated. All in all, the set of binary causal correlations is entropically characterized by the 10 inequalities in Eq. (A.2) that are bounded by 0.

Amongst the extremal points violating each of these inequalities, those that give the maximal violation all satisfy $H(X) = H(Y) = 1$ and $H(XY) = H(XYAB)$ and thus, if realizable, correspond to deterministic conditional distributions taken with uniformly distributed inputs X and Y . In fact, all but one of these 10 inequalities are maximally violated (by which we henceforth mean with respect to the Shannon cone augmented with the independence constraint $X \perp Y$) by one of the three following deterministic distributions taken with uniform inputs:

$$\begin{aligned} P(ab|xy) &= \delta_{a,y} \delta_{b,x \oplus y} \\ P(ab|xy) &= \delta_{a,x \oplus y} \delta_{b,x} \\ P(ab|xy) &= \delta_{a,x \oplus y} \delta_{b,x \oplus y}, \end{aligned} \tag{6.10}$$

where x, y, a, b take the binary values 0, 1, and \oplus denotes addition modulo 2. For example, Eq. (6.8) is violated by the third distribution with a value for the left-hand side of 1. The one exception not violated by the distributions in Eq. (6.10) is the second inequality in (A.2),

$$I(A : B) - I(A : B|X) - I(A : B|Y) - 2H(AB|XY) \leq 0, \tag{6.11}$$

which, in turn, is violated by the deterministic distribution (again taken with uniform inputs)

$$P(ab|xy) = \delta_{a,x \oplus xy} \delta_{b,y \oplus xy}. \tag{6.12}$$

³Note that the nontriviality of the inequalities implies that such extremal rays indeed exist.

However, unlike for the other inequalities, this distribution does not give the maximal possible violation of inequality (6.11) (which is $1/2$), as the corresponding extremal point h_{ext} that does maximally violate it is not reachable by a valid probability distribution with binary inputs and outputs. This is easily verified by making use of the previous observation that this extremal point must correspond to a deterministic distribution taken with uniform inputs, the set of which can easily be enumerated for binary inputs and outputs. Amongst such distributions, the one in Eq. (6.12) gives the best violation of $1 - \frac{3}{2} \log_2 \frac{3}{2} \approx 0.123 > 0$.

The distributions in Eq. (6.10) are particularly interesting, as they all violate maximally some symmetries of the GYNI inequality (2.40) (i.e. under relabeling of the parties, inputs, and outputs), but not Eq. (2.40) itself. Interestingly, it turns out that *all* binary deterministic noncausal distributions, when taken with uniform inputs, violate at least one of our entropic inequalities *except* the distribution $P^{\text{GYNI}}(ab|xy) = \delta_{a,y}\delta_{b,x}$ (which violates maximally Eq. (2.40)) and its four symmetries under relabeling of outputs only. Note, however, that if Alice and Bob have a noncausal resource producing the distribution P^{GYNI} , they can produce any of the distributions in Eq. (6.10) by appropriately XORing their input with their output, and thus still obtain an operational violation of an entropic causal inequality.⁴ It is interesting to observe that distributions maximally violating GYNI-type inequalities have such a crucial role in violating the entropic causal inequalities given that the entropic inequalities superficially bear little resemblance to these, and are valid for arbitrary numbers of inputs and outputs.

Returning to the more general situation with no upper bound imposed on $H(X)$, $H(Y)$, $H(A)$ and $H(B)$, we see that all the remaining entropic causal inequalities can be violated by entropy vectors that are parallel to the realizable entropy vectors giving violations in the restricted scenario – more precisely, those obtained from the distributions Eq. (6.10) (for all but one of the remaining inequalities) and Eq. (6.12) (for the remaining one). This shows that, given large enough alphabets for the input and output variables, all the entropic causal inequalities we obtained can indeed be violated by noncausal probability distributions, since if the distribution $P(xyab)$ has entropy vector h then the distribution

$$P(xyab) = P(x_1y_1a_1b_1) \times \cdots \times P(x_ny_na_nb_n), \quad (6.13)$$

⁴This illustrates an important difference between the probabilistic and entropic frameworks: while all symmetries of a correlation obtained by flipping inputs and outputs (possibly conditioned on the local inputs for the latter) are equivalent in the probabilistic case (in the sense that if one violates a causal inequality, then all other ones violate a symmetry of that inequality) this is not the case in the entropic approach. The entropy vectors of two different symmetries of a correlation may be inequivalent, with one violating an entropic causal inequality while the other does not.

where $\mathbf{x} = (x_1, \dots, x_n)$ and similarly for \mathbf{y} , \mathbf{a} and \mathbf{b} , has entropy vector $n \cdot \mathbf{h}$. One should be careful, however, to note that the operation of sharing multiple independent correlations among the same parties is not a free operation either in the framework of causal correlations (since, for example, two independent copies of a causal distribution may give rise to a noncausal one), or in the process matrix framework (where two independent copies of a process matrix does not, in general, produce a valid process matrix). Nevertheless, $P(\mathbf{ab}|\mathbf{xy}) = P(\mathbf{xyab})/P(\mathbf{xy})$ obtained from Eq. (6.13) still represents a valid (possibly noncausal) distribution.

It is interesting also to ask how sensitive the entropic causal inequalities are for detecting noncausality. Since it does not appear possible to saturate the inequalities (A.3)–(A.5) with non-zero bounds using causal distributions, these inequalities are not tight and, consequentially, unable to detect noncausal correlations that are very close to being causal. For the other inequalities in Eq. (A.2) this is nonetheless a pertinent question. More precisely, one may ask whether there exists a distribution P^ε of the form

$$P^\varepsilon(\mathbf{ab}|\mathbf{xy}) = \varepsilon P^{\text{NC}}(\mathbf{ab}|\mathbf{xy}) + (1 - \varepsilon)P^{\text{C}}(\mathbf{ab}|\mathbf{xy}), \quad (6.14)$$

where P^{NC} is a noncausal distribution and P^{C} is causal, that violates any of these entropic inequalities for arbitrarily small $\varepsilon > 0$.

We looked in detail at this question for the case of binary inputs and outputs, where the inequalities in Eq. (A.2) can all both be saturated by causal distributions, and violated by noncausal ones. By trying exhaustively all deterministic distributions P^{NC} and P^{C} , we found that such behaviour was exhibited (for such distributions) only by the two inequalities

$$I(A : B|X) - I(Y : B) - 2H(B|XY) \leq 0 \quad (6.15)$$

and

$$I(XA : Y) + I(YB : X) - H(X|YA) - H(A) \leq 0. \quad (6.16)$$

Equation (6.15), for example, is violated by P^ε for all $\varepsilon > 0$ when taking $P^{\text{NC}}(\mathbf{ab}|\mathbf{xy}) = \delta_{a,x \oplus y} \delta_{b,x \oplus y}$ and $P^{\text{C}}(\mathbf{ab}|\mathbf{xy}) = \delta_{a,0} \delta_{b,x \oplus y}$ along with uniformly distributed inputs X and Y , which also gives a violation of the GYNI-type causal inequality

$$\frac{1}{4} \sum_{x,y,a,b} \delta_{a,x \oplus y} \delta_{b,x \oplus y} P(\mathbf{ab}|\mathbf{xy}) \leq \frac{1}{2} \quad (6.17)$$

with a left-hand side value of $\frac{1+\varepsilon}{2} > \frac{1}{2}$.

For the remaining inequalities, such mixtures that violate a standard causal inequality for arbitrarily small ε only violate an entropic causal inequality when $\varepsilon > \varepsilon_0$ for some ε_0 bounded away from 0. We observed identical behavior when we extended our consideration also to various non-deterministic distributions P^{NC} and P^{C} , and it thus seems that only Eqs. (6.15) and (6.16) exhibit this ability to detect the noncausality of distributions that are arbitrarily close to being causal.

A final point worth discussing relates to the physical interpretation of the distributions violating entropic causal inequalities. One of the motivations in introducing the notion of causal correlations was whether nature permits more general causal structures that might allow such correlations to be realized, for example in quantum gravity [25]. In particular, the authors of Ref. [25] introduced the so-called process matrix formalism, in which quantum mechanics is assumed to hold locally for each party, while no global order is assumed between the parties. They showed that causal inequalities can be violated within this framework, and this helped motivate further studies of causal correlations, where it has been shown that the violation of facet-inducing causal inequalities is ubiquitous within this framework [83, 84, 86, 87, 157, 158]. It is thus interesting to see whether entropic causal inequalities share this property and can also be violated within the process matrix framework.

To look for such violations, we used the optimization techniques of Refs. [86, 87] with qubit systems to try and optimize the violation of the GYNI-type inequalities that the distributions in Eq. (6.10) violate maximally. We also tried minimizing the distance to other deterministic noncausal correlations such as Eq. (6.12), as well as optimizations in random directions in probability space. Unfortunately, we were unable to find any process matrices operating on qubits that violate entropic causal inequalities with such techniques. We additionally attempted to reproduce (as closely as possible) distributions of the form (6.14) for small ε in order to violate inequalities (6.15) and (6.16), but similarly found no violation. Finally, we looked at noncausal correlations obtained by mixing noncausal correlations realizable by process matrices with causal correlations. An analogous mixing procedure was shown to enable all nonlocal distributions to violate the entropic Bell inequalities described in Sec. 2.5.4 [93], but we were unable to find violations of the entropic causal inequalities with this approach.

This lack of violation is perhaps unsurprising given the general lack of sensitivity of the entropic inequalities to nearly-causal distributions, and the fact that the best-known violations of causal inequalities for this scenario with process matrices are relatively small [86]. Nonetheless, it remains possible that violations can be found with higher-dimensional systems or more inputs and outputs.

6.2.2 Characterization based on counterfactual variables

In this section we will consider counterfactual variables as outlined in Sec. 2.5.4. Rather than considering the inputs as random variables X and Y , we take copies of each output variable for all input combinations, i.e. A_{xy} and B_{xy} . In contrast to the method based on causal Bayesian networks, this method fixes the number of inputs that the inequalities apply to but may lead to novel constraints, as is the case in the Bell scenario.

Counterfactual variables for bipartite causal correlations

To keep the discussion simple, we will consider only the case of binary inputs, but the generalization to arbitrary inputs is straightforward. We consider the variables

$$S = \{A_{00}, A_{01}, A_{10}, A_{11}, B_{00}, B_{01}, B_{10}, B_{11}\}. \quad (6.18)$$

Note that, in contrast to the example of Bell inequalities discussed in Sec. 2.5.4, we need to consider copies of each variable for each input pair (x, y) . This is a consequence of the fact that the correlations which we want to characterize may be signaling, e.g., for the causal order $A \prec B$, B_{00} and B_{10} will in general be different.

Since A_{xy} and $B_{x'y'}$ are jointly observable only if $x = x'$ and $y = y'$, the marginal scenario in this case is

$$\mathcal{M} = \left\{ \{A_{00}, B_{00}\}, \{A_{01}, B_{01}\}, \{A_{10}, B_{10}\}, \{A_{11}, B_{11}\} \right\}. \quad (6.19)$$

In contrast to the DAG-based method, several choices of probability structure \mathcal{S} compatible with \mathcal{M} are possible, and the particular choice must be motivated on the basis of physical assumptions. One natural possibility would be to take $\mathcal{S} = \mathcal{M}$, as one may have no *a priori* reason to think that the variables A_{xy} and $A_{x'y'}$ have simultaneous physical meaning for $(x, y) \neq (x', y')$, and hence may not have a well-defined joint probability distribution. On the other hand, in some cases one may imagine that such inputs correspond to the choice of measurements of some physical properties that are simultaneously well-defined, as in a classical theory; hence, one may alternatively take $\mathcal{S} = \{\cup_{M_j \in \mathcal{M}} M_j\} = \{S\}$. In the following, we will adopt the former approach and take $\mathcal{S} = \mathcal{M}$, since this constitutes the minimum assumptions compatible with the marginal scenario. The Shannon cone for \mathcal{S} is thus

$$\Gamma^{\mathcal{S}} = \Gamma_{\{A_{00}, B_{00}\}} \cap \Gamma_{\{A_{01}, B_{01}\}} \cap \Gamma_{\{A_{10}, B_{10}\}} \cap \Gamma_{\{A_{11}, B_{11}\}}, \quad (6.20)$$

as in Eq. (2.46). We note however that this physically motivated choice for \mathcal{S} implies, for this particular scenario, that a global probability distribution does in fact exist.⁵ Taking $\mathcal{S} = \{S\}$ would thus provide an equivalent entropic characterization, and moreover, an equivalent characterization also at the level of Shannon (rather than entropic) cones (see Appendix A, for an extensive discussion).

We follow a method analogous to that used in Sec. 6.2.1. First, we characterize the cones $\Gamma_{A \prec B}$ and $\Gamma_{B \prec A}$ of entropy vectors for fixed-order causal correlations, then, we characterize the convex mixtures of such correlations.

To do this, we note that the no-signaling conditions obeyed by fixed-order correlations (see Sec. 2.3.7) impose constraints on the counterfactual variables. For example, correlations consistent with the order $A \prec B$ obey $P(a|xy) = P(a|xy')$ for all x, y, y', a , which implies $A_{xy} = A_{xy'}$ and thus $H(A_{xy}) = H(A_{xy'})$ also. Similarly, for $B \prec A$, we have $H(B_{xy}) = H(B_{x'y})$ for all x, x', y . The cones $\Gamma_{A \prec B}$ and $\Gamma_{B \prec A}$ are thus given by

$$\Gamma_{A \prec B} = \Gamma^{\mathcal{S}} \cap \text{LC}(\{A_{00} = A_{01}, A_{10} = A_{11}\}) \quad (6.21)$$

and

$$\Gamma_{B \prec A} = \Gamma^{\mathcal{S}} \cap \text{LC}(\{B_{00} = B_{10}, B_{01} = B_{11}\}), \quad (6.22)$$

where $\text{LC}(\cdot)$ again denotes the linear subspace defined by the corresponding constraints.

As in Sec. 6.2.1, we introduce the latent switch variable Q , denote the augmented set of random variables $\tilde{\mathcal{S}} = \mathcal{S} \cup \{Q\}$, and extend the probability structure as

$$\tilde{\mathcal{S}} = \left\{ \{A_{xy}, B_{xy}, Q\} \mid x, y \in \{0, 1\} \right\} \quad (6.23)$$

(in Appendix A we discuss further the implications of different choices of probability structures). With this extra variable we note again that the convex hull $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$ contains the vectors $\mathbf{h}_{\text{conv}} = (H(T|Q))_{T \in \mathcal{S}^c}$ for causal correlations. The system of inequalities $I\mathbf{h} \leq \mathbf{0}$ characterizing $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$ can then again be transformed in a similar way to Eq. (6.5) into a new system $\tilde{I}\tilde{\mathbf{h}} \leq \mathbf{0}$ defining the cone of corresponding entropy vectors $\tilde{\mathbf{h}} = (H(T))_{T \in \tilde{\mathcal{S}}^c}$, which we again denote by $\text{conv}_Q(\Gamma_{A \prec B}, \Gamma_{B \prec A})$. In contrast to the DAG-based method, the only constraint on Q is, now, $H(Q) \leq 1$, since Q need not be independent of the (counterfactual) output variables A_{xy}, B_{xy} . Finally, we need to project onto the marginal scenario \mathcal{M} in Eq. (6.19). The causal polyhedron

⁵ This is the result of the more general fact that different choices of \mathcal{S} may provide equivalent descriptions of marginal probabilities [159] and entropies [27].

is thus given, in analogy to Eqs. (6.6) and (6.7), by

$$\Gamma^{\text{causal}} = \Pi_{\mathcal{M}} \left[\Gamma^{\tilde{\mathcal{S}}} \cap \text{conv}_Q(\Gamma_{A \prec B}, \Gamma_{B \prec A}) \cap \text{Lc}(\{H(Q) \leq 1\}) \right], \quad (6.24)$$

where we have $\Gamma^{\tilde{\mathcal{S}}} = \bigcap_{x,y \in \{0,1\}} \Gamma_{\{A_{xy}, B_{xy}, Q\}}$.

Entropic causal inequalities for counterfactual variables and their violation

As in Sec. 6.2.1, the construction above allows one to obtain the full list of entropic inequalities characterizing Γ^{causal} . After removing the trivial inequalities directly implied by Shannon constraints on \mathcal{M} , we find that there are 6 nontrivial entropic causal inequalities, which can be grouped into two equivalence classes of inequalities under the relabeling of inputs:

$$I(A_{00} : B_{00}) - H(A_{01}) - H(B_{10}) \leq 1 \quad (6.25)$$

and

$$I(A_{00} : B_{00}) + I(A_{11} : B_{11}) - H(A_{01}B_{01}) - H(A_{10}B_{10}) \leq 2. \quad (6.26)$$

The fact that these inequalities have nontrivial bounds is, as for the DAG-based method, a result of the constraint $H(Q) \leq 1$ which means Γ^{causal} is a polyhedron characterized by a set of inhomogeneous inequalities. Indeed, if one chooses not to eliminate Q from the entropic description, one obtains a convex cone characterized by the above equations, except that the right-hand side is multiplied by $H(Q)$ (see the discussion in Appendix A).

In contrast to the case for the DAG-based approach, where violation of the causal inequalities we obtained was possible even with deterministic distributions, it is clear that such distributions provide no interesting behavior in the counterfactual approach since any such distribution will have a null entropy vector. By looking at equal mixtures of deterministic causal distributions, however, we were able to verify that the inequalities in Eqs. (6.25)–(6.26) can indeed be saturated by such (causal) distributions and are thus tight. In order to study the potential violation of these entropic inequalities, we again need to look at nondeterministic distributions. One can easily see, however, that Eqs. (6.25)–(6.26) cannot be violated when restricted to distributions satisfying $H(A_{xy}) \leq 1$ and $H(B_{xy}) \leq 1$ for all $x, y \in \{0, 1\}$, as this also implies that $I(A_{xy} : B_{xy}) \leq 1$. This means that the inequalities for counterfactual variables are unable to detect noncausality when both parties are restricted to binary outputs.

To study possible violations we again look at the extremal rays of the Shannon cone Γ^S of Eq. (6.20) which violate one of the inequalities, and examine whether these rays can be reached by any probability distribution. Considering bounds on $H(A_{xy})$ and $H(B_{xy})$ strictly larger than 1, we find that violations are possible for any such bound. Moreover, the entropy vectors giving maximal violation of Eqs. (6.25) and (6.26) are generally realizable with equal mixtures of causal and noncausal distributions. For example, given the constraints $H(A_{xy}) \leq \log_2 k$ and $H(B_{xy}) \leq \log_2 k$ for some integer $k \geq 2$, the distribution

$$P_k(ab|xy) = \delta_{x,y} \frac{1}{k} \delta_{a,b} + (1 - \delta_{x,y}) \delta_{a,0} \delta_{b,0}, \quad (6.27)$$

where $a, b \in \{0, \dots, k-1\}$, realizes such an extremal point for all $k \geq 2$, and provides a violation of both Eqs. (6.25) and (6.26) for $k > 2$. For $k = 2$ (binary outputs), this distribution can be written as the convex combination

$$P_2(ab|xy) = \frac{1}{2} P^{\text{NC}}(ab|xy) + \frac{1}{2} P^{\text{C}}(ab|xy), \quad (6.28)$$

where $P^{\text{NC}}(ab|xy) = \delta_{a \oplus 1, x \oplus y} \delta_{b \oplus 1, x \oplus y}$ maximally violates a GYNI-type inequality (it is simply a symmetry of the third distribution in Eq. (6.10), obtained by flipping all outputs), and $P^{\text{C}}(ab|xy) = \delta_{a,0} \delta_{b,0}$ is causal. Even though it does not violate Eq. (6.25) or (6.26), P_2 is noncausal. The distribution P_k can be seen as a possible generalization of a GYNI-violating distribution.

This link to the GYNI-type inequalities and correlations can be made more explicit by considering the related distribution

$$P'_k(ab|xy) = \delta_{x,y} \frac{1}{k-1} \delta_{a,b} (1 - \delta_{a,0} \delta_{b,0}) + (1 - \delta_{x,y}) \delta_{a,0} \delta_{b,0}, \quad (6.29)$$

with again $a, b \in \{0, \dots, k-1\}$. We have $P'_2 = P^{\text{NC}}$, and, for $k \geq 3$, P'_k has the same entropy vector as P_{k-1} . P'_k can be clearly simulated from $P'_2 = P^{\text{NC}}$ by making use of shared randomness and by letting both parties replace the output 1 obtained from P'_2 by a shared random value $a = b \in \{1, \dots, k-1\}$. It is interesting to see, then, that the GYNI-maximally-violating distributions also provide the best behavior entropically, when augmented with shared randomness, even though they fail to violate the inequalities when the parties have only binary outputs.

As for the DAG-based method, it is also interesting to look at the sensitivity of the inequalities with respect to the detection of noncausality. Unfortunately, by looking at distributions of the form given in Eq. (6.14), but now where P^{NC} and P^{C} equal mixtures of 3-outcome deterministic noncausal and causal distributions, respectively, we

were unable to find such distributions $P^\varepsilon(ab|xy)$ which violate the entropic inequalities (6.25) and (6.26) for arbitrary small ε .

Finally, one may again ask whether one can violate any of the entropic inequalities for counterfactuals within the process matrix formalism, or whether any noncausal correlation can be mixed with a causal one to violate an entropic inequality, as is the case for entropic Bell inequalities obtained from the counterfactual approach [93]. We leave this as an open question, but note only that we were not able to find a way to do so: for example, we were unable to find a violation (with or without the use of shared randomness) for noncausal distributions realizable within the process matrix framework.

6.3 Multipartite entropic causal inequalities

The notion of causal correlations can be extended to more than two parties in a recursive manner [85, 87]. Consider N parties A_1, \dots, A_N , with inputs $\mathbf{x} = (x_1, \dots, x_N)$ and outputs $\mathbf{a} = (a_1, \dots, a_N)$. In any given run, one party, say A_k , must act first, and none of the other parties can signal to them, which implies $P(a_k|\mathbf{x}) = P(a_k|x_k)$. The correlations shared by the remaining $N - 1$ parties, conditioned on the input and output of the first, must also in turn be causal. However, note that the causal order itself (and not only the response functions) of the remaining parties may depend on the input and output of the first, a phenomenon called dynamical causal order [85, 87, 156], and which goes beyond the standard model of fixed causal Bayesian networks.

An N -partite correlation $P(\mathbf{a}|\mathbf{x})$ is thus called causal if it can be decomposed in the following way [85, 87]:

$$P(\mathbf{a}|\mathbf{x}) = \sum_{k=1}^N q_k P_k(a_k|x_k) P_{k,x_k,a_k}(\mathbf{a}_{\setminus k}|\mathbf{x}_{\setminus k}), \quad (6.30)$$

where $\mathbf{x}_{\setminus k} = (x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_N)$ and $\mathbf{a}_{\setminus k} = (a_1, \dots, a_{k-1}, a_{k+1}, \dots, a_N)$, with $q_k \geq 0$, $\sum_k q_k = 1$, and where for each k, x_k, a_k , $P_{k,x_k,a_k}(\mathbf{a}_{\setminus k}|\mathbf{x}_{\setminus k})$ is a causal $(N-1)$ -partite correlation (down to the lowest level of this recursive definition, where any 1-partite correlation is considered to be causal). Note that, for $N = 2$ this reduces to Eq. (2.39). The entropic approach can be generalized to the multipartite scenario using a similar recursive method.

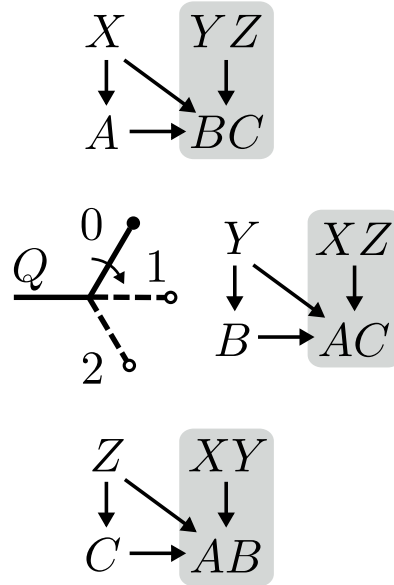


FIGURE 6.2: DAGs for tripartite causal correlations. The latent “switch” variable Q determines which DAG is “activated”. Correlations among variables from shaded rectangles are causal conditionally on the input and output of the party acting first.

6.3.1 Causal Bayesian network method

It is instructive to first look into the details of the tripartite case – in which case we shall denote the parties Alice (A), Bob (B) and Charlie (C), as is standard – before generalizing the method to more parties. The general method follows that used for the bipartite case in Sec. 6.2.1, and the relevant conditional DAG is shown in Fig. 6.2. The set of observable variables to be considered here is $S = \{X, Y, Z, A, B, C\}$.

The polytope of tripartite causal correlations (i.e., of the form Eq. (6.30)) can be written as

$$\mathcal{P}_{ABC}^{\text{causal}} = \text{conv}(\mathcal{P}^A, \mathcal{P}^B, \mathcal{P}^C), \quad (6.31)$$

where \mathcal{P}^A is the polytope of causal distributions consistent with Alice acting first and such that the remaining conditional correlation shared by Bob and Charlie is causal, and analogously for \mathcal{P}^B and \mathcal{P}^C . As a consequence, in order to define the polyhedron characterizing tripartite causal correlations, which we denote $\Gamma_{ABC}^{\text{causal}}$, we first need to define the corresponding Shannon polyhedra, namely $\Gamma^A, \Gamma^B, \Gamma^C$, associated with each party acting first.

Let us thus consider Γ^A . According to the recursive definition given in Eq. (6.30), for any x, a , the conditional entropy vector $\mathbf{h}_{BC}^{xa} = (H(T|X = x, A = a))_{T \subset \{Y, Z, B, C\}}$ for a correlation in \mathcal{P}^A must be contained in the bipartite causal polyhedron $\Gamma_{BC}^{\text{causal}}$, defined for Bob and Charlie as in Eqs. (6.6)–(6.7). By convexity this also implies that

$\mathbf{h}_{BC} = (H(T|XA))_{T \subset \{Y,Z,B,C\}} = \sum_{x,a} P(x,a) \mathbf{h}_{BC}^{xa}$ is in $\Gamma_{BC}^{\text{causal}}$. We can then use a similar transformation to Eq. (6.5) to obtain constraints on Γ^A : if entropy vectors \mathbf{h}_{BC} in $\Gamma_{BC}^{\text{causal}}$ satisfy the inequalities $\mathbf{I} \mathbf{h}_{BC} \leq \beta$, then the corresponding (unconditional) entropy vector $\mathbf{h} = (H(T))_{T \subset S}$ must satisfy the inequalities $\mathcal{T}_{XA}(\mathbf{I}) \mathbf{h} \leq \beta$. Writing \mathcal{T}_{XA}^* for the dual transformation on the space of entropy vectors, we thus have that $\mathbf{h} \in \mathcal{T}_{XA}^*(\Gamma_{BC}^{\text{causal}})$. Together with the facts that \mathbf{h} must lie in the Shannon cone Γ_S for the relevant variables, that all the inputs must be independent from each other, and that Alice's output must be independent from Bob and Charlie's inputs (conditioned on her input), we obtain the characterization

$$\Gamma^A = \Gamma_S \cap \mathcal{T}_{XA}^*(\Gamma_{BC}^{\text{causal}}) \cap \text{L}_C(\{X \perp Y \perp Z, A \perp YZ | X\}), \quad (6.32)$$

with similar expressions for Γ^B and Γ^C .

Following the same approach as in Sec. 6.2.1, we introduce a (now three-valued) switch variable Q (see Fig. 6.2). Similarly to what we observed in the bipartite case, the convex hull $\text{conv}(\Gamma^A, \Gamma^B, \Gamma^C)$ contains the conditional entropy vectors $(H(T|Q))_{T \subset S}$ for tripartite causal correlations. The inequalities characterizing $\text{conv}(\Gamma^A, \Gamma^B, \Gamma^C)$ can again be transformed into inequalities satisfied by the entropy vector $\tilde{\mathbf{h}} = (H(T))_{T \subset \tilde{S}}$, for variables in $\tilde{S} = S \cup \{Q\}$, by introducing a transformation \mathcal{T}_Q as in Eq. (6.5), thus defining the polyhedron $\text{conv}_Q(\Gamma^A, \Gamma^B, \Gamma^C)$ as before. Taking into account the Shannon constraints for all variables in \tilde{S} , the independence constraints $\text{CI}_Q = (X \perp Y \perp Z \perp Q)$ and the bound $H(Q) \leq \log_2 3$, and finally projecting onto the observable variables in S , we see that the entropy vectors for tripartite causal correlations belong to the polyhedron

$$(\Gamma_{ABC}^{\text{causal}})_0 = \Pi_S \left[\Gamma_{\tilde{S}} \cap \text{conv}_Q(\Gamma^A, \Gamma^B, \Gamma^C) \cap \text{L}_C(\{\text{CI}_Q, H(Q) \leq \log_2 3\}) \right]. \quad (6.33)$$

While this characterization is certainly valid, some subtleties arising from the differences between the probabilistic and entropic descriptions allow one to actually make it tighter. Specifically, certain conditions implied by the definition (6.30) need not be implied by the corresponding entropic definition outlined above. For example, if $P(abc|xyz)$ is a causal correlation, then the bipartite marginal distributions $P_x(bc|yz) = \sum_a P(abc|xyz)$ and $P(bc|yz) = \sum_x P(x)P_x(bc|yz)$ are both causal (as are the corresponding marginals for each other pair of parties) [87]. This implies that the entropy vectors $(H(T|X))_{T \subset \{Y,Z,B,C\}}$ and $(H(T))_{T \subset \{Y,Z,B,C\}}$ corresponding to a tripartite causal correlation must also satisfy all the inequalities characterizing the bipartite causal polyhedron $\Gamma_{BC}^{\text{causal}}$ – which may not necessarily be implied by the characterization of $(\Gamma_{ABC}^{\text{causal}})_0$ above. We can thus tighten the previous characterization, and define

the tripartite causal polyhedron as⁶

$$\Gamma_{ABC}^{\text{causal}} = (\Gamma_{ABC}^{\text{causal}})_0 \cap \Gamma_{BC}^{\text{causal}} \cap \mathcal{T}_X^*(\Gamma_{BC}^{\text{causal}}) \cap [\text{perms.}], \quad (6.34)$$

where [perms.] denotes the permutations of the preceding two terms for the other parties. Note that such extra constraints do not need to be imposed in the bipartite case since the causality of all one-party marginals is equivalent to them being valid probability distributions, which is already assured by the elemental inequalities.

To extend the above idea to the general multipartite case of Eq. (6.30), we simply define recursively (here the notation should be self-evident)

$$\Gamma^{A_k} = \Gamma_{\{X,A\}} \cap \mathcal{T}_{X_k A_k}^* \left(\Gamma_{\mathbf{A} \setminus k}^{\text{causal}} \right) \cap \text{L}_C(\text{CI}_{A_k}), \quad (6.35)$$

where CI_{A_k} denotes the set of independence constraints resulting from the assumption that all parties' inputs are independent, i.e. $X_1 \perp \dots \perp X_N$, and that party k acts first, which implies $A_k \perp \mathbf{X} \setminus k | X_k$. The causal polyhedron is then defined as

$$\Gamma_{\mathbf{A}}^{\text{causal}} = \Pi_{X,A} \left[\Gamma_{\{X,A,Q\}} \cap \text{conv}_Q(\{\Gamma^{A_k}\}_k) \cap \text{L}_C(\{\text{CI}_Q, H(Q) \leq \log_2 N\}) \right] \bigcap_k \left[\Gamma_{\mathbf{A}_k}^{\text{causal}} \cap \mathcal{T}_{X_k}^* \left(\Gamma_{\mathbf{A}_k}^{\text{causal}} \right) \right], \quad (6.36)$$

where CI_Q denotes the independence relation between all input variables and Q , i.e. $X_1 \perp \dots \perp X_N \perp Q$.

6.3.2 Counterfactual variable method

A similar generalization is possible also for the counterfactual method. Again, it is instructive to look first at the tripartite case. We start by defining the polyhedron for the case in which Alice acts first,

$$\Gamma^A = \bigcap_{xyz} \left[\Gamma_{\{A_{xyz}, B_{xyz}, C_{xyz}\}} \cap \mathcal{T}_{A_{xyz}}^* \left(\Gamma_{BC}^{\text{causal}} \right) \cap \text{L}_C(\{A_{xyz} = A_{xy'z'}\}_{y'z'}) \right], \quad (6.37)$$

which is the analogue, for the counterfactual method, of the polyhedron in Eq. (6.32). Similar definitions hold for Γ^B and Γ^C . The tripartite polyhedron of causal counterfactual inequalities can then be defined, following a similar reasoning to the previous

⁶In Eq. (6.34) we abuse the notation slightly and denote by $\Gamma_{BC}^{\text{causal}}$ the set of entropy vectors $(H(T))_{T \subset S}$ – instead of $(H(T))_{T \subset \{Y,Z,B,C\}}$ – which satisfy the constraints characterizing $\Gamma_{BC}^{\text{causal}}$ as defined in Eqs. (6.6)–(6.7). The transformation \mathcal{T}_X , of which \mathcal{T}_X^* is the dual, is again defined in a similar way as in Eq. (6.5).

case, as

$$\Gamma_{ABC}^{\text{causal}} = \Pi_{\mathcal{M}} \left[\text{conv}_Q(\Gamma^A, \Gamma^B, \Gamma^C) \cap \left(\bigcap_{xyz} \Gamma_{\{A_{xyz}, B_{xyz}, C_{xyz}, Q\}} \right) \cap \text{Lc}(\{H(Q) \leq \log_2 3\}) \right] \\ \bigcap_x \Gamma_{BC|x}^{\text{causal}} \bigcap_y \Gamma_{AC|y}^{\text{causal}} \bigcap_z \Gamma_{AB|z}^{\text{causal}}, \quad (6.38)$$

where $\mathcal{M} = \{\{A_{xyz}, B_{xyz}, C_{xyz}\}\}_{xyz}$ and $\Gamma_{BC|x}^{\text{causal}}$ is defined by imposing the constraints characterizing Γ_{BC} (*a priori* defined for some variables B_{yz}, C_{yz}) to the variables B_{xyz}, C_{xyz} , and with similar definitions for $\Gamma_{AC|y}^{\text{causal}}$ and $\Gamma_{AB|z}^{\text{causal}}$.

As for the case based on causal Bayesian networks, the construction in Eq. (6.38) can then be generalized to an arbitrary number of parties in a recursive way.

6.4 Information bounds in causal games

One of the advantages of the entropic approach is that it allows information theoretic constraints to be naturally imposed, derived, and interpreted [22, 147]. As an illustration, we consider a simple application of our approach to understanding the role of bounded communication in causal games.

Consider the generalization of the GYNI game described in Sec. 2.3.7 to arbitrary numbers of inputs and outputs, in which two parties try to maximize the winning probability $p_{\text{succ}} = P(a = y, b = x)$. If the parties operate causally, then in any given round of the game only one-way communication may occur. One may be interested in the effect of limiting the amount communication that can occur in any such round. In the entropic framework, this can easily be taken into account by adding an additional constraint of the form $I(X : B) \leq H(M)$ to $\Gamma_{A \prec B}$ in order to restrict B 's dependency on X , and similarly imposing $I(Y : A) \leq H(M)$ to $\Gamma_{B \prec A}$, where the variable M represents the message that is sent. For example, if the parties are permitted, in each round, to exchange a classical d -dimensional system, then $H(M) = \log_2 d$. In general, the amount of one-way communication $H(M)$ does not need to be specified in advance, it will appear as parameter in our inequalities. By applying the approach of Sec. 6.2.1 to this scenario one finds that causal correlations must then obey the inequality

$$I(X : B) + I(Y : A) \leq H(M), \quad (6.39)$$

i.e., the two-way communication is similarly bounded by $H(M)$. Although this is perhaps not unexpected, it shows the ease with which such bounds can be derived in the entropic framework.

A more subtle variant is obtained by considering a slight generalization of the causal game proposed by Oreshkov, Costa, and Brukner (OCB) in Ref. [25]. In this game, the goal is also for one party to guess the other party's input; in contrast to the GYNI game, however, an additional input random bit Y' is given,⁷ which determines whether it is Bob who should guess Alice's input (if $Y' = 0$) or vice versa (if $Y' = 1$). The parties thus now attempt to maximize the winning probability

$$p_{\text{succ}} = \frac{1}{2} \left(P(b = x | Y' = 0) + P(a = y | Y' = 1) \right). \quad (6.40)$$

An analogous entropic inequality can be obtained via a combination of the methods discussed in Sec. 6.2. Since the relevant direction of communication in each round of this game depends on the additional input Y' , we will combine the DAG-based method for the variables A, B, X, Y with the counterfactual approach to condition on Y' . More precisely, one may take $\tilde{\mathcal{S}} = \{ \{A_{y'}, B_{y'}, X, Y, Q\} \}_{y'}$ and $\mathcal{M} = \{ \{A_{y'}, B_{y'}, X, Y\} \}_{y'}$; the relevant causal constraints for the cones $\Gamma_{A \prec B}$ and $\Gamma_{B \prec A}$ and the polyhedron $\tilde{\Gamma}^{\text{causal}}$ are the same as those imposed on A, B, X, Y, Q in the DAG-based method, except that now they are applied to each copy of the conditional variables $A_{y'}$ and $B_{y'}$, and the communication bounds

$$I(X : B_{y'}) \leq H(M), \quad \text{and} \quad I(Y : A_{y'}) \leq H(M) \quad (6.41)$$

are imposed on the corresponding cones. Notice that, in this way, we are assuming that $Q \perp X \perp Y \perp Y'$. Combining the above constraints with the analysis in Sec. 6.2, one finds that causal correlations must obey

$$I(X : B | Y' = 0) + I(Y : A | Y' = 1) \leq H(M). \quad (6.42)$$

This inequality, for the special case of binary inputs and outputs and with $H(M) = 1$, was proposed in Ref. [160] as a potential principle to bound the set of correlations obtainable within the process matrix formalism,⁸ in analogy with the celebrated information causality principle [147] that provides bounds on the strength of bipartite quantum correlations. Our approach allowed us to show that Eq. (6.42) indeed holds for causal processes, but it remains to be seen whether such a constraint on communication for causal correlations can be violated within the process matrix framework.

⁷In the original OCB game, only one party receives the input Y' , whereas in the variant we consider here, both parties have access to it.

⁸Ref. [160] proposed this inequality in the framework of the original OCB game. However, one can easily see that our derivation of Eq. (6.42) in the more general scenario implies that it must hold in that framework too. Indeed, if only Bob receives Y' , then this implies the additional constraint $H(A_0) = H(A_1)$ when $A \prec B$. The set of correlations obtainable is thus a subset of those obtainable in the more general version of the game, and thus Eq. (6.42) must again hold true.

This example, however, highlights the potential of the entropic approach to causal correlations for studying information-theoretic principles.

6.5 Discussion

Since Bell first formulated his eponymous theorem, understanding the role of causality within quantum mechanics has been a central yet thorny goal. Complicating matters further, the very idea of a definite causal order itself has begun to be questioned. While sophisticated frameworks have been introduced in an effort to free physical theories from the shackles of a rigid causal framework, the issue of whether nature permits violations of causal inequalities remains an elusive question.

Against this backdrop, our aim in this work was to introduce an entropic approach to studying causal correlations, and to this end we presented two complementary methods: the first based on the consideration of the entropies of the variables appearing in the causal Bayesian networks describing causal scenarios, and the second based on a counterfactual description of the outcome variables appearing in such networks. Focusing on bipartite causal scenarios, we described in detail the successful application of both methods to derive nontrivial entropic causal inequalities, before showing how the characterizations can be generalized to multipartite scenarios. In contrast to the usual approach to causal correlations based on probability distributions, the entropic causal inequalities we derived using both methods are valid for any finite number of possible outcomes, as well as for any number of inputs for the first method based on causal Bayesian networks, and thus provide a very concise description of causal correlations. We discussed the ability for the derived entropic causal inequalities to witness the noncausality of several classes of interesting noncausal correlations, but were nonetheless unable to find violations of the inequalities by correlations obtainable within the process matrix formalism [25] using qubit systems. In light of the coarse-grained description provided by entropic inequalities and the fact that the known violations of standard causal inequalities are in general rather small [86], that is arguably an unsurprising negative result. The question of whether entropic causal inequalities can be violated within the process matrix formalism and (more importantly) by quantum correlations thus remains open. More generally, our construction can be used to characterize arbitrary convex combinations of different causal Bayesian networks, and thus provides, for example, a natural tool to investigate stronger notions of multipartite Bell nonlocality [153, 154, 161, 162] from the entropic perspective.

In view of this new framework for the study of causal correlations we believe that several other directions of research can naturally be pursued. Here we focused on using

the Shannon entropies of the relevant variables, but it is known that, at least in particular scenarios, the same approach can be used to derive constraints using certain generalized entropies [163, 164] and even with non-statistical information measures such as the Kolmogorov complexity [95]. Can our framework be extended to these other information measures, and if so, are they more sensitive to violations of causality? Similarly, one may wonder whether the addition of non-Shannon-type inequalities to the entropic descriptions of causal correlations considered might lead to tighter constraints [92, 132, 165]. More generally, it remains an open question whether the definition of causal correlations implies any additional constraints within the entropic description that might allow a tighter characterization, particularly in the multipartite case, similar to the additional constraints on marginal and conditional entropies imposed in Sec. 6.3.

Another important direction to consider would be the ability to formulate, and perhaps violate, information-theoretical principles [22] of causality. We provided, as a simple application, an idea at one possible approach, showing how simple bounds on two-way communication can be derived for causal games where communication is limited in each direction. It would be interesting to see, in particular, whether such principles could be violated within the process matrix formalism and, if so, the connection to the violation of causal inequalities. For example, does the violation of causal inequalities imply the violation of some principle implied by quantum mechanics? We expect our results to motivate these and many more future investigations.

Conclusions

In this thesis we have considered different problems in quantum mechanics. Although these problems look very different there is a common idea of inferring the information about the whole system from the data about the subsystems which connects all of them. These problems can be all treated as parts of the larger problem, namely the marginal problem. In this way the results of this thesis may benefit our understanding of more general relations between the parts and the whole in physics.

In Chapter 3 we have considered a phenomenon of emergence of multiparticle entanglement. In this phenomenon multiparticle entanglement of the global state can be proven from the separable marginals. We have shown that quantum states with such properties exist for arbitrary number of particles and we have presented a method to obtain states of few particles for which this phenomenon is likely to be the most noticeable. In addition we have considered various extensions of this problem including the situations where only some of the marginals are known, or where entanglement cannot be localized in the marginals.

In the next Chapter 4 we have introduced a generalization of the class of hypergraph state from the qubit register to d -level systems. In this part of the thesis we have concentrated on the problem of distinguishing different local equivalence classes of tripartite qudit hypergraph states for dimension 3 and 4. The main criteria, which was used to classify these states, is based on the ranks of the reduced states, i.e. in this case the information about the marginals is the most relevant one.

Starting from Chapter 5 we have altered our focus from problems in the entanglement theory to the problems of causal reasoning in quantum mechanics. In Chapter 5 we considered the problem of inferring causal relations under the restrictions on the observed data. More formally, we have derived a necessary condition for general causal structures to be distinguishable from correlations in the given marginal scenario. This result can be applied to the analysis of scenarios for testing nonlocality or presence of signaling in quantum experiments where we have access only to some marginal correlations.

Finally, in Chapter 6 we considered quantum processes for which the causal order might be indefinite. In this framework, the order, like space-time, between events in different laboratories is not fixed, but rather is considered to be a dynamical parameter of the theory. Indefiniteness of the causal order is then can be proven by violation of so-called causal inequalities, which are analogous of Bell inequalities but for signaling correlations. In in Chapter 6 we derived entropic causal inequalities and discussed their violation. In particular, we considered two different variations of the entropic approach and obtained two families of nontrivial constraints on entropic region, associated with causal processes. We concluded the chapter by discussing the information bounds on communication in theories where the causal order is definite.

Appendix A

An Appendix to Chapter 6

Causal correlations not contained in $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$

Starting with the systems of inequalities $I_0 \mathbf{h} \leq \mathbf{0}$ and $I_1 \mathbf{h} \leq \mathbf{0}$ characterizing the cones $\Gamma_{A \prec B}$ and $\Gamma_{B \prec A}$ defined in Eqs. (6.1) and (6.2), the characterization $I \mathbf{h} \leq \mathbf{0}$ of $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$ can be found by first solving the extremal ray enumeration problem for the extremal rays of $\Gamma_{A \prec B}$ and $\Gamma_{B \prec A}$, taking the union of these rays and finally solving the facet enumeration problem for the inequalities characterizing $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$.

We find that there are six nontrivial inequalities (i.e., non Shannon-type inequalities) for $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$, which correspond to four equivalence classes of inequalities under exchange of parties:¹

$$\begin{aligned}
 I(X : YA) + I(Y : XB) - I(XY : AB) &\leq 0 \\
 I(A : B) - I(A : B|X) - I(A : B|Y) &\leq 0 \\
 I(X : A|B) - I(XB : A|Y) &\leq 0 \\
 I(A : B|X) - I(A : B|XY) - I(Y : B) &\leq 0.
 \end{aligned} \tag{A.1}$$

In order to see that there are causal bipartite correlations that have entropy vectors not contained in $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$, consider the following counterexample. Take $P^{A \prec B}(ab|xy) = \delta_{a,x} \delta_{b,x}$ and $P^{B \prec A}(ab|xy) = \delta_{a,y} \delta_{b,y}$ and consider the inputs x, y to be uniformly distributed so that $P^{A \prec B}(xyab) = \frac{1}{4} P^{A \prec B}(ab|xy)$ and $P^{B \prec A}(xyab) = \frac{1}{4} P^{B \prec A}(ab|xy)$. The distribution $P(ab|xy) = \frac{1}{2} (P^{A \prec B}(ab|xy) + P^{B \prec A}(ab|xy))$ is thus

¹For compactness we generically write entropic inequalities not just in terms of Shannon entropies (as defined in Eq. (2.41)), but also in terms of conditional entropies (of the form $H(A|B) := H(AB) - H(B)$), of mutual information ($I(A : B) := H(A) + H(B) - H(AB)$) and of conditional mutual information ($I(A : B|C) := H(AC) + H(BC) - H(ABC) - H(C)$). The expressions given for the inequalities are of course not unique.

also causal, but one can verify that the entropy vector for the distribution $P(xyab) = \frac{1}{4}P(ab|xy)$ violates the first and last inequalities in (A.1) with a value for the left-hand sides of $1 - \frac{3}{2} \log_2 \frac{3}{2} \approx 0.123 > 0$.

A similar conclusion can also be reached for the method based on counterfactual variables: starting from the definitions of $\Gamma_{A \rightarrow B}$ and $\Gamma_{B \rightarrow A}$ in Eqs. (6.21) and (6.22) one finds that the inequalities characterizing $\text{conv}(\Gamma_{A \rightarrow B}, \Gamma_{B \rightarrow A})$ are precisely the same as the causal inequalities in Eqs. (6.25) and Eq. (6.26) except with bounds on the right-hand side of 0. One can easily verify that Eqs. (6.25) and Eq. (6.26) can be saturated by causal correlations (for some equal mixtures of correlations $P^{A \rightarrow B}$ and $P^{B \rightarrow A}$), thus providing such a counterexample.

Bipartite entropic causal inequalities from the DAG method

The following is the full list of (equivalence classes of) inequalities from the DAG method, up to their symmetries under the exchange of parties.

Ten (of the twenty) families of inequalities have bounds of 0 and can be violated by binary distributions:

$$\begin{aligned}
I(X : YA) + I(Y : XB) - H(AB) &\leq 0 \\
I(A : B) - I(A : B|X) - I(A : B|Y) - 2H(AB|XY) &\leq 0 \\
I(X : YA) + I(Y : AB) - H(B|X) - H(A) &\leq 0 \\
I(A : B|X) - I(Y : B) - 2H(B|XY) &\leq 0 \\
I(A : B|X) - I(A : B) - H(A|YB) - 2H(B|XY) &\leq 0 \\
I(XA : Y) + I(YB : X) - H(X|YA) - H(A) &\leq 0 \tag{A.2} \\
I(XA : Y) + I(YB : X) - H(B|YA) - H(A) &\leq 0 \\
I(XA : Y) + I(YB : X) - I(A : B) - H(B|YA) - H(YA|X) &\leq 0 \\
I(XA : Y) + I(YB : X) - I(A : B) - H(B|YA) - H(AB|X) &\leq 0 \\
I(XA : Y) + I(YB : X) - I(A : B) + I(X : A|Y) - H(XAB) &\leq 0.
\end{aligned}$$

Two more have non-zero bounds but, under the constraints that $H(A) \leq 1$, $H(B) \leq 1$, $H(X) \leq 1$, $H(Y) \leq 1$, turn out to be implied by the previous inequalities in Eq. (A.2):

$$\begin{aligned}
H(X|B) + H(Y|A) - I(A : B|XY) - 2H(X|YB) - 2H(Y|XA) &\leq 1 \\
I(XB : A) - 3I(X : A) - 3I(Y : B) - 4I(A : B|XY) - 2H(XB|YA) + 2H(B) &\leq 2. \tag{A.3}
\end{aligned}$$

Four correspond to “corrected” versions of the inequalities (A.1) characterizing the cone $\text{conv}(\Gamma_{A \prec B}, \Gamma_{B \prec A})$, and cannot be violated by binary distributions:

$$\begin{aligned}
I(X : YA) + I(Y : XB) - I(XY : AB) &\leq 1 \\
I(A : B) - I(A : B|X) - I(A : B|Y) &\leq 2 \\
I(X : A|B) - I(XB : A|Y) &\leq 1 \\
I(A : B|X) - I(A : B|XY) - I(Y : B) &\leq 1,
\end{aligned} \tag{A.4}$$

while a further four can also not be violated by binary distributions:

$$\begin{aligned}
I(A : B|X) - I(A : B) - I(X : A|YB) - H(B|XY) &\leq 1 \\
I(A : B|X) - I(A : B) - I(A : B|XY) - H(X|YB) &\leq 1 \\
I(A : B|X) - I(A : B) - I(A : B|XY) - H(A|YB) &\leq 1 \\
I(A : B|X) - I(A : B) - I(A : B|XY) + I(X : YA) + H(B|Y) - H(XB) &\leq 1.
\end{aligned} \tag{A.5}$$

We note that, instead of projecting $\tilde{\Gamma}^{\text{causal}}$ (as defined in Eq. (6.6)) onto the marginal scenario $\mathcal{M} = \{X, Y, A, B\}$ to obtain these entropic causal inequalities, one could start by projecting it onto the marginal scenario $\mathcal{M}' = \{X, Y, A, B, Q\}$ which would amount to eliminating all entropies $H(T, Q)$ for all nonempty subsets $T \subset \{X, Y, A, B\}$ from the description while keeping $H(Q)$. By doing so, one obtains the same inequalities given in Eqs. (A.2) to (A.5), except with the right-hand side multiplied by $H(Q)$. The inequalities in Eq. (A.2) thus have no dependence on $H(Q)$ (i.e., the extent to which correlations of different causal orders are mixed), while the remaining inequalities have a nontrivial dependence on it. By eliminating $H(Q)$ using the constraint $H(Q) \leq 1$ one then obtains the entropic causal inequalities above.²

The inequalities containing $H(Q)$ may be of interest if, for some reason, one puts a nontrivial bound on $H(Q)$ (e.g., if one knows that one fixed causal order is more probable than the other). In the extreme case, if we know that $H(Q) = 0$, then the inequalities we obtain (namely Eqs. (A.2)–(A.5), with all upper bounds replaced by 0) are valid for fixed-order causal correlations. All of the inequalities in Eqs. (A.4)–(A.5) except the second one in (A.4) can be violated by binary noncausal correlations for any $H(Q) < 1$, giving novel constraints in such situations; for the second inequality in Eq. (A.4) we were only able to find a violation for $H(Q) < \frac{1}{2}(1 + \frac{3}{2} \log_2 \frac{3}{2}) \approx 0.939$.

²A similar procedure can also be followed for the approach with counterfactual variables, in which case one obtains upper-bounds of $H(Q)$ and $2H(Q)$ in Eqs. (6.25) and (6.26) (or 0 for fixed-order correlations when $H(Q) = 0$), before eliminating $H(Q)$ and obtaining Eqs. (6.25)–(6.26) again.

Relations between different probability structures

In the application of the counterfactual method to causal correlations discussed in Sec. 6.2.2, as a result of the structure of the marginal scenario one can prove that different choices of probability structure \mathcal{S} give rise to the same marginal distributions. This is due to the fact that since all the marginals $M_j \in \mathcal{M}$ are disjoint, they are always consistent with the global product probability distribution

$$P(a_{00}, \dots, b_{11}) = \prod_{xy} P(a_{xy}, b_{xy}). \quad (\text{A.6})$$

Hence, whichever probability structure \mathcal{S} we choose (consistent with \mathcal{M}), the observed marginal probabilities can always be interpreted as arising from a global probability distribution. Similarly, the choice of extended probability structure $\tilde{\mathcal{S}}$ including the switch variable Q in Eq. (6.23) implies also the existence of a global probability distribution

$$P(a_{00}, \dots, b_{11}, q) = P(q) \prod_{xy} P(a_{xy}, b_{xy} | q). \quad (\text{A.7})$$

(Such a construction is also possible in some other types of scenarios; see Ref. [159] for more general results.) It thus follows that the probability structures $\tilde{\mathcal{S}}$ that we chose and $\tilde{\mathcal{S}}' = \{\tilde{\mathcal{S}}\}$ again give rise to the same marginal distributions on \mathcal{M} . A similar analysis can also be applied to the recursive method presented for the multipartite case in Sec. 6.3.2.

At the level of entropic inequalities, however, the fact that we are considering Shannon inequalities that provide only an outer approximation of the entropy cone means that one may *a priori* obtain different constraints depending on which of these equivalent probability structures one assumes. For the specific case of a marginal scenario with disjoint elements, i.e., $M_i \cap M_j = \emptyset$ for all $M_i, M_j \in \mathcal{M}$ a result by Matúš (see Remark 1 in Ref. [128]) implies that an equivalent description arises also for the Shannon cone by choosing $\mathcal{S} = \mathcal{M}$ or $\mathcal{S} = \{S\}$, with $S = \cup_{M_i \in \mathcal{M}} M_i$. More precisely, we have

$$\Pi_{\mathcal{M}}(\Gamma_S) = \Pi_{\mathcal{M}}(\Gamma_S \cap L_C(\{\mathbf{X}_{M_i} \perp \mathbf{X}_{M_j}\}_{M_i, M_j \in \mathcal{M}, i \neq j})) = \Pi_{\mathcal{M}}\left(\bigcap_{M_i \in \mathcal{M}} \Gamma_{M_i}\right) = \bigcap_{M_i \in \mathcal{M}} \Gamma_{M_i} \quad (\text{A.8})$$

where \mathbf{X}_{M_i} denotes the joint random variable associated with the subset $M_i \in \mathcal{M}$. The linear constraints in Eqs. (6.21) and (6.22), can then be imposed after the projection. Hence, the use of $\mathcal{S} = \mathcal{M}$ or $\mathcal{S}' = \{S\}$ is irrelevant, in this case, even at the level of Shannon cone description.

A similar analysis can be applied to compare $\tilde{\mathcal{S}}$ and $\tilde{\mathcal{S}}' = \{\tilde{\mathcal{S}}\}$, where $\tilde{S}_i \cap \tilde{S}_j = \{Q\}$ for all distinct $\tilde{S}_i, \tilde{S}_j \in \tilde{\mathcal{S}}$. Even though the marginal scenario is the same as above,

Eq. (6.24) involves extra constraints given by $\text{conv}_Q(\Gamma_{A \prec B}, \Gamma_{B \prec A})$ and $H(Q) \leq 1$, so the previous result does not directly apply. However, one can look at an intermediate projection $\Pi_{\tilde{\mathcal{S}}}(\Gamma^{\tilde{\mathcal{S}'}})$, and then impose the constraints involving Q . Hence, to prove the equivalence of the two probability structures at the level of the Shannon description, it would be sufficient to prove that $\Gamma^{\tilde{\mathcal{S}}} = \Pi_{\tilde{\mathcal{S}}}(\Gamma^{\tilde{\mathcal{S}'}})$. Using again the result by Matúš [128], we have

$$\Pi_{\tilde{\mathcal{S}}}(\Gamma^{\tilde{\mathcal{S}'}}) = \Pi_{\tilde{\mathcal{S}}}(\Gamma^{\tilde{\mathcal{S}'}} \cap L_C(\{X_{\tilde{s}_i \setminus \{Q\}} \perp X_{\tilde{s}_j \setminus \{Q\}} | Q\}_{\tilde{s}_i, \tilde{s}_j \in \tilde{\mathcal{S}}, i \neq j})). \quad (\text{A.9})$$

However, we do not know whether the following equality

$$\Pi_{\tilde{\mathcal{S}}}(\Gamma^{\tilde{\mathcal{S}'}} \cap L_C(\{X_{\tilde{s}_i \setminus \{Q\}} \perp X_{\tilde{s}_j \setminus \{Q\}} | Q\}_{\tilde{s}_i, \tilde{s}_j \in \tilde{\mathcal{S}}, i \neq j})) = \Pi_{\tilde{\mathcal{S}}}(\bigcap_{\tilde{s} \in \tilde{\mathcal{S}}} \Gamma_{\tilde{s}}) \quad (\text{A.10})$$

holds in general, so we were unable to prove the equivalence.

Nonetheless, we stress that any differences in tightness between the entropic inequalities arising here from the choice of a particular probability structure do not arise as a consequence of stricter physical assumptions (i.e., the existence of joint probability distributions), but rather as a consequence of different outer approximation of the true entropy cone via Shannon inequalities. We remark, however, that the choice of a minimal probability structure is computationally easier to handle due to the much lower number of variables; for example, compare the case $\mathcal{S} = \mathcal{M}$ in Eq. (6.20), where $\Gamma^{\mathcal{S}} \in \mathbb{R}^{12}$, with the corresponding case for $\mathcal{S}' = \{S\}$, where $\Gamma_S \in \mathbb{R}^{2^8} = \mathbb{R}^{256}$. For an extensive discussion of the role of such constraints in the computation of tighter approximations to the entropy cone we refer the reader to Ref. [27].

Bibliography

- [1] E. Schrödinger, “Die gegenwärtige Situation in der Quantenmechanik.,” *Die Naturwissenschaften*, vol. 23, pp. 807–812, 823–828, 844–849, 1935.
- [2] A. Einstein, B. Podolsky, and N. Rosen, “Can quantum-mechanical description of physical reality be considered complete?,” *Physical review*, vol. 47, no. 10, p. 777, 1935.
- [3] J. S. Bell, “On the Einstein–Podolsky–Rosen paradox,” *Physics*, vol. 1, p. 195, 1964.
- [4] A. Aspect, P. Grangier, and G. Roger, “Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A new violation of Bell’s inequalities,” *Physical review letters*, vol. 49, no. 2, p. 91, 1982.
- [5] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. Vermeulen, R. N. Schouten, C. Abellán, *et al.*, “Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres,” *Nature*, vol. 526, no. 7575, pp. 682–686, 2015.
- [6] M. Giustina, M. A. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-Å. Larsson, C. Abellán, *et al.*, “Significant-loophole-free test of Bell’s theorem with entangled photons,” *Physical review letters*, vol. 115, no. 25, p. 250401, 2015.
- [7] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, “Strong loophole-free test of local realism,” *Physical review letters*, vol. 115, no. 25, p. 250402, 2015.

- [8] C. J. Wood and R. W. Spekkens, "The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning," *New Journal of Physics*, vol. 17, no. 3, p. 033002, 2015.
- [9] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," *Physical review letters*, vol. 70, no. 13, p. 1895, 1993.
- [10] C. H. Bennett, P. Hayden, D. W. Leung, P. W. Shor, and A. Winter, "Remote preparation of quantum states," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 56–74, 2005.
- [11] A. K. Ekert, "Quantum cryptography based on Bell's theorem," *Physical review letters*, vol. 67, no. 6, p. 661, 1991.
- [12] 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," *Physical Review Letters*, vol. 83, no. 5, p. 1054, 1999.
- [13] A. Miyake and M. Wadati, "Geometric strategy for the optimal quantum search," *Physical Review A*, vol. 64, no. 4, p. 042317, 2001.
- [14] R. Jozsa and N. Linden, "On the role of entanglement in quantum-computational speed-up," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 459, 2036, pp. 2011–2032, The Royal Society, 2003.
- [15] R. Raussendorf, D. E. Browne, and H. J. Briegel, "Measurement-based quantum computation on cluster states," *Physical review A*, vol. 68, no. 2, p. 022312, 2003.
- [16] C. H. Bennett and S. J. Wiesner, "Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states," *Physical review letters*, vol. 69, no. 20, p. 2881, 1992.
- [17] M. Pawłowski and M. Żukowski, "Entanglement-assisted random access codes," *Physical Review A*, vol. 81, no. 4, p. 042326, 2010.
- [18] M. Hein, J. Eisert, and H. J. Briegel, "Multiparty entanglement in graph states," *Physical Review A*, vol. 69, no. 6, p. 062311, 2004.
- [19] M. Rossi, M. Huber, D. Bruß, and C. Macchiavello, "Quantum hypergraph states," *New Journal of Physics*, vol. 15, no. 11, p. 113022, 2013.

- [20] M. Gachechiladze, C. Budroni, and O. Gühne, "Extreme violation of local realism in quantum hypergraph states," *Physical review letters*, vol. 116, no. 7, p. 070401, 2016.
- [21] R. Chaves and T. Fritz, "Entropic approach to local realism and noncontextuality," *Physical Review A*, vol. 85, no. 3, p. 032113, 2012.
- [22] R. Chaves, C. Majenz, and D. Gross, "Information-theoretic implications of quantum causal structures," *Nature Communications*, vol. 6, p. 5766, 2015.
- [23] M. Weilenmann and R. Colbeck, "Inability of the entropy vector method to certify nonclassicality in linelike causal structures," *Physical Review A*, vol. 94, no. 4, p. 042112, 2016.
- [24] R. W. Yeung, "A framework for linear information inequalities," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1924–1934, 1997.
- [25] O. Oreshkov, F. Costa, and Č. Brukner, "Quantum correlations with no causal order," *Nature communications*, vol. 3, p. 1092, 2012.
- [26] N. Miklin, T. Moroder, and O. Gühne, "Multiparticle entanglement as an emergent phenomenon," *Physical Review A*, vol. 93, no. 2, p. 020104, 2016.
- [27] C. Budroni, N. Miklin, and R. Chaves, "Indistinguishability of causal relations from limited marginals," *Physical Review A*, vol. 94, no. 4, p. 042127, 2016.
- [28] F. E. Steinhoff, C. Ritz, N. Miklin, and O. Gühne, "Qudit hypergraph states," *Physical Review A*, vol. 95, no. 5, p. 052340, 2017.
- [29] M. Paraschiv, N. Miklin, T. Moroder, and O. Gühne, "Proving genuine multiparticle entanglement from separable nearest-neighbor marginals," *arXiv preprint arXiv:1705.02696*, 2017.
- [30] N. Miklin, A. A. Abbott, C. Branciard, R. Chaves, and C. Budroni, "The entropic approach to causal correlations," *arXiv preprint arXiv:1706.10270*, 2017.
- [31] T. Heinosaari and M. Ziman, *The mathematical language of quantum theory: from uncertainty to entanglement*. Cambridge University Press, 2011.
- [32] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information*. Cambridge University Press, 2002.
- [33] O. Gühne and G. Tóth, "Entanglement detection," *Physics Reports*, vol. 474, no. 1, pp. 1–75, 2009.

- [34] A. Peres, "Separability criterion for density matrices," *Physical Review Letters*, vol. 77, no. 8, p. 1413, 1996.
- [35] M. Horodecki, P. Horodecki, and R. Horodecki, "Separability of mixed states: necessary and sufficient conditions," *Physics Letters A*, vol. 223, no. 1, pp. 1–8, 1996.
- [36] B. M. Terhal, "Bell inequalities and the separability criterion," *Physics Letters A*, vol. 271, no. 5, pp. 319–326, 2000.
- [37] B. Jungnitsch, T. Moroder, and O. Gühne, "Taming multiparticle entanglement," *Physical review letters*, vol. 106, no. 19, p. 190502, 2011.
- [38] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement," *Reviews of Modern Physics*, vol. 81, no. 2, p. 865, 2009.
- [39] W. Dür, G. Vidal, and J. I. Cirac, "Three qubits can be entangled in two inequivalent ways," *Physical Review A*, vol. 62, no. 6, p. 062314, 2000.
- [40] B. Kraus, "Local unitary equivalence of multipartite pure states," *Physical review letters*, vol. 104, no. 2, p. 020504, 2010.
- [41] D. M. Greenberger, M. A. Horne, and A. Zeilinger, "Going beyond Bell's theorem," in *Bell's theorem, quantum theory and conceptions of the universe*, pp. 69–72, Springer, 1989.
- [42] F. Verstraete, J. Dehaene, B. De Moor, and H. Verschelde, "Four qubits can be entangled in nine different ways," *Physical Review A*, vol. 65, no. 5, p. 052112, 2002.
- [43] L. Lamata, J. León, D. Salgado, and E. Solano, "Inductive classification of multipartite entanglement under stochastic local operations and classical communication," *Physical Review A*, vol. 74, no. 5, p. 052336, 2006.
- [44] R. H. Dicke, "Coherence in spontaneous radiation processes," *Physical Review*, vol. 93, no. 1, p. 99, 1954.
- [45] H. J. Briegel and R. Raussendorf, "Persistent entanglement in arrays of interacting particles," *Physical Review Letters*, vol. 86, no. 5, p. 910, 2001.
- [46] M. Van den Nest, J. Dehaene, and B. De Moor, "Graphical description of the action of local Clifford transformations on graph states," *Physical Review A*, vol. 69, no. 2, p. 022316, 2004.

- [47] O. Gühne, M. Cuquet, F. E. Steinhoff, T. Moroder, M. Rossi, D. Bruß, B. Kraus, and C. Macchiavello, "Entanglement and nonclassical properties of hypergraph states," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 33, p. 335303, 2014.
- [48] M. Gachechiladze, N. Tsimakuridze, and O. Gühne, "Graphical description of unitary transformations on hypergraph states," *Journal of Physics A: Mathematical and Theoretical*, vol. 50, no. 19, p. 19LT01, 2017.
- [49] N. Tsimakuridze and O. Gühne, "Graph states and local unitary transformations beyond local clifford operations," *Journal of Physics A: Mathematical and Theoretical*, vol. 50, no. 19, p. 195302, 2017.
- [50] A. Vourdas, "Quantum systems with finite hilbert space," *Reports on Progress in Physics*, vol. 67, no. 3, p. 267, 2004.
- [51] D. Schlingemann, "Cluster states, algorithms and graphs," *arXiv preprint quant-ph/0305170*, 2003.
- [52] E. Hostens, J. Dehaene, and B. De Moor, "Stabilizer states and Clifford operations for systems of arbitrary dimensions and modular arithmetic," *Physical Review A*, vol. 71, no. 4, p. 042315, 2005.
- [53] M. Bahramgiri and S. Beigi, "Graph states under the action of local Clifford group in non-binary case," *arXiv preprint quant-ph/0610267*, 2006.
- [54] S. Y. Looi, L. Yu, V. Gheorghiu, and R. B. Griffiths, "Quantum-error-correcting codes using qudit graph states," *Physical Review A*, vol. 78, no. 4, p. 042303, 2008.
- [55] A. Keet, B. Fortescue, D. Markham, and B. C. Sanders, "Quantum secret sharing with qudit graph states," *Physical Review A*, vol. 82, no. 6, p. 062315, 2010.
- [56] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Physical review letters*, vol. 23, no. 15, p. 880, 1969.
- [57] I. Pitowsky, *Quantum probability, quantum logic*, vol. 321. Springer, 1989.
- [58] B. S. Cirel'son, "Quantum generalizations of Bell's inequality," *Letters in Mathematical Physics*, vol. 4, no. 2, pp. 93–100, 1980.
- [59] R. F. Werner, "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model," *Physical Review A*, vol. 40, no. 8, p. 4277, 1989.

- [60] A. Fine, "Hidden variables, joint probability, and the Bell inequalities," *Physical Review Letters*, vol. 48, no. 5, p. 291, 1982.
- [61] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2009.
- [62] H. P. Williams, "Fourier's method of linear programming and its dual," *The American mathematical monthly*, vol. 93, no. 9, pp. 681–695, 1986.
- [63] C. Branciard, N. Gisin, and S. Pironio, "Characterizing the nonlocal correlations created via entanglement swapping," *Physical review letters*, vol. 104, no. 17, p. 170401, 2010.
- [64] T. Fritz, "Beyond Bell's theorem: correlation scenarios," *New Journal of Physics*, vol. 14, no. 10, p. 103001, 2012.
- [65] C. Kang and J. Tian, "Polynomial constraints in causal Bayesian networks," in *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, pp. 200–208, 2007.
- [66] R. Chaves, "Polynomial Bell inequalities," *Physical review letters*, vol. 116, no. 1, p. 010402, 2016.
- [67] D. Geiger and C. Meek, "Quantifier elimination for statistical problems," in *Proceedings of the 15th conference on Uncertainty in Artificial Intelligence*, pp. 226–235, 1999.
- [68] J. Pearl, *Causality*. Cambridge University Press, Cambridge, 2009.
- [69] H. Reichenbach, *The direction of time*. Berkeley, University of Los Angeles Press, 1956.
- [70] R. W. Yeung, *Information theory and network coding*. Information technology–transmission, processing, and storage, Springer, 2008.
- [71] S. L. Lauritzen, *Graphical models*. Clarendon Press, 1996.
- [72] E. G. Cavalcanti and R. Lal, "On modifications of Reichenbach's principle of common cause in light of Bell's theorem," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 42, p. 424018, 2014.
- [73] C. J. Wood and R. W. Spekkens, "The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning," *New Journal of Physics*, vol. 17, no. 3, p. 033002, 2015.

- [74] M. S. Leifer and R. W. Spekkens, "Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference," *Physical Review A*, vol. 88, no. 5, p. 052130, 2013.
- [75] J. Henson, R. Lal, and M. F. Pusey, "Theory-independent limits on correlations from generalized Bayesian networks," *New Journal of Physics*, vol. 16, no. 11, p. 113043, 2014.
- [76] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, and K. J. Resch, "A quantum advantage for inferring causal structure," *Nature Physics*, vol. 11, no. 5, pp. 414–420, 2015.
- [77] F. Costa and S. Shrapnel, "Quantum causal modelling," *New Journal of Physics*, vol. 18, no. 6, p. 063032, 2016.
- [78] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, "Quantum common causes and quantum causal models," *arXiv preprint arXiv:1609.09487*, 2016.
- [79] C. Portmann, C. Matt, U. Maurer, R. Renner, and B. Tackmann, "Causal boxes: Quantum information-processing systems closed under composition," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 3277–3305, 2017.
- [80] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron, "Quantum computations without definite causal structure," *Physical Review A*, vol. 88, no. 2, p. 022318, 2013.
- [81] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, and Č. Brukner, "Witnessing causal nonseparability," *New Journal of Physics*, vol. 17, no. 10, p. 102001, 2015.
- [82] C. Branciard, "Witnesses of causal nonseparability: an introduction and a few case studies," *Scientific reports*, vol. 6, 2016.
- [83] A. Baumeler and S. Wolf, "Perfect signaling among three parties violating predefined causal order," in *Information Theory (ISIT), 2014 IEEE International Symposium on*, pp. 526–530, IEEE, 2014.
- [84] Ä. Baumeler, A. Feix, and S. Wolf, "Maximal incompatibility of locally classical behavior and global causal order in multiparty scenarios," *Physical Review A*, vol. 90, no. 4, p. 042106, 2014.
- [85] O. Oreshkov and C. Giarmatzi, "Causal and causally separable processes," *New Journal of Physics*, vol. 18, no. 9, p. 093020, 2016.

- [86] C. Branciard, M. Araújo, A. Feix, F. Costa, and Č. Brukner, "The simplest causal inequalities and their violation," *New Journal of Physics*, vol. 18, no. 1, p. 013008, 2015.
- [87] A. A. Abbott, C. Giarmatzi, F. Costa, and C. Branciard, "Multipartite causal correlations: Polytopes and inequalities," *Physical Review A*, vol. 94, no. 3, p. 032131, 2016.
- [88] S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom," *Foundations of Physics*, vol. 24, no. 3, pp. 379–385, 1994.
- [89] G. V. Steeg and A. Galstyan, "A sequence of relaxations constraining hidden variable models," in *Proceedings of the 27th conference on Uncertainty in Artificial Intelligence*, 2011.
- [90] B. Bonet, "Instrumentality tests revisited," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 48–55, Morgan Kaufmann Publishers Inc., 2001.
- [91] A. J. Coleman, "Structure of fermion density matrices," *Reviews of modern Physics*, vol. 35, no. 3, p. 668, 1963.
- [92] T. Fritz and R. Chaves, "Entropic inequalities and marginal problems," *IEEE transactions on information theory*, vol. 59, no. 2, pp. 803–817, 2013.
- [93] R. Chaves, "Entropic inequalities as a necessary and sufficient condition to noncontextuality and locality," *Physical Review A*, vol. 87, no. 2, p. 022102, 2013.
- [94] R. Chaves, L. Luft, and D. Gross, "Causal structures from entropic information: geometry and novel scenarios," *New Journal of Physics*, vol. 16, no. 4, p. 043001, 2014.
- [95] R. Chaves, L. Luft, T. Maciel, D. Gross, D. Janzing, and B. Schölkopf, "Inferring latent structures via information inequalities," *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, pp. 112–121, 2014.
- [96] R. Chaves and C. Budroni, "Entropic nonsignaling correlations," *Physical review letters*, vol. 116, no. 24, p. 240501, 2016.
- [97] C. Budroni and A. Cabello, "Bell inequalities from variable-elimination methods," *Journal of Physics A: Mathematical and Theoretical*, vol. 45, no. 38, p. 385304, 2012.
- [98] S. L. Braunstein and C. M. Caves, "Information-theoretic Bell inequalities," *Physical review letters*, vol. 61, no. 6, p. 662, 1988.

- [99] B. Steudel and N. Ay, "Information-theoretic inference of common ancestors," *Entropy*, vol. 17, no. 4, pp. 2304–2327, 2015.
- [100] M. Navascués, S. Pironio, and A. Acín, "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations," *New Journal of Physics*, vol. 10, no. 7, p. 073013, 2008.
- [101] M. F. Pusey, "Negativity and steering: A stronger Peres conjecture," *Physical Review A*, vol. 88, no. 3, p. 032313, 2013.
- [102] J. Lofberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, pp. 284–289, IEEE, 2004.
- [103] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3—a MATLAB software package for semidefinite programming, version 1.3," *Optimization methods and software*, vol. 11, no. 1-4, pp. 545–581, 1999.
- [104] G. Sagnol, "Picos, a python interface to conic optimization solvers," tech. rep., Technical Report 12-48, ZIB, 2012. <http://picos.zib.de>, 2012.
- [105] G. Tóth, "Entanglement witnesses in spin models," *Physical Review A*, vol. 71, no. 1, p. 010301, 2005.
- [106] G. Tóth, C. Knapp, O. Gühne, and H. J. Briegel, "Spin squeezing and entanglement," *Physical Review A*, vol. 79, no. 4, p. 042334, 2009.
- [107] L. E. Würflinger, J.-D. Bancal, A. Acín, N. Gisin, and T. Vértesi, "Nonlocal multipartite correlations from local marginal probabilities," *Physical Review A*, vol. 86, no. 3, p. 032117, 2012.
- [108] J. Tura, R. Augusiak, A. B. Sainz, T. Vértesi, M. Lewenstein, and A. Acín, "Detecting nonlocality in many-body quantum states," *Science*, vol. 344, no. 6189, pp. 1256–1258, 2014.
- [109] L. Chen, O. Gittsovich, K. Modi, and M. Piani, "Role of correlations in the two-body-marginal problem," *Physical Review A*, vol. 90, no. 4, p. 042314, 2014.
- [110] 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*, vol. 6, no. 12, pp. 943–946, 2010.
- [111] H. Kampermann, O. Gühne, C. Wilmott, and D. Bruß, "Algorithm for characterizing stochastic local operations and classical communication classes of multiparticle entanglement," *Physical Review A*, vol. 86, no. 3, p. 032307, 2012.

- [112] M. Paraschiv, "Aspects of entanglement in multipartite and decaying systems," 2016.
- [113] M. Popp, F. Verstraete, M. A. Martín-Delgado, and J. I. Cirac, "Localizable entanglement," *Physical Review A*, vol. 71, no. 4, p. 042306, 2005.
- [114] D. Schlingemann, "Stabilizer codes can be realized as graph codes," *arXiv preprint quant-ph/0111080*, 2001.
- [115] N. D. Mermin, "Hidden variables and the two theorems of John Bell," *Reviews of Modern Physics*, vol. 65, no. 3, p. 803, 1993.
- [116] B. Jungnitsch, T. Moroder, and O. Gühne, "Entanglement witnesses for graph states: General theory and examples," *Physical Review A*, vol. 84, no. 3, p. 032310, 2011.
- [117] A. Kitaev, "Anyons in an exactly solved model and beyond," *Annals of Physics*, vol. 321, no. 1, pp. 2–111, 2006.
- [118] B. Yoshida, "Topological phases with generalized global symmetries," *Physical Review B*, vol. 93, no. 15, p. 155131, 2016.
- [119] T. Carle, B. Kraus, W. Dür, and J. I. de Vicente, "Purification to locally maximally entangleable states," *Physical Review A*, vol. 87, no. 1, p. 012328, 2013.
- [120] C. E. Mora, H. J. Briegel, and B. Kraus, "Quantum Kolmogorov complexity and its applications," *International Journal of Quantum Information*, vol. 5, no. 05, pp. 729–750, 2007.
- [121] W. Tang, S. Yu, and C. Oh, "Greenberger-Horne-Zeilinger paradoxes from qudit graph states," *Physical review letters*, vol. 110, no. 10, p. 100403, 2013.
- [122] M. A. Nielsen, M. J. Bremner, J. L. Dodd, A. M. Childs, and C. M. Dawson, "Universal simulation of Hamiltonian dynamics for quantum systems with finite-dimensional state spaces," *Physical Review A*, vol. 66, no. 2, p. 022317, 2002.
- [123] M. Horodecki, P. Horodecki, and R. Horodecki, "Separability of n-particle mixed states: necessary and sufficient conditions in terms of linear maps," *Physics Letters A*, vol. 283, no. 1, pp. 1–7, 2001.
- [124] T.-C. Wei and P. M. Goldbart, "Geometric measure of entanglement and applications to bipartite and multipartite quantum states," *Physical Review A*, vol. 68, no. 4, p. 042307, 2003.

- [125] C. Beeri, R. Fagin, D. Maier, and M. Yannakakis, "On the desirability of acyclic database schemes," *Journal of the ACM (JACM)*, vol. 30, no. 3, pp. 479–513, 1983.
- [126] N. N. Vorob'ev, "Coalition games," *Theory Probab. Appl.*, vol. 12, pp. 251–266, 1967.
- [127] N. Vorob'ev, "Markov measures and markov extensions," *Theory of Probability & Its Applications*, vol. 8, no. 4, pp. 420–429, 1963.
- [128] F. Matúš, "Adhesivity of polymatroids," *Discrete Mathematics*, vol. 307, no. 21, pp. 2464–2477, 2007.
- [129] C. Budroni and G. Morchio, "The extension problem for partial Boolean structures in quantum mechanics," *Journal of Mathematical Physics*, vol. 51, no. 12, p. 122205, 2010.
- [130] C. Budroni and G. Morchio, "Bell inequalities as constraints on unmeasurable correlations," *Foundations of Physics*, vol. 42, no. 4, pp. 544–554, 2012.
- [131] P. Kurzyński, R. Ramanathan, and D. Kaszlikowski, "Entropic test of quantum contextuality," *Physical review letters*, vol. 109, no. 2, p. 020404, 2012.
- [132] Z. Zhang and R. W. Yeung, "On characterization of entropy function via information inequalities," *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1440–1452, 1998.
- [133] F. Matúš, "Infinitely many information inequalities," in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pp. 41–44, IEEE, 2007.
- [134] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin, "A new class of non-Shannon-type inequalities for entropies," *Communications in Information and Systems*, vol. 2, no. 2, pp. 147–166, 2002.
- [135] Z. Zhang, "On a new non-shannon type information inequality," *Communications in Information and Systems*, vol. 3, no. 1, pp. 47–60, 2003.
- [136] R. Dougherty, C. Freiling, and K. Zeger, "Six new non-Shannon information inequalities," in *Information Theory, 2006 IEEE International Symposium on*, pp. 233–236, IEEE, 2006.
- [137] H. G. Kellerer, "Maßtheoretische marginalprobleme," *Mathematische Annalen*, vol. 153, no. 3, pp. 168–198, 1964.
- [138] H. G. Kellerer, "Verteilungsfunktionen mit gegebenen marginalverteilungen," *Probability Theory and Related Fields*, vol. 3, no. 3, pp. 247–270, 1964.

- [139] F. M. Malvestuto, "Existence of extensions and product extensions for discrete probability distributions," *Discrete Mathematics*, vol. 69, no. 1, pp. 61–77, 1988.
- [140] P. Heggernes, "Minimal triangulations of graphs: A survey," *Discrete Mathematics*, vol. 306, no. 3, pp. 297–317, 2006.
- [141] M. Araújo, M. T. Quintino, C. Budroni, M. T. Cunha, and A. Cabello, "All noncontextuality inequalities for the n -cycle scenario," *Physical Review A*, vol. 88, no. 2, p. 022118, 2013.
- [142] I. Pitowsky, "Correlation polytopes: their geometry and complexity," *Mathematical Programming*, vol. 50, no. 1, pp. 395–414, 1991.
- [143] F. Matus, "Two constructions on limits of entropy functions," *IEEE Transactions on Information Theory*, vol. 53, no. 1, pp. 320–330, 2007.
- [144] J. Pienaar and Č. Brukner, "A graph-separation theorem for quantum causal models," *New Journal of Physics*, vol. 17, no. 7, p. 073020, 2015.
- [145] T. Fritz, "Beyond bell's theorem ii: Scenarios with arbitrary causal structure," *Communications in Mathematical Physics*, vol. 2, no. 341, pp. 391–434, 2016.
- [146] J. Pienaar, "Which causal structures might support a quantum–classical gap?," *New Journal of Physics*, vol. 19, no. 4, p. 043021, 2017.
- [147] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, "Information causality as a physical principle," *Nature*, vol. 461, no. 7267, pp. 1101–1104, 2009.
- [148] C. M. Lee and R. W. Spekkens, "Causal inference via algebraic geometry: necessary and sufficient conditions for the feasibility of discrete causal models," *arXiv preprint arXiv:1506.03880*, 2015.
- [149] D. Rosset, C. Branciard, T. J. Barnea, G. Pütz, N. Brunner, and N. Gisin, "Nonlinear Bell inequalities tailored for quantum networks," *Physical review letters*, vol. 116, no. 1, p. 010403, 2016.
- [150] S. Raeisi, P. Kurzyński, and D. Kaszlikowski, "Entropic tests of multipartite nonlocality and state-independent contextuality," *Physical review letters*, vol. 114, no. 20, p. 200401, 2015.
- [151] R. Chaves, J. B. Brask, and N. Brunner, "Device-independent tests of entropy," *Physical review letters*, vol. 115, no. 11, p. 110501, 2015.

- [152] F. Zhu, W. Zhang, S. Chen, L. You, Z. Wang, and Y. Huang, "Experimental device-independent tests of classical and quantum entropy," *Physical Review A*, vol. 94, no. 6, p. 062340, 2016.
- [153] G. Svetlichny, "Distinguishing three-body from two-body nonseparability by a Bell-type inequality," *Physical Review D*, vol. 35, no. 10, p. 3066, 1987.
- [154] R. Chaves, D. Cavalcanti, and L. Aolita, "Causal hierarchy of multipartite bell nonlocality," *arXiv preprint arXiv:1607.07666*, 2016.
- [155] S. Lörwald and G. Reinelt, "PANDA: a software for polyhedral transformations," *EURO Journal on Computational Optimization*, vol. 3, no. 4, p. 297, 2015.
- [156] L. Hardy, "Probability theories with dynamic causal structure: a new framework for quantum gravity," *arXiv preprint gr-qc/0509120*, 2005.
- [157] Ä. Baumeler and S. Wolf, "The space of logically consistent classical processes without causal order," *New Journal of Physics*, vol. 18, no. 1, p. 013036, 2016.
- [158] A. Feix, M. Araújo, and Č. Brukner, "Causally nonseparable processes admitting a causal model," *New Journal of Physics*, vol. 18, no. 8, p. 083040, 2016.
- [159] N. N. Vorob'ev, "Consistent families of measures and their extensions," *Theory of Probability & Its Applications*, vol. 7, no. 2, pp. 147–163, 1962.
- [160] I. Ibnouhsein and A. Grinbaum, "Information-theoretic constraints on correlations with indefinite causal order," *Physical Review A*, vol. 92, no. 4, p. 042124, 2015.
- [161] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués, "Operational framework for nonlocality," *Physical review letters*, vol. 109, no. 7, p. 070401, 2012.
- [162] J.-D. Bancal, J. Barrett, N. Gisin, and S. Pironio, "Definitions of multipartite nonlocality," *Physical Review A*, vol. 88, no. 1, p. 014102, 2013.
- [163] A. E. Rastegin, "Formulation of Leggett—Garg inequalities in terms of q -entropies," *Communications in Theoretical Physics*, vol. 62, no. 3, p. 320, 2014.
- [164] M. Wajs, P. Kurzyński, and D. Kaszlikowski, "Information-theoretic Bell inequalities based on Tsallis entropy," *Physical Review A*, vol. 91, no. 1, p. 012114, 2015.
- [165] M. Weilenmann and R. Colbeck, "Non-shannon inequalities in the entropy vector approach to causal structures," *arXiv preprint arXiv:1605.02078*, 2016.

List of publications

- (i) N. Miklin, T. Moroder, and O. Gühne, "Multiparticle entanglement as an emergent phenomenon", *Physical Review A*, vol. 93, no. 2, p. 020104, 2016.
- (ii) F. E. Steinhoff, C. Ritz, N. Miklin, and O. Gühne, "Qudit hypergraph states", *Physical Review A*, vol. 95, no. 5, p. 052340, 2017.
- (iii) C. Budroni, N. Miklin, and R. Chaves, "Indistinguishability of causal relations from limited marginals", *Physical Review A*, vol. 94, no. 4, p. 042127, 2016.
- (iv) M. Paraschiv, N. Miklin, T. Moroder, and O. Gühne, "Proving genuine multiparticle entanglement from separable nearest-neighbor marginals", *arXiv preprint*, arXiv:1705.02696, 2017.
- (v) N. Miklin, A. A. Abbott, C. Branciard, R. Chaves, and C. Budroni, "The entropic approach to causal correlations", *arXiv preprint*, arXiv:1706.10270, 2017.

Acknowledgments

This thesis is a result of work in collaborations with many bright researchers. Many more people helped me with scientific discussions and psychological support. People, who I would like to thank for working with me, are: Prof. Dr. Otfried Gühne, Prof. Dr. Rafael Chaves, Dr. Costantino Budroni, Dr. Frank E.S. Steinhoff, Dr. Alastair A. Abbott, Dr. Cyril Branciard, Dr. Tobias Moroder, Christina Ritz, and Dr. Marius Paraschiv. For the most fruitful scientific discussions I would like to thank Dr. Ali Asadian and Mariami Gachechiladze.

I am especially grateful to:

- Prof. Dr. Otfried Gühne, my supervisor, for giving me the opportunity to do my doctoral studies in his research group, for enabling me to study a broad range of exciting problems, for giving me many chances to attend conferences and meet a lot of bright people, and, of course, for his supervision.
- Dr. Costantino Budroni for his constant support and supervision, his willingness to introduce me to new scientific problems and last but not least, for being a good friend.
- Mariami Gachechiladze for being my constant inspiration and for making my life in Siegen happier than anywhere else.
- my family, who supported me on every stage of my life.

Additionally, I would like to thank Ms. Daniela Lehmann for helping me with legal advices and saving a lot of my time on paperwork. I acknowledge the financial support of the Deutscher Akademischer Austauschdienst (DAAD) who provided a grant for my doctoral studies.