TURUN YLIOPISTO
UNIVERSITY OF TURKU

QUANTUM COMMUNICATION TASKS

Oskari Kerppo

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AI OSA – TOM. 682 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2022

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

QUANTUM COMMUNICATION TASKS

Oskari Kerppo

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AI OSA – TOM. 682 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2022

University of Turku

Faculty of Science
Department of Physics and Astronomy
Theoretical Physics
Doctoral Programme in Exact Sciences (EXACTUS)

Supervised by

Teiko Heinosaari
Department of Physics and Astronomy
University of Turku
Finland

Iiro Vilja
Department of Physics and Astronomy
University of Turku
Finland

Reviewed by

Milán Mosonyi
Institute of Mathematics
Budapest University of Technology and Economics
Hungary

Matthew Leifer
Schmid College of Science and Technology
Chapman University
United States

Opponent

Sarah Croke
School of Physics and Astronomy
University of Glasgow
United Kingdom

ISBN 978-951-29-9126-6 (PRINT)
ISBN 978-951-29-9127-3 (PDF)
ISSN 0082-7002 (PRINT)
ISSN 2343-3175 (ONLINE)
Painosalama Oy, Turku, Finland 2022

Acknowledgements

Writing a doctoral thesis is *not easy* (what an astute observation). In fact, it's surely the greatest achievement I have accomplished during my life so far. These past five years have been such a wild rollercoaster that I have wondered many times if I would ever finish my PhD. Yet here we are. All's well that ends well. My PhD journey wouldn't have been the same without all of the exceptionally great people who I had the privilege of meeting and working with. I am now going to attempt to thank all of you (not in the order of greatest thanks, if such an ordering is even possible).

First of all I want to thank Milán Mosonyi and Matthew Leifer for pre-examining my thesis. Both of you gave wonderful feedback to the work I spent so much time doing. I am happy that the work finally payed off. Secondly, I want to thank the Turku University Foundation for supporting my doctoral research.

Next up, my supervisors Teiko and Iiro. You are both incredibly wise, knowledgeable and unbelievably supportive people. Teiko, you have been my teacher and mentor for many years now, and I have learned so much from you. For that I am grateful. Thank you both for supervising my thesis.

Studying theoretical physics wouldn't have been the same without all of the incredible staff and fellow students at our department. Sabrina, you created a fantastic group and brought incredible people to our university. It was truly a great place to learn and grow as a researcher and person. I want to thank the entire theoretical physics group for inspiring me to be a better physicist. Especially you Tom. You helped me a great deal with proofreading my thesis and I could always rely on you to help with all sorts of problems a PhD student might have. Thank you for that. As a bonus I can say your witty British sense of humor never fails to brighten up my day, and that makes you special amongst all the people I met at Turku.

For the majority of my time as a PhD student my workstation was in the *quiet room*. Thank you Laura, Sina and Johannes for not making too much noise. I also want to collectively thank all of the PhD students I had the pleasure of spending time with, I will never forget my time with you. However, I want to especially thank Laura, Leevi, Juho and Henri for making the coffee breaks memorable and still being my friends to this day (a certain British former post-doc also belongs to this group, don't think you're off the hook just yet Tom!). Moreover, I want to thank both Leevi and Juho for coauthoring papers with me and Juho for scrutinizing my grant applications to the point that I actually got one. My PhD might have looked

very different without all of you, so thank you all from the bottom of my heart.

The gym, or weightlifting more recently, has been an important hobby for me throughout my time as a university and doctoral student. I find it forms a good balance with studying. I am glad I had the gym buddies Henri and Leevi. I watched you both grow from very small to very big and strong in the span of just a couple of years, and you motivated me to train harder.

As we begin to wrap up I think the time has come for me to thank all of the rest of my friends whom I feel incredibly lucky to have met. Jaakko and Miika, thank you both for being true friends, I know I can always count on both of you. And thank you φ. I don't think many words are needed here. Looking back at the very beginning of our time as university students, I find it overwhelming to even start to consider how to thank all of you. Thank you for the constant and ongoing opportunity to share thoughts and laughs.

Now we are truly nearing the end, so I now want to thank my family for all their love and support throughout these long years. Thank you Mom, Dad, Karo and Naffe for always being there for me. And finally thank you Mari for sharing all of these years with me. Looking forward to sharing many more with you!

December 13, 2022
Oskari Kerppo

Table of contents

UNIVERSITY OF TURKU
Faculty of Science
Department of Physics and Astronomy
Theoretical Physics
KERPPO, OSKARI: Quantum communication tasks
Doctoral dissertation, 152 pp.
Doctoral Programme in Exact Sciences (EXACTUS)
December 2022

ABSTRACT

Quantum theory is one of the most important theories in modern physics, yet the physical principles underlying the theory are anything but clear. Nonclassical features, such as entanglement, are often attributed as the key phenomena that make quantum theory special. However, the difference between classical and quantum systems can already be detected by observing the behavior of single systems in various communication setups.

This thesis is based on the original publications **I**–**IV**. A key concept throughout is that of a communication task, by which we mean a description of conditional probabilities in a prepare-and-measure scenario. These conditional probabilities are conveniently collected into row-stochastic matrices, which we call communication matrices.

The concept of a communication task was introduced in Publication **II** where we also studied a preorder on the set of communication matrices. We called this preorder the ultraweak matrix majorization and refined the concept in Publication **III**. A key motivation for introducing this preorder was that the set of communication matrices is closed with respect to the ultraweak matrix majorization. Additionally, ultraweak matrix majorization can be used to give a physical characterization of which communication tasks are harder to implement than others.

We also studied monotone functions of the ultraweak preorder. By studying the different monotones it becomes possible to define different notions of dimension for operational theories. These dimensions each characterize the properties of given operational theories and we are able to capture some key differences between classical and quantum state spaces.

While the preorder of ultraweak matrix majorization is a major part of this thesis, some concrete communication tasks are also studied. One of the main studied communication tasks is antidistinguishability, which plays an important role in the study of the foundations of quantum mechanics. We were able to provide a new algebraic condition for an arbitrary set of quantum states to be antidistinguishable in Publication **I**. We also apply the theory of ultraweak matrix majorization to antidistinguishability in the third chapter of this thesis, where we show that the set of all communication matrices is not convex for classical or quantum state spaces in any dimension.

The other communication tasks studied in this thesis are communication of partial ignorance, studied in Publication **II**, and partial-ignorance communication tasks which was the topic of Publication **IV**. Both of these communication tasks can be seen as communication tasks between two parties, where one party is trying to communicate which choices the other party should avoid. A key observation for these tasks is that they lie between distinguishability and antidistinguishability. Some novel analysis is presented for both of these tasks in the final chapter of this thesis. The quantum implementation for one of the partial-ignorance communication tasks can be shown to break the principle of noncontextuality, thus proving that quantum mechanics holds a contextual advantage in the given task when compared to classical operational theories.

TURUN YLIOPISTO
Matemaattis-luonnontieteellinen tiedekunta
Fysiikan ja tähtitieteen laitos
Teoreettinen fysiikka
KERPPO, OSKARI: Quantum communication tasks
Väitöskirja, 152 s.
Eksaktien tieteiden tohtoriohjelma (EXACTUS)
Joulukuu 2022

TIIVISTELMÄ

Kvanttiteoria on yksi tärkeimmistä modernin fysiikan teorioista. Kuitenkaan fysikaaliset periaatteet, joihin kvanttiteoria perustuu, eivät ole tänäpäivänä täysin selvillä. Ei-klassisia piirteitä, kuten kvanttikietoutumista, pidetään usein ilmiöinä, jotka tekevät kvanttiteoriasta erityisen. Ero kvanttisysteemien ja klassisten systeemien välillä voidaan kuitenkin nähdä jo tarkastelemalla yksittäisten systeemien käyttäytymistä erilaisissa kommunikointijärjestelyissä.

Tämä väitöskirja perustuu alkuperäisjulkaisuihin **I**–**IV**. Avainkonsepti tässä väitöskirjassa on kommunikointitehtävän käsite, jolla tarkoitetaan ehdollisten todennäköisyyksien kuvausta preparoi-ja-mittaa järjestelyssä. Nämä ehdolliset todennäköisyydet voidaan kätevästi kerätä rivistokastisiin matriiseihin, joita kutsutaan kommunikointimatriiseiksi.

Kommunikointitehtävän käsite esitettiin julkaisussa **II**, jossa myös tutkin kommunikointimatriisien esijärjestystä. Kutsuin tätä esijärjestystä ultraheikoksi matriisimajoroinniksi, jota käsitteenä jalostettiin julkaisussa **III**. Tärkeä motivaatio tämän esijärjestyksen tutkimiselle on se, että kommunikointimatriisien joukko on suljettu ultraheikon matriisimajoroinnin suhteen. Tämän lisäksi ultraheikon matriisimajoroinnin avulla voidaan fysikaalisesti karakterisoida ne tilanteet, joissa yksi kommunikointitehtävä on toista helpompi toteuttaa.

Tutkimuksen kohteena ovat myös ultraheikon esijärjestyksen monotoniset funktiot. Näitä funktioita tutkimalla on mahdollista esittää erilaisia dimension käsitteitä eri operationaalisille teorioille. Kukin näistä dimensioista karakterisoi operationaalisen teorian ominaisuuksia. Joitakin oleellisia eroja klassisten systeemien ja kvanttisysteemien välillä pystytäänkin havainnollistamaan näiden dimensioiden avulla.

Ultraheikon matriisimajoroinnin esijärjestyksen lisäksi tutkin tässä väitöskirjassa myös konkreettisia kommunikointitehtäviä. Ensimmäinen näistä tehtävistä on antieroteltavuus, jolla on tärkeitä sovelluksia kvanttimekaniikan perusteiden tutkimuksessa. Pystyin näyttämään uuden algebrallisen ehdon mielivaltaisen kvanttitilajoukon antieroteltavuudelle julkaisussa **I**. Sovelsin myös ultraheikkoa matriisimajorointia antieroteltavuuteen tämän väitöskirjan kolmannessa kappaleessa, jossa näytin ettei klassiset- tai kvanttitila-avaruudet ole konvekseja missään dimensiossa.

Muut tässä väitöskirjassa tutkitut kommunikointitehtävät olivat osittaisen tietämättömyyden kommunikointitehtävät, joita oli kahta oleellisesti erilaista tyyppiä. Ensimmäistä tutkin julkaisussa **II**, kun taas toinen oli julkaisun **IV** aiheena. Kummatkin näistä kommunikointitehtävistä voidaan nähdä kahden osapuolen kommu-

nikointitehtävinä, joissa toinen osapuoli pyrkii välittämään toiselle osapuolelle tietoa siitä, mitä vaihtoehtoja hänen tulisi välttää. Tärkeä havainto on, että nämä kommunikointitehtävät ovat antieroteltavuuden ja eroteltavuuden välissä. Uutta anlyysiä näistä kommunikointitehtävistä esitetään tämän väitöskirjan viimeisessä kappaleessa. Yhden kommunikointitehtävän kvanttitoteutuksen voidaan nähdä rikkovan ei-kontekstuaalisuuden periaatetta. Kvanttimekaanisilla systeemeillä voidaan näin todeta olevan kontekstuaalinen etu klassisiin systeemeihin verrattuna kyseisessä kommunikointitehtävässä.

List of original publications

This dissertation is based on the following original publications [1; 2; 3; 4], which are referred to in the text by their Roman numerals:

I T. Heinosaari and *O. Kerppo*. Antidistinguishability of pure quantum states. J. Phys. A: Math. Theor., 51:365303, 2018.

II T. Heinosaari and *O. Kerppo*. Communication of partial ignorance with qubits. J. Phys. A: Math. Theor., 52:395301, 2019.

III T. Heinosaari, *O. Kerppo* and L. Leppäjärvi. Communication tasks in operational theories. J. Phys. A: Math. Theor., 53:435302, 2020.

IV *O. Kerppo*, Partial-ignorance communication tasks in quantum theory. Phys. Rev. A., 105:062607, 2022.

Other published material

This is a list of the publications produced which have not been chosen as a part of this doctoral thesis:

- T. Bullock, F. Cosco, M. Haddara, S. H. Raja, *O. Kerppo*, L. Leppäjärvi, O. Siltanen, N. W. Talarico, A. De Pasquale, V. Giovannetti and S. Maniscalco. Entanglement protection via periodic environment resetting in continuous-time quantum-dynamical processes. Phys. Rev. A, 98:042301, 2018.
- J. Lankinen, *O. Kerppo* and I. Vilja. Reheating via gravitational particle production in the kination epoch. Phys. Rev. D, 101:063529, 2020.
- J. Lankinen, *O. Kerppo* and I. Vilja. Reheating in the kination epoch via multichannel decay of gravitationally created massive scalars. Phys. Rev. D, 103:083522, 2021.
- G. García-Pérez, *O. Kerppo*, M. A. C. Rossi and S. Maniscalco. Experimentally accessible non-separability criteria for multipartite entanglement structure detection. arXiv:2110.04177 [quant-ph], 2021

Introduction

Communication is fundamental to human society, culture and technology. In a sense, almost anything could be called communication: from the transmission of pressure waves through air that our brains interpret as letters, words, sentences and their semantics, to electromagnetic signals traveling through vast distances between nodes of a network called the internet. Any repeatable action whose purpose is the transmission of information from one place to another is essentially communication.

The most basic unit of information is the *bit*. As an abstract notion, the bit is something that signifies the manifestation of two possibilities. In information theory the two possibilities are usually denoted by 0 and 1, that is, the bit can take on the binary values of zero or one, but never both at the same time. Remarkable inventions have emerged from the manipulation of bits, the most notable being perhaps the Turing machine and modern computers. However, these devices follow the laws of classical information theory, and we are now reaching the point in time where we do not just understand, but can build machines that fundamentally utilize quantum resources and phenomena. The promise of quantum technologies is astounding and possibly disruptive.

Whilst a traditional computer achieves its function by manipulating bits in a register, a quantum computer promises superior computational performance by manipulating quantum analogues of the bit, the *qubit*. A qubit differs from a bit in many important ways. Whilst a bit can never possess the value of 0 and 1 at the same time, the qubit is (in)famous for its ability to do exactly that. A qubit can be in a *superposition* state, where it has a nonzero probability of being found in the “zero” or “one” state when measured. How a superposition state should be interpreted is controversial; it depends on the interpretation of quantum mechanics one adopts whether a superposition state should be thought of as a distinct state that is neither “zero” or “one”, a state that is a “zero” and a “one” at the same time or some kind of a mixture between the two states. Moreover, a collection of many qubits can behave in surprising ways. For example, two bits are always just that, two bits. There are four possibilities for two bits to manifest the values 0 and 1, namely 00, 01, 10 and 11. Once the fact that a qubit can be in a superposition state is taken into account, the combined state of two qubits can become intertwined in a way that neither qubit possesses a definite value while simultaneously the value of both qubits is dependent on the other. This phenomena is called *entanglement*.

The theory of entanglement has remained an active subject of research since its conception. Einstein famously critiqued the phenomena [5] and dismissed the subject as “spooky action at a distance”. Entanglement was not easy to understand for the people who were intimately involved in the foundation of the theory itself, and it certainly is not easy to understand today. Perhaps even now no one truly understands it in physical terms, though that has not stopped people from trying. What can be said is that superposition, entanglement and other nonclassical features of quantum mechanics seem to stem from the differences between the bit and qubit. We should therefore try to understand that difference as well as possible at the lowest level possible, the level of single bits and qubits. Communication with these elementary objects seems like a good place to start.

The most primitive form of communication is the transmission of one bit from a sender to a receiver. In quantum theory these parties are traditionally called *Alice* and *Bob*, respectively. However, just like language is more than the transmission of single letters, the transmission of a bit needs to have a purpose, a reason for why it is being transmitted in the first place. The bit needs a communicational or computational task that is attempted to be realized. An example of such a task would be, for instance, a random access code (RAC) [6; 7; 8]. The basic idea of a RAC is that Alice is trying to encode the contents of a large bit-string into a small message, perhaps just a single bit. Bob must try to retrieve information about a random portion of the original bit-string from Alice’s encoded message. Communication tasks like the RAC are good candidates for studying the differences between the bit and qubit on the most primitive level.

A fundamental result in quantum information theory states that the qubit can only transmit a single bit of information [9]. There are infinitely many different pure states for the qubit but only two pure qubit states can be distinguished with certainty in a single measurement. The qubit is a two-level system and in this sense it is the correct analogue for the classical bit. Although the amount of information transmitted by a single bit and qubit is the same, the qubit can outperform the bit in many communication tasks such as the RAC. In this case the existence of superposition states is the reason for the qubit’s superior performance. The amount of (retrievable) information carried by the bit and qubit is the same but the richer state space of the qubit allows for more sophisticated methods for communication.

Some communication tasks require entanglement as a resource for communication. It is impossible to encode two bits of information into a single bit. However, with the help of entanglement, it becomes possible to encode two bits of information into a single qubit. In this protocol, often called *superdense coding* [10], Alice and Bob share an entangled pair of qubits. By applying clever local transformations to her qubit, Alice can drive the joint state so that upon receiving Alice’s qubit Bob can decode two bits of information by performing a measurement on both qubits. In some sense superdense coding is the opposite of quantum teleportation [11] where

Alice can transmit the state of her qubit to Bob by sending two bits of classical information.

Entanglement is always a property of a single quantum state. That is, the joint state of a quantum system is either entangled or it is not. However, there is another way to look into entanglement as a resource for communication.

Suppose you are in possession of two bits, ab, but you do not know their values. You can check the value of either bit a or b and you can be sure that by checking the value of one bit the other bit's value will not be changed in any way. In general, this is not the case in quantum mechanics. Suppose Alice and Bob share a pair of qubits. If the joint system of the qubits is in an entangled state and Alice performs an operation on her qubit, the state of Bob's qubit might be changed in a process called *steering* [12; 13]. In other words there is a *context* for Bob's measurement. In general, *contextuality* is a principle which states that some things, such as states and measurements, depend on the context in which they occur. Contextuality is a purely nonclassical feature like entanglement, that is, the bit is not contextual. However, in some communication tasks it is possible to prove that a contextual advantage exists for quantum theory. Unlike entanglement, contextuality is generally speaking not a property of a single quantum state, but one of a collection of states, measurements or channels. Therefore, in communication tasks where contextuality is a resource, it might be possible for qubits to display a contextual advantage in the absence of entanglement. Sometimes entanglement and contextuality are connected as contextuality can in some situations be seen as a manifestation of strong spatially separated correlations enabled by entangled states, that is, correlations breaking the principle of local realism [14]. The failure of local realism is sometimes referred to as *nonlocality*.

What is usually meant when it is said that contextuality and entanglement are nonclassical features is that there exists some notion of classicality which is violated by quantum mechanical systems. In the case of nonlocality the underlying principle being violated is that spatially separated classical systems should not interact faster than is allowed by the speed of light, whilst for contextuality it is that the epistemic states of indistinguishable physical systems should be identical irrespective of the context in which they occur. The key question is then whether quantum mechanics admits a hidden variable model that does not violate these principles and would thus explain quantum mechanical behavior in terms of a classical explanation [14; 15]. It turns out that questions like these are not exclusive to *gedankenexperiments*, or thought experiments, but can actually be tested in a controlled environment [16; 17; 18]. Today it is understood that any efforts to explain quantum mechanics through purely classical models are ultimately futile.

This thesis is organized in the following way. In Chapter 1 a brief introduction to operational theories and select topics in quantum mechanics are given. The basic operational setting of quantum theory is explored to the necessary depth required to understand the original research presented in Publications **I** – **IV**. Additionally, the topics of tomography, entanglement and contextuality are discussed briefly.

Chapter 2 explores the concept of *communication matrices* introduced in Publication **II**. A communication matrix is a convenient way to describe a prepare-and-measure setting in any operational theory. The conditional outcome probabilities of a measurement are simply collected into a row-stochastic matrix. A preorder on the set of communication matrices called *ultraweak matrix majorization*, introduced in Publication **II** and refined in **III**, is presented in full detail. Monotones that characterize the operational hierarchy of communication matrices as given by the preorder of ultraweak matrix majorization are studied. Moreover, different notions of dimension are introduced as maximal values of the ultraweak monotones in given operational theories. We collect these dimensions and compare them at the end of Chapter 2.

The methods developed in the previous chapter are applied to various communication tasks in Chapter 3. This is a good place to revisit Publication **I**, which studied *antidistinguishability* of pure quantum states. Publication **I** directly inspired the research for Publication **II** where communication of partial ignorance was introduced along with the concept of communication matrices. *Communication of partial ignorance* is a family of communication tasks where the objective of communication is to avoid certain outcomes. The previously known results on the operational hierarchy of these tasks is improved slightly with the help of an ultraweak monotone function. In addition, we analyze the Shannon entropy in the quantum implementations of these communication tasks. Finally, we present a generalization of communication of partial ignorance, the partial-ignorance communication tasks, which were the subject of Publication **IV**.

1 Operational quantum mechanics

The general objective of this thesis is to study different communication tasks and their properties in the context of quantum theory. However, many of the topics discussed throughout this thesis will not be specific to quantum theory and can be applied in a large variety of operational settings. A general description of an operational setting is called an *operational theory*. Loosely speaking an operational theory can be defined as a collection of the following mathematical structures.

Definition 1.1. An operational theory is any set of mathematical structures that describe the following aspects of a prepare-and-measure setting:

1: a collection of state spaces, each representing a class of physical systems

2: a composition rule that specifies the joint description of physical systems

3: a set of effects that determine the outcome probabilities of measurements for every state

4: a set of operations that describe transformations of physical systems

These mathematical structures are defined in a very specific way in *general probabilistic theories*, or GPTs [19], and in quantum mechanics. There is no unique axiomatic way to derive quantum mechanics, and we will not present an axiomatic derivation [20; 21] of the theory here. Instead we will give a brief introduction to the framework of GPTs and highlight the connection to standard Hilbert space quantum mechanics [22] with examples.

1.1 General probabilistic framework

It is a common scenario in physics, and other sciences, that the outcome of an event cannot be predicted conclusively from theory. Instead such an event has to be described probabilistically, or with statistics. If the event can be reproduced repeatedly in a controlled environment, data can be gathered about the frequency of different possible outcomes. Once enough data has been gathered, the conclusion that can be drawn is that an event is described by a probability distribution over the possible outcomes.

In an idealized version of a laboratory experiment, it is thought that an experimenter can prepare states of some physical system repeatedly. A device that can prepare states is called a *preparation device*. After being prepared, the state of a physical system goes through some possible transformation or time-evolution before a measurement is performed on it. A measurement is understood as a device, the *measurement device*, that takes a state as an input and outputs a measurement outcome with some probability that depends on the input state. After repeating the experiment many times the experimenter can describe how the prepared system behaves in the performed measurement. We can see that this setup is closely related to Definition 1.1.

It is commonly thought that an experimenter can produce mixtures of states by alternating the device used to prepare them probabilistically. This can happen, for instance, by sampling a random variable and choosing a preparation device according to the obtained value, after which the value of the random variable is forgotten. If s_1 and s_2 are valid states prepared by some devices, it is then assumed that the state $\lambda s_1 + (1-\lambda)s_2$, $\lambda \in [0,1]$ should also describe a valid state. This assumption imposes a convex structure for all states described by an operational theory. Based on this convex structure, it is conceivable that some states could be prepared in more than one way. Thus a *state* is really an equivalence class of preparations; two preparations are equivalent if they produce the same outcome probabilities for all measurements. Likewise, an *observable* is an equivalence class of measurements; two measurements are equivalent if they predict the same outcome probabilities for all states.

With the convex structure imposed on states we can now give a definition, or a requirement, for an abstract space of states, or the state space.

Definition 1.2. A state space $\mathcal{S}$ is a convex compact subset of a real finite-dimensional vector space V.

The extreme points of a state space $\mathcal{S}$ are called *pure states*. All other states are called *mixed states*.

Suppose that a state space $\mathcal{S}$ is d-dimensional, i.e., that the dimension of the affine hull $\dim(\mathrm{aff}(\mathcal{S}))$ equals d. Then it is possible to embed $\mathcal{S}$ in a $(d+1)$-dimensional real vector space V in such a way that $\mathcal{S}$ is a compact base for a closed generating proper convex cone V_+. A convex cone $C \subseteq V$ is a subset of a vector space V for which $C + C \subseteq C$ and $aC \subseteq C$ for all $a \in \mathbb{R}_+$. A cone is proper if $C \cap (-C) = \{0\}$, and generating if $\mathrm{span}(C) = V$. A set $B \subset C$ is a base for C if for every $x \in C \setminus \{0\}$ there exists a unique $b \in B$ and a number $\beta > 0$ such that $x = \beta b$. A partial order on V is induced by the proper cone V_+: $x \geq_{V_+} y$ if and only if $x - y \in V_+$. Usually the subscript in $\geq_{V_+}$ is omitted if there is no risk of confusion. With this partial order we can express the state space $\mathcal{S}$ as

$$\mathcal{S} = \{x \in V \mid x \geq 0,\ u(x) = 1\}, \tag{1}$$

where u is a strictly positive functional on V. Actually, u is a very specific functional, which we will define properly a bit later.

Example 1.3. In quantum mechanics the states with the least amount of uncertainty, or the pure states, are given by unit vectors on a complex Hilbert space $\mathcal{H}$. The linear operator associated with a pure state $\varphi \in \mathcal{H}$ is the one-dimensional projection $|\varphi\rangle\langle\varphi|$. Bounded linear operators on $\mathcal{H}$ are denoted with $\mathcal{L}(\mathcal{H})$. As mixtures of states are also considered states, the state space associated with $\mathcal{H}$ is

$$\mathcal{S}(\mathcal{H}) = \{\varrho \in \mathcal{L}_s(\mathcal{H}) \,|\, \varrho \geq 0, \mathrm{tr}\,[\varrho] = 1\}\,,$$

where the selfadjoint operators acting on $\mathcal{H}$ are denoted with $\mathcal{L}_s(\mathcal{H})$. The selfadjoint operators form a real vector space, and the partial order on $\mathcal{L}_s(\mathcal{H})$ is induced by the proper cone of positive semidefinite operators on $\mathcal{H}$. The elements of $\mathcal{S}(\mathcal{H})$ are often called *density operators*. Throughout this thesis Hilbert spaces are assumed to be finite-dimensional.

The simplest kind of measurements have binary outcomes, i.e. they answer 'yes-no' questions. These kinds of measurements are described by *effects*, which are defined as affine functionals and which assign probabilities to states in measurements with binary outcomes.

Definition 1.4. Let $\mathcal{S}$ be a state space. Then the set of effects $\mathcal{E}(\mathcal{S})$ consists of all affine functionals $e : \mathcal{S} \to [0, 1]$, i.e., $0 \leq e(s) \leq 1$ for all $s \in \mathcal{S}$ and

$$e(\lambda s_1 + (1 - \lambda)s_2) = \lambda e(s_1) + (1 - \lambda)e(s_2) \tag{2}$$

for all $s_1, s_2 \in \mathcal{S}$ and $\lambda \in [0, 1]$.

For any $s \in \mathcal{S}$ and $e \in \mathcal{E}(\mathcal{S})$ the number $e(s)$ is interpreted as the probability of detecting an outcome described by the effect e when s was prepared. The convex combination preserving property in Equation (2) means that the probability of a 'yes-no' event can be constructed by considering the probability of the event in the individual elements of a mixture. Typically an effect e is associated with the 'yes' outcome. The probability of the 'no' outcome, which is described by the effect $u - e$, is naturally given by $1 - e(s)$.

Throughout this thesis we assume the *no-restriction* hypothesis to hold true, whereby we consider all valid affine functionals to be physical effects [23]. This allows us to always identify an operational theory by its state space.

Suppose that $\mathcal{S}$ is a compact base for a closed generating proper convex cone V_+. Consider an element $\alpha s \in V_+$, where $\alpha \in [0, 1]$ and $s \in \mathcal{S}$. The number α can be interpreted as the probability of successfully preparing s in an imperfect preparation procedure, so that the element αs can be taken to be a *subnormalized state*. In particular, the probability of detecting an event described by any effect when nothing is prepared should be zero, so that $e(0s) = e(0) = 0$ for all $s \in \mathcal{S}$. This allows us to fix an origin in $\mathcal{E}(\mathcal{S})$, so that the effects can be considered as linear functionals, i.e., elements of the dual space V^* of all linear functionals on V. In this extension the effects can be defined as

$$\mathcal{E}(\mathcal{S}) = \{e \in V^* \,|\, o \leq e \leq u\} = V_+^* \cap (u - V_+^*), \tag{3}$$

where $V_+^* = \{e \in V^* \,|\, e(x) \geq 0 \,\forall x \in V_+\}$ is the dual cone of V_+ and o, u are the zero and unit effects, respectively, defined as $o(s) = 0$ and $u(s) = 1$ for all $s \in \mathcal{S}$.

An effect is called a *pure effect* if it is an extreme point of the (convex) set $\mathcal{E}(\mathcal{S})$. However, effects do not have the analogous concept of subnormalization, i.e., a subnormalized effect is still an effect. Taking into account that effects can be scaled, we call an effect e *indecomposable* if $e = f + g$ implies that $f = \alpha e$ and $g = \beta e$ for some weights α, β, i.e., the effects f, g are proportional to e. All effects can be decomposed into a finite sum of indecomposable effects [24]. The indecomposable effects, together with the pure effects, form the extreme rays of the dual cone V_+^*.

An observable in general probabilistic theories is a collection of measurement outcomes, each of which is associated with a corresponding effect.

Definition 1.5. Let $\mathcal{S}$ be a state space and $\mathcal{E}(\mathcal{S})$ the corresponding set of effects. A mapping $\mathsf{A} : x \mapsto \mathsf{A}_x$ from a finite outcome set Ω to $\mathcal{E}(\mathcal{S})$ is called an observable if

$$\sum_{x \in \Omega} \mathsf{A}_x(s) = 1 \tag{4}$$

for all states $s \in \mathcal{S}$.

Example 1.6. In quantum mechanics the set of affine mappings from $\mathcal{S}(\mathcal{H})$ to the interval $[0, 1]$ can be shown [22] to be isomorphic to the set

$$\mathcal{E}(\mathcal{H}) = \{E \in \mathcal{L}_s(\mathcal{H}) \,|\, 0 \leq E \leq \mathbb{1}\}. \tag{5}$$

That is, an effect $e \in \mathcal{E}(\mathcal{S}(\mathcal{H}))$ is mapped to a positive unit-bounded operator by a mapping defined by the relation $e(\varrho) = \mathrm{tr}\,[\varrho E]$. A quantum mechanical observable consisting of a mapping from a set of outcomes to effects is called a *positive operator-valued measure*, or a POVM.

Transformations of states are described by linear mappings.

Definition 1.7. Let $S \subset V_+ \subset V$ and $S' \subset V'_+ \subset V'$ be two state spaces. A mapping $\mathcal{N} : V \to V'$ is an *operation* (*channel*) if it is linear, $\mathcal{N}(x) \in V'_+$ and $u'(\mathcal{N}(x)) \leq u(x)$ ($u'(\mathcal{N}(x)) = u(x)$) for all x in V_+.

The number $u'(\mathcal{N}(x))$ can be interpreted as the probability that the operation described by the linear mapping $\mathcal{N}$ was successful. That is, it is not guaranteed that every operation is successful. However, a channel always transforms states into states, so the action of a channel is deterministic in this sense. Throughout this thesis it is assumed that a state is always discarded after a measurement. However, sometimes it is desired that a notion of a "post-measurement" state exists so that subsequent measurements and operations can be attempted. In this case the use of instruments is required [22]. Simply put, an instrument is a mapping from an outcome set to a set of operations.

Example 1.8. In quantum mechanics the composite states of subsystems A and B are defined in the tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_A$ and $\mathcal{H}_B$ describe the subsystems A and B, respectively. It is not enough for a quantum mechanical operation to be positive, i.e., it is not enough that a linear mapping $\mathcal{N} : \mathcal{T}(\mathcal{H}) \to \mathcal{T}(\mathcal{H})$ is trace nonincreasing, where $\mathcal{T}(\mathcal{H})$ denotes the trace class operators on $\mathcal{H}$. For instance, it may happen that a mapping $\mathcal{N} : \mathcal{T}(\mathcal{H}_A) \to \mathcal{T}(\mathcal{H}_A)$ maps every state in $\mathcal{S}(\mathcal{H}_A)$ to a valid state, but the mapping $\mathcal{N}_A \otimes \mathcal{I}_B : \mathcal{T}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{T}(\mathcal{H}_A \otimes \mathcal{H}_B)$, where $\mathcal{I}_B$ is the identity operator on $\mathcal{H}_B$, might not be positive. Therefore for a quantum mechanical operation it is required that the extension $\mathcal{N}_A \otimes \mathcal{I}_B : \mathcal{T}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{T}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is positive for all finite-dimensional $\mathcal{H}_B$. This property is called *complete positivity*.

As a last consideration on GPTs we should discuss how composite systems are described in them. The mathematical structure of composite systems should be operationally motivated. Based on empirical information on nature, we can limit the discussion to composite systems that obey the principles of *non-signalling* and *local tomography*. The non-signalling principle states that composite systems should not allow faster-than-light communication between distant parties. i.e., the marginal probability distribution for outcomes on one subsystem should not be affected by measurements on another subsystem. The principle of local tomography is a bit more controversial [1]. It posits that any joint state can be completely and uniquely determined by local measurements on individual subsystems. Under these principles it can be shown [26; 27] that the joint state space $\mathcal{S}^{AB}$ of subsystems given by $S^A \subset V_+^A \subset V^A$ and $S^B \subset V_+^B \subset V^B$ can be considered as a subset on the tensor product $V^A \otimes V^B$.

Even though we have now fixed that the joint state space $\mathcal{S}^{AB}$ should live in the vector space $V^A \otimes V^B$, it turns out there is more than one way in which the cone of subnormalized states whose base $\mathcal{S}^{AB}$ is can be formed, and there does not appear to exist a strong operational motivation to choose one over another. The first way in which the composite state space can be formed is with the so-called *minimal tensor product* cone $\left(V^A \otimes_{\min} V^B\right)_+$ which consists of all positive linear combinations of products of positive elements.

[1] This principle is controversial because composite systems can be formed without it [23; 25]. However, this principle is obeyed by standard quantum mechanics.

Definition 1.9. Let $S^A \subset V_+^A \subset V^A$ and $S^B \subset V_+^B \subset V^B$ be two state spaces. The set

$$\mathcal{S}^A \otimes_{\min} \mathcal{S}^B = \left\{ \sum_i \lambda_i s_i^A \otimes s_i^B \,\middle|\, \forall i : s_i^A \in \mathcal{S}^A,\, s_i^B \in \mathcal{S}^B,\, \lambda_i \geq 0,\, \sum_i \lambda_i = 1 \right\}$$

is called the *minimal state space* (of the composite system consisting of subsystems A and B), and it forms the compact base for the cone $\left(V^A \otimes_{\min} V^B\right)_+$.

The operational motivation behind the minimal state space is that it should be possible to prepare states independently, i.e., mixtures of product states should be valid states. However, it would be equally valid to insist states be normalized elements on product effects. This leads to the *maximal tensor product* cone, defined as $\left(V^A \otimes_{\max} V^B\right)_+ := \left(\left(V^A\right)^* \otimes_{\min} \left(V^B\right)^*\right)_+^*$.

Definition 1.10. Let $S^A \subset V_+^A \subset V^A$ and $S^B \subset V_+^B \subset V^B$ be two state spaces. The set

$$\mathcal{S}^A \otimes_{\max} \mathcal{S}^B = \left\{ s \in V^A \otimes V^B \,\middle|\, (u^A \otimes u^B)(s) = 1,\, (e^A \otimes e^B)(s) \geq 0 \right.$$
$$\left. \forall\, e^A \in \mathcal{E}(\mathcal{S}^A),\, e^B \in \mathcal{E}(\mathcal{S}^B) \right\}$$

is called the *maximal state space* (of the composite system consisting of subsystems A and B), and it forms the compact base for the cone $\left(V^A \otimes_{\max} V^B\right)_+$.

Clearly $\mathcal{S}^A \otimes_{\min} \mathcal{S}^B \subseteq \mathcal{S}^A \otimes_{\max} \mathcal{S}^B$. It has been shown that the minimal and maximal state spaces are equivalent only if one of the state spaces is a simplex [28]. Otherwise the proper cone of joint states has to be determined, but generally speaking any positive proper cone $V_+^{AB} \subset V^A \otimes V^B$ that satisfies $\left(V^A \otimes_{\min} V^B\right)_+ \subseteq V_+^{AB} \subseteq \left(V^A \otimes_{\max} V^B\right)_+$ can be considered to be a valid cone of subnormalized states for the composite system of A and B.

Example 1.11. In quantum mechanics a selfadjoint operator $W \in \mathcal{L}_s(\mathcal{H}_A \otimes \mathcal{H}_B)$ is called an *entanglement witness* if $\langle \psi \otimes \varphi | W \psi \otimes \varphi \rangle \geq 0$ for all factorized vectors $\psi \otimes \phi \in \mathcal{H}_A \otimes \mathcal{H}_B$ but W is not positive [22]. It is then clear that by scaling the operator W appropriately so that $\mathrm{tr}\,[W'] = 1$, $W' = \frac{W}{\mathrm{tr}[W]}$ (the trace of W is positive), the operator W' is in fact an element of the maximal state space $\mathcal{S}(\mathcal{H}_A) \otimes_{\max} \mathcal{S}(\mathcal{H}_B)$. Since W' is not positive it is not an element of $\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and therefore $\mathcal{S}(\mathcal{H}_A) \otimes_{\min} \mathcal{S}(\mathcal{H}_B) \subsetneq \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B) \subsetneq \mathcal{S}(\mathcal{H}_A) \otimes_{\max} \mathcal{S}(\mathcal{H}_B)$ as the state space $\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ does nevertheless contain some entangled states and so it is not equivalent to the minimal tensor product $\mathcal{S}(\mathcal{H}_A) \otimes_{\min} \mathcal{S}(\mathcal{H}_B)$.

Although most of the results throughout this thesis can be considered without specifying the underlying theory, there are two classes of theories that are often considered as examples. Actually these classes also form the only "physical" theories, and while other theories also exist, all other theories can be thought of as toy theories or backup theories in case some of the postulates of quantum mechanics ever fail. Let us now define the two classes of physical theories.

Example 1.12 (Operational classical theory)**.** Although there exists various notions of classicality, a state space is said to be classical if and only if every state has a unique convex decomposition into pure states. As a compact convex set this means that a classical state space should be a simplex, as simplices are the only sets that satisfy this requirement [29]. Some of the first simplices are a point, a line segment (*the bit*), a triangle (*the trit*) and a tetrahedron.

Suppose that a classical state space is d-dimensional, or a $d-1$-simplex. We denote this state space with $\mathcal{S}_d^{cl}$. The cone of subnormalized states and the set of effects are defined as before. Notably the extremal effects are the indecomposable effects which map different pure states to probability one, together with the zero and unit effects. Essentially this means that a classical theory only has a single measurement, namely the measurement that distinguishes pure states with certainty. All of the other measurements in classical theory are post-processings of the distinguishing measurement. The joint state space of two classical state spaces is defined by the minimal tensor product. The set of channels in a classical theory can be shown to coincide with pre- and post-processings of preparations and outcomes [19].

Example 1.13 (Operational quantum theory)**.** Operational quantum theory is defined by fixing a set of Hilbert spaces $\{\mathcal{H}_A, \mathcal{H}_B, \ldots\}$. Each Hilbert space $\mathcal{H}_A$ describes a class of physical systems whose states are defined in $\mathcal{S}(\mathcal{H}_A)$. A d-dimensional quantum state space is denoted with $\mathcal{Q}_d$. A two-dimensional quantum system is called a *qubit*, a three-dimensional a *qutrit*, while a general d-dimensional quantum system is called a *qudit*. Composite systems are described by the tensor product of Hilbert spaces, i.e., a state of a composite system of A and B is described by a state in $\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Measurements are described by POVMs and each effect of a POVM maps quantum states to probabilities via the Born rule. Transformations of physical systems are described by channels, or completely positive trace preserving maps.

Prepare-and-measure scenarios

In the basic operational setting two parties, Alice and Bob, are trying to establish a communication channel between them. Suppose Alice possesses a device that can prepare quantum states with n different labels. Then we denote with $\{\varrho_i\}_{i=1}^n$ the set of states that Alice can prepare. In the case of generic preparations in an arbitrary operational theory the symbols s or P are typically used. Communication between

Alice and Bob consists of Alice transmitting a state of some physical medium to Bob through a channel. In a typical prepare-and-measure setting the channel can be thought of as part of the preparation. Throughout this thesis we will implicitly assume the channel to be the identity channel. Upon receiving the state sent by Alice, Bob will perform a measurement on the state, after which the state is discarded. In quantum theory measurements are described by POVMs. A POVM with (discrete) outcome set $\{1, 2, \ldots, k\}$ is a collection $\mathsf{M} = \{\mathsf{M}(i)\}_{i=1}^{k}$ of positive effects $\mathsf{M}(i) \geq 0$ such that $\sum_{i=1}^{k} \mathsf{M}(i) = \mathbb{1}$. The probability of obtaining outcome i in a measurement of M on a state ϱ is given by the Born rule:

$$p(i|\varrho, \mathsf{M}) = \mathrm{tr}\left[\varrho \mathsf{M}(i)\right]. \tag{6}$$

Outcome statistics of several measurements are often collected into a data table, or a *behavior*, $p(k|\varrho_i, \mathsf{M}_j) \equiv \mathrm{tr}\left[\varrho_i \mathsf{M}_j(k)\right]$. Behaviors can be defined in any operational theory. To distinguish a general operational setting from quantum theory we denote preparations with P and measurements with M in a general operational theory, while ϱ and M are always associated with quantum theory. The measurements in $p(k|P_i, M_j)$ are naturally defined in the same operational theory as the preparations. Sometimes behaviors are also used to store relative frequencies of different outcomes in a sequence of measurements. In this case the term *data table* is often used to specify that the probabilities are not predicted by theory but are given by an experiment. The same notation is used for behaviors and data tables but the origin of probabilities should always be clear from context.

1.2 Distinguishability of states

After Alice and Bob have established a communication channel between them, the most basic question regarding communication is the following: how many distinct messages can Alice send to Bob? A definition is in order.

Definition 1.14. A set of preparations $\{P_i\}_{i=1}^{n}$ is distinguishable if there exists a measurement M such that:

$$\forall i : \; p(i|P_i, M) = 1. \tag{7}$$

In quantum mechanics it is well-known that only orthogonal states can be distinguished with certainty in a single measurement.

The number of distinguishable states can be used as the definition for the *operational dimension* of the related operational theory. As an example, bits and qubits have operational dimension $d_{op} = 2$ while trits and qutrits have $d_{op} = 3$. In general, a d-dimensional quantum system is analogous to a classical unit of information with base d, i.e., approximately $\log_2 d$ bits of information.

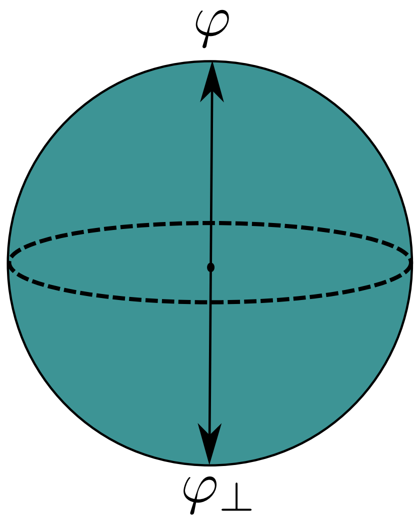

Figure 1. State space of the qubit, the Bloch sphere. Antipodal points on the Bloch sphere define pairs of orthogonal states.

The state space of qubits is the *Bloch sphere*, i.e., a three-dimensional ball with radius 1, as illustrated in Figure 1. Three real numbers are needed to encode the state of a qubit. Each point on the surface of the Bloch sphere represents a distinct pure state, i.e., there is an uncountable number of different pure qubit states. How is it that in a single measurement of any POVM at most two qubit states can be distinguished with certainty? There are several possible ways to answer this question.

The first possible answer is that a qubit is a two-level system by definition. The Hilbert space of a qubit is two-dimensional and therefore there can be at most two orthogonal pure states. This is admittedly a bit anticlimactic. An alternative answer comes in the form of a famous theorem by Holevo [9].

***Theorem* 1.15.** *Suppose Alice has a random variable X with values $\{1, 2, \ldots, n\}$ and corresponding probabilities $\{p_1, p_2, \ldots, p_n\}$. Upon obtaining a value $x \in X$ Alice will prepare a quantum state ϱ_x from the set $\{\varrho_i\}_{i=i}^n$ and send it to Bob. Let* M *be any POVM with an outcome set Y. After Bob has performed a measurement of* M*, the accessible information on X given an outcome $y \in Y$ is bounded from above by:*

$$I(X : Y) \leq S(\varrho) - \sum_i p_i S(\varrho_i), \tag{8}$$

where

$$I(X : Y) = \sum_{y \in Y} \sum_{x \in X} P_{(X,Y)}(x, y) \log \left(\frac{P_{(X,Y)}(x, y)}{P_X(x) P_Y(y)} \right) \tag{9}$$

is the mutual information of random variables X and Y, $P(X,Y)$ is the joint probability mass function of X and Y, P_X and P_Y are the corresponding marginal probability mass functions, $S(\varrho) = -\mathrm{tr}[\varrho \log \varrho]$ is the von Neumann entropy and $\varrho = \sum_i p_i \varrho_i$.

The choice of base for the logarithm in Equation (9), and for the von Neumann entropy, defines the unit of information. Throughout this thesis the choice of base 2 is used, which gives the unit of information as bits. Other common choices are the natural logarithm and base 10, which give the units nats and dits, respectively. The quantum state of maximum entropy is the maximally mixed state, i.e. $\frac{1}{d}\mathbb{1}$ for states in $\mathcal{S}(\mathcal{H}_d)$. For qubits the maximally mixed state has an entropy of one bit. Pure states have the minimum entropy of zero bits. Therefore the accessible information on any random variable X given an outcome of a qubit measurement is at most one bit.

Holevo's theorem gives an information-theoretical explanation of why only two qubit states can be distinguished in a single measurement. A qubit is a quantum system which can transmit at most one bit of information. If more than two qubit states could be distinguished, the amount of information they could transmit would be greater than one bit. In some sense this is surprising as three real numbers are needed to specify an arbitrary quantum state on the Bloch sphere. Therefore an arbitrary amount of information can be encoded into a qubit in a superposition state, but no more than one bit of this information can be retrieved in a single measurement.

1.3 Tomography in quantum mechanics

It is impossible to identify the state of an unknown quantum state in a single measurement without prior knowledge on the preparation procedure [22]. Suppose Bob's measurement is described by the POVM $\mathsf{M} = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ with outcome set $\{0,1\}$, where $\{|0\rangle, |1\rangle\}$ is an orthonormal basis for the Hilbert space $\mathcal{H}_2$. If Alice prepares a superposition state $\alpha|0\rangle + \beta|1\rangle$, where α and β are complex numbers with $|\alpha|^2 + |\beta|^2 = 1$, the probability for Bob to obtain the outcome 0 equals $|\alpha|^2$. In other words there is an infinite number of states that give a nonzero probability for outcome 0. Therefore, upon obtaining outcome 0 after performing a measurement on an unknown state, it is impossible for Bob to conclude anything else except that the probability of obtaining outcome 0 was not zero.

Suppose Bob receives many copies of a d-dimensional unknown quantum state ϱ and he collects the outcome statistics of a POVM M into a data table $p(k|\varrho, \mathsf{M})$. Suppose also that M is minimal and informationally complete so that the effects of M span $\mathcal{L}_s(\mathcal{H})$ and M has exactly d^2 elements, i.e., the effects of M form a basis for $\mathcal{L}_s(\mathcal{H})$. In this case it is possible to write every selfadjoint operator, including ϱ, as a linear combination of the effects of M. Let $\varrho = \sum_x r_x \mathsf{M}(x)$ and define $\mathcal{U}_{xy} \equiv \mathrm{tr}[\mathsf{M}(x)\mathsf{M}(y)]$. Then the unknown quantum state ϱ can be reconstructed from the data table $p(k|\varrho, \mathsf{M})$ with the reconstruction formula:

$$r_y = \sum_x \mathcal{U}_{yx}^{-1} p(x|\varrho, \mathsf{M}). \tag{10}$$

The reconstruction formula is a naive approach to quantum tomography. It is overly optimistic that enough data has been gathered such that the data table $p(k|\varrho, \mathsf{M})$ is a good representation of the outcome probabilities of M in measurements of ϱ. Outcome statistics are always based on a finite amount of data in any real experiment. Suppose Bob has access to N copies of ϱ and he performs a measurement of M on each one. The best he can do is set $p(k|\varrho, \mathsf{M}) = N_k/N$, where N_k is the frequency of outcome k among all N measurement runs, as this is the probability that most likely gives rise to the observed frequencies. This often leads to an unphysical reconstructed state, i.e. a "state" with small negative eigenvalues, simply due to the fact that there was not enough data to determine the correct behavior [30].

A better approach to tomography, although still somewhat problematic, is the *maximum likelihood estimate* (MLE) [31; 32; 33]. Instead of trying to directly reconstruct an unknown quantum state from a data table, the MLE tries to estimate which quantum state is the most likely given a sequence of outcomes $\vec{k}$. The conditional probability of observing an outcome sequence $\vec{k} = k_1, \ldots, k_N$ given that the unknown state was ϱ is easily calculated:

$$p(\vec{k}|\varrho) = \prod_{j=1}^{N} p(k_j|\varrho) = \prod_{k=1}^{n} p(k|\varrho)^{N_k}, \tag{11}$$

where $p(k|\varrho) = \mathrm{tr}\left[\varrho\mathsf{M}(k)\right]$ and N_k is the absolute frequency of outcome k. The different measurement runs are assumed to be independent so that the order of outcomes does not matter. The likelihood function which describes the conditional probability that the unknown state was ϱ given that the outcome sequence $\vec{k} = k_1, \ldots, k_N$ was observed is given by

$$\ell(\varrho|\vec{k}) = \frac{1}{N} \log p(\vec{k}|\varrho) = \sum_{k=1}^{n} \frac{N_k}{N} \log p(k|\varrho). \tag{12}$$

By maximizing the likelihood function over $\mathcal{S}(\mathcal{H})$ we obtain the maximum likelihood estimate, or the MLE, of which state was the most probable given an outcome sequence:

$$\hat{\varrho} = \arg \max_{\varrho \in \mathcal{S}(\mathcal{H})} \ell(\varrho|\vec{k}). \tag{13}$$

The MLE has a very compelling and intuitive logic behind it. For that reason it is one of the more popular estimation methods in statistical inference and by extension in quantum tomography. However, the MLE has a significant drawback worth pointing out. In many cases the MLE estimates that the most probable quantum state lies on the boundary of $\mathcal{S}(\mathcal{H})$, i.e., the most probable state has at least one zero eigenvalue. This is hard to justify based on finite statistics as it would imply that some measurement outcomes would have a zero probability of occurring. A full-rank estimate would in most cases be preferable, and some estimates are designed with this in mind. Examples of such estimates would be, for instance, Bayesian mean estimation [34] and maximum entropy estimation [35]. Moreover, it must be said that finding the MLE is not a trivial task numerically. The likelihood function might have many local maxima, and therefore iterative methods are needed to find the global optimum. The diluted maximum-likelihood algorithm is an example of a method with good convergence properties for the MLE [36].

In conclusion, quantum tomography is a process where the goal is to estimate an unknown quantum state based on finite statistics of a fixed set of measurements. The reconstruction of the unknown quantum state can be attempted directly from the observed statistics through the reconstruction formula given by Eq. (10). However, the direct reconstruction often gives an unphysical result. Often it is best to use an estimation method to obtain an estimate for the unknown quantum state. The estimation method should be chosen with specific goals in mind. For instance, if a full-rank estimate is desired, the MLE cannot be used. However, the MLE remains a popular choice because of its intuitive logic and the existence of algorithms with good convergence properties.

So far we have explained some of the most fundamental topics in quantum theory: distinguishability of quantum states, information transmission and quantum tomography. In the remaining sections of this chapter we will explore entanglement and contextuality as examples of nonclassical features of quantum mechanics.

1.4 Entanglement

The state of a composite quantum system ϱ_{AB} of two subsystems is represented by a quantum state in the tensor product space of the individual Hilbert spaces. Such a state is called *separable* if and only if the composite quantum state can be written as a convex sum of product states:

$$\rho_{A|B} = \sum_i p_i \rho_A^{(i)} \otimes \rho_B^{(i)}, \tag{14}$$

where p_i's form a probability distribution, $\rho_A^{(i)}$'s are density operators on the Hilbert space of system A and $\rho_B^{(i)}$'s are density operators on the Hilbert space of system B. If the state of a composite quantum system is not separable then it is entangled. The following example shows that entangled states do in fact exist.

Example 1.16. Let $\mathcal{H} = \mathcal{H}_2 \otimes \mathcal{H}_2$ and

$$\varphi_\pm = \frac{1}{\sqrt{2}} \left(|01\rangle \pm |10\rangle \right), \tag{15}$$

$$\psi_\pm = \frac{1}{\sqrt{2}} \left(|00\rangle \pm |11\rangle \right), \tag{16}$$

where the vectors $\{|0\rangle, |1\rangle\}$ form a basis for $\mathcal{H}_2$. The pure states defined by the vectors $\varphi_\pm$ and $\psi_\pm$ are generally referred to as the Bell states, or maximally entangled states. Suppose that any of the vectors $\varphi_\pm$ or $\psi_\pm$, for instance φ_+, can be written as a product $\varphi_+ = \phi \otimes \xi$, where $\phi = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $\xi = \beta_0|0\rangle + \beta_1|1\rangle$. It follows that $\varphi_+ = \alpha_0\beta_0|00\rangle + \alpha_0\beta_1|01\rangle + \alpha_1\beta_0|10\rangle + \alpha_1\beta_1|11\rangle$. This means that the coefficient of $|00\rangle$ times the coefficient of $|11\rangle$ has to be equal to the coefficient of $|01\rangle$ times the coefficient of $|10\rangle$ in order for φ_+ to be of the product form $\phi \otimes \xi$, which is a clear contradiction with Equation (15). Therefore all four Bell states are entangled.

Entanglement is a phenomena that happens when the states of two or more subsystems become intertwined in such a way that the entire system cannot be accurately described by the properties of the individual subsystems. Moreover, the state of a composite system can become entangled in several different ways [37]. Suppose there exists three distinct systems: A, B and C. The combined state is fully separable if it can be written as

$$\varrho_{A|B|C} = \sum_i p_i \varrho_A^{(i)} \otimes \varrho_B^{(i)} \otimes \varrho_C^{(i)}. \tag{17}$$

If the combined system is not fully separable, there must exist some entanglement in the system. The simplest form of entanglement is so-called level-I entanglement that forbids the combined state from being partitioned in a certain way. The possible bipartitions for a tripartite state are as follows:

$$\varrho_{A|BC} = \sum_i p_i \varrho_A^{(i)} \otimes \varrho_{BC}^{(i)}, \quad \varrho_{B|AC} = \sum_i p_i \varrho_B^{(i)} \otimes \varrho_{AC}^{(i)} \tag{18}$$

$$\varrho_{C|AB} = \sum_i p_i \varrho_C^{(i)} \otimes \varrho_{AB}^{(i)}, \quad \varrho_{ABC} = \sum_i p_i \varrho_{ABC}^{(i)}. \tag{19}$$

The state ϱ_{ABC} for which no partition is allowed is called genuinely tripartite entangled. States that are not genuinely tripartite entangled can always be divided to some separable subsystems. Some states can be divided into separable subsystems in more than one way. For instance, there exists tripartite states that can be divided according to all bipartitions $\varrho_{A|BC}$, $\varrho_{B|AC}$ and $\varrho_{C|AB}$ but yet are not fully separable [38]. The term *delocalized* entanglement is sometimes used when the state can be partitioned in more than one way.

Level-I entanglement gives a clear picture how a composite system can be divided into separable subsystems. However, the state of a quantum system can be entangled in a way that it is not genuinely tripartite entangled and yet the state cannot be divided into any partitions according to level-I characterization of entanglement. For instance, the state of a quantum system could be written as a convex sum

$$\varrho_{ABC} = p_1 \varrho_{A|BC} + p_2 \varrho_{B|AC} + p_3 \varrho_{C|AB} \tag{20}$$

where the state is not genuinely tripartite entangled and yet the state cannot be divided into separable subsystems according to a single bipartition. This kind of entanglement is called level-II entanglement.

Detecting entanglement is a difficult task. In fact, no general method to determine whether a quantum state is entangled has been presented. Moreover, many different approaches for detecting entanglement exist. We will introduce two methods that are commonly used.

The first approach is to attempt to find out whether a given quantum state is entangled based on some entanglement measure [39]. An additional goal is to find out the state's entanglement structure, i.e., the partitions according to which the state is entangled. A crucial observation is that this approach assumes the quantum state is known. That is, either the question of whether the quantum state is entangled concerns a purely mathematical quantum state, or the quantum state is known to a very high degree through tomography. The point that the quantum state has to be known to a very high degree cannot be stressed enough. The set of separable states is clearly convex. If the true quantum state lies on the boundary of the convex set of separable states, a minuscule disturbance in the outcome statistics could be enough to make the state appear entangled.

Once the quantum state is known, the question of separability can be resolved by finding a representation for the state as a convex sum of product states. Unfortunately, this is a very difficult task and no necessary and sufficient efficiently checkable criteria to test separability exist. The most common test is to calculate the so-called *partial transpose* and check if the partial transpose is positive. This test is known as the *positive partial transpose* (PPT) criterion. A negative partial transpose is a sufficient condition for the state to be entangled [40]. This can be demonstrated rather straightforwardly. Let us start with the definition of the partial transpose.

Suppose $S \in \mathcal{L}(\mathcal{H})$ is defined in a tensor product space of two individual Hilbert spaces of finite dimension, $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. Then the partial transpose of S with respect to the subsystem A, denoted S^{τ_A}, is defined by the operator with matrix elements

$$\langle i_A, j_B | S^{\tau_A} | k_A, l_B \rangle = \langle k_A, j_B | S | i_A, l_B \rangle. \tag{21}$$

For a bounded linear operator $S \in \mathcal{L}(\mathcal{H})$ the *trace norm* is defined as $\|S\|_{\mathrm{tr}} = \mathrm{tr}\,[|S|]$, where $|S| = \sqrt{S^* S}$ is the absolute value of S. However, for a selfadjoint trace class operator the trace norm is equal to the sum of the absolute values of its eigenvalues. In particular, for a state $\varrho \in \mathcal{S}(\mathcal{H})$ it holds that $\mathrm{tr}\left[\varrho^{\tau_A}\right] = 1$, and the trace norm of ϱ^{τ_A} is equal to

$$\|\varrho^{\tau_A}\|_{\mathrm{tr}} = 1 + 2\left|\sum_i \mu_i\right| \equiv 1 + 2\mathcal{N}(\varrho), \tag{22}$$

where μ_i's are the negative eigenvalues of ϱ^{τ_A} and where we have defined the absolute value of the sum of μ_i's to be the *negativity* $\mathcal{N}(\varrho)$ of ϱ.

It is straightforward to verify that the partial transpose of a separable state is always positive. Namely, if $\varrho = \sum_i p_i \varrho_A^{(i)} \otimes \varrho_B^{(i)}$, then $\varrho^{\tau_A} = \sum_i p_i \varrho_A^{(i)T} \otimes \varrho_B^{(i)}$ which is a convex sum of positive operators. Therefore a negativity greater than zero is a clear sign that the state is entangled. Unfortunately there exists entangled states with zero negativity. However, for small Hilbert spaces, $\mathbb{C}^2 \otimes \mathbb{C}^2$ and $\mathbb{C}^2 \otimes \mathbb{C}^3$ to be exact, the PPT criterion is also sufficient [41; 42].

The PPT criterion is easy to use, but its application is limited because the quantum state has to be known and because entangled states with zero negativity exist. For unknown quantum states it may be more efficient to measure an observable that optimally gives you an answer whether the quantum state is entangled or not. Such observables are known as entanglement witnesses.

Suppose $W \in \mathcal{L}_s(\mathcal{H}_A \otimes \mathcal{H}_B)$ is an entanglement witness as defined in Example 1.11. Then W is not positive but $\mathrm{tr}\left[\varrho_A \otimes \varrho_B W\right] \geq 0$ for all separable states. For all states ϱ for which $\mathrm{tr}\left[\varrho W\right] < 0$ it is said that the entangled state ϱ is detected by W. Using entanglement witnesses has the advantage that it is not necessary to do full state tomography, as only a single observable has to be measured in the best case scenario. However, a single entanglement witness cannot detect all entangled states. Instead in this approach detecting entanglement becomes a question of finding a suitable entanglement witness and measuring it. It is possible to detect all entangled states in this way. Figure 2 depicts a schematic diagram for an entanglement witness.

1.5 Contextuality

Noncontextuality is a principle according to which systems that are operationally indistinguishable should also be ontologically indistinguishable. To be more precise, with noncontextuality we will refer to the notion of generalized noncontextuality introduced by Spekkens [43]. This is not the only way noncontextuality is defined in the literature [44; 45], but Spekkens' generalized noncontextuality does appear frequently in the literature, although perhaps the original notion of noncontextuality introduced by Kochen and Specker [15] is still the most commonly seen notion.

Historically, Kochen–Specker (KS) noncontextuality was the first attempt to formalize the idea that there exists a concept of context which is very relevant when

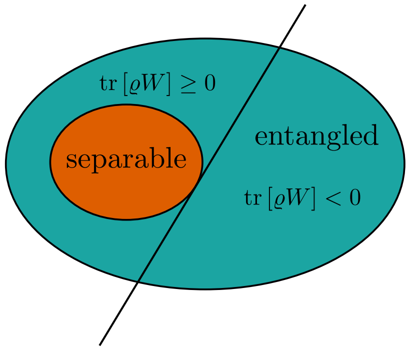

Figure 2. A depiction of an entanglement witness W for which the expectation value $\mathrm{tr}\,[\varrho W]$ is nonnegative for all separable states. The separating hyperplane, as illustrated by the solid line, can be defined by setting $\mathrm{tr}\,[\varrho W] = 0$.

distinguishing quantum theory from classical physical theories. However, the original theorem by Kochen and Specker only applied to quantum theory and projective measurements. Spekkens' generalized noncontextuality attempts to extend the concept of contextuality to arbitrary operational theories. It is worth mentioning that the notion of noncontextuality can also be characterized in the framework of generalized probabilistic theories [46].

In the framework of Spekkens' generalized contextuality it is assumed that all aspects of physical systems can be described by ontological models. That is, it is assumed that a complete description of reality exists and that this description is at least partly accessible to experimenters through the details laid down by a specific ontological model. The primitives of an operational description of such a physical theory are preparation procedures, transformation procedures[2] and measurement procedures. The ontic state of a physical system is denoted typically with λ, while the state space of all possible physical configurations of physical systems is denoted with Λ. An experimenter does not in general have access to λ upon preparing a state. Instead an experimenter will describe the physical state of a system after a preparation procedure P probabilistically with a probability distribution μ_P.

In addition to the ontic states of physical systems, an ontological model is expected to explain the probabilities that arise in measurements. Thus each measurement operator, or effect, should have a corresponding element that is specified by the ontological model and that explains the probability of each possible outcome in any given measurement. Thus, in a measurement of M with an outcome set $\{1, 2, \ldots, n\}$, the probability of obtaining outcome i following a preparation pro-

[2] We will not deal with transformations separately. Instead we will treat them as part of the preparation procedures.

cedure P is given by

$$p(i\,|\,P,M) = \int_\Lambda \xi_M(i\,|\,\lambda)\mu_P(\lambda)d\lambda,$$

where $\xi_M(i\,|\,\lambda)$ is the response function associated with outcome i in a measurement of M. These probabilities should always be positive and normalized so that summing over the outcomes gives a total of one.

Two preparation procedures P_1 and P_2 are operationally equivalent if and only if $p(i\,|\,P_1,M) = p(i\,|\,P_2,M)$ for all measurement procedures M and outcomes i. Likewise, two measurement procedures M_1 and M_2 are operationally equivalent if and only if $p(i\,|\,P,M_1) = p(i\,|\,P,M_2)$ for all preparation procedures P and outcomes i. Operationally equivalent preparation and measurement procedures are denoted with the symbol "$\simeq$", i.e., we write $P_1 \simeq P_2$ if the preparation procedures are operationally equivalent.

As stated in the beginning of this section, noncontextuality is understood as a principle according to which operationally equivalent procedures should be described equivalently at the ontological level. For preparation procedures this means that a noncontextual ontological model will assign the same probability distribution to operationally equivalent preparation procedures, i.e., $\mu_{P_1}(\lambda) = \mu_{P_2}(\lambda)$ for all $\lambda \in \Lambda$ whenever $P_1 \simeq P_2$. Likewise, for measurement procedures it is understood that a noncontextual ontological model will assign the same response functions to operationally equivalent measurement procedures.

Let us then assume that the outcome statistics for a given number of measurement procedures have been measured in a laboratory. Alternatively the outcome statistics could be the predictions of some operational theory, quantum theory for instance. Once the operational equivalences have been determined from the statistics or from theory, the question of whether a noncontextual ontological model can be used to explain the statistics is a problem that turns out to be solvable as a linear program [47]. That is, having collected experimental data it is a relatively straightforward task to check if the given data can be explained with a classical model that obeys the principle of noncontextuality. A violation of the principle of noncontextuality can be witnessed if it is not possible to construct a noncontextual model for a given set of experimental data. Analogously to Bell inequalities for entanglement, the certificate of primal infeasibility for the primal linear program can be understood as a noncontextuality inequality [47]. Whenever an operational theory is capable of violating a noncontextuality inequality the conclusion is that the operational theory in question is contextual. Multiple instances where quantum mechanical systems violate the principle of noncontextuality, either in theory or experimentally, can be found in the literature. See e.g. [48; 49] for recent experimental tests. One example of a theoretical violation is presented in Publication **IV**.

An ontological model that is not universally noncontextual in Spekkens' generalized framework is said to be contextual. However, it turns out an ontological model can be noncontextual in several different, nonequivalent ways. An ontological model is measurement noncontextual if the response functions associated with operationally equivalent measurement procedures are unique. This is taken to hold on the level of individual outcomes, so that whenever $p(i \,|\, P, M) = p(i \,|\, P, M')$ for all P, then $\xi_M(i \,|\, \lambda) = \xi_{M'}(i \,|\, \lambda)$ for all $\lambda \in \Lambda$. If an ontological model is KS noncontextual, then it should be measurement noncontextual and outcome deterministic, i.e., the response functions $\xi_M(i \,|\, \lambda) \in \{0, 1\}$ for all measurements and outcomes and for all $\lambda \in \Lambda$. An ontological model can also be noncontextual with respect to preparations, i.e., for a preparation noncontextual model it holds that $\mu_P(\lambda) = \mu_{P'}(\lambda)$ for all $\lambda \in \Lambda$ whenever $P \simeq P'$.

It turns out that preparation noncontextuality is a more general notion than KS noncontextuality as it has been shown that preparation noncontextuality implies measurement noncontextuality and outcome determinism [50]. A proof of KS theorem therefore implies that preparations should be contextual in quantum theory. Many proofs of the KS theorem exist in the literature, see e.g. [51; 52; 53; 54]. Let us present a simple proof of preparation contextuality from [43] to conclude this Chapter.

Example 1.17. Consider the following set of density operators:

$$\begin{aligned} \varrho_a = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ \varrho_A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \ \varrho_b = \frac{1}{4}\begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix} \\ \varrho_B = \frac{1}{4}\begin{bmatrix} 3 & -\sqrt{3} \\ -\sqrt{3} & 1 \end{bmatrix}, \ \varrho_c = \frac{1}{4}\begin{bmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & 3 \end{bmatrix}, \ \varrho_C = \frac{1}{4}\begin{bmatrix} 3 & \sqrt{3} \\ \sqrt{3} & 1 \end{bmatrix}. \end{aligned} \tag{23}$$

These are the projections onto the following vectors, respectively: $\begin{bmatrix} 1 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 \end{bmatrix}$, $\begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$, $\begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$, $\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}$ and $\begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$. It is straightforward to verify that $\varrho_a\varrho_A = \varrho_b\varrho_B = \varrho_c\varrho_C = 0$ holds. Moreover, we have to following decompositions of the maximally mixed state:

$$\begin{aligned} \frac{1}{2}\mathbb{1} = \frac{1}{2}(\varrho_a + \varrho_A) = \frac{1}{2}(\varrho_b + \varrho_B) = \frac{1}{2}(\varrho_c + \varrho_C) \\ = \frac{1}{3}(\varrho_a + \varrho_b + \varrho_c) = \frac{1}{3}(\varrho_A + \varrho_B + \varrho_C). \end{aligned} \tag{24}$$

Any preparation noncontextual ontological model must therefore adhere to the following set of restrictions for all $\lambda \in \Lambda$:

$$\mu_a(\lambda)\mu_A(\lambda) = 0, \tag{25}$$

$$\mu_b(\lambda)\mu_B(\lambda) = 0, \tag{26}$$

$$\mu_c(\lambda)\mu_C(\lambda) = 0, \tag{27}$$

where the orthogonality of the epistemic states follows from the perfect distinguishability of orthogonal quantum states. Additionally, denoting the epistemic state associated with $\frac{1}{2}\mathbb{1}$ with $\nu(\lambda)$:

$$\nu(\lambda) = \frac{1}{2}\left(\mu_a(\lambda) + \mu_A(\lambda)\right) \tag{28}$$
$$= \frac{1}{2}\left(\mu_b(\lambda) + \mu_B(\lambda)\right) \tag{29}$$
$$= \frac{1}{2}\left(\mu_c(\lambda) + \mu_C(\lambda)\right) \tag{30}$$
$$= \frac{1}{3}\left(\mu_a(\lambda) + \mu_b(\lambda) + \mu_c(\lambda)\right) \tag{31}$$
$$= \frac{1}{3}\left(\mu_A(\lambda) + \mu_B(\lambda) + \mu_C(\lambda)\right). \tag{32}$$

These equalities follow from the fact that the corresponding mixtures of states are all equal to the maximally mixed state, and thus indistinguishable. Therefore they should be represented identically on the ontological level. It is now a relatively straightforward task to derive a contradiction. Namely, consider the Equations (25)-(27) for a fixed $\lambda \in \Lambda$. Clearly one of μ_a and μ_A must equal zero, one of μ_b and μ_B must equal zero and finally one of μ_c and μ_C must equal zero. There are 8 possible ways to realize these restrictions, and they should be considered one by one. Assume first that $\mu_a = \mu_b = \mu_c = 0$. Then by Equation (31) we have that $\nu(\lambda) = 0$ for all $\lambda \in \Lambda$ which is a contradiction as $\nu(\lambda)$ should be a probability distribution over λ. Consider then that $\mu_a = \mu_b = \mu_C = 0$. From Equations (30) and (31) we get that $\frac{1}{2}\mu_c = \frac{1}{3}\mu_c$ for which the only solution is $\mu_c = 0$ which implies again that $\nu(\lambda) = 0$ for all λ.

The remaining six cases can be considered individually, but they also follow from the symmetric layout of the states on the Bloch sphere, as illustrated in Fig 3. The conclusion is that the maximally mixed state is unavoidably contextual in quantum theory.

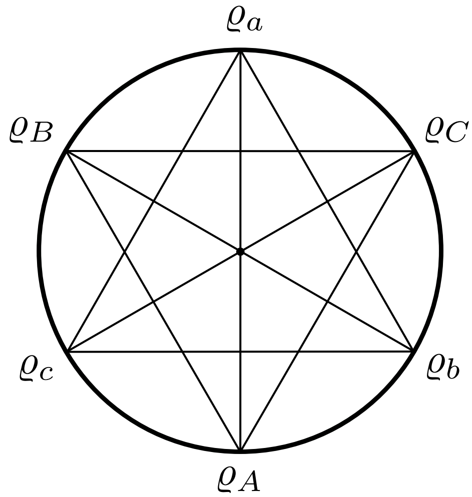

Figure 3. Three pairs of orthogonal pure qubit states that can be used to prove preparation contextuality for quantum theory. The maximally mixed state lies at the center.

2 Operational hierarchy of communication tasks

Communication is usually understood as the preparation and measurement of some physical system. The most general way to describe communication is through black boxes. The sender, typically Alice, has in her possession a device that can prepare distinct states of some physical medium. The receiver, usually Bob, has a device that produces measurement outcomes based on the input state given into the device. This kind of setup is called a prepare-and-measure scenario.

The simplest kind of a prepare-and-measure scenario is one where Bob does not receive any additional input besides the state sent by Alice, i.e., Bob is using a single measurement device. This leads to the concept of *communication matrices*, while the case where Bob is free to choose his measurement based on some additional input is described with behaviors. A behavior is obviously the more general of the two cases, but in order to analyze behaviors with the best possible detail knowledge about communication matrices is of great importance. Thus this Chapter is dedicated to the study of communication matrices in a theory-independent setting.

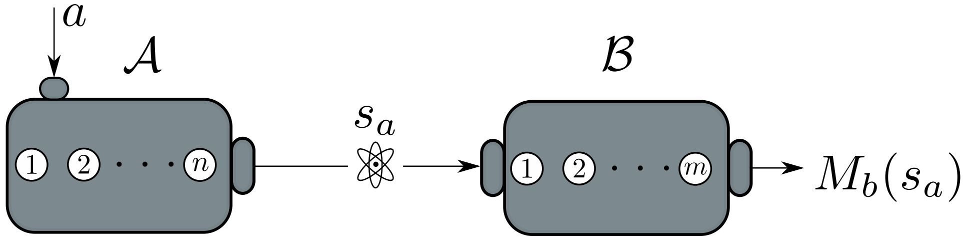

Figure 4. Basic prepare-and-measure setting. Alice prepares a state corresponding to label a, while Bob observes outcome b in his measurement with probability $M_b(s_a)$.

2.1 Communication tasks

A preparation device is understood as a device that can prepare states from a state space $\mathcal{S}$. The only requirements for the abstract state space are compactness[1] and convexity – a probabilistic mixture of two or more states should always be a valid state. Thus we always assume $\mathcal{S}$ to be a convex set. The user of the preparation device, Alice, can prepare any state from the device and send it to Bob at will. Alice's *state ensemble*[2] thus consists of the states the preparation device can prepare. Mathematically this is described as a map $s : a \mapsto s_a$ from a finite set of labels $\{1, \ldots, n\}$ into $\mathcal{S}$.

A measurement device is a device that outputs an outcome from a finite outcome set $\{1, \ldots, m\}$ upon receiving an input state. We take the outcome set to be finite in order to avoid any possible complications arising from infinite dimensional matrices and other technicalities. Descriptions of individual outcomes, also called effects, are given by affine functionals $e : \mathcal{S} \to [0, 1]$ that map states of $\mathcal{S}$ to probabilities. A measurement is always normalized over the outcome set, i.e.,

$$\sum_{b=1}^{m} M_b(s) = 1 \tag{33}$$

for all states $s \in \mathcal{S}$ where M_b is the effect corresponding to outcome b. This setup is illustrated in Figure 4.

All conditional probabilities for a fixed preparation and measurement device can be collected into an $n \times m$ *communication matrix*

$$C_{ab} = M_b(s_a). \tag{34}$$

We then say that the communication matrix C is implemented with the pair s, M. The matrix C is row-stochastic because of the normalization of effects. In fact, in a theory-independent setting we take all row-stochastic matrices to be valid communication matrices.

The difference between a communication task and a communication matrix is largely negligible, although generally speaking the terms mean very different things. By *communication task* we mean a communication setting where Alice and Bob are trying to implement a specific prepare-and-measure scenario between them according to some predefined criteria. A communication matrix, on the other hand, always describes the outcome probabilities of a specific communication task. Thus we can

[1] Compactness means that the set should be closed and bounded.

[2] Typically a state ensemble also includes the prior probabilities according to which the states are prepared. In the framework of communication tasks we choose to exclude these probabilities, and hence the state ensemble would perhaps be more appropriately described just as an ordered set of states.

use the two terms interchangeably in practice. The question of whether Alice and Bob can implement a specific communication task becomes a question of whether the corresponding communication matrix can be implemented with given resources.

One of the most important questions related to communication tasks is the following: if we fix the state space $\mathcal{S}$, what is the set of all communication matrices that can be implemented with states and measurements coming from the theory determined by $\mathcal{S}$? An additional relevant question is whether we can compare different communication tasks in the following sense: are some communication tasks easier to implement than others?

Let us make concrete the sets we are studying. We shall denote with $\mathcal{C}_{n,m}(\mathcal{S})$ the set of all $n \times m$ communication matrices that have an implementation of the form given by Eq. (34) for some pairs s, M determined by the theory $\mathcal{S}$. Additionally we take $\mathcal{C}(\mathcal{S})$ to be the union of all finite-sized $\mathcal{S}$-implementable communication matrices, i.e., $\mathcal{C}(\mathcal{S}) = \cup_{n,m}\mathcal{C}_{n,m}(\mathcal{S})$. Likewise we denote with $\mathcal{M}^{\text{row}}_{a,b}$ the set of $a \times b$ row-stochastic matrices and with $\mathcal{M}^{\text{row}}$ the set of all row-stochastic matrices of finite size. General real-valued matrices of size $a \times b$ we denote with $\mathcal{M}_{a,b}$.

Throughout this Chapter we will at times fix the state space $\mathcal{S}$ to be of a specific type. The usual examples are d-dimensional quantum theory ($\mathcal{Q}_d$) and d-dimensional classical theory ($\mathcal{S}^{cl}_d$). Notably $\mathcal{S}^{cl}_d$ is characterized by the property that every pure state is distinguishable and every mixed state has a unique convex decomposition into pure states. Geometrically it means that a classical state space $\mathcal{S}^{cl}_d$ is equivalent to a $d-1$-simplex [28]. The *operational dimension* d_{op} of a state space $\mathcal{S}$ is always equal to the maximal number of distinguishable states in the given state space. Sometimes we simply state that a state space $\mathcal{S}$ is d-dimensional, by which we mean that the affine hull of $\mathcal{S}$ is d-dimensional.

2.2 Ultraweak matrix majorization

Suppose a communication matrix C has an implementation of the form (34) for some set of physical states and measurements given by s, M. If Alice and Bob have access to the devices that implement C, then there is a natural way to describe all other communication tasks that Alice and Bob can implement with the same devices.

Definition 2.1. Let $C \in \mathcal{M}_{a,b}$ and $D \in \mathcal{M}_{c,d}$. We say that C is *ultraweakly majorized* by D if there exists row-stochastic matrices $L \in \mathcal{M}^{\text{row}}_{a,c}$ and $R \in \mathcal{M}^{\text{row}}_{d,b}$ such that $C = LDR$. In this case we denote the relation between C and D with $C \preceq D$. Whenever $C \preceq D \preceq C$ we say that the matrices C and D are *ultraweakly equivalent* and denote them with $C \simeq D$. However, if $C \npreceq D$ and $D \npreceq C$ we say that C and D are *ultraweakly incomparable*.

Ultraweak matrix majorization was first introduced in the current context in Publication **II** and later refined in Publication **III**, although as a mathematical concept it was known previously as I/O-degradation [55]. The concepts of matrix majorization [56] and weak matrix majorization [57] are closely related to the ultraweak matrix majorization. Matrix majorization corresponds to setting $L = \mathbb{1}$ in Def. 2.1 while weak matrix majorization corresponds to setting $R = \mathbb{1}$ in Def. 2.1. Clearly ultraweak matrix majorization is a weaker concept than either matrix majorization or weak matrix majorization.

Mathematically ultraweak matrix majorization defines a preorder on the set of communication matrices as clearly it is both reflexive and transitive. However, not all communication matrices are comparable in the ultraweak preorder and, moreover, the ultraweak preorder is not antisymmetric as $C \simeq D$ does not imply $C = D$.

The preorder of ultraweak majorization makes precise the sense in which some communication tasks are easier to implement than others. The following example illustrates this concept, which is depicted in Figure 5.

Example 2.2. Suppose Alice has a device that can prepare states $\{s_1, s_2, \ldots, s_n\}$ and Bob has a measurement device implementing a measurement M with the outcome set $\{1, 2, \ldots, m\}$. With the given devices Alice and Bob can implement the communication matrix $D_{ij} = M_j(s_i)$. Suppose now that there is another communication matrix C that is ultraweakly majorized by D, or $C \preceq D$, so that there exists row-stochastic matrices L and R such that $C = LDR$. Alice can now define new states through the following preprocessing on the states:

$$s'_i = \sum_{k=1}^{n} L_{ik} s_k. \tag{35}$$

Likewise, Bob can construct a new measurement M' as the following post-processing of the existing measurement M:

$$M'_j(s) = \sum_{k=1}^{m} R_{kj} M_k(s). \tag{36}$$

Clearly now the pair s', M' implements the communication matrix $C = LDR$.

Example 2.2 has an important consequence, namely that the set of communication matrices is closed with respect to ultraweak matrix majorization. Additionally, Example 2.2 gives a clear physical interpretation for ultraweak majorization. If Alice and Bob can implement the communication matrix D, then the set of all other communication matrices that they can implement with the same devices is those communication matrices that are ultraweakly majorized by D. We can now compare different communication tasks in a completely theory-independent way – a communication task C is easier to implement than communication task D whenever $C \preceq D$. If C and D are ultraweakly incomparable we cannot say which one of them is more

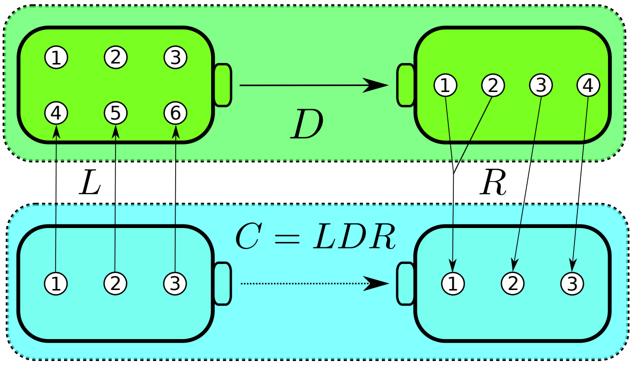

Figure 5. The physical interpretation of ultraweak matrix majorization for communication tasks. The communication task C, implemented by the blue devices, is easier to implement than communication task D.

difficult to implement as their implementation requires a different set of physical devices.

Our goal for the remainder of this Chapter is to characterize ultraweak matrix majorization to the best of our abilities. The following proposition lists some simple conditions for two matrices to be ultraweakly equivalent.

***Proposition* 2.3.** *Two matrices C and D are ultraweakly equivalent in the following cases.*

1. *C is obtained from D as a permutation of the rows and columns of D.*
2. *C is obtained from D by duplicating some of the rows of D.*
3. *C is obtained from D by adding a zero column to D.*
4. *C is obtained from D by adding a row that is a convex mixture of the existing rows of D.*
5. *C is obtained from D by splitting a column into two or more columns according to some convex weights.*

The proof for Prop. 2.3 is very straightforward and is therefore omitted. We can give the following physical interpretations to conditions $1. - 5.$ Notably conditions $4.$ and $5.$ are generalizations of conditions $2.$ and $3.$, respectively. However, we list them explicitly because the physical interpretations are slightly different.

1. A bijective relabeling of the states and outcomes.
2. Mapping a new preparation label into an existing preparation in the preparation device.

3. Adding a new outcome that never occurs into the measurement device.
4. Adding a new preparation that is a mixture of existing preparations into the preparation device.
5. Splitting an existing outcome into several new outcomes according to some convex weights.

As ultraweak equivalence is a symmetric condition, all of the conditions $1. - 5.$ are reversible, e.g. zero columns can be added but they can also be removed.

Clearly each communication matrix defines an equivalence class in the ultraweak preorder, namely the equivalence class of $C \in \mathcal{M}^{\text{row}}$ is $\{A \in \mathcal{M}^{\text{row}} \mid A \simeq C\}$. The ultraweak preorder can be extended to a partial order between such equivalence classes. A natural question of maximal and minimal elements in the partial order then arises. In Publication **II** it was shown that communication matrices with all elements equal form a unique minimal equivalence class. Completely characterizing a set $\mathcal{C}(\mathcal{S})$ amounts to finding all of the ultraweakly maximal equivalence classes.

2.3 Ultraweak monotone functions

To prove that $C \preceq D$, it obviously suffices to find row-stochastic matrices L and R such that $C = LDR$. However, the task of finding suitable matrices L and R might prove to be a difficult task. In the case where $C \npreceq D$ all efforts to find such matrices are in vain. In order to avoid unnecessary work it would be desirable to have mathematical conditions that detect whenever $C \npreceq D$. This motivates the following definition.

Definition 2.4. A function $f : \mathcal{M}^{\text{row}} \to \mathbb{R}$ is an *ultraweak monotone* if $C \preceq D$ implies $f(C) \leq f(D)$ for all matrices $C, D \in \mathcal{M}^{\text{row}}$.

It is noteworthy that the matrices C and D can be of any finite size. Therefore any ultraweak monotone function f can potentially be used to compare any pair of communication matrices[3]. Indeed, whenever $f(C) > f(D)$ for some $C, D \in \mathcal{M}^{\text{row}}$ we can conclude that $C \npreceq D$. Most importantly, whenever $f(C) \neq f(D)$ it follows that $C \not\simeq D$ and we say that the ultraweak monotone f detects the inequivalence of C and D. A single monotone will not detect all inequivalent communication matrices, though. Ultimately the goal would be to have a complete set of monotones so that we would always know if two communication matrices are ultraweakly equivalent, incomparable or if one majorizes the other. It is not known currently if the set of complete monotones is finite.

[3] It would be justified to call these functions functionals, as there is a function for each finite matrix size, or equivalently the ultraweak monotone functions cold be thought of as functions from matrix sizes to functions.

The set of ultraweak monotone functions potentially has some inner structure in the following way.

Definition 2.5. Let f and g be ultraweak monotones and $C, D \in \mathcal{M}^{\mathrm{row}}$. If $f(C) = f(D)$ implies $g(C) = g(D)$ we say that f is *finer* than g.

The relation of Definition 2.5 defines a preorder in the set of ultraweak monotones. Supposing that one monotone is finer than another, it would be tempting to say that we should always use the finer one and ignore the other. While generally speaking this is true, in practice the value of a monotone might be very difficult to calculate. In any case, it was shown in Publication **III** that none of the currently known monotones are finer than any other.

On a final note, before moving on to give actual examples, of special interest is the maximal value of a given ultraweak monotone within a given theory. Later we will see how each known monotone defines a notion of dimension or property that is a characterizing feature of the given theory. This will give us the necessary tools to actually compare different operational theories as well as some intuition on the physical features of each theory.

Rank monotones

The first example of an ultraweak monotone is the rank of a matrix as the rank cannot increase in matrix multiplication. Although rank is perhaps the simplest monotone, it is very useful nonetheless in detecting the "obvious" cases where two communication matrices are inequivalent.

***Proposition* 2.6.** *If $\mathcal{S}$ is a d-dimensional state space the maximal rank of $C \in \mathcal{C}(\mathcal{S})$ equals $d+1$.*

The proof for Proposition 2.6 can be found in Publication **III**. The proof relies in finding $d+1$ affinely independent states and $d+1$ linearly independent effects and showing that the communication matrix constructed this way has maximal rank and that no other communication matrix has a higher rank. The dimension of the

The linear dimension of a theory can now be defined with the rank monotone:

$$d_{lin} := \sup \left\{ \mathrm{rank}(C) \,|\, C \in \mathcal{C}(\mathcal{S}) \right\}. \tag{37}$$

As Proposition 2.6 shows, the supremum is always attained so that if $\mathcal{S}$ is d-dimensional then $d_{lin}(\mathcal{S}) = d+1$.

The identity matrix $\mathbb{1}_k$ is a particularly important communication matrix. The largest identity matrix implementable within a given theory is an extremal element of ultraweak matrix majorization and, in fact, it is the only extremal element for classical theories. Let $C \in \mathcal{M}_{n,m}$ be a nonnegative matrix so that $C_{ij} \geq 0$ for all

i, j. The *nonnegative rank*, denoted rank_+, of C is defined as the smallest integer k such that there exists nonnegative matrices $L \in \mathcal{M}_{n,k}$ and $R \in \mathcal{M}_{k,m}$ so that $C = LR$. Importantly, for row-stochastic matrices $C \in \mathcal{M}^{\text{row}}_{n,m}$ the matrices L and R can be chosen to be stochastic as well [58]:

$$\text{rank}_+(C) \coloneqq \min\left\{k \in \mathbb{N} \,|\, \exists L \in \mathcal{M}^{\text{row}}_{n,k}, R \in \mathcal{M}^{\text{row}}_{k,m} : C = LR\right\}. \tag{38}$$

Alternatively, we can define the nonnegative rank of a row-stochastic matrix with respect to the ultraweak matrix majorization:

$$\text{rank}_+(C) \coloneqq \min\left\{k \in \mathbb{N} \,|\, C \preceq \mathbb{1}_k\right\}. \tag{39}$$

From (39) it is apparent that the nonnegative rank is an ultraweak monotone. We also get the following inequalities from (38):

$$\text{rank}(C) \leq \text{rank}_+(C) \leq \min(n, m). \tag{40}$$

The nonnegative rank allows us to search for the smallest classical system that can be used to implement a given communication matrix.

***Proposition* 2.7.** *The following are equivalent for all $C \in \mathcal{M}^{\text{row}}$:*

(i) $\text{rank}_+(C) \leq n$

(ii) $C \preceq \mathbb{1}_n$

(iii) $C \in \mathcal{C}(\mathcal{S}^{cl}_n)$

The implication $(ii) \Rightarrow (iii)$ follows from the fact that $\mathbb{1}_n \in \mathcal{C}(\mathcal{S}^{cl}_n)$. In the other direction $(iii) \Rightarrow (ii)$ it should be noted that an n-dimensional classical theory has n distinguishable (pure) states $s_1, \ldots s_n$ and all other states have unique convex decompositions into these states. Additionally an n-dimensional classical theory has a unique measurement $M_b(s_a) = \delta_{ab}$ which all other measurements are post-processing of.

If $C \in \mathcal{M}^{\text{row}}_{n,m}$, then obviously $C \preceq \mathbb{1}_n$ and $C \preceq \mathbb{1}_m$. However, the actual nonnegative rank of a matrix is highly nontrivial to calculate. Determining whether $\text{rank}_+(C) = \text{rank}(C)$ for a general nonnegative matrix C has been proven to be NP-hard [59]. The following simple conditions hold for small matrices [58]. If $A \in \mathcal{M}_{n,m}$ is nonnegative, then $\text{rank}(A) = \text{rank}_+(A)$ if:

(i) $\text{rank}(A) \leq 2$

(ii) $n \in \{1,2,3\}$ or $m \in \{1,2,3\}$.

Even if the nonnegative rank is difficult to calculate, the corresponding ultraweak dimension is of crucial importance as the classical dimension of an operational theory can be defined with it:

$$d_{cl} := \inf\left\{k \in \mathbb{N} \,|\, \mathcal{C}(\mathcal{S}) \subseteq \mathcal{C}(\mathcal{S}_k^{cl})\right\}. \tag{41}$$

Alternatively, as a consequence of Proposition 2.7, the classical dimension can be defined as:

$$\begin{aligned} d_{cl} &= \inf\{k \in \mathbb{N} \,|\, \forall C \in \mathcal{C}(\mathcal{S}) : C \preceq \mathbb{1}_k\} \\ &= \inf\{k \in \mathbb{N} \,|\, \forall C \in \mathcal{C}(\mathcal{S}) : \text{rank}_+(C) \leq k\} \\ &= \sup\{\text{rank}_+(C) \,|\, C \in \mathcal{C}(\mathcal{S})\}. \end{aligned}$$

Whenever the classical dimension of a theory is finite, the supremum in the last expression is attained.

Analogously to the nonnegative rank, we can define the *positive semidefinite* (PSD) rank, denoted rank_{psd}, of a nonnegative matrix $C \in \mathcal{M}_{n,m}$ to be the smallest integer k such that there exists positive semidefinite $k \times k$ matrices $A_1, \dots A_n$ and $B_1, \dots B_m$ such that $C_{ij} = \text{tr}\,[A_i B_j]$. The factorization $C_{ij} = \text{tr}\,[A_i B_j]$ is called a *positive semidefinite decomposition* of C. For quantum implementations this already looks promising and, in fact, in [60] it was proved that the PSD rank of a communication matrix equals the minimal quantum dimension required to implement the given communication matrix.

***Proposition* 2.8.** *[60, Lemma 5] Let $C \in \mathcal{M}^{\text{row}}$. Then $C \in \mathcal{C}(\mathcal{Q}_d)$ if and only if $\text{rank}_{psd}(C) \leq d$.*

Intuitively it is clear that the PSD rank is an ultraweak monotone. The following proposition proves this for the sake of completeness.

***Proposition* 2.9.** *The positive semidefinite rank is an ultraweak monotone.*

Proof. Let $C \in \mathcal{M}_{a,b}^{\text{row}}, D \in \mathcal{M}_{n,m}^{\text{row}}$ and $C \preceq D$ so that there exists row-stochastic matrices $L \in \mathcal{M}_{a,n}^{\text{row}}, R \in \mathcal{M}_{m,b}^{\text{row}}$ such that $C = LDR$. Suppose that $\text{rank}_{psd}(D) = d$. If $D_{ij} = \text{tr}\,[A_i B_j]$ where $\{A_i\}, \{B_j\}$ are $d \times d$ positive semidefinite matrices, then

$$\begin{aligned} C_{kl} &= \sum_{i=1}^{n}\sum_{j=1}^{m} L_{ki} D_{ij} R_{jl} = \sum_{i=1}^{n}\sum_{j=1}^{m} L_{ki} \text{tr}\,[A_i B_j]\, R_{jl} \\ &= \text{tr}\left[\left(\sum_{i=1}^{n} L_{ki} A_i\right)\left(\sum_{j=1}^{m} R_{jl} B_j\right)\right] = \text{tr}\left[A_i' B_j'\right]. \end{aligned}$$

The PSD decomposition for C with $A'_i = \sum_{i=1}^{n} L_{ki} A_i$ and $B'_j = \sum_{j=1}^{m} R_{jl} B_j$ proves that $\mathrm{rank}_{psd}(C) \leq d$. □

In general it is known that

$$\sqrt{\mathrm{rank}(C)} \leq \mathrm{rank}_{psd}(C) \leq \mathrm{rank}_{+}(C) \tag{42}$$

for all nonnegative matrices [61]. The second inequality follows from the extra requirement that the PSD factorization be diagonal, while the first inequality is easy to confirm for communication matrices with the help of Proposition 2.6.

Just like the nonnegative rank, the PSD rank is very hard to compute, NP-hard in fact [62]. As an ultraweak monotone the maximal PSD rank within a theory defines the quantum dimension of that theory, i.e.,

$$d_q = \inf \{ d \in \mathbb{N} \mid \forall C \in \mathcal{C}(\mathcal{S}) : \mathrm{rank}_{psd}(C) \leq d \} . \tag{43}$$

Alternatively, $d_q = \sup \{ \mathrm{rank}_{psd}(C) \mid C \in \mathcal{C}(\mathcal{S}) \}$, where the supremum is attained for some C if the quantum dimension is finite.

Discrimination monotones

The remaining examples of ultraweak monotones are all related to some form of discrimination tasks.

The maximal number of distinguishable states is a fundamental property of all state spaces. It tells directly how many distinguishable messages can be sent. If a state from a set of d distinguishable states is sent, then (at most) $\log_2 d$ bits of information can be retrieved by measuring that state. Recall that the Basic Decoding Theorem [63] states that whenever there are more messages than distinguishable states, the error in this kind of communication is at least $1 - \frac{d}{n}$, where n is the number of messages and d the operational dimension of the state space. Motivated by these facts we define the following function for $C \in \mathcal{M}^{\mathrm{row}}$:

$$\iota(C) := \max \{ n \in \mathbb{N} \mid \mathbb{1}_n \preceq C \} . \tag{44}$$

It is clear that ι is an ultraweak monotone, as it equals the size of the largest identity matrix that is ultraweakly majorized by C. We simply call it the distinguishability monotone.

***Proposition* 2.10.** *Let $C \in \mathcal{M}^{\mathrm{row}}$. Then $\iota(C) = k$ if and only if the maximal number of orthogonal rows of C equals k.*

The proof of the previous Proposition was presented in Publication **III**. Clearly the minimum value of ι within any $\mathcal{C}(\mathcal{S})$ is 1 and the maximum value is $d_{op}(\mathcal{S})$.

The distinguishability monotone is related to entropy in the following way. Suppose Alice and Bob have devices that implement $C \in \mathcal{M}^{\text{row}}$ and $\iota(C) = k$. Then Alice and Bob can actually physically implement the communication matrix $\mathbb{1}_k$, i.e., Alice can send k distinct messages to Bob. If Alice prepares the k distinguishable states with equal relative frequencies, then the amount of information that Alice can transmit to Bob in a single run of the devices is at least $\log_2 k$ bits.

The last two monotones are very similar to each other. They are introduced in the following Proposition.

***Proposition* 2.11.** *Let*

$$\lambda_{max}(C) := \sum_j \max_i C_{ij}, \quad \lambda_{min}(C) := -\sum_j \min_i C_{ij}. \tag{45}$$

The functions λ_{max} and λ_{min} are ultraweak monotones.

The proof for Proposition 2.11 is straightforward. We call these monotones simply the *min and max* monotones. The function λ_{max} is related to minimum error discrimination tasks in the following way. First of all, let us set

$$\lambda_{max}(\mathcal{S}) := \sup \left\{ \lambda_{max}(C) \mid C \in \mathcal{C}(\mathcal{S}) \right\}. \tag{46}$$

Alternatively, if we denote by $\mathcal{O}(\mathcal{S})$ the set of all measurements in $\mathcal{S}$ and with $\Omega_M \subset \mathbb{N}$ the finite outcome set of $M \in \mathcal{O}(\mathcal{S})$, we can define $\lambda_{max}(M) := \sum_{j\in\Omega_M} \max_{s\in\mathcal{S}} M_j(s)$ and $\lambda_{max}(\mathcal{S}) = \sup_{M\in\mathcal{O}(\mathcal{S})} \lambda_{max}(M)$. The maximal value of λ_{max} for a theory determined by $\mathcal{S}$ is sometimes called the information storability of the state space $\mathcal{S}$ [64; 65]. Additionally, let $\lambda_{max,n}(\mathcal{S})$ denote the supremum of λ_{max} over measurements with the number of outcomes fixed as $\#\Omega_M \leq n$. Then the error probability in a minimum error discrimination task between Alice and Bob, assuming equal a priori probabilities for n preparations, is bounded by

$$P^n_{error} \geq 1 - \frac{\lambda_{max,n}(\mathcal{S})}{n}.$$

The term $\frac{\lambda_{max}(\mathcal{S})}{n}$ can be interpreted as the maximal decoding power of all measurements in $\mathcal{S}$, as clearly $\lambda_{max}(\mathcal{S}) = \sup_{n\in\mathbb{N}} \lambda_{max,n}(\mathcal{S})$. The decoding power of a measurement is typically associated with noise robustness, i.e., the maximal amount of noise the measurement tolerates while not becoming a trivial observable [66; 67; 68].

For $C \in \mathcal{C}(\mathcal{Q}_d)$ we have a specific form for the max monotone:

$$\lambda_{max}(C) = \sum_j \max_i \text{tr}\left[\varrho_i \mathsf{M}(j)\right] \leq \sum_j \text{tr}\left[\mathsf{M}(j)\right] = d,$$

where d is the operational dimension $\mathcal{Q}_d$. Hence we obtain the Basic Decoding Theorem from the maximal value of λ_{max} within $\mathcal{Q}_d$: $\lambda_{max}(\mathcal{Q}_d) \leq d = d_{op}(\mathcal{Q}_d)$ and $P^n_{error} \geq 1 - \frac{d}{n}$.

Remark 2.12. In quantum theory the information storability of the state space is the same as the operational dimension. However, in general this is not true. In [64] it was shown that the information storability of a state space is related to the point-symmetry of $\mathcal{S}$. More specifically, the information storability of $\mathcal{S}$ can exceed the operational dimension if the state space is point-asymmetric, such as in some polygon state spaces [69]. Furthermore, it was shown in [64] that the information storability of a state space gives an upper bound on the classical channel capacity. Thus a generalized version of the Holevo bound is required in general probabilistic theories [70; 71].

The min monotone is just as easy to compute as the max monotone. However, calculating the minimal and maximal values of λ_{min} within a theory is not very useful. Clearly $\lambda_{min}(C) \in [-1, 0]$ for all $C \in \mathcal{M}^{\mathrm{row}}$. The minimum value -1 is obtained by any C with all entries equal, while the maximum value of 0 is obtained by any identity matrix, for instance. The physical interpretation for the min monotone is the following: the probability of observing the least probable measurement events cannot decrease in ultraweak majorization. This is somehow contrary to the max monotone, whose physical interpretation is that the decoding power of measurements cannot increase in ultraweak majorization, i.e., the probability of observing the most probable measurement events cannot increase.

2.4 Dimensions induced by the monotones

To conclude this Chapter we will collect all of the dimensions induced by the ultraweak monotones and compare them. Let us first recall all of the relations between the different monotones.

***Proposition* 2.13.** *The following holds for all $C \in \mathcal{M}^{\mathrm{row}}$:*

$$\sqrt{\mathrm{rank}(C)} \leq \mathrm{rank}_{psd}(C), \quad \mathrm{rank}(C) \leq \mathrm{rank}_{+}(C). \tag{47}$$

Further,

$$\iota(C) \leq \lambda_{max}(C) \leq \mathrm{rank}_{psd}(C) \leq \mathrm{rank}_{+}(C). \tag{48}$$

The inequality $\iota(C) \leq \lambda_{max}(C)$ follows from the fact that $\mathbb{1}_{\iota(C)} \preceq C$ and $\lambda_{max}\left(\mathbb{1}_{\iota(C)}\right) = \iota(C)$.

Taking the supremum within $\mathcal{C}(\mathcal{S})$ we find that, in general,

$$d_{op}(\mathcal{S}) \leq \lambda_{max}\left(\mathcal{S}\right) \leq d_q(\mathcal{S}) \leq d_{cl}(\mathcal{S}). \tag{49}$$

Additionally,

$$d_{lin}(\mathcal{S}) \leq d_q(\mathcal{S})^2, \quad d_{lin}(\mathcal{S}) \leq d_{cl}(\mathcal{S}) \tag{50}$$

An important remark is that, for $\mathcal{S}_d^{cl}$, all of these values coincide. Whether these coinciding values are enough to uniquely fix the classical theory remains an open question, one which we could present as a conjecture here. However, it can be confirmed that it is a necessary condition for all of these dimension to be equal for a state space to be classical.

In contrast to classical theory, for d-dimensional quantum theory $\mathcal{Q}_d$ we have that $d_{lin}(\mathcal{Q}_d) = d^2 \leq d_{cl}(\mathcal{Q}_d)$ while $d_{op}(\mathcal{Q}_d) = d_q(\mathcal{Q}_d) = \lambda_{max}\left(\mathcal{Q}_d\right) = d$. It follows that $\mathcal{C}(\mathcal{Q}_d)$ must have other ultraweakly maximal elements than $\mathbb{1}_d$. In Publication **III** it was conjectured that $d_{cl}(\mathcal{Q}_d) = d^2$. However, to the best of our knowledge, this conjecture remains open.

In conclusion, the dimensions induced by the different ultraweak monotones offer a convenient way to compare different operational theories. The ultraweak monotones themselves can be used to check whether two communication matrices are ultraweakly inequivalent. Moreover, the monotones can be used to check whether a communication matrix can be implemented in a given theory. It is not know whether a complete set of monotones has finite cardinality.

3 Communication tasks in quantum theory

The original Publications **I**, **II** and **IV** all considered very specific communication tasks, namely antidistinguishability, communication of partial ignorance and partial-ignorance communication tasks, respectively. All of these communication tasks can be classified as some kind of "exclusion" tasks. Notably antidistinguishability and communication of partial ignorance are communication tasks with a single measurement device so the entirety of Chapter 2 applies. On the other hand, in partial-ignorance communication tasks, the major discrepancy to the other two tasks is that the measurement device is not fixed but can be changed according to some additional input, so the use of behaviors is required. As such the partial-ignorance communication tasks cannot be completely analyzed by simply looking at individual communication matrices.

In this Chapter a general introduction to the previously mentioned three types of communication tasks is given. Previously known and some new results from the literature are highlighted where applicable. Specifically the section dealing with antidistinguishability attempts to provide an up-to-date overview on the topic as antidistinguishability has attracted some attention in the research community in recent years.

3.1 Antidistinguishability

Antidistinguishability, also known as conclusive exclusion [72] and post-Peierls incompatibility [73] in the literature, is simply a property of a set of states.

Definition 3.1. A set $\{\varrho_i\}_{i=1}^n$ of n quantum states is called *antidistinguishable* if there exists an n-outcome POVM M such that

$$\mathrm{tr}\left[\varrho_i \mathsf{M}(i)\right] = 0 \tag{51}$$

for all $i \in \{1, 2, \ldots, n\}$. Additionally it is required that $\sum_i \mathrm{tr}\left[\varrho_i \mathsf{M}(j)\right] \neq 0$ for all $j \in \{1, 2, \ldots, n\}$ so that there does not exist outcomes that never occur.

The additional requirement that none of the outcomes occur with zero probability is a necessary addition. Otherwise some of the states could be "antidistinguished" by the zero effect while the entire set of states might not be antidistinguishable as defined in the previous definition.

While Definition 3.1 may look conspicuously simple, there is an intriguing reason to study antidistinguishability in foundations of quantum mechanics. Supposing that there exists a complete description of nature in the form of an ontological model (see e.g. [43; 74; 75; 76; 77; 78]), whenever a set of quantum states is antidistinguishable the underlying epistemic states are known to have the property that the union of their supports has measure zero [50]. Applied to the case where two identical preparation devices operate independently, this result yields the (in)famous Pusey–Barrett–Rudolph theorem [79]. Additionally, antidistinguishability can be used to restrict the amount of pair-wise overlap epistemic states can display [80; 81; 82; 83]. Thus the study of antidistinguishability is well motivated within foundations of quantum mechanics.

The main goal of Publication **I** was to find algebraic conditions for a set of (pure) quantum states to be antidistinguishable according to Definition 3.1. Before presenting any such conditions, it must be noted that the problem of finding an antidistinguishing POVM for a set of quantum states is actually a *semidefinite program* (SDP) (see e.g. [84; 85; 86; 87] for usage of SDPs in a quantum mechanical setting). SDPs are a class of optimization problems which can be solved almost as efficiently as linear programs [88; 89]. The SDP for deciding whether a set of quantum states $\{\varrho_i\}_{i=1}^n$ is antidistinguishable is presented below.

$$\begin{aligned} &\min \sum_{i=1}^{n} \mathrm{tr}\left[\varrho_i \mathsf{M}(i)\right] \\ \text{s.t.} \quad & \mathsf{M}(j) \geq 0 \, \forall j \in \{1, 2, \ldots, n\} \\ &\sum_{i=1}^{n} \mathsf{M}(i) = \mathbb{1} \end{aligned} \tag{52}$$

Whenever a numerical solver finds the minimum for the objective function of Equation (52) to be 0, the corresponding set of quantum states is surely antidistinguishable, at least within numerical accuracy. However, it should be checked that the solution does not contain any outcomes that never occur. This could also be taken to be a part of the problem definition, e.g. by requiring that $\sum_{i=1}^{n} \mathrm{tr}\left[\varrho_i \mathsf{M}(j)\right] \geq \epsilon$ for all j and for some suitably chosen $\epsilon > 0$.

While numerically solving a problem of form (52) can be considered easy, at least when the cardinality of the set of states is small and the corresponding Hilbert space has low dimension, there are a few shortcomings with this methodology. First of all, most numerical methods rarely give exact results, i.e., the results are prone to small numerical imperfections. The degree to which this is a problem depends on the situation. Secondly, a mere numerical solution to a problem scarcely gives any intuition on the physical phenomena behind the problem. While it may be possible to gain some intuition by formulating the dual problem of the SDP and considering weak and strong forms of duality, an analytic solution is often more descriptive of the original problem.

Conditions for antidistinguishability

To the best of our knowledge, there currently only exists one necessary and sufficient formulation for an arbitrary set of pure quantum states to be antidistinguishable (in addition to the numerical formulation as an SDP). This general criterion was presented in Publication **I**. While the general condition is indeed necessary and sufficient, unfortunately it is not explicit in the sense that in most cases it is not possible to construct the antidistinguishing POVM with the help of the criterion. Therefore a complete solution for a set of quantum states to be antidistinguishable is only known for qubits. In the following proposition *post-Peierls incompatibility* is used as a synonym for antidistinguishability.

***Proposition* 3.2.** *A set of n pure qubit states $\{\varrho_i\}_{i=1}^n$ is post-Peierls incompatible if and only if the maximally mixed state $\frac{1}{2}\mathbb{1}$ belongs to the convex hull of $\{\varrho_i\}_{i=1}^n$.*

Proposition 3.2 was first presented in [73]. However, the previous proposition has a major flaw. Namely, it allows for some of the effects in the antidistinguishing POVM to be zero effects. It is unclear whether the authors intended to allow for this. In their writing they specify the effects should be rank-1. Nonetheless they explicitly allow for the weights of some effects to equal zero, which would contradict our definition for antidistinguishability. This can be rectified by requiring that the maximally mixed state be an interior point of the convex set generated by the pure states. Thus we arrive at an alternative characterization of antidistinguishability for qubits, which was presented in Publication **I**.

***Proposition* 3.3.** *A set of n pure qubit states $\{\varrho_i\}_{i=1}^n$ is antidistinguishable if and only if there exist positive real numbers $t_i > 0$ such that*

$$\sum_{i=1}^{n} t_i \varrho_i = \mathbb{1}. \tag{53}$$

The proof for Proposition 3.3 essentially relies on the fact that a pure qubit state has a unique pure state that is orthogonal to it. This is also the reason the proof method only works for qubits – when $d \geq 3$ there no longer is an unique antipodal state, but rather an orthogonal subspace of the state space. It is unclear if it is possible to always choose one state from this subspace such that the original proof method works. However, Proposition 3.3 has the following generalization.

***Proposition* 3.4.** *A set of n pure quantum states $\{\varrho_i\}_{i=1}^n$ is antidistinguishable if there exist positive real numbers $t_i > 0$ such that*

$$\sum_{i=1}^{n} t_i \varrho_i = R, \tag{54}$$

where R is a projection.

While Proposition 3.4 is fully general in that it applies in all dimensions and to any number of states, unfortunately it is no longer a necessary condition.

As a last consideration on algebraic conditions for antidistinguishability in Publication **I** we discussed a fully general criterion that is also sufficient and necessary. The general criterion can be outlined as follows. Suppose a set of n pure quantum states $\{\varrho_i\}_{i=1}^n$ is antidistinguishable. Then there exists an antidistinguishing POVM M such that $\sum_{i=1}^n \mathrm{tr}\left[\varrho_i \mathsf{M}(j)\right] > 0$ for all j and $\mathrm{tr}\left[\varrho_i \mathsf{M}(i)\right] = 0$. As $\mathrm{tr}\left[\varrho_i \mathsf{M}(i)\right] = 0$ is equivalent with $\varrho_i \mathsf{M}(i) = 0$ and ϱ is a pure state, it follows that the spectral decomposition for $\mathsf{M}(i)$ must not contain a projection that is equal to ϱ_i. Therefore we must be able to write $\mathsf{M}(i)$ as

$$\mathsf{M}(i) = \sum_{k=2}^{d} \alpha_i^k P_i^k,$$

where P_i^k are pair-wise orthogonal one-dimensional projections, each of which are orthogonal to ϱ_i, and d equals the dimension of the Hilbert space in which the states are defined. Additionally, it is required that none of the outcomes occur with probability zero, so that

$$\sum_{i=1}^{n} \sum_{k=2}^{d} \alpha_j^k \mathrm{tr}\left[\varrho_i P_j^k\right] \neq 0,$$

as $\sum_{i=1}^n \mathrm{tr}\left[\varrho_i \mathsf{M}(j)\right]$ must not equal zero.

While it is clear that the method outlined above must always work whenever a set of pure states is antidistinguishable, unfortunately the method does not offer any way to reliably construct the antidistinguishing POVM. It is only guaranteed that such a POVM is possible to construct in this way.

Apart from previously known results of [73], which contained the original qubit results and some results regarding sets of three or less quantum states, and those of [72] which presented a result based on pair-wise fidelities of quantum states, there was another recent paper that contained an interesting conjecture [90].

***Conjecture* 3.5.** *Let $\{|\varphi_i\rangle\}_{i=1}^d$ be the vectors of d pure quantum states. If*

$$|\langle \varphi_i | \varphi_j \rangle| \leq \frac{d-2}{d-1} \tag{55}$$

for all $i \neq j$, then the set $\{|\varphi_i\rangle\}_{i=1}^d$ is antidistinguishable.

While [90] presented convincing arguments in addition to numerical evidence for Conjecture 3.5, what makes the conjecture truly intriguing is that very recently a counterexample was provided as a preprint on arXiv [91]. Let us analyze this counterexample.

First let us define the dual problem to the SDP defined in Equation (52) as:

$$\begin{aligned} &\max \operatorname{tr}[Y] \\ \text{s.t.} \quad & Y \leq \varrho_i \, \forall i \in \{1, 2, \ldots, n\}, \end{aligned} \tag{56}$$

where the variable Y is understood to be Hermitian. In [72] it was proved that there is a strong duality between the dual and primal problems defined in Equations (52) and (56). That is, the value that is obtained for both of those problems should always coincide. Therefore we can present the following numerical characterization for antidistinguishability (adapted from [72] and [91]).

***Lemma* 3.6.** *Let $\{\varrho_i\}_{i=1}^n$ be a set of quantum states (not necessarily pure). Then $\{\varrho_i\}_{i=1}^n$ is not antidistinguishable if and only if there exists a Hermitian matrix Y such that* $\operatorname{tr}[Y] > 0$ *and* $\varrho_i - Y \geq 0$ *for all* $i \in \{1, 2, \ldots, n\}$.

Example 3.7 (Counterexample from [91]). Let us define four states in $\mathbb{C}^4$:

$$\varphi_1 = \begin{bmatrix} 0.50127198 - 0.037607i \\ -0.00698152 - 0.590973i \\ 0.08186514 - 0.4497548i \\ -0.01299883 + 0.43458491i \end{bmatrix}, \varphi_2 = \begin{bmatrix} -0.07115345 - 0.27080326i \\ 0.82047712 + 0.26320823i \\ 0.22105089 - 0.2091996i \\ -0.23575591 - 0.1758769i \end{bmatrix}$$

$$\varphi_3 = \begin{bmatrix} 0.31360906 + 0.46339313i \\ -0.0465825 - 0.47825017i \\ -0.10470394 - 0.11776404i \\ 0.60231515 + 0.26154959i \end{bmatrix}, \varphi_4 = \begin{bmatrix} -0.53532122 - 0.03654632i \\ 0.40955941 - 0.15150576i \\ -0.05741386 + 0.23873985i \\ -0.4737113 - 0.48652564i \end{bmatrix}$$

Then we indeed find that $\max_{i \neq j} |\langle \varphi_i | \varphi_j \rangle| \approx 0.64514235 < \frac{4-2}{4-1} = \frac{2}{3}$. However, solving the dual SDP of (56) we find that $\max_{Y \in \text{Herm.}} \operatorname{tr}[Y] \approx 0.00039382$ so the four pure states are not antidistinguishable[1].

It is stated in [91] that the counterexample was found after generating a considerable amount of random quantum states. The power of numerical methods here is that they were able to prove that Conjecture 3.5 does not hold. The counterexample found in this way, however, does not seem to give us much insight on *why* the conjecture fails.

[1] The value presented here was obtained using the same libraries as in [91], namely `cvxopt`, `picos` and `numpy` in `Python`. However, we were using slightly newer versions of the libraries. The result we got matches the one in [91] with 7 decimals, so the results are in good agreement in any regard.

The main result of [90] was to show that there exists a communication complexity separation between classical and quantum communication that is based on antidistinguishability. Their result is somewhat weakened by the failure of their conjecture. Nonetheless their main result still applies. The literature contains other examples of applications for antidistinguishability as well. Namely, antidistinguishability can also be used to generate noncontextuality inequalities [92]. We mention this here for the sake of completeness, but the connection between antidistinguishability and contextuality, although intriguing, was not a topic in the original Publications. Hence we will not develop this topic further here.

Generating antidistinguishable sets of states

In the literature, such as [90; 91], a typical method of generating sets of antidistinguishable states is to generate a random sample of states according to some measure, typically the Haar measure, and then checking whether the generated set is antidistinguishable or not with numerical methods. While this method may work well enough in practice (and considering what the goal is this might be exactly what is desired), there are better ways of generating antidistinguishable sets of states if the goal involves avoiding generating sets that are not antidistinguishable.

As was discussed in Publication **I**, a simple method of generating a set of antidistinguishable states is to consider a finite group G and its irreducible unitary representations $U : g \mapsto U_g$ on a Hilbert space $\mathbb{C}^d$. Starting from a pure state P, the orbit of P under G, i.e. $\{P_g = U_g P U_g^* : g \in G\}$ where P is not a fixed point, is guaranteed to be antidistinguishable by Schur's lemma as long as $P_g \neq P_h$ for $g \neq h$. This follows from the fact that $\sum_h P_h$ is easily seen to commute with every U_g, hence $\sum_h P_h$ is proportional to $\mathbb{1}_d$ and therefore Proposition 3.4 applies. Even if $P_g = P_h$ for some $g \neq h$ the above method still works as the distinct elements of the orbit of P can be labeled by the elements of the quotient group $G/\mathcal{I}(P)$, where $\mathcal{I}(P) = \{h \in G \,|\, U_h P U_h^* = P\}$. The antidistinguishing POVM is given by

$$\mathsf{M}(g) = \frac{d}{\#G(d-1)} U_g P^{\perp} U_g^*,$$

where $P^{\perp} = \mathbb{1} - P$, in the case where all of P_g are different. Otherwise an extra factor of $\#\mathcal{I}(P)$ has to be included and the effects are then labeled by elements in the quotient group $G/\mathcal{I}(P)$.

Another way to generate antidistinguishable states is to start by generating a random POVM, provided the generated POVM does not contain full-rank effects. The theory of random POVMs is well-developed [93]. Each effect in the random POVM defines at least one state that it antidistinguishes. A suitable state can then be chosen from the kernel of the effect by any desired means.

There is a special class of POVMs that can be used to generate antidistinguishable sets of states. Namely, a *symmetric informationally complete* (SIC) POVM is defined as a POVM M with d^2 outcomes such that $\operatorname{tr}[\mathsf{M}(j)]$ is a constant for all j and $\operatorname{tr}[\mathsf{M}(i)\mathsf{M}(j)]$ is a constant for all $i \neq j$. It turns out these constants are fixed by the dimension d, i.e., for $j \neq k$:

$$\operatorname{tr}[\mathsf{M}(j)] = \frac{1}{d}, \quad \operatorname{tr}[\mathsf{M}(j)\mathsf{M}(k)] = \frac{1}{d^2(d+1)}.$$

Additionally $\operatorname{tr}\left[\mathsf{M}(j)^2\right] = \frac{1}{d^2}$ for all j. SIC POVMs are known to exist for up to $d = 151$ [94]. A SIC POVM can always be used to define an antidistinguishable set of states by setting

$$\varrho_i = \frac{1}{d-1}(\mathbb{1} - d\mathsf{M}(i)). \tag{57}$$

It is then straightforward to check that $\operatorname{tr}[\varrho_i \mathsf{M}(i)] = 0$, $\operatorname{tr}[\varrho_i \mathsf{M}(j)] = \frac{1}{d^2-1}$, $\operatorname{tr}[\varrho_i] = 1$ and $\varrho_i \geq 0$.

Uniform antidistinguishability as a communication task

There is a particular class of communication matrices that rely on a stronger form of antidistinguishability. Let us define a class of $n \times n$ communication matrices A_n of the form

$$A_n = \frac{1}{n-1} \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 0 & & 1 \\ \vdots & & & \ddots & \\ 1 & \cdots & \cdots & 1 & 0 \end{bmatrix}. \tag{58}$$

A set of n states are called *uniformly antidistinguishable* if they can be used to implement A_n. It is easy to verify that $\operatorname{rank}(A_n) = n$ so that $A_n \preceq \mathbb{1}_n$ as A_n is a square matrix and hence also $\operatorname{rank}_+(A_n) = n$. Clearly a SIC POVM and states defined as in (57) always implement A_n for $n = d^2$, as long as a SIC POVM exists in dimension d.

Uniform antidistinguishability was used in Publication **II** to solve a specific type of communication game (which we will define in the next section). It is also interesting to note that in some sense the communication matrix A_n is 'as far away' from $\mathbb{1}_n$ as possible. Where one has only ones on the diagonal and zeros everywhere else, the other has zeros on the diagonal and equal non-zero off-diagonal elements. We can thus present the following family of $n \times n$ communication matrices:

$$D_{n,p} = \begin{bmatrix} 1-p & \frac{p}{n-1} & \frac{p}{n-1} & \cdots & \frac{p}{n-1} \\ \frac{p}{n-1} & 1-p & \frac{p}{n-1} & \cdots & \frac{p}{n-1} \\ \frac{p}{n-1} & \frac{p}{n-1} & 1-p & & \frac{p}{n-1} \\ \vdots & & & \ddots & \\ \frac{p}{n-1} & \cdots & \cdots & \frac{p}{n-1} & 1-p \end{bmatrix}. \tag{59}$$

Clearly $D_{n,1} = A_n$ and $D_{n,0} = \mathbb{1}_n$. The communication task defined by $D_{n,p}$ is related to either minimum error discrimination tasks when $0 \leq p \leq 1 - \frac{1}{n}$ or to some noisy uniform antidistinguishability task when $1 - \frac{1}{n} \leq p \leq 1$. The limiting value $p = 1 - \frac{1}{n}$ defines the $n \times n$ communication matrix with all elements equal to $\frac{1}{n}$, which was established to be a minimal element in the ultraweak preorder in Chapter 2. This leads us to ponder if there exists some order within the family $D_{n,p}$ defined by the ultraweak matrix majorization. This is indeed the case, and in Publication **III** the following was shown:

(i) The noiseless uniform antidistinguishability matrix $D_{n,1} = A_n$ is ultraweakly majorized only by $D_{n,0} = \mathbb{1}_n$.

(ii) If $D_{n,p}$ relates to distinguishability, i.e. $p \in \left[0, 1 - \frac{1}{n}\right]$, then $D_{n,p} \succeq D_{n,q}$ if and only if $q \in \left[p, 1 - \frac{p}{n-1}\right]$.

(iii) If $D_{n,p}$ relates to antidistinguishability, i.e. $p \in \left[1 - \frac{1}{n}, 1\right]$, then $D_{n,p} \succeq D_{n,q}$ if and only if $q \in \left[1 - \frac{p}{n-1}, p\right]$.

Interestingly it is possible to cross the limiting value of pure noise $1-\frac{1}{n}$ from both sides. However, it is not possible to get back to the original communication matrix by crossing the limit value twice. Publication **III** contains a graphical illustration of the ultraweak relations within the family $D_{n,p}$.

As a last consideration on antidistinguishability we may wonder what the PSD rank of the matrices A_n are, but currently the answer is not completely known. In Publication **II** it was shown that there cannot be more than d^2 uniformly antidistinguishable states, and it is known that $\text{rank}_{psd}(A_{d^2}) = d$ when d is odd [60]. When d is even it is known that $A_{d^2-1} = d$. Whether $A_{d^2} = d$ for even d is thus not known, but is surely either d or $d+1$. This can be seen in a straightforward way. Suppose that $\text{rank}_{psd}(A_n) = k$. Then the quantum implementation for A_n gives a PSD decomposition for A_n. If we drop one state and effect from this decomposition, we obtain a PSD decomposition for A_{n-1} which is of the same size as the decomposition for A_n. Therefore, by Proposition 2.8 $\text{rank}_{psd}(A_{n-1}) \leq k$.

Lemma 5 in [60] details a method for normalizing the states and effects of the new PSD decomposition into an actual POVM and new states that implement A_{n-1}.

Interestingly, this normalization cannot be achieved through the ultraweak matrix majorization. This is because $\lambda_{max}(A_n) = \frac{n}{n-1}$ and so $\lambda_{max}(A_n)$ decreases as n grows, i.e., $A_{n-1} \not\succ A_n$. This means that A_{n-1} and A_n are ultraweakly incomparable as the rank and λ_{max} monotones can be used to establish their inequivalence.

Remark 3.8. In [95] it was shown that $\bar{\mathcal{C}}(\mathcal{Q}_d) = \bar{\mathcal{C}}(\mathcal{S}_d^{cl})$, i.e., that the convex hull of communication matrices implementable in d-dimensional classical and quantum systems are the same. This means that, with the additional resource of shared randomness, there is no advantage to use quantum systems in favor of classical systems in simple prepare-and-measure scenarios. It is not difficult to see that the set $\mathcal{C}(\mathcal{Q}_d)$ is not convex. Therefore shared randomness is truly an additional and powerful resource in communication. This is exemplified in the following proposition and examples.

***Proposition* 3.9.** *Let $B \in \mathcal{M}_{n,n}^{\mathrm{row}}$ be any $n \times n$ communication matrix with only zeros on the diagonal. Then there exists a set of $n!$ communication matrices, each of which is ultraweakly majorized by B, whose convex sum with equal weights equals A_n.*

Proof. Let $\{P_i\}_{i=1}^{n!}$ be the set of all permutation matrices of n elements. In general there are $n!$ such permutation matrices. Let us define $C = \frac{1}{n!}\sum_{k=1}^{n!} P_i B P_i^T$, where $P_i B P_i^T$ is the communication matrix whose rows and columns are relabeled according to permutation i. Note that $(P_i B P_i^T)_{jj} = 0$ for all $i \in \{1, 2, \ldots, n!\}$ and $j \in \{1, 2, \ldots, n\}$. Therefore $C_{ii} = 0$ for all $i \in \{1, 2, \ldots, n\}$.

Let us then calculate C_{ij} for $i \neq j$. Clearly there are $(n-2)!$ permutations that leave i and j unchanged. Likewise, for each integer $i' \neq i$ and $j' \neq j$ such that $i' \neq j'$, there are $(n-2)!$ permutations that map $i \mapsto i'$ and $j \mapsto j'$. Therefore

$$C_{ij} = \frac{1}{n!}(n-2)! \sum_{\substack{k,l=1 \\ k \neq l}}^{n} B_{kl} = \frac{(n-2)!}{n!} n = \frac{1}{n-1}.$$

Hence $C = A_n$. □

Proposition 3.9 shows the power of shared randomness, as it becomes possible to implement any A_n with just qubits (or even the bit for that matter). Particularly noteworthy is that $\mathrm{rank}(A_n) = n$, so with shared randomness the rank of implementable communication matrices is no longer bounded within $\bar{\mathcal{C}}(\mathcal{Q}_2)$. Likewise the dimensions $d_q(\mathcal{S})$ and $d_{cl}(\mathcal{S})$ no longer make sense as there no longer exists a finite upper bound for the corresponding monotones within the set $\bar{\mathcal{C}}(\mathcal{Q}_2)$.

Example 3.10 (A_3 with the bit)**.** The communication matrices implementable with the bit are precisely those that are ultraweakly majorized by $\mathbb{1}_2$. Let

$$B = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix},$$

so that Proposition 3.9 applies. Let us choose the following three permutation matrices: $P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ and $P_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$. Then it is straightforward to confirm that

$$\frac{1}{3}\left(P_1BP_1 + P_2BP_2 + P_3BP_3\right) = \frac{1}{3}\left(\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right)$$
$$= \frac{1}{2}\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = A_3.$$

Example 3.11 (A_5 with qubits)**.** Let $\{\vec{r}_i\}_{i=1}^5, \vec{r}_i = \left[\cos\left((i-1)\frac{2\pi}{5}\right) \quad 0 \quad \sin\left((i-1)\frac{2\pi}{5}\right)\right]$ be the Bloch vectors of five pure qubit states so that $\varrho_i = \frac{1}{2}\left(\mathbb{1} + \vec{r}_i \cdot \vec{\sigma}\right)$. An antidistinguishing measurement for $\{\varrho_i\}$ can be constructed as $A_i = \frac{1}{5}\left(\mathbb{1} - \vec{r}_i \cdot \vec{\sigma}\right)$. Then the states $\{\varrho_i\}$ and the POVM $\{A_j\}$ implement the communication matrix

$$B = \frac{1}{20}\begin{bmatrix} 0 & 5-\sqrt{5} & 5+\sqrt{5} & 5+\sqrt{5} & 5-\sqrt{5} \\ 5-\sqrt{5} & 0 & 5-\sqrt{5} & 5+\sqrt{5} & 5+\sqrt{5} \\ 5+\sqrt{5} & 5-\sqrt{5} & 0 & 5-\sqrt{5} & 5+\sqrt{5} \\ 5+\sqrt{5} & 5+\sqrt{5} & 5-\sqrt{5} & 0 & 5-\sqrt{5} \\ 5-\sqrt{5} & 5+\sqrt{5} & 5+\sqrt{5} & 5-\sqrt{5} & 0 \end{bmatrix}.$$

Let us choose the following permutation matrix: $P = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$. Then it is straightforward to verify that

$$PBP^T = \frac{1}{20}\begin{bmatrix} 0 & 5+\sqrt{5} & 5-\sqrt{5} & 5-\sqrt{5} & 5+\sqrt{5} \\ 5+\sqrt{5} & 0 & 5+\sqrt{5} & 5-\sqrt{5} & 5-\sqrt{5} \\ 5-\sqrt{5} & 5+\sqrt{5} & 0 & 5+\sqrt{5} & 5-\sqrt{5} \\ 5-\sqrt{5} & 5-\sqrt{5} & 5+\sqrt{5} & 0 & 5+\sqrt{5} \\ 5+\sqrt{5} & 5-\sqrt{5} & 5-\sqrt{5} & 5+\sqrt{5} & 0 \end{bmatrix}$$

so that $\frac{1}{2}\left(B + PBP^T\right) = A_5$.

Examples 3.10 and 3.11 show that the sets $\mathcal{C}(\mathcal{S}_2^{cl})$ and $\mathcal{C}(\mathcal{Q}_2)$ are not convex. However, with the help of Proposition 3.9 it is in fact possible to see that $\mathcal{C}(\mathcal{S}_d^{cl})$ and $\mathcal{C}(\mathcal{Q}_d)$ are not convex for any d. The only thing that is required is to construct the communication matrix A_n for sufficiently large n so that $A_n \notin \mathcal{C}(\mathcal{S}_d^{cl})$ and $A_n \notin$

$\mathcal{C}(\mathcal{Q}_d)$. This is clearly always possible as $\text{rank}_+(A_n) = n$ and $\text{rank}_{psd}(A_{d^2}) \leq d+1$. Perhaps a bit paradoxically it is thus possible to prove that $\mathcal{C}(\mathcal{Q}_d)$ is not convex by using communication matrices that are implementable with the bit.

***Corollary* 3.12.** *As a consequence of Proposition 3.9, the sets $\mathcal{C}(\mathcal{S}_d^{cl})$ and $\mathcal{C}(\mathcal{Q}_d)$ are not convex for any d.*

While Proposition 3.9 used $n!$ permutation matrices to construct the matrix A_n, generally speaking it is possible to construct the matrix A_n with far less permutations. This can be seen from Examples 3.10 and 3.11 where A_3 was constructed with the identity and two permutations, and A_5 was constructed with just a single permutation in addition to the identity. It is not easy to say how many permutations exactly are required and clearly it depends a lot on the situation. Fewer permutations may suffice if there is some symmetry in the starting communication matrix.

3.2 Communication of partial ignorance

The family of matrices $D_{n,p}$ establishes a family of communication tasks that lie in between perfect discrimination and perfect uniform antidistinguishability. A key observation in Publication **II** was that there is another family of communication tasks between discrimination and antidistinguishability that is different from the previously defined matrix family. This generalization is evident from the following characterization of discrimination and antidistinguishability:

(i) Discrimination: each measurement outcome should exclude all except one preparation.

(ii) Antidistinguishability: each measurement outcome should exclude one preparation.

Based on the previous characterization we arrive at the following family of communication matrices. We define the communication matrix $G_{n,t}$ as the $\binom{n}{t} \times n$ matrix with $\frac{1}{n-t}\begin{bmatrix}1 & \cdots & 1 & 0 & \cdots & 0\end{bmatrix}$ as the first row. Each row contains $n-t$ ones and t zeros. The other rows of $G_{n,t}$ are the distinct permutations of the first row, written in lexicographical order. For instance,

$$G_{4,2} = \frac{1}{2}\begin{bmatrix}1 & 1 & 0 & 0\\ 1 & 0 & 1 & 0\\ 1 & 0 & 0 & 1\\ 0 & 1 & 1 & 0\\ 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 1\end{bmatrix}.$$

It is evident that $G_{n,n-1}$ corresponds to $\mathbb{1}_n$ and $G_{n,1}$ to A_n, although the zeros in $G_{n,1}$ are located on the antidiagonal of the matrix (they are ultraweakly equivalent).

The matrices $G_{n,t}$ were introduced in Publication **II** as an optimal solution to the following communication game. Suppose Charlie has in his possession a set of n boxes. He decides to hide a prize inside one of the boxes and then invites Alice and Bob to play a guessing game according to the following rules.

(i) Charlie will reveal t empty boxes to Alice. Hence in general $1 \leq t < n$.

(ii) Alice is allowed to communicate with Bob via preparing a single state from a preparation device. Thus she will optimally need $\binom{n}{t}$ distinct preparations from a state space $\mathcal{S}$.

(iii) Bob is not given any additional information on the location of the prize. The only information he will receive is the state sent by Alice, which he will measure according to a fixed measurement device.

(iv) After measuring the state Bob will produce a guess on the location of the prize. Alice and Bob win if Bob guesses correctly.

This communication game was called *communication of partial ignorance* in Publication **II**. There are some noteworthy remarks to be made about the given rules. First of all, Charlie is not required to distribute the prize according to any fixed probability distribution. In fact, he is free to choose the location of the prize even after he has already revealed empty boxes to Alice and Alice and Bob's communication has taken place. However, he is not allowed to cheat. In addition, Alice and Bob are free to build their preparation and measurement devices before starting the game. The only rule is that their devices are limited to preparing states from a given state space $\mathcal{S}$. They are not allowed to change their setups mid-game, hence it is possible for Charlie to seek deficiencies in Alice and Bob's guessing strategy. In general we define Alice and Bob's success probability to equal the "worst case" probability of winning the game, as Charlie could always choose the inputs in order to minimize the winning chance.

It is clear that the matrices $G_{n,t}$ are exactly the communication matrices that maximize Alice and Bob's winning probability in a guessing game where there are n boxes and Charlie reveals t empty boxes to Alice. This can be seen from the example matrix $G_{4,2}$. Suppose Charlie reveals the last two boxes to be empty to Alice. Then Alice will use the first preparation so that Bob will receive outcomes 1 and 2 with equal probability. This maximizes their winning probability because Charlie cannot exploit any weakness in the guessing strategy.

From now on we will identify optimal communication of partial ignorance with the communication matrices $G_{n,t}$. A key question is then if the optimal communication matrix can be implemented by Alice and Bob, i.e., whether $G_{n,t} \in \mathcal{C}(\mathcal{S})$ or not. Another key question is whether some of the $G_{n,t}$ are easier to implement than others according to ultraweak matrix majorization.

In Publication **II** it was shown that $\mathrm{rank}(G_{n,t}) = n$. Because $G_{n,t}$ is a $\binom{n}{t} \times n$ matrix, this also implies that $\mathrm{rank}_+(G_{n,t}) = n$ through the inequalities in (40). In general we can say that $\mathrm{rank}_{psd}(G_{n,t}) \leq n$ with the help of the last inequality in (42). In Publication **II** it was shown that $\mathrm{rank}_{psd}(G_{4,2}) = 3$. Clearly $\mathrm{rank}_{psd}(G_{n,n-1}) = n$ while $\mathrm{rank}_{psd}(G_{d^2,1})$ is either d or $d+1$ according to Section 3.1. In general the PSD rank of $G_{n,t}$ is not known when $1 < t < n-1$.

Regarding the question of whether some of the $G_{n,t}$ are easier to implement than others, the following was shown in Publication **II**.

***Proposition* 3.13.** $G_{n,t-1} \preceq G_{n,t} \preceq G_{n+1,t+1}$.

The proof of Proposition 3.13 basically relies first on the fact $G_{n+1,t+1}$ contains $G_{n,t}$ as a submatrix. Secondly it is straightforward to see that the rows of $G_{n,t-1}$ are contained in the convex hull of the rows of $G_{n,t}$. The majorizing matrices L and R are then straightforward to construct explicitly.

While Proposition 3.13 does a good job in characterizing the ultraweak relations of the matrices $G_{n,t}$, we are not yet done completely determining these relations. The matrix rank can be used to clearly see that $G_{n+1,t} \npreceq G_{n,t'}$ for any n, t and t' because the rank cannot increase in ultraweak majorization, hence there are two "obvious" ultraweak majorization relations that were left open in Publication **II**. Namely, we do not know yet if $G_{n,t} \preceq G_{n,t-1}$ and we do not know if $G_{n-1,t} \preceq G_{n,t}$ for all t. We have already seen that $G_{n-1,1} \npreceq G_{n,1}$ in Section 3.1, and with the help of the ultraweak monotone function λ_{max} we can now present additional ultraweak majorization relations between the communication matrices $G_{n,t}$.

***Proposition* 3.14.** *$G_{n-q,t+s} \npreceq G_{n,t}$ for any $q > 0$, $s \geq 0$ and $G_{n,t} \npreceq G_{n,t-s}$ for any $s > 0$.*

Proof. From the definition of $G_{n,t}$ we have

$$\lambda_{max}\left(G_{n,t}\right) = \frac{n}{n-t}.$$

Let us first consider $G_{n-q,t+s}$ and $G_{n,t}$.

Firstly, notice that in general $t+s \leq n-q-1$, which means $0 < q \leq n-t-s-1$. This also bounds s by $0 \leq s \leq n-q-t-1$. Simple calculation then yields

$$\begin{aligned}\lambda_{max}\left(G_{n,t}\right) - \lambda_{max}\left(G_{n-q,t+s}\right) &= \frac{n}{n-t} - \frac{n-q}{n-q-t-s} \\ &= \frac{n(n-q-t-s)}{(n-t)(n-q-t-s)} - \frac{(n-q)(n-t)}{(n-t)(n-q-t-s)} \\ &= \frac{-ns-qt}{(n-t)(n-q-t-s)},\end{aligned}$$

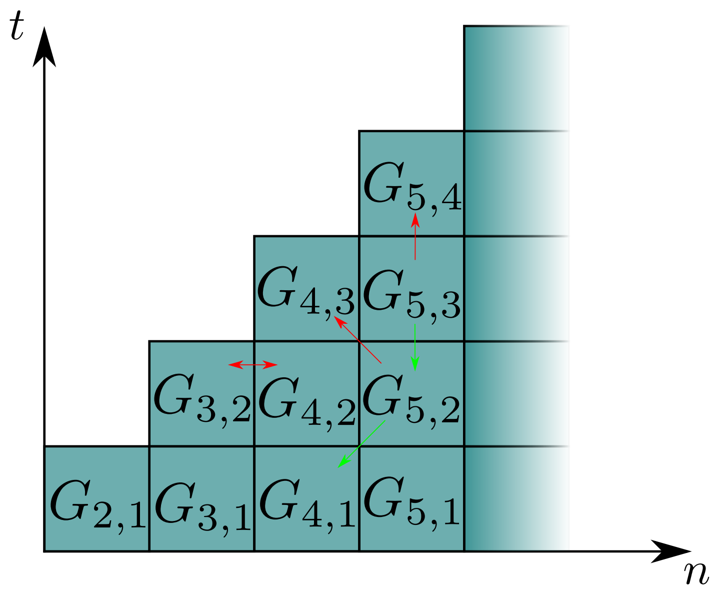

Figure 6. The optimal communication matrices $G_{n,t}$ for different communication of partial ignorance tasks, presented in a chart by the parameters n and t. The green arrows indicate the directions in which it is possible to move adjacently within the family $G_{n,t}$ with the ultraweak matrix majorization. The red arrows indicate directions in which ultraweak majorization is impossible.

which is clearly always negative as $n - t > 0$ and $n - q - t - s \geq 1$. Therefore $G_{n-q,t+s} \not\preceq G_{n,t}$ for any $q > 0$, $s \geq 0$ for which $G_{n-q,t+s}$ is a valid communication matrix.

Let us then consider $G_{n,t}$ and $G_{n,t-s}$. Clearly $\lambda_{max}(G_{n,t}) = \frac{n}{n-t} > \frac{n}{n-t+s} = \lambda_{max}(G_{n,t-s})$. Therefore $G_{n,t} \not\preceq G_{n,t-s}$ for any $0 < s \leq t-1$. □

Propositions 3.13 and 3.14 now completely determine the ultraweak preordering of adjacent $G_{n,t}$ matrices, i.e., those matrices in the family $G_{n,t}$ which differ only by one in n or t. This is illustrated in Figure 6. Additionally, we have the following corollary.

***Corollary* 3.15.** *If $m = n$, then $G_{m,s} \preceq G_{n,t}$ if and only if $s \leq t$. Otherwise a necessary condition for $G_{m,s} \preceq G_{n,t}$ is that $m < n$ and $s < t$.*

Corollary 3.15 is a direct consequence of Propositions 3.13 and 3.14.

The only remaining case left to determine is exactly when $G_{m,s} \preceq G_{n,t}$ in the case where $m < n$ and $s < t$. In Publication **II** we proved the following proposition.

***Proposition* 3.16.** *A sufficient condition for $G_{m,m-1} \preceq G_{n,t}$ is $\lfloor \frac{n}{n-t} \rfloor \geq m$.*

	Initial	Final
X_C	$\frac{1}{n}[1\ 1\ \cdots\ 1]$	$\frac{1}{n-t}[1\ 1\ \cdots\ 1\ 0\ \cdots\ 0]$
$H(X_C)$	$-\log_2 \frac{1}{n}$	$-\log_2(\frac{1}{n-t})$

Table 1. Entropy of complete communication of partial ignorance. Note that there are t zeros and $n-t$ ones in the final probability distribution. Hence the entropy is $-(n-t)\frac{1}{n-t}\log_2(\frac{1}{n-t})$, or $\log_2(n-t)$. Each row of $G_{n,t}$ is a permutation of the first row, so the entropy is the same for all rows.

As an example, we can consider the following:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = G_{2,1},$$

which shows that $G_{2,1} \preceq G_{4,2}$. Proposition 3.16 relies on finding orthogonal rows in the larger matrix so that a smaller identity can be constructed. This gives an example of one situation where $G_{m,s} \preceq G_{n,t}$ with $m < n$ and $s < t$ is possible. By Proposition 3.13 we also have that $G_{n-1,t-1} \preceq G_{n,t}$ always holds. To prove precisely when $G_{m,s} \preceq G_{n,t}$ seems to be a difficult combinatorial problem as a complete set of ultraweak monotones is not known. It is left as an open problem here.

Shannon entropy in communication of partial ignorance

As a final consideration on communication of partial ignorance we can consider the Shannon entropy of optimal communication. First of all, it is clear that if Alice knows the location of the prize, then she will have to communicate $\log_2 n$ bits of information to Bob. This can be seen as Bob's initial probability distribution on the location of the prize being totally uniform, and after communication Bob will know the location of the prize. In general, if Alice communicates everything she knows to Bob in the communication task $G_{n,t}$, the entropy in the communication is transformed according to Table 1. Thus we can see that, in general, the transmission of $\log_2 n - \log_2(n-t) = \log_2(\frac{n}{n-t})$ bits of information is required for Alice to tell everything she knows to Bob.

In communication of partial ignorance Bob will produce a guess after he has measured the state sent by Alice. Alice and Bob win if Bob's guess is correct. Otherwise they lose. Because the rules were defined in this way, it is not necessary for

	Initial	Final
X_Q	$\frac{1}{n}[1\ 1\ \cdots\ 1]$	$\begin{bmatrix} \frac{1}{n-t} & \frac{1-\frac{1}{n-t}}{n-1} & \cdots & \frac{1-\frac{1}{n-t}}{n-1} \end{bmatrix}$
$H(X_Q)$	$-\log_2 \frac{1}{n}$	$-\frac{1}{n-t}\log_2(\frac{1}{n-t}) - \left(1-\frac{1}{n-t}\right)\log_2\left(\frac{1-\frac{1}{n-t}}{n-1}\right)$

Table 2. Entropy of Bob's probability distributions before and after a measurement when Alice and Bob's quantum devices implement $G_{n,t}$.

	$G_{4,1}$	$G_{4,2}$	$G_{4,3}$
$H(X_U) - H(X_Q^f)$	0.0250625	0.207519	2
$H(X_U) - H(X_C^f)$	0.415037	1	2

Table 3. The amount of information transmitted in the communication tasks $G_{4,1}$, $G_{4,2}$ and $G_{4,3}$ in the quantum implementation ($H(X_Q)$) and in the case where Alice communicates everything she knows to Bob ($H(X_C)$). X_U denotes the uniform probability distribution.

Alice to communicate everything she knows about the empty boxes to Bob in order to reach the maximal probability of winning. In fact, suppose that Alice and Bob have quantum devices that implement $G_{n,t}$. Alice and Bob have agreed upon a strategy that, upon obtaining outcome m, Bob will guess that the prize was in box m. Alice should hence prepare the state from the row with zeros in the corresponding indices that Charlie has revealed to be empty. Suppose Charlie reveals the last t boxes to be empty to Alice. Then Bob's probability distribution will be updated according to Table 2 upon obtaining outcome 1.

Let us denote with X_C^f the final probability distribution in Table 1 and with X_Q^f the corresponding probability distribution in Table 2. We can see that the difference of the final Shannon entropies can be written as:

$$H(X_Q^f) - H(X_C^f) = \left(1 - \frac{1}{n-t}\right)\left(\log_2\left(\frac{1}{n-t}\right) - \log_2\left(\frac{1-\frac{1}{n-t}}{n-1}\right)\right),$$

which is generally speaking a positive function. The difference is zero only when $t = n-1$, i.e., when the implemented communication matrix is $\mathbb{1}_n$. Thus the actual information that is transmitted in the quantum implementation of $G_{n,t}$ is less than or equal to what would be transmitted if Alice told Bob everything she knows. The amount is only equal if the task is $G_{n,n-1}$.

As an example, let us consider the communication tasks $G_{4,1}$, $G_{4,2}$ and $G_{4,3}$. Table 3 lists the classical and quantum entropies for these tasks.

From Table 3 we can see that, generally speaking, the amount of information transmitted in the quantum implementation is much less than perfect communication whenever $t < n - 1$. However, both of these methods of communication achieve the same probability of winning in a communication game of partial ignorance. This is not really paradoxical, however. It just means that Alice has communicated with Bob in a way that maximizes the winning probability in the first guess Bob produces. If additional guesses were granted upon an erroneous guess by Bob, the complete transmitted information would allow Bob to guess with better chances of finding the prize compared to the quantum implementation. If Bob's guess is wrong after measuring a quantum state, he will have no additional information on the location of the prize and he has to update his probability distribution to the uniform distribution again, as there are $n - 1$ remaining possibilities of where the prize could be. In contrast, upon guessing wrong and having the complete information available, Bob's probability of guessing correctly would be greater as he would know some of the boxes to be empty, leaving $n - t - 1$ choices.

The analysis of the Shannon entropies shows that the communication game defined by communication of partial ignorance would be significantly changed if additional guesses were granted after an erroneous guess. The probability of guessing wrong and then updating Bob's probability distribution according to available information would have to be taken into account. It is unclear what the best possible d-dimensional classical and quantum strategies would be in this case.

3.3 Partial-ignorance communication tasks

Publication **IV** introduced a generalized version of communication of partial ignorance. In the generalized version there also exists an input for Bob who is operating the measurement device. This means that in general Bob will have multiple measurements to choose from based on his input. As a consequence the use of behaviors is now required.

Formally we define a partial-ignorance communication task of type $T_{n,m}$ to be a communication task with the following rules.

Definition 3.17 (Partial-ignorance communication task $T_{n,m}$)**.** Charlie, who acts as a game master, chooses an n-bit string s with exactly one 1. He reveals the indices of m zeros to Alice by sending her an n-bit string a with m 1's. Each index of a 1 in a will reveal that the corresponding bit in s was zero. Charlie will reveal the remaining zeros to Bob as an n-bit string input b. Therefore the combined knowledge of Alice and Bob will determine the location of the 1 in s completely. Alice and Bob are not allowed to communicate freely: Alice is allowed to send Bob a single state from a state space $\mathcal{S}$. Bob can perform any measurement on the state Alice sends to him, after which he will have to guess the index of the 1 in s. Alice and Bob win if Bob's guess is correct.

Just like in communication of partial ignorance, there is no extra guesses granted in a partial-ignorance communication task after an erroneous guess. While in communication of partial ignorance even Alice did not necessarily know the location of the prize, a communication task of partial-ignorance is now a communication task with complete knowledge in the sense that Alice and Bob would know the location of the prize, or the index of the 1 in the bit string s, if they combined their knowledge. Going forward a partial-ignorance communication task of type $T_{n,m}$ will be simply referred to as the communication task $T_{n,m}$.

While it would be possible to consider the communication tasks $T_{n,m}$ in the general framework introduced in Chapter 2, i.e. with state spaces other than quantum or classical, it turns out the analysis of these tasks is pretty complicated. So far only the tasks $T_{3,1}$, $T_{4,1}$ and $T_{4,2}$ have been analyzed in Publication **IV** for the bit ($\mathcal{S}_2^{cl}$) and up to 4-dimensional quantum systems. It is pretty straightforward to see that in $\mathcal{Q}_4$ and $\mathcal{S}_4^{cl}$ the optimal success probabilities are already saturated for these tasks, while the bit has been used as a comparison for the qubit. The correct classical analogue for the qutrit would be a three-level classical system, or the trit, but this analysis has not been done. The analysis for the trit will be presented here in order to get the correct comparison for the qutrit. Let us first recall some mathematical tools that were used in Publication **IV** to obtain lower and upper bounds on success probabilities in the communication tasks $T_{n,m}$.

Semidefinite programming

We have already seen how the question of whether a set of quantum states is antidistinguishable is an SDP. The optimal quantum implementation of a partial-ignorance communication task can be formulated as an optimization problem of the following form:

$$
\begin{aligned}
\max \sum_{i,j,k} & p_{ijk} \operatorname{tr}\left[\varrho_i \mathsf{M}_j(k)\right] \\
\text{s.t.} \quad & \varrho_i \geq 0 \,\forall i, \\
& \operatorname{tr}\left[\varrho_i\right] = 1 \,\forall i \\
& \mathbb{1} \geq \mathsf{M}_j(k) \geq 0 \,\forall j,k \\
& \sum_k \mathsf{M}_j(k) = \mathbb{1} \,\forall j,
\end{aligned}
\tag{60}
$$

where the weights p_{ijk} define the success metric of the problem. Additionally, operational constraints of the form $\sum_i \alpha_i^r \varrho_i = \sum_j \beta_j^r \varrho_j$ can be added to the preparations, where the variable r indexes these constraints and α_i^r, β_j^r define a set of convex weights. Operational constraints for measurements can be added in a similar fashion.

While Equations (60) clearly do not define an SDP (the success metric is not a linear function), the used formulation can still be optimized with methods from semidefinite programming. This is done by first fixing a dimension for the Hilbert space of the states and then generating random rank-1 states. After the states are fixed as constants the problem of (60) can be optimized as an SDP over the effects. The optimal effects can then be fixed as constants and the problem can be optimized as an SDP over the states. This procedure can be continued until additional steps no longer produce an improved result. An algorithm like this is usually called a *see-saw* method in the literature [96; 97; 98; 99; 100]. While there are no performance guarantees, in practice the see-saw algorithm works very well. In Publication **IV** the see-saw method was able to produce optimal lower bounds on all considered success metrics.

The see-saw method can be always used to generate a dimension-dependent lower bound on success metrics of the form (60). However, this is just a lower bound. Optimality has to be proven in most cases, if the goal is to determine the optimal success probability of some communication task. Generally speaking there are many possible methods to produce upper bounds on success metrics. A typical method is to produce a converging hierarchy of SDPs where each level of the hierarchy produces an upper bound on the success metric and the hierarchy can be guaranteed to converge on the optimal quantum value. See e.g. [84; 85; 86; 101; 102; 103; 104; 105; 106] for more information on the theoretical framework of SDPs. In Publication **IV** the unitary SDP hierarchy from [87] was used to produce the upper bounds.

Noncontextuality as a principle bounding correlations

A partial-ignorance communication task involves an input for both Alice and Bob. Thus it is possible to analyze the effects of contextuality in these tasks. A brief introduction to contextuality was given in Publication **IV**. Additional resources on the topic of hidden variables in quantum mechanics can be found e.g. from [14; 15; 43; 46; 47; 50; 79; 87; 107; 108; 109; 110; 111; 112]. We will recall some of the basic concepts here in general terms.

In all simplicity noncontextuality is the physical principle that operationally indistinguishable things should also be indistinguishable on the ontological level. For instance, whenever two quantum states are represented by the same density matrix those two quantum states should be described by the same underlying ontological state as they cannot be discriminated by any conceivable quantum measurement. It is also immediately obvious that contextuality is a purely nonclassical feature. A classical state space $\mathcal{S}_d^{cl}$ is uniquely defined as the convex hull of d distinguishable pure states, or a $d-1$-simplex. Each state has a unique convex decomposition into pure states. Therefore the principle of noncontextuality applies on all states and there can be no contextual effects in communication.

A key question regarding the principle of noncontextuality is then the following:

given a behavior $p(k|\varrho_i, \mathsf{M}_j(k))$ and a set of operational equivalences between the states and effects, does the behavior admit a noncontextual ontological model? It turns out the problem of querying for a noncontextual model constitutes a linear program (LP) [47]. However, there is some technicality involved as the polytope of noncontextual correlations has to be characterized first by constructing the so-called noncontextual measurement-assignment polytope.

Whenever the observed behavior is contextual, i.e. the behavior does not admit a noncontextual ontological model, it must be true by Farkas' lemma [113] that a certificate of primal feasibility is negative. By formulating the Farkas dual of the primal LP it is then possible to find the noncontextuality inequality that is most violated by the observed behavior. This is especially useful as characterizing the noncontextual polytope by the vertices and checking all the different noncontextuality inequalities can be a very laborious task.

Grassmannian frames

Frame theory offers an alternative way to derive some upper bounds on success metrics in the case where operational equivalences are not defined between states or effects. A partial-ignorance communication task $T_{n,m}$ typically involves $\binom{n}{m}$ distinct preparations for Alice. However, Bob can exclude some of these based on his own input. Therefore Bob only has to attempt to discriminate between some subset of states that Alice can prepare. In the spirit of communication of partial ignorance an optimal strategy should be as uniform as possible with respect to erroneous guesses so that there would not be any weakness in the guessing strategy that Charlie could exploit. With the help of frame theory it is possible to limit the maximal overlap between Alice's states.

Recall that a sequence $\{f_i\}_{i=1}^n$ of vectors in an inner product space V is called a frame if there exists frame bounds A and B such that

$$A\,\|v\|^2 \leq \sum_{i=1}^{n} |\langle\, f_i \,|\, v \,\rangle|^2 \leq B\,\|v\|^2 \tag{61}$$

for all $v \in V$. There are some important special cases for frames. A frame is tight whenever it is possible to choose $A = B$. If $\|f_i\| = 1$ for all i, then the frame is uniform. A frame is equiangular if $|\langle\, f_i \,|\, f_j \,\rangle| = c$ for all $i \neq j$ and $c \geq 0$ is a constant. A frame that has the same number of elements as the dimension of the inner product space V is also a basis. In many cases an overcomplete frame is preferable [114; 115; 116].

Definition 3.18. The maximal frame correlation of a uniform frame $\{f_i\}_{i=1}^n$ is defined as

$$\mathcal{M}\left(\{f_i\}_{i=1}^n\right) = \max_{\substack{j,k \\ j \neq k}} |\langle f_j \,|\, f_k \rangle|$$

A uniform frame is called a *Grassmannian frame* if it minimizes the maximal frame correlation [117; 118].

Grassmannian frames seem very promising to our current investigation. This is because we can always identify pure states with vectors of length one in the corresponding Hilbert space. Any bounds that can be derived on Grassmannian frames then also automatically apply to pure states. The following was shown in [119].

***Proposition* 3.19.** *If $\{f_i\}_{i=1}^n$ is a uniform frame in $\mathbb{R}^d$ or $\mathbb{C}^d$, then*

$$\mathcal{M}\left(\{f_i\}_{i=1}^n\right) \geq \sqrt{\frac{n-d}{d(n-1)}}. \tag{62}$$

Additionally, only an equiangular frame can achieve equality in (62). *Equality can only be achieved if $n \leq \frac{d(d+1)}{2}$ for $\mathbb{R}^d$ or $n \leq d^2$ for $\mathbb{C}^d$.*

Sometimes it is important to consider sets of vectors that do not span the entire inner product space V. In these situations the previous bound can still be used, as a collection of unit vectors is always a frame for their span in V [120]. The dimension d just has to be adjusted according to $\dim(\text{span}(\{f_i\}_{i=1}^n))$.

The way in which we can utilize the bound (62) is through the Helstrom bound [121; 122]. In the case of two pure quantum states prepared with equal probability the Helstrom bound gives the optimal minimum-error discrimination probability:

$$P_{success} = \frac{1}{2}\left(1 + \sqrt{1 - |\langle \varphi_1 \,|\, \varphi_2 \rangle|^2}\right). \tag{63}$$

If we replace the inner product with the bound from (62) we obtain

$$P_{success} \leq \frac{1}{2}\left(1 + \sqrt{1 - \frac{n-d}{d(n-1)}}\right), \tag{64}$$

which should be interpreted as the worst ambiguous pair-wise discrimination probability among a set of n pure states in a d-dimensional Hilbert space. Clearly it would be desirable to have a similar formula for more than two pure states and for mixed states as well. However, the problem of discriminating an arbitrary set of quantum states with minimal error is a complicated topic [123; 124; 125; 126; 127; 128]. As far as we know there does not exist an equation like (63) in this case. Instead some bounds that limit success probability are known.

Average success probabilities in partial-ignorance communication tasks

The simplest partial-ignorance communication task is $T_{3,1}$. In that task Charlie chooses one string from 100, 010 and 001. He reveals the index of one zero to Alice and the other zero is revealed to Bob. Therefore in each run of the communication task Bob is trying to distinguish two distinct preparations of Alice. For instance, if Bob receives input 100, then he knows that Alice will prepare a state corresponding to either input 010 or 001. We can thus use Equation (64) to bound Bob's success chance to discriminate any two out of the three states that Alice can prepare. The success metric in this task can be written as:

$$
\begin{aligned}
&\mathrm{tr}\left[\varrho_3 \mathsf{M}_1(1)\right] + \mathrm{tr}\left[\varrho_2 \mathsf{M}_1(2)\right] + \mathrm{tr}\left[\varrho_3 \mathsf{M}_2(1)\right] + \\
&\mathrm{tr}\left[\varrho_1 \mathsf{M}_2(2)\right] + \mathrm{tr}\left[\varrho_2 \mathsf{M}_3(1)\right] + \mathrm{tr}\left[\varrho_1 \mathsf{M}_3(2)\right] \\
=&\mathrm{tr}\left[\varrho_3 \mathsf{M}_1(1)\right] + \mathrm{tr}\left[\varrho_2 (\mathbb{1} - \mathsf{M}_1(1))\right] + \mathrm{tr}\left[\varrho_3 \mathsf{M}_2(1)\right] + \mathrm{tr}\left[\varrho_1 (\mathbb{1} - \mathsf{M}_2(1))\right] \\
&\mathrm{tr}\left[\varrho_2 \mathsf{M}_3(1)\right] + \mathrm{tr}\left[\varrho_1 (\mathbb{1} - \mathsf{M}_3(1))\right] \\
\leq& 3(P_{success} - P_{fail}) + 3 = 3(P_{success} - (1 - P_{success})) + 3 = 6P_{success},
\end{aligned}
$$

where $P_{success}$ is the optimal pair-wise discrimination probability between three quantum states. From (64) we get that $P_{success} \leq \frac{1}{2}\left(1 + \sqrt{1 - \frac{1}{4}}\right) = \frac{1}{2}\left(1 + \frac{\sqrt{3}}{2}\right) \approx 0.933913$ for the average success probability. This value matches exactly the value presented in Publication **IV**, where the value was obtained with heuristic methods. Since the average is obtained here from an equiangular frame for $\mathbb{C}^2$ it can be argued that the average success probability is equal to the worst case success probability where it is assumed that Charlie is malicious and tries to find a weakness in Bob's guessing strategy. In comparison to the qubit, in Publication **IV** it was showed that the average success probability for the bit is $\frac{5}{6}$ in $T_{3,1}$. Thus the qubit achieves a much higher average success probability. Additionally, the worst case success probability for the bit is 0 as there is always at least one combination of inputs that lead to Bob's guess being wrong.

Let us now collect the known optimal success probabilities for $T_{4,1}$ and $T_{4,2}$ in Table 4. The classical bounds were obtained by first listing a general deterministic guessing strategy in table form. A maximal average success probability was then derived from the general table by noticing that some of Alice's messages should not be equal. The maximal success probability was then achieved by a concrete strategy formed by filling the general table. We will replicate this process for the trit to complete Table 4.

The lower bounds for qubits and qutrits were obtained by using the see-saw method. This lower bound was proved to be optimal in $T_{4,1}$ by considering the optimal ambiguous guessing probability given by (64). For $T_{4,2}$ the lower bound was proved to be optimal by using the ultraweak monotone λ_{max} on individual com-

	$T_{4,1}$	$T_{4,2}$
$\mathcal{S}_2^{cl}$	$\frac{5}{6}$	$\frac{2}{3}$
$\mathcal{Q}_2$	$\frac{1}{2}\left(1+\sqrt{\frac{2}{3}}\right) \approx 0.908248$	$\frac{2}{3}$
$\mathcal{Q}_3$	$\frac{1}{2}\left(1+\frac{2\sqrt{2}}{3}\right) \approx 0.971405$	1

Table 4. Optimal average success probabilities for the communication tasks $T_{4,1}$ and $T_{4,2}$.

munication matrices (there is one communication matrix for each of Bob's measurements). There was no need to consider 4-dimensional quantum systems as it is straightforward to see that four distinguishable states is enough to achieve maximum success probability in $T_{4,1}$. Notably the worst case success probability in $T_{4,2}$ for the qubit was left as an open problem in Publication **IV**.

The communication tasks $T_{4,1}$ and $T_{4,2}$ were not considered for the trit in Publication **IV**. We now perform this analysis to provide an illustrative example and in order to get the correct analogue for the qutrit.

Example 3.20 ($T_{4,1}$ with the trit)**.** Recall from Publication **IV** that a general classical strategy for $T_{4,1}$ can be written according to Table 5.

The variables i, j, k and l determine which state Alice should send to Bob given each of her inputs. Likewise, each of the variables x_i, y_i and z_i determines what Bob will guess given that he received input b and the message $r(a)$. By looking at the rows of the lower table where values of b are identical we can deduce that Alice's states should comply with the following constraints:

$$k \neq l,\ j \neq l,\ j \neq k,\ i \neq l,\ i \neq k,\ i \neq j.$$

Each constraint was derived from rows with a different value of s, so each violation of the constraints leads to at least one mistake in the strategy. Clearly the above constraints mean that each of Alice's states should be different. For the trit this is impossible, so at least one mistake is bound to happen. An optimal strategy is presented in Table 6. The average success probability for the trit in the communication task $T_{4,1}$ is therefore $\frac{11}{12} \approx 0.916667$. The worst case success probability is still zero for the trit in all deterministic strategies.

Example 3.21 ($T_{4,2}$ with the trit)**.** We will follow a similar strategy as with $T_{4,1}$. Let us first write down the general form of the strategy in Table 7.

Again looking at the rows of the lower table we can deduce that Alice's states should comply with the following constraints:

$$i \neq j \neq l,\ i \neq k \neq m,\ j \neq k \neq n,\ l \neq m \neq n.$$

a	$r(a)$	b	$g(b,0)$	$g(b,1)$	$g(b,2)$
1000	i	1100	x_1	y_1	z_1
0100	j	1010	x_2	y_2	z_2
0010	k	1001	x_3	y_3	z_3
0001	l	0110	x_4	y_4	z_4
		0101	x_5	y_5	z_5
		0011	x_6	y_6	z_6
s	a	b	$r(a)$	$g[b,r(a)]$	
1000	0100	0011	j	$g(0011,j)$	
1000	0010	0101	k	$g(0101,k)$	
1000	0001	0110	l	$g(0110,l)$	
0100	1000	0011	i	$g(0011,i)$	
0100	0010	1001	k	$g(1001,k)$	
0100	0001	1010	l	$g(1010,l)$	
0010	1000	0101	i	$g(0101,i)$	
0010	0100	1001	j	$g(1001,j)$	
0010	0001	1100	l	$g(1100,l)$	
0001	1000	0110	i	$g(0110,i)$	
0001	0100	1010	j	$g(1010,j)$	
0001	0010	1100	k	$g(1100,k)$	

Table 5. A general classical strategy for the trit in the communication task $T_{4,1}$.

a	$r(a)$	b	$g(b,0)$	$g(b,1)$	$g(b,2)$
1000	0	1100	–	–	3
0100	1	1010	–	4	2
0010	2	1001	–	3	2
0001	2	0110	4	–	1
		0101	3	–	1
		0011	2	1	–

s	a	b	$r(a)$	$g[b,r(a)]$
1000	0100	0011	1	1
1000	0010	0101	2	1
1000	0001	0110	2	1
0100	1000	0011	0	2
0100	0010	1001	2	2
0100	0001	1010	2	2
0010	1000	0101	0	3
0010	0100	1001	1	3
0010	0001	1100	2	3
0001	1000	0110	0	4
0001	0100	1010	1	4
0001	0010	1100	2	**3**

Table 6. An optimal classical strategy for the trit in the communication task $T_{4,1}$. The erroneous guess by Bob is highlighted in bold.

a	$r(a)$	b	$g(b,0)$	$g(b,1)$	$g(b,2)$
1100	i	1000	x_1	y_1	z_1
1010	j	0100	x_2	y_2	z_2
1001	k	0010	x_3	y_3	z_3
0110	l	0001	x_4	y_4	z_4
0101	m				
0011	n				

s	a	b	$r(a)$	$g[b,r(a)]$
1000	0110	0001	l	$g(0001,l)$
1000	0101	0010	m	$g(0010,m)$
1000	0011	0100	n	$g(0100,n)$
0100	1010	0001	j	$g(0001,j)$
0100	1001	0010	k	$g(0010,k)$
0100	0011	1000	n	$g(1000,n)$
0010	1100	0001	i	$g(0001,i)$
0010	1001	0100	k	$g(0100,k)$
0010	0101	1000	m	$g(1000,m)$
0001	1100	0010	i	$g(0010,i)$
0001	1010	0100	j	$g(0100,j)$
0001	0110	1000	l	$g(1000,l)$

Table 7. A general classical strategy for the trit in the communication task $T_{4,2}$.

a	$r(a)$	b	$g(b,0)$	$g(b,1)$	$g(b,2)$
1100	0	1000	2	3	4
1010	1	0100	1	4	3
1001	2	0010	4	1	2
0110	2	0001	3	2	1
0101	1				
0011	0				

s	a	b	$r(a)$	$g[b,r(a)]$
1000	0110	0001	2	1
1000	0101	0010	0	1
1000	0011	0100	0	1
0100	1010	0001	1	2
0100	1001	0010	2	2
0100	0011	1000	0	2
0010	1100	0001	0	3
0010	1001	0100	2	3
0010	0101	1000	1	3
0001	1100	0010	0	4
0001	1010	0100	1	4
0001	0110	1000	2	4

Table 8. An optimal classical strategy for the trit in the communication task $T_{4,2}$.

It turns out that with the trit it is possible to not break any of these constraints. Hence the success probability with the trit should equal one. An optimal strategy is presented in Table 8.

After the examples with the trit we can now write down the complete table of average success probabilities for the partial-ignorance communication tasks $T_{4,1}$ and $T_{4,2}$ in Table 9. Note that the trit outperforms the qubit by quite a large margin. In Publication **IV** it was shown that the qubit is capable of violating a noncontextual bound in the partial-ignorance communication task $T_{4,1}$. The corresponding optimal values for noncontextual models and the violation found by SDP methods for qubits with an operational constraint between preparations are also included in the table.

	$T_{4,1}$	$T_{4,2}$
$\mathcal{S}_2^{cl}$	$\frac{5}{6}$	$\frac{2}{3}$
$\mathcal{S}_3^{cl}$	$\frac{11}{12} \approx 0.916667$	1
$\mathcal{Q}_2$	$\frac{1}{2}\left(1+\sqrt{\frac{2}{3}}\right) \approx 0.908248$	$\frac{2}{3}$
$\mathcal{Q}_3$	$\frac{1}{2}\left(1+\frac{2\sqrt{2}}{3}\right) \approx 0.971405$	1
Noncontextual models	$\frac{5}{6}$	$\frac{2}{3}$
Q_2 with operational equivalence	0.902369	$\frac{2}{3}$

Table 9. Optimal average success probabilities for the communication tasks $T_{4,1}$ and $T_{4,2}$ with probabilities listed for the trit and noncontextual models. The qubit can violate the noncontextual bound for $T_{4,1}$.

4 Conclusions

In this thesis I have studied various communication tasks in a quantum mechanical and classical setting. This ultimately led to the study of the operational hierarchy of such tasks in the framework of operational theories. The key questions related to the operational hierarchy were if some communication tasks are easier to implement than others and whether a given communication task can be implemented by parties preparing and measuring states with given resources. By comparing which communication tasks can be implemented in specific operational theories it becomes possible to compare different theories in a robust way. As quantum theory exhibits nonclassical and even bizarre features, the comparison between quantum theory and the analogous classical theory is especially intriguing.

The motivation behind the research throughout this thesis has been to understand the peculiarities behind quantum mechanics. Nature has ultimately decided to arrange herself in such a way as to hide the true ontology of physical systems behind probabilities and statistics. Various puzzling phenomena then arise from said statistics, such as entanglement and contextuality. It is no wonder that there does not exist a universally accepted interpretation of quantum mechanics; even the question of whether quantum mechanics needs an interpretation is contested. While the topic of interpretation, and its disputed necessity in the current theoretical framework, has so far resulted in little more than enormous confusion and headaches amongst physicists and philosophers alike, the study of communication tasks is at least capable of offering some conceptual concreteness in the analysis.

Chapter 1 was largely intended to give an introduction to the framework of operational theories without delving too much into technicalities. A short description of select topics in quantum mechanics was also included. The distinguishability of states was a natural choice as it is also perhaps the most fundamental communication task and appears frequently throughout the other Chapters. Tomography and entanglement were also discussed briefly as at the time of writing the introduction the thesis was going to include the publication [129] as well. However, the decision was made later to not include this paper into the main thesis, instead listing it in the other published material section. Chapter 1 works nonetheless as a general introduction to the topics that are relevant in the main articles.

Chapter 2 was devoted to introducing the operational hierarchy of communication tasks that was introduced in Publications **II** and **III**. As stated, the main motivations were to characterize which communication tasks are easier to implement than others and to understand what resources are necessary to implement a given communication task. The preorder of ultraweak matrix majorization was introduced to answer these questions. Importantly, the set of communication matrices implementable within a given theory is closed with respect to the operations defined by ultraweak matrix majorization. This ensures that the mathematical framework is built on a legitimate basis.

To characterize the operational hierarchy defined by the ultraweak preorder I introduced and studied various monotone functions. It was shown that each monotone can be used to define a dimension that characterizes different operational theories. The main studied examples were d-dimensional classical and quantum theory, for which we saw that the linear, classical and quantum dimensions, induced by the rank, rank_+ and rank_{psd} monotones respectively, were very different. In Publication **III** it was shown that the classical dimension of a quantum state space is at least the operational quantum dimension squared; our conjecture is that it is exactly equal to this value.

Apart from the dimensions induced by ultraweak monotone functions, a key problem in the operational hierarchy of communication tasks is which communication tasks are the maximal elements of the ultraweak preorder. In the case of a classical state space this is clear, as there is a unique maximal equivalence class of communication tasks defined by the identity matrix. For quantum state spaces and other general state spaces it remains an open problem to characterize all of the different maximal elements.

Finally, in Chapter 3 the knowledge on the operational hierarchy of communication tasks was applied to concrete communication tasks. The first communication task was antidistinguishability, which has important applications in the foundations of quantum mechanics. Most notably antidistinguishability is used in the proof of the Pusey–Barrett–Rudolph theorem. This connection was the main motivation behind Publication **I** where I wanted to understand if antidistinguishability can be characterized with algebraic conditions. A new algebraic condition was presented for a set of pure quantum states to be antidistinguishable which also generalizes previously known results. I also studied so called uniform antidistinguishability and used it to show that the set of communication matrices implementable with classical or quantum states is not convex in any dimension.

The second communication task I studied was communication of partial ignorance, which was introduced in Publication **II**. A key observation of Publication **II** was that there exists a family of communication tasks that lie between distinguishability and antidistinguishability. However, at the time of writing Publication **II** I did not have a good understanding on the ultraweak preorder, as the concept had just been introduced. In Publication **III** the knowledge on the ultraweak preorder was improved greatly, but the communication tasks of partial ignorance were not really considered in that publication. With the help of the ultraweak monotone function λ_{max} it was possible to improve the results of Publication **II** slightly with regards to the operational hierarchy of the communication tasks of partial ignorance.

As a final observation on communication of partial ignorance I performed basic analysis of the Shannon entropies in these tasks. It was concluded that the quantum implementations of these communication tasks transmit relatively small amounts of information, whilst nevertheless obtaining an equal success probability as if complete information was transmitted.

The last communication task that was studied was partial-ignorance communication tasks. Introduced in Publication **IV**, a partial-ignorance communication task differentiates from communication of partial ignorance in the feature that a partial-ignorance communication task also includes an input for the person operating measurement devices. Thus a partial-ignorance communication task is very different from communication of partial ignorance. Generally speaking the partial-ignorance communication tasks have to be analyzed with entirely different mathematical tools. A brief description of these tools was included in the final Chapter of this thesis. While the exact results of Publication **IV** were not replicated in the final Chapter, two novel examples with the trit were presented as the correct analogue for the qutrit.

List of references

[1] T. Heinosaari and O. Kerppo. Antidistinguishability of pure quantum states. *J. Phys. A: Math. Theor.*, 51:365303, 2018.

[2] T. Heinosaari and O. Kerppo. Communication of partial ignorance with qubits. *J. Phys. A: Math. Theor.*, 52:395301, 2019.

[3] T. Heinosaari, O. Kerppo, and L. Leppäjärvi. Communication tasks in operational theories. *J. Phys. A: Math. Theor.*, 53:435302, 2020.

[4] O. Kerppo. Partial-ignorance communication tasks in quantum theory. *Phys. Rev. A*, 105:062607, 2022.

[5] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777, 1935.

[6] S. Wiesner. Conjugate coding. *SIGACT News*, 15:78, 1983.

[7] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and quantum finite automata. *J. ACM*, 49:496, 2002.

[8] A. Ambainis, D. Leung, L. Mancinska, and M. Ozols. Quantum random access codes with shared randomness. arXiv:0810.2937 [quant-ph], 2008.

[9] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Probl. Peredachi Inf.*, 9:3, 1973.

[10] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69:2881, 1992.

[11] 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. *Phys. Rev. Lett.*, 70:1895, 1993.

[12] E. Schrödinger. Probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society*, 32:446, 1936.

[13] H. M. Wiseman, S. J. Jones, and A. C. Doherty. Steering, entanglement, nonlocality, and the Einstein-Podolsky-Rosen paradox. *Phys. Rev. Lett.*, 98:140402, 2007.

[14] J. S. Bell. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38:447, 1966.

[15] S. Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. *J. Math. Mech.*, 17:59, 1967.

[16] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880, 1969.

[17] Hensen et al. Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526:682, 2015.

[18] Giustina et al. Significant-loophole-free test of bell's theorem with entangled photons. *Phys. Rev. Lett.*, 115:250401, 2015.

[19] M. Plávala. General probabilistic theories: An introduction. arXiv:2103.07469 [quant-ph], 2021.

[20] P. Busch, P. Lahti, J.-P. Pellonpää, and Kari Ylinen. *Quantum Measurement*. Springer International Publishing, Switzerland, 2016.

[21] G. D'Ariano, G. Chiribella, and P. Perinotti. *Quantum Theory from First Principles: An Informational Approach*. Cambridge University Press, 2017.

[22] T. Heinosaari and M. Ziman. *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*. Cambridge University Press, Cambridge, 2012.

[23] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Probabilistic theories with purification. *Phys. Rev. A*, 81:062348, 2010.

[24] G. Kimura, K. Nuida, and H. Imai. Distinguishability measures and entropies for general probabilistic theories. *Rep. Math. Phys.*, 66:175, 2010.

[25] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Informational derivation of quantum theory. *Phys. Rev. A*, 84:012311, 2011.

[26] M. Kläy, C. Randall, and D. Foulis. Tensor products and probability weights. *Int. J. Theor. Phys.*, 26:199, 1986.

[27] A. Wilce. Tensor products in generalized measure theory. *Int. J. Theor. Phys.*, 31:1915, 1992.

[28] M. Plávala. All measurements in a probabilistic theory are compatible if and only if the state space is a simplex. *Phys. Rev. A*, 94:042108, 2016.

[29] R. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[30] Z. Hradil. Quantum-state estimation. *Phys. Rev. A*, 55:R1561(R), 1997.

[31] K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi. Maximum-likelihood estimation of the density matrix. *Phys. Rev. A*, 61:010304(R), 1999.

[32] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White. Measurement of qubits. *Phys. Rev. A*, 64:052312, 2001.

[33] J. A. Smolin, J. M. Gambetta, and G. Smith. Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise. *Phys. Rev. Lett.*, 108: 070502, 2012.

[34] R. Blume-Kohout. Optimal, reliable estimation of quantum states. *New J. Phys.*, 12:043034, 2010.

[35] N. Canosa, A. Plastino, and R. Rossignoli. Ground-state wave functions and maximum entropy. *Phys. Rev. A*, 40:519, 1989.

[36] J. Řeháček, Z. Hradil, E. Knill, and A. I. Lvovsky. Diluted maximum-likelihood algorithm for quantum tomography. *Phys. Rev. A*, 75:042108, 2007.

[37] S. Szalay. Multipartite entanglement measures. *Phys. Rev. A*, 92:042329, 2015.

[38] D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal. Unextendible product bases, uncompletable product bases and bound entanglement. *Commun. Math. Phys.*, 238:379, 2003.

[39] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865, 2009.

[40] G. Vidal and R. F. Werner. Computable measure of entanglement. *Phys. Rev. A*, 65:032314, 2002.

[41] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223:1, 1996.

[42] A. Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413, 1996.

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

[44] A. Cabello. Kochen-specker theorem for a single qubit using positive operator-valued measures. *Phys. Rev. Lett.*, 90:190401, 2003.

[45] P. Busch. Quantum states and generalized observables: A simple proof of Gleason's theorem. *Phys. Rev. Lett.*, 91:120403, 2003.

[46] D. Schmid, J. H. Selby, E. Wolfe, R. Kunjwal, and R. Spekkens. Characterization of noncontextuality in the framework of generalized probabilistic theories. *PRX Quantum*, 2:010331, 2021.

[47] D. Schmid, R. Spekkens, and E. Wolfe. All the noncontextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed set of operational equivalences. *Phys. Rev. A*, 97:062103, 2018.

[48] P. Wang, J. Zhang, C. Luan, M. Um, Y. Wang, M. Qiao, T. Xie, J. Zhang, A. Cabello, and K. Kim. Significant loophole-free test of kochen-specker contextuality using two species of atomic ions. *Science Advances*, 8:eabk1660, 2022.

[49] D. Qu, K. Wang, L. Xiao, X. Zhan, and P. Xue. State-independent test of quantum contextuality with either single photons or coherent light. *npj Quantum Information*, 7:154, 2021.

[50] M. S. Leifer. Is the quantum state real? an extended review of ψ-ontology theorems. *Quanta*, 3: 67, 2014.

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

[52] A. Peres. Two simple proofs of the kochen-specker theorem. *J. Phys. A: Math. Gen.*, 24:L175, 1991.

[53] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky. Simple test for hidden variables in spin-1 systems. *Phys. Rev. Lett.*, 101:020403, Jul 2008.

[54] S. Yu and C. H. Oh. State-independent proof of kochen-specker theorem with 13 rays. *Phys. Rev. Lett.*, 108:030402, 2012.

[55] J. E. Cohen, J. H. B. Kempreman, and Gh. Zbăganu. *Comparisons of stochastic matrices, with applications in information theory, statistics, economics, and population sciences*. Birkhäuser, 1998.

[56] G. Dahl. Matrix majorization. *Linear Algebra Appl.*, 288:53, 1999.

[57] F. D. Martínez Pería, P. G. Massey, and L. E. Silvestre. Weak matrix majorization. *Linear Algebra Appl.*, 403:343, 2005.

[58] J. E. Cohen and U. G. Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra Appl.*, 190:149, 1993.

[59] S. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM J. Optim.*, 20:1364, 2009.

[60] T. Lee, Z. Wei, and R. de Wolf. Some upper and lower bounds on psd-rank. *Math. Program.*, 162:495, 2017.

[61] J. Gouveia, P. A. Parrilo, and R. R. Thomas. Lifts of convex sets and cone factorizations. *Math. Oper. Research*, 38:248, 2013.

[62] Y. Shitov. The complexity of positive semidefinite matrix factorization. *SIAM J. Optim.*, 27: 18981909, 2016.

[63] B. Schumacher and M. Westmoreland. *Quantum Processes, Systems, and Information*. Cambridge University Press, Cambridge, 2010.

[64] K. Matsumoto and G. Kimura. Information storing yields a point-asymmetry of state space in general probabilistic theories. arXiv:1802.01162 [quant-ph], 2018.

[65] T. Heinosaari and L. Leppäjärvi. Random access test as an identifier of nonclassicality. *J. Phys. A: Math. Theor.*, 55:174003, 2022.

[66] P. Skrzypczyk and N. Linden. Robustness of measurement, discrimination games, and accessible information. *Phys. Rev. Lett.*, 122:140403, 2019.

[67] M. Oszmaniec and T. Biswas. Operational relevance of resource theories of quantum measurements. *Quantum*, 3:133, 2019.

[68] Y. Kuramochi. Compact convex structure of measurements and its applications to simulability, incompatibility, and convex resource theory of continuous-outcome measurements. arXiv:2002.03504 [math.FA], 2020.

[69] T. Heinosaari, L. Leppäjärvi, and M. Plávala. No-free-information principle in general probabilistic theories. *Quantum*, 3:157, 2019.

[70] G. Kimura, J. Ishiguro, and M. Fukui. Entropies in general probabilistic theories and their application to the Holevo bound. *Phys. Rev. A*, 94:042113, 2016.

[71] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke. Entropy and information causality in general probabilistic theories. *New J. Phys.*, 12:033024, 2010.

[72] S. Bandyopadhyay, R. Jain, J. Oppenheim, and C. Perry. Conclusive exclusion of quantum states. *Phys. Rev. A*, 89:022336, 2014.
[73] C. M. Caves, C. A. Fuchs, and R. Schack. Conditions for compatibility of quantum-state assignments. *Phys. Rev. A*, 66:062111, 2002.
[74] R. W. Spekkens. Evidence for the epistemic view of quantum states: A toy theory. *Phys. Rev. A*, 75:032110, 2007.
[75] N. Harrigan and R. W. Spekkens. Einstein, incompleteness, and the epistemic view of quantum states. *Found. Phys.*, 40:125, 2010.
[76] R. Colbeck and R. Renner. A systems wave function is uniquely determined by its underlying physical state. *New Journal of Physics*, 19:013016, 2017.
[77] L. Hardy. Are quantum states real? *International Journal of Modern Physics B*, 27:1345012, 2013.
[78] M. S. Leifer and O. J. E. Maroney. Maximally epistemic interpretations of the quantum state and contextuality. *Phys. Rev. Lett.*, 110:120401, 2013.
[79] M. F. Pusey, J. Barrett, and T. Rudolph. On the reality of the quantum state. *Nature Physics*, 8: 475, 2012.
[80] J. Barrett, E. G. Cavalcanti, R. Lal, and O. J. E. Maroney. No ψ-epistemic model can fully explain the indistinguishability of quantum states. *Phys. Rev. Lett.*, 112:250403, 2014.
[81] C. Branciard. How ψ-epistemic models fail at explaining the indistinguishability of quantum states. *Phys. Rev. Lett.*, 113:020409, 2014.
[82] M. Ringbauer, B. Duffus, C. Branciard, E. G. Cavalcanti, A. G. White, and A. Fedrizzi. Measurements on the reality of the wavefunction. *Nature Physics*, 11:249, 2015.
[83] G. C. Knee. Towards optimal experimental tests on the reality of the quantum state. *New J. Phys.*, 19:023004, 2017.
[84] M. Navascués, S. Pironio, and A. Acín. Bounding the set of quantum correlations. *Phys. Rev. Lett.*, 98:010401, 2007.
[85] M. Navascués, S. Pironio, and A. Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New J. Phys.*, 10:073013, 2008.
[86] A. Tavakoli, E. Cruzeiro, and R. Uola. Bounding and simulating contextual correlations in quantum theory. *PRX Quantum*, 2:020334, 2021.
[87] A. Chaturvedi, M. Farkas, and V. Wright. Characterising and bounding the set of quantum behaviours in contextuality scenarios. *Quantum*, 5:484, 2021.
[88] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, England, 2004.
[89] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49, 1996.
[90] V. Havlíček and J. Barrett. Simple communication complexity separation from quantum state antidistinguishability. *Phys. Rev. Research*, 2:013326, 2020.
[91] V. Russo and J. Sikora. A note on the inner products of pure states and their antidistinguishability. arXiv:2206.08313 [quant-ph], 2022.
[92] M. S. Leifer and C. Duarte. Noncontextuality inequalities from antidistinguishability. *Phys. Rev. A*, 101:062113, 2020.
[93] T. Heinosaari, M. A. Jivulescu, and I. Nechita. Random positive operator valued measures. *J. Math. Phys.*, 61:042202, 2020.
[94] C. A. Fuchs, M. C. Hoang, and B. C. Stacey. The sic question: History and state of play. *Axioms*, 6:21, 2017.
[95] P. E. Frenkel and M. Weiner. Classical information storage in an n-level quantum system. *Commun. Math. Phys.*, 340:563, 2015.
[96] R. F. Werner and M. M. Wolf. Bell inequalities and entanglement. *Quantum Inf. Comput.*, 1:1, 2001.
[97] Y. C. Liang and A. C. Doherty. Better Bell-inequality violation by collective measurements. *Phys. Rev. A*, 73:052116, 2006.

[98] Y. C. Liang and A. C. Doherty. Bounds on quantum correlations in Bell-inequality experiments. *Phys. Rev. A*, 75:042103, 2007.

[99] Y. C. Liang, C. W. Lim, and D. L. Deng. Reexamination of a multisetting Bell inequality for qudits. *Phys. Rev. A*, 80:052116, 2009.

[100] A. Ambainis, M. Banik, A. Chaturvedi, D. Kravchenko, and A. Rai. Parity oblivious d-level random access codes and class of noncontextuality inequalities. *Quantum Inf. Process.*, 18:111, 2019.

[101] M. Navascués and T. Vértesi. Bounding the set of finite dimensional quantum correlations. *Phys. Rev. Lett.*, 115:020501, 2015.

[102] M. Navascués, A. Feix, M. Araújo, and T. Vértesi. Characterizing finite-dimensional quantum behavior. *Phys. Rev. A*, 92:042117, 2015.

[103] P. Wittek. Algorithm 950: Ncpol2sdpa-sparse semidefinite programming relaxations for polynomial optimization problems of noncommuting variables. *ACM Trans. Math. Softw.*, 41:1, 2015.

[104] P. Mironowicz. *Applications of semi-definite optimization in quantum information protocols, Ph.D. thesis*. Gdansk University of Technology, 2015.

[105] Y. Wang, I. W. Primaatmaja, E. Lavie, A. Varvitsiotis, and C. C. W. Lim. Characterising the correlations of prepare-and-measure quantum networks. *npj Quantum Inf.*, 5:17, 2019.

[106] A. Tavakoli, E. Zambrini Cruzeiro, E. Woodhead, and S. Pironio. Informationally restricted correlations: a general framework for classical and quantum systems. *Quantum*, 6:620, 2022.

[107] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde. Preparation contextuality powers parity-oblivious multiplexing. *Phys. Rev. Lett.*, 102:010401, 2009.

[108] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens. An experimental test of noncontextuality without unphysical idealizations. *Nat. Commun.*, 7:11780, 2016.

[109] D. Schmid and R. Spekkens. Contextual advantage for state discrimination. *Phys. Rev. X*, 8: 011015, 2018.

[110] D. Saha and A. Chaturvedi. Preparation contextuality as an essential feature underlying quantum communication advantage. *Phys. Rev. A*, 100:022108, 2019.

[111] R. Kunjwal, M. Lostaglio, and M. F. Pusey. Anomalous weak values and contextuality: Robustness, tightness, and imaginary parts. *Phys. Rev. A*, 100:042116, 2019.

[112] A. Tavakoli and R. Uola. Measurement incompatibility and steering are necessary and sufficient for operational contextuality. *Phys. Rev. Research*, 2:013011, 2020.

[113] E. D. Andersen. Certificates of primal or dual infeasibility in linear programming. *Comp. Opt. Applic.*, 20:171, 2001.

[114] R. J. Duffin and A. C. Schaeffer. A class of nonharmonic fourier series. *Trans. Am. Math. Soc.*, 72:341, 1952.

[115] I. Daubechies. *Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1992.

[116] R. Balan, P. Casazza, C. Heil, and Z. Landau. Density, overcompleteness, and localization of frames. *Electron. Res. Announc. Am. Math. Soc.*, 12:71, 2006.

[117] T. Strohmer and Jr. R. W. Heath. Grassmannian frames with applications to coding and communication. *Appl. Comput. Harmon. Anal.*, 14:257, 2003.

[118] N. Leonhard. *Correlation minimizing frames, Ph.D. thesis*. University of Houston, 2016.

[119] L. R. Welch. Lower bounds on the maximum cross-correlation of signals. *IEEE Trans. Inform. Theory*, 20:397, 1974.

[120] O. Christensen. *Frames and Bases: An Introductory Course*. Birkhäuser, Basel, Switzerland, 2008.

[121] C. W. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, New York, 1976.

[122] M. Paris and J. Řeháček. *Quantum State Estimation, Vol. 649 of Lecture Notes in Physics*. Springer-Verlag, Berlin, 2004.

[123] D. Qiu and L. Li. Minimum-error discrimination of quantum states: Bounds and comparisons. *Phys. Rev. A*, 81:042329, 2010.

[124] E. R. Loubenets. General lower and upper bounds under minimum-error quantum state discrimination. *Phys. Rev. A*, 105:032410, 2022.

[125] K. Nakahira, T. S. Usuda, and K. Kato. Upper and lower bounds on optimal success probability of quantum state discrimination with and without inconclusive results. *Phys. Rev. A*, 97:012103, 2018.

[126] S. Yang, J. Lee, and H. Jeong. Entropic lower bound for distinguishability of quantum states. *Adv. Math. Phys.*, 2015:683658, 2015.

[127] D. Spehner. Quantum correlations and distinguishability of quantum states. *J. Math. Phys.*, 55: 075211, 2014.

[128] D. Qiu and L. Li. Relation between minimum-error discrimination and optimum unambiguous discrimination. *Phys. Rev. A*, 82:032333, 2010.

[129] G. García-Pérez, O. Kerppo, M. A. C. Rossi, and S. Maniscalco. Experimentally accessible non-separability criteria for multipartite entanglement structure detection. arXiv:2110.04177 [quant-ph], 2021.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama Oy, Turku, Finland 2022

ISBN 978-951-29-9126-6 (PRINT)
ISBN 978-951-29-9127-3 (PDF)
ISSN 0082-7002 (PRINT)
ISSN 2343-3175 (ONLINE)