Perspectives on the Formalism of Quantum Theory

by

Cozmin Ududec

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2012

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Authorship Statement

The contents of Chapter Three are based on a collaboration led by Markus Müller, and in particular the publication:

- M. P. Müller and **C. Ududec**. "The structure of reversible computation determines the self-duality of quantum theory." Phys. Rev. Lett., 108(13):130401 (2012),

with the manuscript largely prepared by Müller. Passages and figures have been adapted for inclusion here with his consent. In particular, the results in Section 3.1 are translations from reference [102] made by myself with the help of Google Translate. The results in Section 3.2 were proved by myself. The concept of "bit symmetry" was introduced Müller. Results 3.23 and 3.24 were largely proved by Müller.

The contents of Chapter Four are based on a collaboration with Howard Barnum and Joseph Emerson, which was led by myself. Sections 4.1–4.7 are partially based on the publication:

- **C. Ududec**, H. Barnum, and J. Emerson. "Three Slit Experiments and the Structure of Quantum Theory." Foundations of Physics [Special Issue: Advances in Quantum Theory], 41(3), 396-405 (2010),

which was largely prepared by myself. Passages and figures have been adapted for inclusion here with the consent of the other authors. In particular, the analysis and discussions in Sections 4.1, 4.2, and 4.3 are my own. The translation of Sorkin's hierarchy of interference expressions into the formalism of operational probabilistic theories was done by myself. The idea to use filters (Definition 4.4) and the projective state space assumption was made by Barnum. The definitions in Section 4.5 were proposed in collaboration with Barnum, with result 4.17 proved by myself. The results in sections 4.6, 4.7,and 4.8 were proved by myself, with the exception of result 4.28, which was proved by Barnum. The results in section 4.9 are my own as well.

The contents of Chapter Five are based on an even collaboration with Howard Barnum and Markus Müller. Principles 1 and 2 were suggested by Müller, and Principle 3 by myself. The results in Section 5.3 are largely by Müller, with the exceptions of results 5.27, 5.31, and 5.33–5.40 (unless otherwise cited) which were largely proved by myself. The proof of result 5.48 was inspired by comments in Chapter 9 of [7]. The remaining results in Section 5.4, as well as the results in Section 5.5 are my own, unless otherwise cited.

The contents of Chapter Six are based on a collaboration with Joseph Emerson and Nathan Wiebe. In particular, result 6.3 and the contents of Appendix D are my work. Result 6.4 was proved in collaboration with Wiebe. Section 6.4 is based on discussions with Emerson and Wiebe.

Abstract

Quantum theory has the distinction among physical theories of currently underpinning most of modern physics, while remaining essentially mysterious, with no general agreement about the nature of its principles or the underlying reality. Recently, the rise of quantum information science has shown that thinking in operational or information-theoretic terms can be extremely enlightening, and that a fruitful direction for understanding quantum theory is to study it in the context of more general probabilistic theories. The framework for such theories will be reviewed in the Chapter Two.

In Chapter Three we will study a property of quantum theory called self-duality, which is a correspondence between states and observables. In particular, we will show that self-duality follows from a computational primitive called bit symmetry, which states that every logical bit can be mapped to any other logical bit by a reversible transformation.

In Chapter Four we will study a notion of probabilistic interference based on a hierarchy of interference-type experiments involving multiple slits. We characterize theories which do not exhibit interference in experiments with k slits, and give a simple operational interpretation. We also prove a connection between bit symmetric theories which possess certain natural transformations, and those which exhibit at most two-slit interference.

In Chapter Five we will focus on reconstructing the algebraic structures of quantum theory. We will show that the closest cousins to standard quantum theory, namely the finite-dimensional Jordan-algebraic theories, can be characterized by three simple principles: (1) a generalized spectral decomposition, (2) a high degree of symmetry, and (3) a generalization of the von Neumann-Luders projection postulate. Finally, we also show that the absence of three-slit interference may be used as an alternative to the third principle.

In Chapter Six, we focus on quantum statistical mechanics and the problem of understanding how its characteristic features can be derived from an exact treatment of the underlying quantum system. Our central assumptions are sufficiently complex dynamics encoded as a condition on the complexity of the eigenvectors of the Hamiltonian, and an information theoretic restriction on measurement resources. We show that for almost all Hamiltonian systems measurement outcome probabilities are indistinguishable from the uniform distribution.

Acknowledgements

First I would like to thank my advisor Joseph Emerson for all his guidance, patronage, and encouragement, as well as his friendship over the course of my studies. I am sincerely grateful for the broad freedom he gave me over the years, and for his patience with my snail-like writing speed.

Thanks very much to my collaborators Howard Barnum and Markus Mueller for the time and effort they have put into our joint work, for the many useful discussions, for listening to my many vague suggestions and trying to make something of them. Thank you also to Nathan Wiebe for doing things with Matlab and probabilities that I couldn't do.

To Lucien Hardy and Chris Fuchs for sparking my interest in quantum foundations seven years ago, and for helping maintain it with their encouragement and stimulating conversations. To the whole Quantum Foundations group for the wonderful and inspiring atmosphere, as well as all the other colleagues and friends from the Perimeter Institute and the Institute for Quantum computing who I've had the pleasure of sharing offices, discussions, and lunches with. To all the other members of the scientific community who have played crucial roles in supporting me over the years.

My parents are pretty great as well. Thanks Mom and Dad for having me, without you none of this would have been possible. Thank you for the boundless love and support, and for appropriately prodding me during my lazier moments and encouraging me to achieve something.

To all the wonderful friends from Waterloo as well as from Winnipeg, for the uncountably many beers and great conversations, for putting up with my insatiable desire for danceoffs at Starlight, all the ridiculous nights, and most of all for your true friendship.

A special thank you Nathan Wiebe for providing useful feedback on Chapter Six, and to to Markus Mueller for reviewing endless drafts of most of the thesis and providing invaluable feedback.

Financial support for my doctoral studies has been provided by the government of Canada through NSERC, the province of Ontario through OGS, the University of Waterloo, the Perimeter Institute for Theoretical Physics, and the Institute for Quantum Computing.

Lastly, I would like to thank the members of my advisory and examination committees for the time they took to examine my research critically and help me to present my work in its best light, as well as for going easy on me during my defense.

Dedication

To everyone dedicated to changing the course of physics.

Table of Contents

List of Figures

Chapter 1

Introduction

Quantum theory currently underpins most of modern physics, and is also essential in many other fields such as chemistry, biology, and cosmology, and has countless technological and engineering applications. Its development over the last 112 years has been along broadly similar lines to other major upheavals in scientific thought, such as the Copernican revolution. Beginning from an inability to explain certain experimental data from a single scientific specialty, new simple models were developed which were based on various ad hoc postulates, concepts, and techniques. The researchers who proposed and – eventually – accepted the new conceptions did so because they economically described what was known, and in spite of the conflict with physical intuition and the basic concepts of other sciences and philosophy. Eventually, the new concepts and techniques were synthesized into a mathematical formalism which became a basic tool for explaining and exploring phenomena far outside the original scientific specialty, and had a large impact on the broader intellectual world and culture.

However, unlike in the Copernican revolution where the new conceptions were eventually understood to follow from simple principles and a plausible and intuitive picture of the underlying reality, by most accounts quantum phenomena remain rather mysterious and counterintuitive, and the formalism appears ad hoc. It would be easy to fill a PhD thesis with statements from physicists as well as philosophers to the effect that they do not understand the world-view underlying quantum mechanics, but only know how to use the theory. Every current description of quantum theory is simply a statement of its mathematical formalism – a very specific and abstract set of principles for how to represent experiments and make predictions – without any fundamental explanation other than its usefulness. Why does this particular formalism work so well for such a broad range of physical systems and phenomena?

We tend to think that we understand some phenomenon or physical theory when we can take it apart (physically or conceptually) and precisely describe the properties and functioning of all the constituent parts, and then put them back together. Going back to the analogy with the Copernican revolution, Kepler's three laws of planetary motion can also be seen as empirically adequate but essentially ad hoc. They were only truly understood with the development of the Newtonian laws of motion and gravitation. Another illustrative example is the history of the Lorentz transformations. These were discovered through the study of Maxwell's equations, with no forthcoming explanation for why these – rather than the Galilean transformations – should appear. Einstein's contribution was to account for these transformations by deriving them from the simple postulates that the laws of physics are the same in every inertial frame, and that the speed of light is independent of its source.

Providing this form of explanation for quantum theory is important for several reasons. First, deeper and more reasonable principles can dissolve the mysteries of quantum phenomena and make them more intuitive. Second, it can be argued that this is essential to making progress on problems such as formulating a quantum theory of gravity, as well as for developing other potentially more accurate and more fundamental theories. More practically, this approach can shed light on what is responsible for the power of quantum information processing and cryptography. Finally, it can help the philosophically minded get a better night's sleep knowing the nature of reality underlying quantum phenomena.

A more modest approach is to first understand some phenomenon or physical theory by placing it in a relevant context and comparing it with other similar phenomena or theories. Recently the rise of quantum information science has shown that thinking in operational or information-theoretic terms can be extremely enlightening, and further, that a fruitful direction for understanding quantum theory is by studying it in the context of more general probabilistic theories (sometimes called *operational probabilistic theories*). These are theories which describe sets of experiments constructed from various hypothetical or real devices, and which assign probabilities to events in these experiments. Essentially any aspect or property of the quantum formalism can be generalized and studied in this framework, and it becomes possible to ask which characteristics of classical or quantum theory are typical of the framework, and how these characteristics relate to each other. The various parts of the framework for such theories will be introduced and reviewed in Chapter Two.

In Chapter Three we will study a mathematical property of the quantum formalism, often called *self-duality*, which is essentially a correspondence between states and observables. This correspondence lies at the very heart of quantum theory, and can be understood as one of the main ingredients in the Born rule. It is interesting to ask why there is such a correspondence, and what it means operationally. We will study self-duality by placing it

in the larger context of general probabilistic theories and looking for natural or operational principles which imply it. In particular, we will show that self-duality follows from a property called *bit symmetry*, which states that every logical bit can be mapped to any other logical bit by a reversible transformation, and which we argue is necessary for powerful computation.

In Chapter Four we will study a concept which has played a central role in the development and interpretation of quantum theory and is one of the most characteristically quantum phenomena, that is, interference. We will study *probabilistic interference* (rather than classical wave, or wave-function interference), and how it relates to other properties of a probabilistic theory. We will take as a point of departure the work of Raphael Sorkin who introduced a hierarchy of interference-type experiments involving multiple slits. Our first result characterizes theories which do not exhibit interference in experiments with k slits, to which we give a simple operational interpretation. Then we give a sharper characterization of those theories which do not exhibit interference in experiments with three or more slits. In the process we also prove that the state spaces of finite-dimensional Jordan-algebras exhibit at most two slit interference. Finally, we prove a connection between bit symmetric theories which possess certain natural transformations, and those bit symmetric theories which exhibit at most two slit interference.

As mentioned above, the rise of quantum information science has shown that the mathematical principles of quantum theory have many interesting operational or information theoretic consequences. This has led to a renewed interest in the axiomatization problem, and in particular whether the quantum formalism can be derived from principles with an information-theoretic flavor. In Chapter Five we will focus in particular on reconstructing the algebraic structures of quantum theory for individual systems, giving an answer to the question "why should the set of observables of a system form any algebra at all, let alone of the C^*-algebra variety?". While not recovering the exact (complex) structure of standard quantum theory, we will show that its closest mathematical cousins, namely the finite-dimensional formally real Jordan-algebraic theories, can be characterized by three simple principles: (1) a generalized spectral decomposition, (2) a high degree of symmetry, and (3) a generalization of the von Neumann-Luders projection postulate. Finally, we also show that the absence of interference in experiments with three or more slits may be used as an alternative to the third principle.

In Chapter Six, we change direction and focus on quantum statistical mechanics and the problem of understanding how its characteristic features can be derived rigorously from an exact treatment of the underlying quantum mechanical system. In recent years there has been a resurgence of interest in this problem and progress in providing rigorous justification for the existence and time scale of equilibration of small subsystems of a

larger system. We will demonstrates an alternative, complementary mechanism for micro-canonical equilibration of isolated quantum systems which are at all times in a pure state, without invoking decoherence or coarse-graining of observables. Our central assumptions are (1) "sufficiently complex" but reversible dynamics, and (2) an information theoretic restriction on the resources required to distinguish the dynamical quantum state from the relevant micro-canonical state. More precisely, by sufficiently complex we mean that if the Hamiltonian is represented in the basis of an observable of interest, its eigenstates can be accurately modeled by those typical of a unitary chosen from the Haar measure. The form of equilibration we will study is an *information theoretic* one, meaning that it is based on well-motivated operational constraints on the difficulty of predicting the detailed evolution of the system and the infeasibility of collecting an astronomically large amount of measurement data. We will show that such equilibration is mathematically generic in the sense that it holds for almost all Hamiltonian systems, where "almost all" is defined with respect to the Gaussian Unitary Ensemble (GUE). We will conclude by suggesting less restrictive assumptions on the Hamiltonian eigenvectors under which information theoretic equilibration is expected to hold.

Chapter 2

General Probabilistic Theories

If asked what physics is about, the typical physicist may respond that the fundamental goal is to give an account of what is real and how it behaves at the deepest level possible. Further, theoretical tools which postulate unobservable underlying mechanisms in order to explain observations can be very useful, and give us hints about what is real. This may be called the ontological view of physical theories. There is also an opposing point of view common throughout the history of physics, that the goal of physics is to efficiently organize our experiences and to enable us to predict and control phenomena with ever increasing precision and scope. Postulating unobservable underlying mechanisms should be done very carefully, and generally theoretical constructions should not be taken literally. This may be called the *operational* point of view. This view can be understood as a general methodology rather than as a basic philosophical position that we should not speak of any reality beyond our direct experience of instrument settings and detector clicks. In particular, in the absence of some fundamental picture of the world, this methodology allows us to proceed toward the next fundamental physical theory in a careful, conceptual fashion. Important ideas can be formulated in operational terms, without committing to a particular mathematical framework.

Recently the rise of quantum information science has shown that thinking in operational or information-theoretic terms can be extremely enlightening. An enormous number of protocols and tasks, such as the possibility of teleportation, the no cloning theorem, secure key distribution, new and powerful algorithms, and many more besides [137], have been shown to follow from the mathematical structure of quantum theory. It is then interesting to ask what is responsible for the power of quantum information processing and cryptography, and whether there are general connections between information processing and physical principles. Is anything canonical, or otherwise special, about the particular

mathematical framework of quantum theory? The question is then, compared to what? We need a way of studying the various properties of quantum theory so that these can be generalized, modified, contrasted and related to each other without the whole package of quantum theory necessarily coming along for the ride.

In the following sections we will review the various parts of the framework of *operational probabilistic theories* (sometimes also called general probabilistic theories). These are theories which describe sets of experiments constructed from various hypothetical or real devices, and efficiently assign probabilities to events in these experiments. This largely canonical mathematical framework has been developed and rediscovered several times over the last 60 years [120, 119, 100, 55, 65, 66, 31, 85, 95, 122, 26, 53, 45]. Almost any conceivable statistical physical theory, including quantum theory and classical probability theory as special cases, can be described within this framework.[1]The main innovation over, say, standard statistical theory or probability theory is the existence of many distinct and essentially incompatible or incomparable tests/contexts/instruments/operations (or 'sample spaces' in the standard statistical language) which can be used to study some phenomenon. Further, these frameworks allow us to bypass the interpretational issue, because the mathematical structure is designed with a goal in mind, after we have decided what we are talking about. This is in contrast to the usual approach of setting up a mathematical formalism and then attempting to interpret it.

Over the last decade there have been many studies of information theoretic protocols and tasks within this framework. Some examples are: the relationship between powerful computation and dynamics [26, 84], properties of information theoretic quantities like entropy [20, 156], general no-cloning and no-broadcasting theorems [21, 22], steering [25], teleportation [23], correlations between systems [24, 104, 139], symmetry [179], theories with purifications [45], and state discrimination [111, 109]. There have also been many other studies with a similar theme, but not within the general probabilistic theories framework, such as on non-locality [145, 170, 41, 141, 27, 126, 160], and other types of modifications of quantum theory [1, 136, 183].

We begin in Section 2.1 by introducing the language of operational theories. Then in Section 2.2 we introduce probabilities into this language, and discuss the notion of an operational probabilistic theory. In Sections 2.3 and 2.4 we begin the mathematical development of this framework and discuss the notions of states and effects, and various assumptions related to these, and finally in Section 2.5 we discuss transformations.

[1]One important exception is that the presentations cited above generally only deal with experiments represented in a fixed space-time or causal structure. Recently, new frameworks have been developed in order to describe and deal with situations in which there are no such fixed structures [96, 98, 99], while still maintaining an operational point of view.

2.1 Operational Theories

The presentation and terminology of this section is largely based on that of [45, 98, 99], where a graphical language inspired by the circuit model for quantum computation and the category theoretic point of view of physical theories [3, 48] is used. As we will not require many of the details of this graphical language going forward, we only give a taste of it in order to anchor the discussion of what an operational theory is, and then move on to the mathematical framework for general probabilistic theories.

Imagine that we are in a laboratory where we have available a large set of experimental apparatuses or devices (such as Stern-Gerlach magnets, beam splitters, photon counters, antennas, ovens, etc.) which we can use to study some hypothetical physical systems. These devices can be connected together in sequences and in parallel, making up larger apparatuses and complete experimental arrangements. Each device is assumed to be uniquely specified by a list of instructions on how it can be used and connected to other devices.[2] A single use of such an apparatus (during some finite time interval generally specified by an external clock) in a single run of an experiment will be called a *test* [178, 45]. If two of the same type of apparatus are used in a particular experimental arrangement, or the same apparatus is used twice at different times in a single experimental run, these will be labeled as unique tests. Further, each device has labeled *input* and *output* 'ports', which can come in various *types*. These labels can be thought of as parts of the instruction set for how apparatuses can be connected together. In particular, two apparatuses can be connected in sequence only if the output of the first device is of the same type as the input of the second. Alternatively, the input/output labels can be taken to represent *systems* (such as photons, electrons, or cats) which can pass between the apparatuses. [3] Finally, each apparatus has a finite set of possible *outcomes*, which are distinct and easily identifiable by graduate students (such as a set of distinct lights, or a pointer with some range of possible positions). The occurrence of a particular outcome (or set of outcomes) in a particular experiment corresponds to an *event*, and is usually taken to signal the occurrence of a particular *operation* on the input systems.[4] Collecting the above pieces, we have the following

[2] We are also assuming that various physical parameters which specify an apparatus (the orientation of a Stern-Gerlach apparatus, for example) are described classically as real-valued numbers.

[3] Two implicit assumptions here are: (1) the universe can (effectively, or for experimental purposes) be partitioned into such apparatuses, the systems passing between them, and the background of the rest of the universe, and (2) many (effectively, or for experimental purposes) independent copies of each subsystem can be produced.

[4] Sometimes devices are also taken to have settings, but we can also take anything we might call a setting on some device, and make that part of the definition of a particular operation, i.e., devices with different settings give different operations.

definition:

Terminology 2.1 (**Tests**). *A* test*, $\mathcal{O}$, is defined as a single use of a particular apparatus in a particular experimental arrangement. Each test has an associated set of* inputs *and* outputs*, labeled $A, B, C, ...,$. Further, each test $\mathcal{O}$ has an* outcome set *$o(\mathcal{O})$ with distinct outcomes i, and associated operations $\{\mathcal{O}_i\}$. We will often refer to a test $\mathcal{O}$ by its associated set of operations $\{\mathcal{O}_i\}$.*

Generally, in the frameworks discussed above, preparations, transformations, and observations are taken as the primitive objects, but following [45, 98, 99] we will define preparations and observations as follows:

Terminology 2.2 (**Preparations and Observations**).
Tests $\{\mathcal{P}_i\}$ with no inputs (or more generally, connected sets of tests with no open inputs) will be called preparations*, and the corresponding outcomes* preparation-events*.*
Tests $\{\mathcal{M}_i\}$ with no outputs (or more generally, connected sets of tests with no open outputs) will be called observations*, and the corresponding outcomes will be called* observation-events*.*

Full experimental setups are built up by connecting together many tests. These connections must be ordered or *directed* in the sense that the output of one test can only be connected to the input of a test following it. In particular, starting from some apparatus and following connections from its outputs to the inputs of the next apparatus, and so on, we must never be able to arrive back where we started as this would correspond to a closed time-like curve (recall a test corresponds to a single use of an apparatus). For more detail and a more formal treatment of sequential and parallel composition and their rules see [45, 98, 99].

Terminology 2.3 (**Operational theory**). *An operational theory is specified by a collection of well defined tests (and the associated systems) and rules for connecting them together in sequence and in parallel. This collection must be closed in the sense that any well connected set of tests gives another valid test in the theory.*

2.2 Operational Probabilistic theories

In a sense an operational theory is a kind of language within which we can discuss and describe experiments, or more generally some physical phenomena of interest. This is not yet a complete language for physical theories, as whatever else a physical theory is or does, it should at least give us the ability to assign probabilities to the outcomes of instruments in at least some well defined set of situations or experiments.

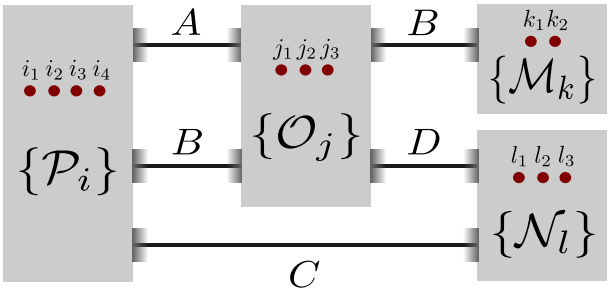

Figure 2.1: An abstraction of an experimental setup with a preparation $\{\mathcal{P}_i\}$, a test $\{\mathcal{O}_j\}$ and two observations $\{\mathcal{M}_k\}$, $\{\mathcal{N}_l\}$. Also labeled are the possible outcomes of these devices, as well as their inputs and outputs. Note that the lines labeled A, B, C, D connecting the tests represent how these devices are connected, not the physical connections themselves.

The question of what probabilities are, or how they should be interpreted and assigned to events has a long history [76]. While it is possible that the nature of probabilities will play an important role in resolving the foundational problems of quantum theory [71], we will remain agnostic on this issue (possibly at our own peril) and take classical probability theory and its standard well known formalism, as well as the assignment of probabilities to events as primitives. The reader can fill in his favorite interpretation and assignment method at will.

A natural requirement is that the probabilities assigned should be well defined (or *well conditioned* [99]) in the sense that they do not depend on events occurring outside the experiment of interest. For example, if we have an experiment which consists of some connected set of tests with some open inputs, then the probabilities of the outcomes in this experiment may depend on whether there is something influencing these inputs. Generally it is good experimental practice to ensure that there are no uncontrolled influences (at least within experimental error) on the phenomena of interest. In operational language, this means that the tests and connections assumed are accurately specified, and that there are no open inputs or outputs in the experiment of interest.

Assumption 1. *The probabilities of the outcomes in any connected set of tests with no open inputs or outputs depend only on the tests and connections given, and not on external tests and events.*

In effect, any observed dependency on external influences can be interpreted as an incorrect specification of the tests of interest and their inputs and outputs. This assumption

was explicitly stated in [99, 98], as well as being implicit in how probabilities are assigned in [45]. See also [80] for further discussion of related issues.

Given a well specified and connected set of tests $\{\mathcal{M}_i\}, \{\mathcal{N}_j\}, \{\mathcal{O}_k\}, \{\mathcal{P}_l\}, ...$, with no open inputs or outputs (such as in Figure 2.1), we will write the joint probabilities of the events $(i, j, k, l, ...)$ as

$$Pr\left(i, j, k, l, ..., \mid \mathcal{M}, \mathcal{N}, \mathcal{O}, \mathcal{P}\right).$$

Terminology 2.4 (**Operational probabilistic theory**). *An* operational probabilistic theory *(sometimes also called a* general probabilistic theory*) is an operational theory, together with an assignment of valid joint probabilities, $Pr(i, j, k, l, ..., \mid \mathcal{M}, \mathcal{N}, \mathcal{O}, \mathcal{P})$, to the events in any well specified and connected set of tests $\{\mathcal{M}_i\}, \{\mathcal{N}_j\}, \{\mathcal{O}_k\}, \{\mathcal{P}_l\}, ...$, with no open inputs or outputs.*

A further assumption which is often implicit in the framework of general probabilistic theories is the arrow defined by the output $\mapsto$ input relations between tests in an experiment corresponds exactly to the arrow defined by 'causal' influence in the experiment (or more precisely, the arrow defined by the ability of one test to detectably influence the outcome probabilities of another following it in an experimental setup). In particular, given an observation $\{\mathcal{M}_j\}$, we can connect one of several distinct preparations to the inputs of $\{\mathcal{M}_j\}$, and the probabilities of the outcomes $\{j\}$ will in general depend on which preparation we select. However, given a preparation $\{\mathcal{P}_i\}$, where none of the outcomes $\{i\}$ are certain to occur, we can ask whether their probabilities depend on the observation which is connected to the output of $\{\mathcal{P}_i\}$. The possibility of this kind of "signaling from the future" was pointed out in [45, 98, 99]. We will explicitly state this assumption below – and later use an equivalent characterization of it proven in the above references – without going into any greater detail.

Assumption 2 (**No-Signaling from the future**). *For every preparation $\{\mathcal{P}_i\}$, and every observation $\{\mathcal{M}_j\}$ on a system A, the marginal probabilities $p_i := \sum_{j \in o(\mathcal{M})} Pr(i, j|\mathcal{P}, \mathcal{M})$ are independent of the observation $\{\mathcal{M}_j\}_{j \in o(\mathcal{M})}$. In particular, if $\{\mathcal{M}_j\}$ and $\{\mathcal{N}_k\}$ are two distinct observations, then*

$$\sum_{j \in o(\mathcal{M})} Pr(i, j|\mathcal{P}, \mathcal{M}) = \sum_{k \in o(\mathcal{N})} Pr(i, k|\mathcal{P}, \mathcal{N}). \tag{2.1}$$

2.3 States, Effects, and Convexity

We now have all the prerequisites to discuss the mathematical formalism of operational probabilistic theories which will use extensively in the remaining chapters. Note that in

this section (and the remainder of the thesis) we will refer to various concepts and results stated in various appendices. For example, Appendix A contains a review of well known notions from the theory of convex geometry in finite dimensional real vector spaces which are useful for our development.

An important issue in probabilistic theories is whether two devices which have different constructions or instruction sets, but nevertheless have the same inputs, outputs, and outcomes are distinguishable probabilistically. More precisely, we will say that two tests $\mathcal{O}, \mathcal{O}'$ are *probabilistically indistinguishable* if for any experimental setup containing $\mathcal{O}$ in a given spot, the joint probabilities of all events of the experiment do not change when $\mathcal{O}$ is replaced with $\mathcal{O}'$. This allows us to define *equivalence classes* of tests, where each class contains all tests which are probabilistically indistinguishable [119, 100].

Terminology 2.5 (**States, effects, and transformations**).
An equivalence class of probabilistically indistinguishable preparation-events $\mathcal{P}_i$ for a system A is called a state. *The set of all states for a system A is denoted by $\mathsf{S}(A)$.*
An equivalence class of probabilistically indistinguishable observation-events $\mathcal{M}_j$ for a system A is called an effect. *The set of all effects for a system A is denoted by $\mathsf{E}(A)$.*
An equivalence class of probabilistically indistinguishable operation-events $\mathcal{O}_k$ is called a transformation. *The set of all transformations from A to B is denoted by $\mathsf{T}(A, B)$.*

In the following we will identify events and their equivalence classes, and use the same notation for both. Next, we will use these equivalence classes to encode the conditional probabilities $Pr(\mathcal{M}_j|\mathcal{P}_i)$ of observation-events given preparation-events in a pair of real, finite dimensional vectors. One method for carrying out this encoding is to use states $\mathcal{P}_i$ to define functionals $\omega_{\mathcal{P}_i}$ which map effects to probabilities:

$$\omega_{\mathcal{P}_i} : \mathsf{E}(A) \to [0, 1]; \quad \mathcal{M}_j \mapsto Pr(\mathcal{M}_j|\mathcal{P}_i). \tag{2.2}$$

Similarly, each effect $\mathcal{M}_j$ can be used to define a functional $e_{\mathcal{M}_j}$ mapping states to probabilities:

$$e_{\mathcal{M}_j} : \mathsf{S}(A) \to [0, 1]; \quad \mathcal{P}_i \mapsto Pr(\mathcal{M}_j|\mathcal{P}_i). \tag{2.3}$$

Essentially, the functional $\omega_{\mathcal{P}_i}$ associated with the state (i.e., the equivalence class of the preparation-event) $\mathcal{P}_i$ is exactly the mathematical object needed to calculate the probability of any observation event $\mathcal{M}_j$ (which can be validly connected to the preparation device $\{\mathcal{P}_i\}$).[5] From a probabilistic or operational point of view this functional contains

[5]Note that in this encoding procedure we have implicitly assumed that the functionals for states and effects are finite dimensional objects so that we do not have to worry about boundedness and other mathematical technicalities. If the total number of preparation and observation tests is finite, then this is trivial. We will state this assumption formally below.

all the relevant information about the state. For this reason, from now on we will identify states $\mathcal{P}_i$ with their corresponding functionals $\omega_{\mathcal{P}_i}$. Similarly, from now on we will identify effects $\mathcal{M}_j$ with their corresponding functionals $e_{\mathcal{M}_j}$. Under this identification, it becomes clear that the conditional probabilities $Pr(\mathcal{M}_j|\mathcal{P}_i)$ can be calculated from the appropriate functionals as

$$Pr(\mathcal{M}_j|\mathcal{P}_i) = e_{\mathcal{M}_j}(\mathcal{P}_i) = \omega_{\mathcal{P}_i}(\mathcal{M}_j). \tag{2.4}$$

In the following we will generally use letters $\omega, \phi, \psi, ...$, for states, and $\mathsf{S}(A)$ for the set of these functionals, and further denote effects by letters $e, f, g, ...$, and the set of these functionals by $\mathsf{E}(A)$.

We should emphasize at this point that the notion of probabilistic indistinguishability and the resulting notions of states, effects, and transformations are heavily dependent on the assumed starting set of tests. For example, adding or removing an observation device from the valid set may change the equivalence classes of the preparation devices, and the structure of the resulting set of states. As we will see, this will generally have a large impact on the predictions of the operational probabilistic theory.

We can now formally take real linear combinations of the state functionals and effect functionals. The sets of these linear combinations define two real vector spaces which we denote by

$$\begin{aligned} A &:= \text{lin}(\mathsf{S}(A)), \\ A^* &:= \text{lin}(\mathsf{E}(A)). \end{aligned} \tag{2.5}$$

From the fact that probabilities are calculated as $e(\omega)$ for $e \in A^*$, $\omega \in A$, it is clear that A^* is the dual vector space to A (see Appendix A.13).

An alternative but conceptually similar route to states and effects as real vectors is discussed in [122]. The idea is that we can form a probability table where each column corresponds to a preparation event $\mathcal{P}_i$, and each row corresponds to an observation event $\mathcal{M}_j$. The elements of the table are the conditional probabilities, $Pr(\mathcal{M}_j|\mathcal{P}_i)$. With a little linear algebra it becomes clear that this this table can be compressed into two sets of real vectors: each element of the first set corresponds to a preparation-event (more precisely, an equivalence class of probabilistically indistinguishable preparation events), and each element of the second set corresponds to an observation event (more precisely, an equivalence class of probabilistically indistinguishable observation events). Each element of the probability table is then calculated by taking the inner product of one vector from the first set and one from the second, i.e., an inner product of a state and an effect.

Next, because most physical theories have some structure and constrain probabilities of observations which are related to each other (by some symmetries, for example), it is

plausible that we will not need states to list all probabilities for all possible observation events. We therefore make the following assumption:

Assumption 3 (**Finite dimension**). *Each state requires only a* finite *number of real parameters to specify, so* $A \simeq \mathbb{R}^K$ *for some* $K < \infty$. K *is the minimum number of probabilities needed to uniquely specify a state.*

In the literature on operational probabilistic theories, this is either a background assumption, or follows from some more basic assumption [95, 45, 124, 51].

Now suppose that we have two states $\varphi, \omega \in \mathsf{S}(A)$ which correspond to events from distinct preparations, and a coin biased with probabilities $(p, 1-p)$ (or some other random process with two outcomes and corresponding biased probability distribution). If we can build devices with preparation-events φ and ω, then we can build a preparation corresponding to tossing the coin, preparing either φ or ω depending on the outcome of the coin, and then forgetting or ignoring the outcome. From the point of view of all possible observations connected to the outputs of this device and their outcome probabilities, our constructed preparation will behave exactly like the state $p\varphi + (1-p)\omega$ [100]. This state is a convex combination of φ and ω (see Appendix A.2). It is generally assumed in the operational probabilistic theories framework that the above kind of convex combination can be made for any two states, or effects, or transformations.[6] For an example of a theory in which no convexity assumption is made, see [163]. See [121] for more discussion of convexity.

Further, the probability of any observation event, $e \in \mathsf{E}(A)$, on a state such as $p\varphi + (1-p)\omega$, must be of the form $e\left(p\varphi + (1-p)\omega\right) = pe(\varphi) + (1-p)e(\omega)$, i.e., effects must be *convex linear* on states. This is because the information about which state, φ or ω, was actually prepared in each run of the experiment could be subsequently revealed to the observer, and he could check whether his data was consistent which each sub-ensemble of prepared states. By a similar argument, states must be convex linear on effects, and further, transformations must be convex linear on states as well as effects. It can be shown that convex linearity of effects implies that they must in fact be fully linear on A [26], justifying our embedding of $\mathsf{E}(A)$ in the dual vector space to A. Similar statements hold for states and transformations.

Assumption 4 (**Convexity**). *All state spaces, effect spaces, and transformation spaces are convex. Further, states, effects, and transformations act convex linearly.*

[6] The kind of justification given for convexity is standard in the literature on general probabilistic theories, but can be criticized for being too focused on preparations in a laboratory, and apparently not dealing with preparations (such as a star exploding in a far away galaxy) outside of experimental control. However, a more generally applicable justification can be given based on the idea of 'situations of uncertainty'; see [121] and references therein.

2.4 Abstract state spaces

It will be convenient in the following to include arbitrary positive real multiples of states (and effects) in our mathematical formalism, that is, elements of the form $\lambda \cdot \omega$ for $\lambda \geq 0$ and $\omega \in \mathsf{S}(A)$, and similarly for effects. To this end, define

$$\begin{aligned} S_+(A) &:= \{\lambda \cdot \omega \mid \lambda \geq 0,\ \omega \in \mathsf{S}(A)\}, \\ E_+(A) &:= \{\lambda \cdot e \mid \lambda \geq 0,\ e \in \mathsf{E}(A)\}. \end{aligned} \tag{2.6}$$

These sets are closed with respect to sums and convex combinations, i.e., they are *cones* (see Appendix A.10). By construction $S_+(A)$ spans the whole space A, and similarly, $E_+(A)$ spans the whole space A^*. Further, because each pairing of a state with an effect gives a positive real number, we have the inclusions

$$\begin{aligned} S_+(A) &\subseteq E_+(A)^*, \\ E_+(A) &\subseteq S_+(A)^*, \end{aligned} \tag{2.7}$$

where the sets $S_+(A)^*$, and $E_+(A)^*$ are the dual cones of $S_+(A)$ and $E_+(A)$ (see Appendix A.14). More precisely, the set $S_+(A)^*$ is the set of all mathematically conceivable functions taking states to non-negative real numbers. In particular, we may have theories where the set of physically possible effects is strictly smaller than the set of mathematically consistent functions mapping the set of states $S_+(A)$ to positive reals, i.e., $E_+(A) \subsetneq S_+(A)^*$. Physically, this may be interpreted as arising from something like a 'super-selection rule' which restricts the set of physical effects. In fact, the equality $E_+(A) = S_+(A)^*$ of what is physically possible and what is mathematically consistent is an important issue in many reconstructions of quantum theory [124, 51, 45, 99], and is usually taken as a simplifying assumption in the general probabilistic theories framework.

Definition 2.6 (**No-restriction hypothesis**). *A general probabilistic theory satisfies the* no-restriction hypothesis *if* $E_+(A) = S_+(A)^*$ *for all systems.*

In the following we will explicitly state whenever we use the no-restriction hypothesis. Note that this hypothesis is not exactly an operational one, as it refers to the set of "positive functionals" (see Appendix A.14), which are a mathematical concept. In [46], Chiribella et al. use a closely related but weaker and more operational axiom called *perfect distinguishability* in place of the no-restriction hypothesis. We conjecture that any use of the no-restriction hypothesis can be replaced by perfect distinguishability, but we do not pursue this further.

Next we look at how the sets of states and effects are embedded in the cones which they generate. Recall that an observation $\mathcal{M}$ can be considered as a set of effects $\{e_j\}_{j \in \mathsf{o}(\mathcal{M})}$.

Now take a state $\omega \in S_+(A)$, and consider the probabilities $e_j(\omega)$, and in particular the sum $\sum_{j\in \mathsf{o}(\mathcal{M})} e_j(\omega)$, which must be a valid probability. Operationally, this corresponds to carrying out the observation $\mathcal{M}$ and forgetting or ignoring which event occurred in some particular run of the experiment. In particular, there will be states such that this sum is equal to 1 for all observation.

Definition 2.7 (**Normalized states**). *A state $\omega \in S_+(A)$ is* normalized *if for all observations $\mathcal{M} = \{e_j\}_{j\in \mathsf{o}(\mathcal{M})} \subset \mathsf{E}(A)$, we have $\sum_{j\in \mathsf{o}(\mathcal{M})} e_j(\omega) = 1$. The set of normalized states will be denoted by $\Omega(A)$.*

Next recall the no-signaling from the future Assumption 2, which states that if $\mathcal{M} = \{e_j\}$ and $\mathcal{N} = \{f_k\}$ are two distinct observations, then $\sum_{j\in \mathsf{o}(\mathcal{M})} e_j(\omega) = \sum_{k\in \mathsf{o}(\mathcal{N})} f_k(\omega)$ for all states $\omega \in \mathsf{S}(A)$. By linearity, we can write

$$\sum_{j\in \mathsf{o}(\mathcal{M})} e_j(\omega) = \Big(\sum_{j\in \mathsf{o}(\mathcal{M})} e_j \Big)(\omega) = u_A(\omega), \tag{2.8}$$

where $u_A := \sum_{j\in \mathsf{o}(\mathcal{M})} e_j$ is often called the *order unit* (see Appendix A.18). Given the no-signaling from the future assumption, it is not difficult to see that u_A is unique, i.e., it is independent of the observation being summed over. Further, u_A is strictly positive on non-zero elements of the cone $S_+(A)$, and can be used to write the set of normalized states more simply as

$$\Omega(A) = \{\omega \in \mathsf{S}(A) \mid u_A(\omega) = 1\}. \tag{2.9}$$

In fact, u_A defines an affine hyperplane $H_1 = \{x \in A \mid u_A(x) = 1\}$ in the vector space A, and we have

$$\Omega(A) = S_+(A) \cap H_1 = \mathsf{S}(A) \cap H_1. \tag{2.10}$$

It is then easy to see that the set of allowed states, $\mathsf{S}(\mathrm{A})$, is the convex hull of the set of normalized states and the 0 state, namely,

$$\mathsf{S}(A) = \mathrm{conv}\left(\Omega(A), 0\right). \tag{2.11}$$

The 0 state is literally the 0 vector in A, and corresponds to a preparation of 'no system', or 'the preparation device failed'. Often the set of effects $\mathsf{E}(A)$ is written as $\mathsf{E}(A) = [0, u_A]$ and called the *order interval* (see Appendix A.17). It is clear that for all $e \in \mathsf{E}(A)$, $e \leq u_A$ in the sense of Appendix A.16.

Next we make one substantive and one simplifying assumption for the sets of states we consider.

Assumption 5 (**Boundedness**). *For any set of states $\mathsf{S}(A)$, the associated set of normalized states $\Omega(A)$ is bounded, i.e., it is contained in a ball of finite radius.*

It is not difficult to see that $\Omega(A)$ being bounded is equivalent to the property that the cone $S_+(A)$ generated by $\Omega(A)$ is a *pointed* cone, namely, $S_+(A) \cap (-S_+(A)) = \{0\}$ (see Appendix A.11).

Assumption 6 (**Closure**). *All state spaces, effect spaces, and transformation spaces are topologically closed (closed for short) in their embedding real vector spaces.*[7]

For the precise definition of what we mean by topologically closed see Appendix A.5. Generally, the vectors in the topological boundary of a state space $\mathsf{S}(A)$ can be approximated arbitrarily accurately (under an appropriate norm [7, 45]) by valid states in $\mathsf{S}(A)$. The physical interpretation of the topological closure assumption is that these vectors in the boundary are in fact valid states themselves, and can in principle be prepared. There is no observable physical difference between one of these boundary vectors and an arbitrarily close state, so it does no harm to consider them part of the state space.

Definition 2.8 (**Pure and mixed states**). *A state $\omega \in \mathsf{S}(A)$ will be called* mixed *if it can be written as a convex combination $\omega = p\varphi_1 + (1-p)\varphi_2$ for some $0 < p < 1$ and $\varphi_1 \neq \varphi_2 \in \mathsf{S}(A)$, and otherwise* extreme. *A normalized extreme states will be called* pure. *The set of extreme states will be denoted by* $\mathrm{ext}(\mathsf{S}(A))$, *and the set of pure states by* $\mathrm{ext}(\Omega(A))$.

In the language of Appendix A.3, pure states are extreme points of the convex set $\Omega(A)$, and a state is extreme in the cone of states if and only if the smallest face it is contained in is an extremal ray of the state cone (see Appendix A.7 and A.12). Given the finite-dimensionality, boundedness, and closedness assumptions, it is not difficult to see that every normalized state can be written as a convex combination of a finite number of pure states (see Appendix A.4 and following discussion).

Definition 2.9 (**Abstract state space**). *An* abstract state space, $(A, S_+(A), E_+(A), u_A)$, *describing a type of physical system is given by a finite dimensional real vector space A, a closed pointed cone $S_+(A) \subset A$, which spans A, a second closed pointed cone $E_+(A) \subseteq S_+(A)^* \subset A^*$, and a linear functional $u_A \in E_+(A)$ which is strictly positive on $S_+(A) \setminus \{0\}$. In the context of the no-restriction hypothesis, we will write $(A, S_+(A), u_A)$.*

We can also start with some bounded and closed convex set of normalized states $\Omega(A)$ and derive full cones of states and effects. Take $\Omega(A) \subset \mathbb{R}^{K-1}$, and embed it in $A = \mathbb{R}^K$

[7] We have not yet embedded transformations into a real vector space, but the method, which we outline in Section 2.5, is similar to that for states and effects. Nevertheless, we present the closure assumption for transformations in this section for the sake of flow.

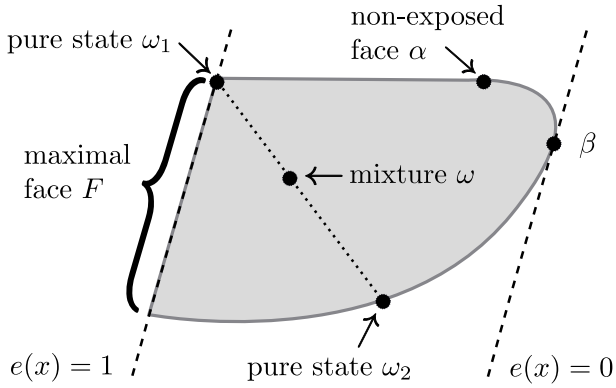

Figure 2.2: An example of a two dimensional convex set of normalized states. Displayed is a pure state α which is not exposed (Appendix A.21), two exposed pure states ω_1, ω_2 and their convex combination ω, a maximal face F (Appendix A.9) and its supporting hyperplane $e(x) = 1$ (Appendix A.20), and another extreme point β and its supporting hyperplane $e(x) = 0$. In the language of Definition 2.12, all states in F are perfectly distinguishable from β. The boundary (Appendix A.5) of the set is displayed in dark grey; all displayed points except ω are on the boundary.

in such a way that the affine span of $\Omega(A)$ in A does not contain the zero vector. The positive cone of states is then given by

$$S_+(A) := \bigcup_{\lambda \geq 0} \lambda \cdot \Omega(A), \tag{2.12}$$

and u_A is the functional on A which evaluates to 1 on $\Omega(A)$. In fact, there is a rigorous one-to-one correspondence between compact convex subsets of finite-dimensional vector spaces and cones of states [143].

Example 2.10. *The simplest and best known example of the structures described above is the formalism of classical probability theory, with a finite sample space. In particular, let $J = \{1, 2, ..., d\}$ be a d element sample space, which can be interpreted as the possible discrete outcomes of a single test on a system. The cone of states is given by*

$$S_+(d) = \{(p_1, ..., p_d)^T \in \mathbb{R}^d \mid 0 \leq p_i\}. \tag{2.13}$$

Taking the Euclidean inner product on $\mathbb{R}^d$ and using it to make the identification $A \simeq A^*$, we can write the set of all mathematically valid effects as

$$\mathsf{E}(d) = \{(e_1, ..., e_d) \in \mathbb{R}^d \mid 0 \leq e_i \leq 1\}, \tag{2.14}$$

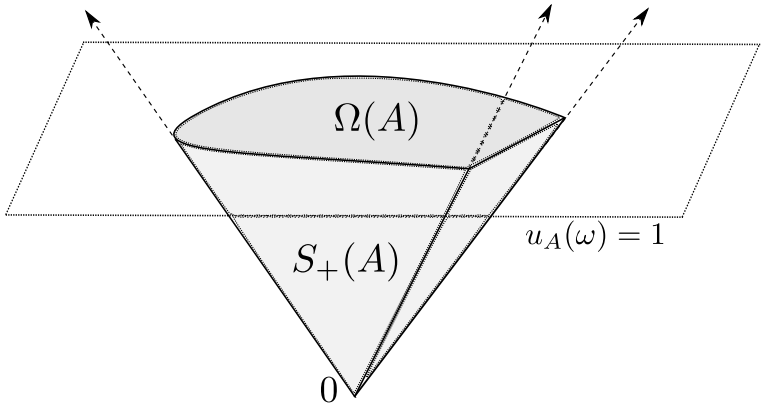

Figure 2.3: An example of a three dimensional closed pointed cone. Displayed is the hyperplane $H_1 = \{\omega \in A \mid u_A(\omega) = 1\}$ defined by an order unit u_A, as well as its intersection with the cone, namely $\Omega_A = H_1 \cap S_+(A)$. Further, the three dashed arrows are examples of extremal rays (Appendix A.12) of $S_+(A)$.

which has the shape of a hypercube. Observations are then given by collections of effects $\{e_i\}$*, such that* $\sum_i e_i = u_d = (1, 1, ..., 1)$*, where* u_d *is the order unit. This condition arises from the fact that the normalization condition is given by* $\sum_i p_i = 1$*, i.e., the order unit must the functional given by summing the components of a state. The set of normalized states therefore forms a* $(d-1)$*-simplex [34]. There are* d *pure states, each a permutation of* $(1, 0, ..., 0)^T$*.*

In order to help the reader navigate the landscape of general probabilistic theories, we now introduce a *running quantum example*. In various places throughout the following chapters we will briefly discuss the quantum representation of the concepts and mathematical objects we are using, and point out if there is any further special or interesting structure in the quantum case. We will assume the reader is familiar with the abstract structure of finite dimensional quantum theory [34, 137], so we will not dwell on the mathematical details.

Quantum Example 2.11. *Let* $\mathbb{C}^d$ *be a* d *dimensional complex Hilbert space, and* $\mathcal{H}_{(d)}$ *be the real* (d^2 *dimensional) vector space of Hermitian operators on* $\mathbb{C}^d$*. This is exactly the vector space* A *of a* d*-level quantum system. In the following we will write* $A_{(d)} = \mathcal{H}_{(d)}$ *to emphasize the type of system we are dealing with. The full cone of states of a* d*-level quantum system is then described by the set of positive semi-definite (Hermitian with non-*

negative eigenvalues) operators on $\mathbb{C}^d$, namely,

$$S_+(\mathcal{H}_{(d)}) = \{\hat{\rho} \in \mathcal{H}_{(d)} |\ \hat{\rho} \geq 0\}. \tag{2.15}$$

The dual cone of $S_+(\mathcal{H}_{(d)})$ lives in the vector space of linear functionals on $\mathcal{H}^{(d)}$, but as we will see in more detail in Quantum Example 3.4, we can represent every functional $f \in (\mathcal{H}_{(d)})^*$ by an Hermitian operator $\hat{F}$ as $f(\hat{T}) = \mathrm{Tr}[\hat{F}\hat{T}]$. It is then not difficult to see that an effect functional f is non-negative on $S_+(\mathcal{H}_{(d)})$ in the sense of Appendix A.14 if and only if its associated operator $\hat{F}$ is positive semi-definite, i.e., $S_+(\mathcal{H}_{(d)})^* \simeq S_+(\mathcal{H}_{(d)})$. Note that it is generally assumed that finite-dimensional quantum systems satisfy the no-restriction hypothesis, i.e., $E_+(\mathcal{H}_{(d)}) = S_+(\mathcal{H}_{(d)})^*$.

In particular, this representation allows us to write the probability of some event represented by the effect f in the state represented by the density matrix $\hat{\rho}$ as $Pr(f|\rho) = \mathrm{Tr}[\hat{F}\hat{\rho}]$. This is the Born rule - or at least the natural generalization of it to the full set of states and effects. The full set of effects (often called positive operator valued measurement (POVM) elements) is then given by

$$\mathsf{E}(\mathcal{H}_{(d)}) = \{\hat{M} \in \mathcal{H}_{(d)} \mid 0 \leq \hat{M} \leq \hat{\mathbb{I}}_d\} = E_+(\mathcal{H}_{(d)}) \cap (\hat{\mathbb{I}}_d - E_+(\mathcal{H}_{(d)})). \tag{2.16}$$

Observations are given by collections $\{\hat{M}_i\}$ of effects such that $\sum_i \hat{M}_i = \hat{\mathbb{I}}_d$. The order unit is therefore given by the trace linear functional, namely, $u_{\mathcal{H}_{(d)}}(\hat{T}) = \mathrm{Tr}[\hat{\mathbb{I}}_d\hat{T}]$.

The normalized states are the density matrices:

$$\Omega(\mathcal{H}_{(d)}) = \{\hat{\rho} \in \mathcal{H}_{(d)} \mid \hat{\rho} \geq 0,\ \mathrm{Tr}[\hat{\rho}] = 1\}, \tag{2.17}$$

and the set of pure states of $\Omega(\mathcal{H}_{(d)})$ is isomorphic to the set of rank-1 projection operators on $\mathbb{C}^d$. For $d = 2$ the set of density matrices is the well known Bloch ball [34], and for $d > 2$, this set is much more complicated than a ball [33].

Note that an alternative and more principled starting point is to assume that observation events on a quantum system should be represented by projectors on $\mathcal{H}_{(d)}$ (or more generally POVM elements), and then to use Gleason's Theorem [77] (or generalizations thereof [42, 70]) to derive that the states must be density matrices.

For more examples of these structures see Sections 3.1, 4.10 as well as [100, 54, 28, 26, 143, 104, 121, 33, 7].

All the concepts we have introduced so far, and the many more we will introduce in what follows, are essentially convex-geometric in nature. As we will see, it is the various

properties of the geometries of the state, effect, and transformation spaces of a particular theory which define and determine the predictions of the theory and its information processing features and capabilities. Essentially, for any theory we have

convex geometry $\Leftrightarrow$ *physical predictions and information processing capabilities.*

In particular, in the context of quantum theory, we will take the point of view that the Hilbert space structures and concepts of the usual presentation are simply a compact representation of the relevant convex structures. For example, the quantum trace rule, $Pr(e|\rho) = Tr[\hat{E}\hat{\rho}]$, is a particular representation of the general state and effect pairing we have introduced, which results from the geometry of states and effects which quantum theory postulates.

Next we briefly discuss the important notion of perfectly distinguishable states, and the effects which distinguish them. Suppose we have a device which prepares one of a set $\omega_1, \ldots, \omega_n \in \mathsf{S}(A)$ of n normalized states, and we wish to determine which one. In general, all we can do is perform an m-outcome observation $\{e_j\}_{j=1}^m$ (generally many times on many copies of the prepared state), and look at the relative frequencies of the outcomes. Clearly these will depend on the set of states we started with, and also on the observation chosen, and in general there may be no conceivable observation that can tell us with certainty in a single observation which state was prepared. If there is such an observation, we will say that the states $\omega_1, \ldots, \omega_n$ are *perfectly distinguishable*:

Definition 2.12 (**Perfectly distinguishable states**). *A set $\omega_1, \ldots, \omega_n$ of normalized states are* perfectly distinguishable *if there exists an n-outcome observation $\{e_j\}_{j=1}^n$ such that*

$$e_j(\omega_i) = \delta_{ij}. \tag{2.18}$$

The observation $\{e_j\}_{j=1}^n$ will be called a discriminating observation *for the states $\omega_1, \ldots, \omega_n$.*

As a simple example, suppose we take two normalized states φ and ω in $\Omega(A)$. These are perfectly distinguishable if there is an effect $e \in \mathsf{E}(A)$ such that $e(\varphi) = 0$ and $e(\omega) = 1$ (and vice versa for the effect $u_A - e$). Since all normalized states ψ have $0 \leq e(\psi) \leq 1$, the states φ and ω must lie on 'opposite sides' of the state space $\Omega(A)$, in the sense that the set of vectors $x \in A$ with $e(x) = 1$, respectively $e(x) = 0$, are two parallel supporting hyperplanes (see Appendix A.20), touching the state space in φ and ω, with the full state space $\Omega(A)$ lying in between (see Figure 2.2).

Every state space has a maximum number of states which can be perfectly distinguished, and this depends strongly on the geometry of the space. In general, this number

is much smaller than the dimension of the vector space which the state space is embedded in. An exception is classical probability theory, where the maximal number of perfectly distinguishable state is exactly equal to the linear dimension of the state cone.

Quantum Example 2.13. *Any set of density matrices, $\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_n$, with mutually orthogonal supports are perfectly distinguishable. An example of an observation which is discriminating for this set is the collection of orthogonal projectors $\{\hat{P}_i\}_{i=1}^n$, where $\hat{P}_i$ is the projector on the support of $\hat{\rho}_i$ for all $i < n$, and $\hat{P}_n = \mathbb{I} - \sum_{i=1}^{n-1} \hat{P}_i$. For a d-level system, an example of a maximal set of states which can be perfectly distinguished is a set of rank-one projectors on an orthonormal basis (recall that for a d-level system, the state space is dimension d^2).*

Quantum theory further has the interesting properties that for a d-level system, for all $n \leq d$, every set of n perfectly distinguishable pure states can be transformed into every other, and further, every mixed state can be decomposed as a convex combination of d perfectly distinguishable pure states (see [34, 33] for more discussion). These properties will be important in Chapter 5.

2.5 Transformations

In the previous sections we have focused on two particular types of instruments, namely states and effects. We will now go into more detail on transformations. Recall that transformations were defined in Section 2.3, as equivalence classes of probabilistically indistinguishable operation-events, and the set of all transformations from A to B was denoted by $\mathsf{T}(A, B)$. Transformations can be seen as describing on the one hand possible physical time evolutions, and on the other hand possible computations that can be accomplished in a particular theory [137, 26, 134].

In a similar fashion to the way functionals for states and effects were defined in Equations (2.2) and (2.3), we can encode transformations $\mathcal{O}_k$ into linear maps $T_{\mathcal{O}_k}$ from the vector space A to B, which are uniquely defined by their action on states:

$$T_{\mathcal{O}_k} : \mathsf{S}(A) \to \mathsf{S}(B); \quad \omega_A \mapsto T_{\mathcal{O}_k}\omega_A. \tag{2.19}$$

The linearity of the action on $\mathsf{S}(A)$ can be seen to follow from a similar argument to that given for the convex linearity for effects [26]. From a probabilistic or operational point of view this linear map contains all the relevant information about the associated transformation. For this reason, from now on we will identify transformations $\mathcal{O}_k$ with their corresponding maps $T_{\mathcal{O}_k}$. In the following transformations will be denoted by letters,

$T, U, V, ...$, the set of transformations (from A to B) by $\mathsf{T}(A, B)$, and the vector space of all real linear combinations of transformations by $L(A, B)$, i.e.,

$$L(A, B) := \mathrm{lin}(\mathsf{T}(A, B)). \tag{2.20}$$

Next recall that each operation $\mathcal{O}$ consists of a set of transformations, i.e., a set of linear maps $\{T_k\}_{k\in\mathsf{o}(\mathcal{O})}$. In particular, for a set of linear maps $\{T_k\} \subset L(A, B)$ to constitute a valid operation certain *necessary* conditions must be satisfied:

Definition 2.14 (**Conditions on transformations and operations**). *A set of linear maps $\{T_k\}_{k\in\mathcal{K}} \subset L(A, B)$ constitute a valid operation only if the following conditions are satisfied for all $k \in \mathcal{K}$:*

(i) $T_k(S_+(A)) \subseteq S_+(B)$, *($T_k$ is* positive*),*
(ii) $u_B(T_k\omega_A) \leq u_A(\omega_A)$, *($T_k$ is* normalization non-increasing*)*
(iii) $\sum_k u_B(T_k\omega_A) = u_A(\omega_A)$, *($\sum_k T_k$ is* normalization preserving*).*

A linear map $T \subset L(A, B)$ is a valid transformation, i.e., $T \in \mathsf{T}(A, B)$, only if it satisfies conditions (i) *and* (ii).

The set of linear maps from a system A to B which are simply positive in the above sense are often denoted by $L_+(A, B)$, and form a proper convex cone in $L(A, B)$, which by Assumption 6 is closed. Further, the set of physically valid transformations and operations is generally smaller than the set of mathematically well defined ones. Since the occurrence of a transformation is always identified by some event, we can define an effect induced by a transformation T_k as follows:

Definition 2.15 (**Induced effect**). *The* effect induced by a transformation $T \in \mathsf{T}(A, B)$ *is defined by the action*

$$f_T(\omega_A) = u_B(T\omega_A), \quad \forall\ \omega_A \in S_+(A). \tag{2.21}$$

We regard $f_T(\omega_A)/u_A(\omega_A)$ as the probability that the transformation T occurs when the input state is ω_A.

It is simple to see that for a valid operation $\{T_k\}$, the induced set of effects $\{f_k\}$ satisfy $\sum_k f_k = u_A$ and are thus a valid observation on system A. Note that distinct transformations may occur with the same probability in all allowed states. In such a case the events associated with the transformations are treated as probabilistically equivalent, and represented by the same effect functional. Finally, we define a notion of a state conditional on an event:

Definition 2.16 (**Conditional state**). *Given an operation, $\{T_k\} \subset \mathsf{T}(A,B)$, and an initial state $\omega_A \in S_+(A)$ the* state conditioned on the event k *having occurred is given by:*

$$\varphi_k = \begin{cases} \frac{u_A(\omega_A)}{f_k(\omega_A)} T_k(\omega_A), & \text{if } f_k(\omega_A) \neq 0 \\ 0, & \text{if } f_k(\omega_A) = 0. \end{cases} \tag{2.22}$$

In the following we will generally focus on transformations from a system A to itself. In particular, an important class of transformations are the reversible ones:

Definition 2.17 (**Reversible transformation**). *A transformation $T \in \mathsf{T}(A,A)$ is called a* reversible transformation *if T is invertible and T^{-1} is valid transformation. The set of all (physically allowed) reversible transformations on a system A will be denoted by $\mathcal{G}(A)$.*

In particular, if T is a reversible transformation, we must have $T(\Omega(A)) = \Omega(A)$ (T preserves the set of normalized states), and $E_+(A) \circ T = E_+(A)$ (the valid physical effects are preserved under the dual action of T). Further, the set of all reversible transformations, $\mathcal{G}(A)$, is a compact group since the set of normalized states is compact. Reversible transformations will play an important role in Chapters 3, and 5, so we define the following richer type of state space:

Definition 2.18 (**Dynamical state space**). *A tuple $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$, where $(A, S_+(A), E_+(A), u_A)$ is an abstract state space, and $\mathcal{G}(A)$ is a (possibly finite) group of reversible transformations, is called a* dynamical state space.

A final issue related to transformations deserves mention. From the fact that every operation induces an observation, there is an issue of consistency between the set of effects, $\mathsf{E}(A)$ on a system A, and the set of transformations $\mathsf{T}(A,B)$ from A to some other system B. This is most relevant when the set of transformations appears to be restricted in some way, or when the no-restriction hypothesis (Definition 2.6) for effects is not satisfied.

Quantum Example 2.19. *Transformations are given by linear maps $\Phi : \mathcal{H}_{(d)} \to \mathcal{H}_{(d')}$, which are trace non-increasing and not only positivity preserving but completely positivity preserving (also called completely positive) [34]. Operations (which are typically referred to as* quantum instruments *[55]) are collections of transformations $\{\Phi_k\}$ such that the map $\sum_k \Phi_k$ is trace preserving on all positive operators.*

Given an input normalized state $\hat{\rho}$, the normalized state conditional on an event k of an operation $\{\Phi_k\}$ is given by $\hat{\sigma}_k = \frac{\Phi_k(\hat{\rho})}{Tr[\Phi_k(\hat{\rho})]}$. The von Neumann update rule is a special case of this, when $\Phi_k(\hat{\rho}) = \hat{P}_k \hat{\rho} \hat{P}_k$, with $\{\hat{P}_k\}$ is a set of mutually orthogonal projection operators.

The reversible transformations of a d-level system act as $\hat{\rho} \to V\hat{\rho}V$, where $V \in SU(d)$, the special unitary group.

2.6 Composite Systems

There is one remaining important element in the framework of operational probabilistic theories: how to describe systems made up of multiple parts, or sub-systems. It is well known that the structure of multiple-system state and effect spaces have important consequences for the information theoretic properties of a theory. However, in the following chapters we focus exclusively on single systems, so for the sake of brevity, we will not review the relevant material here. For more on this aspect of the framework see [34, 25, 23, 22, 21, 51, 98, 99, 3, 45, 26, 104, 84, 139, 126, 179].

Chapter 3

Self-Duality

A central element of every general probabilistic theory is the distinction between preparations and observations, or more precisely, between the set of states and the set of effects. In principle, states and effects are distinct types of mathematical objects with different roles in the theory, and only have operational meaning when combined to obtain probabilities. An interesting aspect of quantum theory is that states and effects are in fact represented by identical types of mathematical objects. A simple way to see this is by looking at the well known transition probability: for any two pure states $|\varphi\rangle$ and $|\psi\rangle$, this is given by the 'overlap' between them

$$\text{Prob}(\psi \to \varphi) = |\langle\varphi|\psi\rangle|^2 = \text{Tr}\left[|\varphi\rangle\langle\varphi|\,|\psi\rangle\langle\psi|\right]. \tag{3.1}$$

This expression is often interpreted as the probability of finding a system in state $|\varphi\rangle$, given that it was previously prepared in state $|\psi\rangle$. More generally, the probability of obtaining an outcome described by the POVM element $\hat{E}$, measured on a system described by the density matrix $\hat{\rho}$, is given by $\text{Tr}[\hat{\rho}\hat{E}]$. The state $\hat{\rho}$ and effect $\hat{E}$ are described by the same mathematical objects: up to normalization, they are both arbitrary positive semi-definite operators. This property, often called *self-duality*, lies at the very heart of quantum theory, and can be understood as one of the main ingredient in the Born rule. Why should there be such a correspondence between states and effects, and what does it mean operationally?

In fact there is a long tradition of reconstructions of quantum theory which use various postulates that effectively require certain states to be closely related to certain effects. For example, Hardy [99] has recently taken as a postulate that "associated with any given pure state is a unique [extremal] effect giving probability equal to one. This [extremal] effect does not give probability equal to one for any other pure state." See also Wilce [177], Guz

[90], and Gunson [87] for very similar axioms regarding the relationship between pure states and extremal effects. There are also several approaches (see Chapter 9 of Alfsen and Shultz [7], and Araki [9]), which in fact derive such a relationship from other assumptions, and then go on to postulate a "symmetry of transition probabilities". In these works transition probabilities are understood in an analogous fashion to the quantum case, namely, as the probability of finding a system in some state, given that it was previously prepared in another state (see (3.1) and the discussion above). Note that a statement like the above presumes that a positive outcome for some effect is evidence that the system is in a state uniquely associated with the effect. If we think of pure states as the most refined or precise preparations of a system, and extremal effects as the most refined or most basic propositions about a system, then it seems natural that there should be a correspondence between these.

In this chapter we take a different tack: we will study self-duality by placing it in the larger context of general probabilistic theories and looking for natural or operational principles which imply it. After defining self-duality more precisely in Section 3.1, we will review some interesting and useful characterizations and other results related to self-duality. In Section 3.2 we then discuss the problem of optimally distinguishing (or discriminating) a pair of states in self-dual theories. This will highlight how a mathematical property of a state space like self-duality has consequences for the form and features of operationally relevant quantities. The main technical result of this chapter is that self-duality can be understood from a dynamical point of view. In particular, in Section 3.3 we show that self-duality follows from a property called *bit symmetry*: every logical bit (defined precisely below) can be mapped to any other logical bit by a reversible transformation – which we argue is necessary for powerful computation.

3.1 Self-Duality

A straightforward but rather mathematical way to build a correspondence between states and effects is simply to represent the full cone of mathematically valid effects (i.e., all positive functionals on states), $S_+(A)^* \subset A^*$, in the vector space A containing the states $S_+(A)$. In order to do this we choose an inner product $\langle \cdot, \cdot \rangle$ on A, and use it to identify functionals $f : A \to \mathbb{R}$, with *vectors* $\vec{f} \in A$ via

$$\langle \vec{f}, x \rangle = f(x). \tag{3.2}$$

This is simply the Riesz representation theorem at the level of the underlying vector space A and its dual space A^*. For notational convenience, in the following we will often treat

A and A^* as *identical* rather than merely being *isomorphic*, and write f for a functional as well as its vector representative (it being understood that there is a choice of an inner product in the background). With this in mind, we can write

$$S_+(A)^* = \{f \in A \mid \langle f, \rho\rangle \geq 0,\ \forall\ \rho \in S_+(A)\}. \tag{3.3}$$

An interesting question is how $S_+(A)$ and $S_+(A)^*$ are related to each other, and how this changes with the inner product we have used. For any given abstract state space, there may not even be a linear relationship between $S_+(A)$ and $S_+(A)^*$, for any inner product. If there is in fact a linear relationship between $S_+(A)$ and $S_+(A)^*$, then $S_+(A)$ is said to be *weakly self-dual* [25, 104, 180].

Definition 3.1 (**Weak self-duality**). *A closed and proper cone $S_+(A)$ in the real vector space A is* weakly self-dual *if and only if there exists a linear bijection $T : A^* \to A$ such that $T(S_+(A)^*) = S_+(A)$ (such a map is often called an* order isomorphism*).*

As an example of a weakly self-dual state space, consider the following square model.

Example 3.2 (**Square state space**). *Let the cone of states be defined as:*

$$S_+(A) = \left\{ (x_1, x_2, x_3)^T \in \mathbb{R}^3 \;\middle|\; -x_3 \leq x_1, x_2 \leq x_3 \right\}. \tag{3.4}$$

This state space contains exactly four pure states:

$$\omega_1 = (1,1,1)^T, \quad \omega_2 = (1,-1,1)^T, \quad \omega_3 = (-1,1,1)^T, \quad \omega_4 = (-1,-1,1)^T, \tag{3.5}$$

with all other states being convex combinations of these. Using the standard inner product on $\mathbb{R}^3$, we can represent the effect cone $S_+(A)^*$ as

$$S_+(A)^* = \left\{ (y_1, y_2, y_3) \in \mathbb{R}^3 \;\middle|\; |y_1| + |y_2| \leq y_3 \right\}. \tag{3.6}$$

The order unit is defined by $u_A(x) := x_3$. It is not difficult to see that for any adjacent pair of pure states there is an observation with two effects such that one of the effects evaluates to 1 on these states and to 0 on the other two pure states. The map T in this case is given by $T = CR$ where R is a rotation about the y_3 axis by $\pi/4$ (or some integer multiple thereof), and C is a dilation given by $C(y_1, y_2, y_3) = (\sqrt{2}y_1, \sqrt{2}y_2, y_3)$.

It is interesting to note that the square state space can be seen as a convexified version of 'half' of a Popescu-Rohrlich box (PR-box), which displays non-local correlations stronger than allowed by quantum theory [26, 145, 27].

It should be clear from this example that all polygons with all sides of equal length (regular polygons) and with an even number of vertices are weakly self-dual. It is also interesting to note that for weakly self-dual state spaces there is a close connection between the order isomorphisms taking effects to states and certain states of bi-partite systems (with each sub-system a weakly self-dual state space) [104, 25, 23].

Definition 3.3 (**Self-duality**). *A closed and proper cone $S_+(A)$ in the real vector space A is* self-dual *if and only if there exists an inner product $\langle \cdot, \cdot \rangle$ on A such that a functional $f \in A^*$ (represented as a vector $f \in A$ through the inner product) is positive on $S_+(A)$ if and only if $f \in S_+(A)$, namely,*

$$S_+(A) = \{f \in A \mid \langle f, \rho \rangle \geq 0, \ \forall \ \rho \in S_+(A)\} =: S_+(A)^*. \tag{3.7}$$

The state cones of quantum theory are all self-dual, but this fact is usually obscured by the particular representation of linear functionals which is generally used.

Quantum Example 3.4. *To see precisely how self-duality is manifested in quantum theory, recall from Quantum Example 2.11 that for a d-level quantum system the cone of states is the set of positive semi-definite operators $S_+(\mathcal{H}_{(d)})$. The dual cone of $S_+(\mathcal{H}_{(d)})$ lives in the vector space of linear functionals on $\mathcal{H}^{(d)}$, but using the Riesz Representation Theorem, and the Hilbert-Schmidt inner product on Hermitian operators (given by $\langle \hat{X}, \hat{Y} \rangle = \mathrm{Tr}[\hat{X}\hat{Y}]$), we can represent every effect functional $f \in (\mathcal{H}_{(d)})^*$ by an Hermitian operator $\hat{F}$ as $f(\hat{T}) = \mathrm{Tr}[\hat{F}\hat{T}]$. This gives us the identification $(\mathcal{H}_{(d)})^* \simeq \mathcal{H}_{(d)}$ (as above, this can be taken to be an equality if we treat f and $\hat{F}$ as identical). It is then not difficult to see that an effect functional f is non-negative on $S_+(\mathcal{H}_{(d)})$ in the sense of Appendix A.14 if and only if its associated operator $\hat{F}$ is positive semi-definite, i.e., $S_+(\mathcal{H}_{(d)})^* \simeq S_+(\mathcal{H}_{(d)})$.*

See [33, 176] for further discussions of self-duality in the context of the quantum state spaces.

In order to see the difference between weak self-duality and self-duality, consider again the square state space from Example 3.2. Using the standard inner product on $A = \mathbb{R}^3$ and the pure state ω_1, we can define a linear map e_{ω_1} by $e_{\omega_1}(x) := \langle \omega_1, x \rangle = x_1 + x_2 + x_3$. Even though ω_1 is a valid state, e_{ω_1} is *not* a valid effect: for example $e_{\omega_1}(\omega_4) = -1 \ngeq 0$. For the square state space, $S_+(A)$ and $S_+(A)^*$ cannot be identified in this way – they will be different no matter which inner product is used [104]. More generally, regular polygons with n vertices are self-dual if and only if n is odd. For more examples of self-dual cones see [102] Example I.1.11, and [104].

3.1.1 Characterizing self-dual state spaces

Next, we discuss some useful results relating to self-duality. See also [108] for another interesting characterization of self-dual cones.

First, note that every order isomorphism $T : A^* \to A$ induces an associated bilinear form $t : A^* \times A^* \to \mathbb{R}$ defined by $t(x, y) := y(Tx)$, which is non-negative on $S_+(A)^* \times S_+(A)^*$, but not necessarily symmetric or positive-definite on $A^* \times A^*$ [104, 25, 23].

Proposition 3.5. *Let $S_+(A)$ be a closed and pointed cone in the real vector space A. Then the following are equivalent:*

(i) *$S_+(A)$ is self-dual.*

(ii) *([102] Lemma I.1.2) For all $x \in A$ there exists a unique decomposition (called the* Jordan decomposition*) $x = x^+ - x^-$, such that $x^\pm \in S_+(A)$, and $\langle x^+, x^- \rangle = 0$. Further, the map $x \to x^\pm$ is continuous (with Lipschitz constant 1).*[1]

(iii) *There exists an order isomorphism $T : A^* \to A$ whose associated bilinear form t is symmetric and positive definite, i.e., $t(x, y) = t(y, x)$ for all $x, y \in A^*$, and $t(x, x) > 0$ for all $x \in A^* \setminus \{0\}$.*

Note the similarity of (iii) with weak self-duality. What makes self-duality 'strong' is exactly the fact that the order isomorphism from weak self-duality induces the inner product. The equivalence (i) $\Leftrightarrow$ (ii) will be important throughout the following chapters, so for completeness we rehearse the proof.

Proof. ([102] Lemma I.1.2) (i) $\Rightarrow$ (ii) Take $x \in A$, and let x^+ be defined by the requirement that $||x - x^+|| = \inf\{||x - y|| \mid y \in S_+(A)\}$ (where the norm is induced by a self-dualizing inner product), namely the closest point in $S_+(A)$ to x. Because $S_+(A)$ is a closed convex set, such a point exists, and is unique (see [175] Theorem 2.4.1).

We will show that $x^- := x^+ - x \in S_+(A)^*$ and $\langle x^+, x^- \rangle = 0$. If $y \in S_+(A)$ then for every positive real λ we have:

$$||x - x^+||^2 \leq ||x - (x^+ + \lambda y)||^2 = ||x - x^+||^2 - 2\lambda\langle x - x^+, y\rangle + \lambda^2||y||^2.$$

[1] The Jordan decomposition can be seen as a special case of the the Moreau decomposition [132]: given any closed cone $A_+ \subset A$ and an inner product $\langle , \rangle$ on A, any element $x \in A$ has an orthogonal decomposition in terms of an element of the cone $S_+(A)$, and another element in $-S_+(A)^*$, where the dual here is that induced by the inner product. This is a powerful generalization of the standard orthogonal decomposition of an element of an inner product space in terms of its projections onto a subspace and its orthogonal complement.

Therefore $\langle x - x^+, y\rangle \leq 0$, which proves $x^- \in S_+(A)^* = S_+(A)$. Using the same reasoning with $y = x^+$, and $-1 \leq \lambda \leq 0$ we get $\langle x - x^+, x^+\rangle \geq 0$, that is to say $\langle x^+, x^-\rangle = 0$. This shows that $x^+ - x^-$ is a decomposition of x with the right properties. To show uniqueness, suppose that $z^+ - z^-$ is another decomposition of x. Then,

$$||x^+ - z^+||^2 = \langle x^+ - z^+, x^- - z^-\rangle = -\langle x^+, z^-\rangle - \langle z^+, x^-\rangle \leq 0,$$

where the inequality follows from $x^+, x^-, z^+, z^- \in S_+(A)$, which is self-dual. This implies that $x^+ = z^+$.

(i) $\Leftarrow$ (ii) Suppose that every element x in A has a unique Jordan decomposition given by $x = x^+ - x^-$, and further suppose that $x \in S_+(A)^*$ and also $x \notin S_+(A)$. Then $0 \leq \langle x^-, x\rangle = -||x^-||^2$, which gives $x = x^+ \in S_+(A)$. This contradicts the assumption that $x \notin S_+(A)$, so $S_+(A)^* \subseteq S_+(A)$. For the inverse inclusion, suppose that $x \in S_+(A)$ and $x \notin S_+(A)^*$. Let y be the closest point in $S_+(A)^*$ to x, which as shown in the first part of the proof satisfies $y - x \in S_+(A)^{**}$ and $\langle y - x, y\rangle = 0$. Since $S_+(A)^{**} = S_+(A)$, we also have $y - x \in S_+(A)$, and because $x \in S_+(A)$, this means that also $y \in S_+(A)$. Therefore $x = y - (y - x)$ is a Jordan decomposition of x, so by hypothesis $y - x = 0$, which implies a contradiction.

Finally, to show continuity, suppose $x, y \in A$, then

$$||x^+ - y^+||^2 = \langle x^+ - x, x^+ - y^+\rangle + \langle x - y, x^+ - y^+\rangle + \langle y - y^+, x^+ - y^+\rangle.$$

Therefore,

$$||x^+ - y^+||^2 = -\langle x^-, y^+\rangle - \langle y^-, x^+\rangle + \langle x - y, x^+ - y^+\rangle \leq ||x - y||||x^+ - y^+||,$$

so $x \to x^+$ is continuous. □

Next we state a few more important results related to self-duality which will become especially important in the next chapter. Recall from Appendix A.19 that the positive annihilator of a subset $B \subseteq S_+(A)$ is defined by $B^\bullet = \{y \in S_+(A)^* \mid y(x) = 0, \forall\ x \in B\}$. For self-dual cones we can use the inner product to write this as

$$B^\perp := B^\bullet = \{y \in S_+(A) \mid \langle y, x\rangle = 0, \forall\ x \in B\}. \tag{3.8}$$

We write $B^\perp$ for this positive annihilator to emphasize that we are working with an inner product. We now introduce and discuss certain important projections which we will use in the following chapters.

Definition 3.6. *Given a face F of a self-dual cone $S_+(A)$, let P_F be the (not necessarily positive on $S_+(A)$) symmetric projection (i.e., $P^* = P$ under the self-dualizing inner product) onto* $\mathrm{lin}(F)$. *We will say P_F is the symmetric projection associated with F.*[2]

Given $x \in S_+(A)$ and its Jordan decomposition $x = x^+ - x^-$, if we let $F = \mathrm{face}(x^+)$ (see Apendix A.8), then it is not difficult to see that $x^+ = P_F x$, and that $x^- \in F^\perp$. Further note that a projection P is symmetric under an inner product, i.e., $P^* = P$, if and only if its image is orthogonal to its kernel.

Lemma 3.7. *([102] Lemma I.1.6) Given a subset M of a self-dual cone $S_+(A) \subset A$, then $M^\perp$ is a closed face in $S_+(A)$ such that $M^\perp = M^{\perp\perp\perp} = \mathrm{face}(M)^\perp$. Further, if F is a face of $S_+(A)$ then the following are equivalent:*

(i) $x \in F^\perp$,
(ii) $x \in S_+(A)$ *and* $P_{F^\perp} x = x$,
(iii) $x \in S_+(A)$ *and* $P_F x = 0$.

Therefore $F^\perp = P_{F^\perp}(A) \cap S_+(A)$ and $P_F P_{F^\perp} = 0$.

Proof. Let $M \subseteq S_+(A)$, and $x \in M$. If $0 \leq x \leq y \in M^\perp$, then for $z \in M$ we have $0 \leq \langle x, z\rangle \leq \langle y, z\rangle = 0$. Because $M^\perp$ is a cone, this shows that it is a face, which is clearly also closed.

Next, for any $M, N \subseteq S_+(A)$, we have the implication $M \subseteq N \Rightarrow N^\perp \supseteq M^\perp$, and we also have $M \subseteq M^{\perp\perp}$ in general. Taking $N = M^{\perp\perp}$, the previous implication and inclusion imply $M^{\perp\perp\perp} \subseteq M^\perp$. Further, taking $N = M^\perp$, and using $N \subseteq N^{\perp\perp}$, we have $M^\perp \subseteq M^{\perp\perp\perp}$, which proves that $M^\perp = M^{\perp\perp\perp}$.

In order to prove that $M^\perp = \mathrm{face}(M)^\perp$, first note that because $M \subseteq \mathrm{face}(M)$, then $\mathrm{face}(M)^\perp \subset M^\perp$. For the opposite inclusion, take $x \in M^\perp$, and note that for $y \in \mathrm{face}(M)$ there exists a $\lambda \in \mathbb{R}^+$, and $z \in M$ such that $y \leq \lambda z$. Therefore, $0 \leq \langle x, y\rangle \leq \lambda\langle x, z\rangle = 0$, which proves the claim.

Now let F be a face of $S_+(A)$.
(i) $\Rightarrow$ (ii) is clear.
(ii) $\Rightarrow$ (iii) The closed subspace $\mathrm{lin}(F)$ is clearly orthogonal to $F^\perp$, which proves $P_F P_{F^\perp} = 0$.
(iii) $\Rightarrow$ (i) If $x \in F$, and $z \in S_+(A)$ is such that $P_F z = 0$, then $\langle z, x\rangle = \langle z, P_F x\rangle = \langle P_F z, x\rangle = 0$ implies $z \in F^\perp$. This further shows that $F^\perp = P_{F^\perp}(A) \cap S_+(A)$. □

[2] Given an arbitrary cone $S_+(A) \subset A$ and an inner product on A we can define a symmetric projection P_F onto each face F of $S_+(A)$. As we will only use these projections in the context of self-dual cones, we therefore only define them in this context. Further, if $S_+(A)$ is not self-dual under the chosen inner product, then these projections will not have the nice properties outlined in the next lemma.

Next we introduce a stronger form of self-duality which has been studied in [4, 15, 18, 19, 28, 29, 102].

Definition 3.8 (**Perfect cone**, [15])**.** *For a subset M of an inner product space A, define* the dual of M in its span *by*

$$M^\wedge := \{y \in \mathrm{lin}(M) \mid \langle x, y\rangle \geq 0,\ \forall\ x \in M\}. \tag{3.9}$$

A self-dual cone $S_+(A) \subset A$ is called perfect *if for every face F of $S_+(A)$, $F^\wedge = F$ (with respect to the self-dualizing inner product).*

Note that the distinction between every face being self-dual in its span (via *some* self-dualizing inner product on $\mathrm{lin}(F)$), and every face being self-dual according to the restriction of the ambient self-dualizing inner product on A is not vacuous. The latter is what defines a perfect cone. Further, it is clear that for all faces F of a self-dual cone, $F \subseteq F^\wedge$.

As mentioned above, the cone with a regular pentagon as the set of normalized states is self-dual. However, it is not perfect: the duals (according to the ambient inner product) of its maximal two dimensional faces are larger than the faces themselves, which can be seen by noting that the extremal rays of these faces are not mutually orthogonal.

Proposition 3.9 ([102] Lemma I.1.13)**.** *Assume $S_+(A) \subset A$ is a self-dual cone. If P is a symmetric projection on A which is positive on $S_+(A)$, then $P(S_+(A)) = P(A) \cap S_+(A)$ is a self-dual cone (under the ambient inner product on A) in the subspace $P(A)$, i.e., $P(S_+(A))^\wedge = P(S_+(A))$. Further, if F is a (closed) face of $S_+(A)$ such that $F^\wedge = F$, then the associated projection P_F is positive on $S_+(A)$.*

Proof. First suppose that P is a symmetric projection which is positive on $S_+(A)$. It is clear that $P(S_+(A)) = P(A) \cap S_+(A)$, and from Lemma 4.6,

$$P(A) = P(S_+(A)) - P(S_+(A)) = \mathrm{lin}(P(S_+(A))).$$

Further, $P(S_+(A))$ is a face of $S_+(A)$, so $P(S_+(A)) \subseteq P(S_+(A))^\wedge$. Next take an $x \in P(A)$, such that $\langle x, Py\rangle \geq 0$ for all $y \in S_+(A)$. Then $\langle x, Py\rangle = \langle Px, y\rangle = \langle x, y\rangle \geq 0$, and since $S_+(A)$ is self-dual, then x must be in $S_+(A)$, and in particular $x \in P(S_+(A))$. This means that

$$P(S_+(A))^\wedge = \{x \in P(A) \mid \langle x, y\rangle \geq 0\ \forall\ y \in P(S_+(A))\} \subseteq P(S_+(A)),$$

which proves that $P(S_+(A))^\wedge = P(S_+(A))$.

Next suppose that F is a face satisfying $F^\wedge = F$, i.e., F is self-dual in $P_F(A)$. Then for all $y \in S_+(A)$, and $x = P_F(y)$, by the self-duality of F we can decompose $x = x^+ - x^-$ with $x^\pm \in F$ and $\langle x^+, x^- \rangle = 0$. Then we have

$$0 \leq \langle x^-, y \rangle = \langle P_F x^-, y \rangle = \langle x^-, P_F y \rangle = -||x^-||^2,$$

which implies that $x^- = 0$, and so P_F is positive on $S_+(A)$. □

Quantum Example 3.10. *The cone $S_+(\mathcal{H}_{(d)})$ of positive semi-definite operators on a complex Hilbert space is perfect. Note that every face of $S_+(\mathcal{H}_{(d)})$ is defined by a symmetric (under the Hilbert-Schmidt inner product on $\mathcal{H}_{(d)}$) projection $P(\hat{B}) := \hat{P}\hat{B}\hat{P}$, where $\hat{B} \in \mathcal{H}_{(d)}$, and $\hat{P}$ is a symmetric projection onto a subspace of $\mathbb{C}^d$ (see [7] Theorem 5.32). These projections are positive on $S_+(\mathcal{H}_{(d)})$, so by Proposition 3.9 the faces they are associated with are self-dual in their span. In fact, the above is also true if the field of the underlying Hilbert space is the reals or quaternions.*

This property is closely related to the "crystalline" structure of quantum state spaces discussed in [34, 131], as well as to Hardy's "subspaces" axiom [95].

3.2 State distinguishability in self-dual models

In this section we discuss the problem of optimally distinguishing (or discriminating) a pair of states in self-dual theories. This will highlight how a mathematical property of a state space like self-duality has consequences for the form and features of operationally relevant quantities.

The question of how to distinguish ensembles of states (under various measures, and restrictions on the resources available) has been extensively studied in classical and quantum information theory. In quantum theory it is well know that there is no measurement which can perfectly distinguish non-orthogonal pure states. This fact plays an important role quantum key distribution, as well as many other aspects of quantum information [34, 137]. This problem has also been studied in generalized probabilistic theories in [21, 45, 111, 109]. In particular, it is known that the set of all pure states of a generalized probabilistic theory can be perfectly distinguished in a single measurement if and only if this theory is classical [21].

Here we will focus on the simplest problem of distinguishing – with minimum error, and in a single shot – two given states. Classically, a commonly used measure for how easy

it is to distinguish two probability distributions $p, q \in \mathbb{R}^d$ is the *variation distance*, defined by:

$$D_1(p,q) = \frac{1}{2}\sum_i^d |p_i - q_i|. \tag{3.10}$$

This quantity can be understood operationally in the following way: if an agent Bob is given a message in the form of an event drawn with *equal* probability from the distribution p or q, and he has to guess which distribution the event was drawn from, then it can be shown [34, 137] that the probability of successfully guessing, denoted by $P_S(p,q)$, is given by

$$P_S(p,q) = \frac{1}{2} + \frac{1}{2}D_1(p,q). \tag{3.11}$$

Turning to the general case, we wish to quantify the probability of success in choosing between two a-priori equally likely states φ, ω of a state space $(A, S_+(A), E_+(A), u_A)$, based on a single measurement. In this setting we must deal with the fact that the experimental probabilities depend both on the state and the observation device, namely, we are given samples from the distribution $p_i(E,\varphi) = e_i(\varphi)$, where $E = \{e_i\}_{i=1}^N \subset \mathsf{E}(A)$ is an N outcome observation. The idea will be to maximize the classical D_1 distance over all possible observations E, given φ and ω. It can be shown that for this problem it is sufficient to use two outcome observations [111, 34, 137]. To this end, we define a measure of distinguishability for general probabilistic theories in the following way.

Definition 3.11 (**Distinguishability**). *Let φ, ω be two arbitrary normalized states of a state space $(A, S_+(A), E_+(A), u_A)$. Then the* distinguishability *between them is defined by*

$$\Delta(\varphi,\omega) = \max_{e \in \mathsf{E}(A)} |e(\varphi) - e(\omega)|. \tag{3.12}$$

Note that the distinguishability between two states explicitly depends on the set of allowed effects $\mathsf{E}(A)$. If the state space does not satisfy the no-restriction hypothesis, i.e., $E_+(A) \neq S_+(A)^*$, then in general states will be less distinguishable than otherwise. The following properties of the distinguishability are well known in the quantum case. We omit the proofs because the same techniques can be used as in the classical and quantum proofs of the analogous statements for the D_1 distance and quantum trace distance [34, 137].

Lemma 3.12. *Let φ, ω be two arbitrary normalized states of a dynamical state space $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$. Then $\Delta(\varphi,\omega)$ has the following properties:*

(i) $0 \leq \Delta(\varphi,\omega) \leq 1$,
(ii) $\Delta(\varphi,\omega) = 0$ *if and only if* $\varphi = \omega$,

(iii) $\Delta(\varphi,\omega)=1$ *if and only if* φ *and* ω *are perfectly distinguishable,*
(iv) $\Delta(T\varphi,T\omega)=\Delta(\varphi,\omega)$ *for all reversible transformations* $T\in\mathcal{G}_A$,
(v) *subadditivity:* $\Delta(\varphi,\rho)\leq\Delta(\varphi,\omega)+\Delta(\omega,\rho)$,
(vi) *convexity:* $\Delta(\lambda\varphi+(1-\lambda)\varphi',\omega)\leq\lambda\Delta(\varphi,\omega)+(1-\lambda)\Delta(\varphi',\omega)$, *and similarly for the second argument.*

Next we will define an analogue of the quantum trace distance [34, 137] for self-dual models.

Definition 3.13 (**Trace distance**). *Given a state space* $(A,S_+(A),E_+(A),u_A)$ *which is self-dual, and two vectors* $x,y\in A$, *let* $\xi^+,\xi^-\in S_+(A)$ *be the elements of the Jordan decomposition (see Proposition 3.5) of* $x-y$, *namely* $x-y=\xi^+-\xi^-$. *Define the* trace distance *between* x *and* y *by*

$$D_{tr}(x,y)=\frac{1}{2}\langle u_A,\xi^++\xi^-\rangle. \tag{3.13}$$

Theorem 3.14. *Let the state space* $(A,S_+(A),E_+(A),u_A)$ *be self-dual, and take two normalized states* $\varphi,\omega\in\Omega(A)$. *Then the distinguishability between* φ,ω *satisfies the inequality*

$$\Delta(\varphi,\omega)\leq D_{tr}(\varphi,\omega). \tag{3.14}$$

Proof. Using the Jordan decomposition of $\varphi-\omega=\xi^+-\xi^-$ in the expression for $\Delta(\varphi,\omega)$, we have without loss of generality (else we can interchange φ and ω) that

$$\Delta(\varphi,\omega)=\max_{e\in\mathsf{E}(A)}\langle e,\xi^+\rangle-\langle e,\xi^-\rangle\leq\max_{e\in\mathsf{E}(A)}\langle e,\xi^+\rangle, \tag{3.15}$$

where we have used $\langle e,\xi^-\rangle\geq 0$. Therefore an effect of the form

$$\tilde{e}:=\arg\max_{e\in\mathsf{E}(A)}\langle e,\xi^+\rangle,$$

is sufficient. Further, because $\tilde{e}\leq u_A$, we have $\langle\tilde{e},\xi^+\rangle\leq\langle u_A,\xi^+\rangle$. Next, note that $\langle u_A,\varphi-\omega\rangle=\langle u_A,\xi^+-\xi^-\rangle=0$, which implies that $\langle u_A,\xi^+\rangle=\langle u_A,\xi^-\rangle$. This finally gives

$$\Delta(\varphi,\omega)=\frac{1}{2}2\langle\tilde{e},\xi^+\rangle\leq\frac{1}{2}\langle u_A,\xi^++\xi^-\rangle=D_{tr}(\varphi,\omega),$$

which proves the claim. □

Quantum Example 3.15. *In quantum theory, the trace distance between two Hermitian operators $\hat{A}, \hat{B}$ is defined as $D^Q_{tr}(\hat{A}, \hat{B}) := \mathrm{Tr}\,|\hat{A} - \hat{B}|/2$. Note that because all quantum state spaces are self-dual, any Hermitian operator such as $\hat{A} - \hat{B}$ can be written as the the difference of two positive operators $\hat{R}, \hat{S}$ with orthogonal supports, i.e., $\hat{A} - \hat{B} = \hat{R} - \hat{S}$ where $\hat{R}, \hat{S} \geq 0$ and $\mathrm{Tr}[\hat{R}\hat{S}] = 0$. Therefore, $\mathrm{Tr}\,|\hat{A} - \hat{B}| = \mathrm{Tr}[\hat{R}] + \mathrm{Tr}[\hat{S}]$.*

Further, for two density matrices $\hat{\rho}, \hat{\sigma}$, by Helstrom's theorem we in fact have the equality $D^Q_{tr}(\hat{\rho}, \hat{\sigma}) = \max_{\hat{E}} D_1(p, q)$, where $\hat{E} = \{\hat{E}_i\}$ is a POVM, and $p_i = \mathrm{Tr}[\hat{E}_i\hat{\rho}]$ and $q_i = \mathrm{Tr}[\hat{E}_i\hat{\rho}]$ (see 13.2 of [34]). Further, it can also be shown that $D^Q_{tr}(\hat{\rho}, \hat{\sigma}) = \max_{\hat{P}} \mathrm{Tr}[\hat{P}(\hat{\rho} - \hat{\sigma})]$, where the maximization is taken over all projection operators $\hat{P}$ [137]. The maximum is achieved by the projector onto the support of the positive part of $\hat{\rho} - \hat{\sigma}$.

Note that in the proof of Theorem 3.14, there is no guarantee that the effect $\tilde{e}$ which maximizes $|\langle e, \varphi - \omega\rangle| = \langle e, \xi^+\rangle$ achieves the maximum possible value $\langle u_A, \xi^+\rangle$. As we have seen above, in quantum theory this value is achieved by the effect which is the projector onto the support of the positive part of the difference of the two states. However, if we further assume either that one of the elements of the Jordan decomposition is ray extremal, or that $S_+(A)$ is perfect (see Definition 3.8), we get a similar behavior as in quantum theory.

Proposition 3.16. *Let the state space $(A, S_+(A), E_+(A), u_A)$ be self-dual and satisfy the no-restriction hypothesis, i.e., $E_+(A) = S_+(A)^*$, and take two distinct normalized states $\varphi, \omega \in \Omega(A)$. If one of the elements ξ^+, ξ^- in the Jordan decomposition $\varphi - \omega = \xi^+ - \xi^-$ is ray extremal in $S_+(A)$, then the distinguishability is given by $\Delta(\varphi, \omega) = D_{tr}(\varphi, \omega)$.*

Proof. Take the Jordan decomposition of $\varphi - \omega = \xi^+ - \xi^-$, and suppose, without loss of generality, that ξ^+ is ray extremal in $S_+(A)$. Next, note that $\langle u_A, \varphi - \omega\rangle = \langle u_A, \xi^+ - \xi^-\rangle = 0$, which implies that $\langle u_A, \xi^+\rangle = \langle u_A, \xi^-\rangle > 0$, where the last inequality follows from the facts that φ and ω are distinct and u_A is positive on $S_+(A) \setminus \{0\}$. Define $\psi_+ := \xi^+/\langle u_A, \xi^+\rangle$ and $\psi_- := \xi^-/\langle u_A, \xi^-\rangle$, and note that ψ_+ and ψ_- are normalized states, and ψ_+ is pure by assumption. We have

$$\Delta(\varphi, \omega) = \max_{e \in \mathsf{E}(A)} |\langle e, \xi^+\rangle - \langle e, \xi^-\rangle| = \langle u_A, \xi^+\rangle \max_{e \in \mathsf{E}(A)} |\langle e, \psi_+\rangle - \langle e, \psi_-\rangle| = \langle u_A, \xi^+\rangle \cdot \Delta(\psi_+, \psi_-).$$

Next, let $\lambda := \max_{\chi \in \Omega(A)} \langle \xi^+, \chi\rangle$, and define $e := \xi^+/\lambda$. From self-duality, the no-restriction hypothesis, and the definition of e, we have that e is a valid effect, i.e., $e \in \mathsf{E}(A)$. Note that by the Cauchy-Schwartz inequality, this maximum is achieved when χ and ξ^+ are co-linear, namely, by the pure state $\chi = \psi_+ = \xi^+/\langle u_A, \xi^+\rangle$. In particular, we have $e = \frac{\langle u_A, \xi^+\rangle}{\langle \xi^+, \xi^+\rangle}\xi^+$. From the fact that $\langle \xi^+, \xi^-\rangle = 0$, and the definition of e, it is clear that $\langle e, \psi_+\rangle = 1$ and $\langle e, \psi^-\rangle = 0$. Therefore $\Delta(\psi_+, \psi_-) = 1$ is achieved by the effect e, which further shows that $\Delta(\varphi, \omega) = \langle u_A, \xi^+\rangle = D_{tr}(\varphi, \omega)$. □

Proposition 3.17. *Let the state space $(A, S_+(A), E_+(A), u_A)$ be perfect and satisfy the no-restriction hypothesis, i.e., $E_+(A) = S_+(A)^*$. Given two normalized states $\varphi, \omega \in \Omega(A)$, then the distinguishability between them is given by $\Delta(\varphi, \omega) = D_{tr}(\varphi, \omega)$. Further, the optimizing effect is given by $\tilde{e} = Pu_A$, where P is the symmetric projection onto* face(ξ^+)*, and ξ^+ is the positive part of the Jordan decomposition of $\varphi - \omega$.*

Proof. Recall that for perfect cones, the symmetric projection onto any face is positive on $S_+(A)$. Let P be the symmetric projection onto face(ξ^+). Then

$$\langle u_A, P(\varphi - \omega)\rangle = \langle u_A, P(\xi^+ - \xi^-)\rangle = \langle u_A, P\xi^+\rangle = \langle u_A, \xi^+\rangle,$$

where we have used the Jordan decomposition, and the fact that ξ^- is contained in the orthogonal face to face(ξ^+), along with Lemma 3.7. Next, because P is symmetric under the self-dualizing inner product, we have $\langle u_A, \xi^+\rangle = \langle u_A, P\xi^+\rangle = \langle Pu_A, \xi^+\rangle$. Finally, by positivity of P, $0 \leq Pu_A \leq u_A$, i.e., Pu_A is a valid mathematical effect, and because $E_+(A) = S_+(A)^*$, Pu_A is a valid physical effect. This shows that the inequality in Theorem 3.14 is saturated by the effect Pu_A. □

Given the above expressions for the distinguishability of states of self-dual and perfect state spaces, it would be interesting to investigate which tasks or protocols generalized from quantum information theory, and in which the trace distance is relevant, have the same structure or bounds on the relevant quantities as in quantum theory.

3.3 Bit Symmetry

As is evident from results in the fields of quantum computation and information processing, the computational or information processing power of some physical device depends on the physical theory which describes the functioning of this device. Put another way, the efficiency with which we can encode information in the states of a physical system and then appropriately transform and read out this information depends on the geometries of the sets of states, effects, and transformations. In particular, the transformations allowed in some theory can be seen abstractly as representing the computations which can be carried out using systems described by the theory. This is further highlighted by recent studies of computation and information processing in general probabilistic theories [170, 141, 41, 20, 23, 20, 45]. Further, there appear to be tradeoffs between the strength of correlations and the richness of the set of allowed reversible transformations in a theory [26, 84, 139]. In fact, in [26] Barrett has conjectured that quantum theory achieves an

optimal balance between the richness of the allowed states and the allowed dynamics so that no other general probabilistic theory has exponentially more information processing power than quantum theory.

In this section we will show that self-duality is a consequence of a certain computational primitive we call *bit symmetry*. This primitive essentially demands that the theory has a rich dynamics, which we argue is necessary for powerful computation. For similar results, see [180]. As we are discussing computation in the context of general probabilistic theories, we need a notion of a *bit* appropriate to this setting.

Definition 3.18 (**Logical bit**). *Let φ and ω be pure and perfectly distinguishable states (see Definition 2.12) of a state space $(A, S_+(A), E_+(A), u_A)$. Define the* logical bit *generated by φ and ω as the smallest face of $\Omega(A)$ containing both states (see Definition A.8), namely,* $\mathrm{face}(\varphi, \omega)$.

Note that when we talk about a logical bit, $\mathrm{face}(\varphi, \omega)$, we are fixing a choice of perfectly distinguishable pure states φ, ω in that face (analogous to a choice of "basis states" in quantum theory), and the rest of the states making up the logical bit come along for the ride. Classically, a logical bit, $\mathrm{face}(\varphi, \omega)$, is exactly the set of all convex combinations of φ and ω, i.e., the line segment joining them. In general non-classical state spaces there will be many other pure states in $\mathrm{face}(\varphi, \omega)$, and many choices of "basis states" which result in the same logical bit (see Quantum Example 3.20 below).

Next recall from Definition 2.17 that a reversible transformation T is one which is invertible with the inverse being a valid transformation as well. Further recall from Definition 2.18 that a dynamical state space $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$ is a state space with a (possibly finite) group $\mathcal{G}(A)$ of reversible transformations. In this section we will only consider reversible transformations, as time evolution in our Universe seems to be fundamentally reversible, and also to make a conceptual analogy to the reversible circuit model of quantum computation [137].

Definition 3.19 (**Bit symmetry**). *A dynamical state space $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$ is called* bit-symmetric *if one of the following equivalent conditions holds:*

(i) *If φ, ω are perfectly distinguishable pure states, and so are φ', ω', then there is a reversible transformation $T \in \mathcal{G}_A$ such that $T\varphi = \varphi'$ and $T\omega = \omega'$.*

(ii) *Every logical bit can be mapped to every other logical bit by some reversible transformation.*

In physical terms, bit symmetry means that the state of any natural two-level system can be transferred to any other two-level system by a suitable reversible interaction. One

may argue that the computational power of a theory would be severely constrained if this property did not hold (see the next Quantum Example for more discussion of this point). It is interesting to compare bit-symmetric state spaces with those that are merely *symmetric*, in the sense that for every pair of pure states $\omega, \varphi \in \Omega(A)$ there exists a reversible transformation $T \in \mathcal{G}_A$ such that $T\varphi = \omega$. Symmetric state spaces have played an important role in recent studies of general probabilistic theories [26, 84, 179], as well as reconstructions of quantum theory [95, 124, 177].

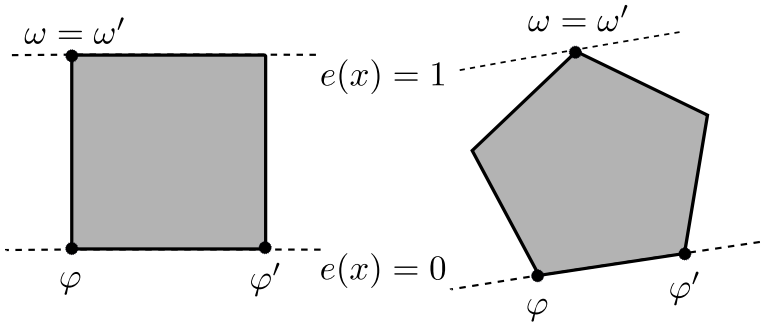

Figure 3.1: The square and pentagon state spaces. Shown are pairs of perfectly distinguishable states ω, φ and ω', φ'. For the square, there is no reversible transformation which maps the pair ω, φ to the pair ω', φ': the square state space is not bit-symmetric. For the pentagon, the pair ω, φ is mapped to ω', φ' by a reflection across a symmetry axis. In fact all pairs of perfectly distinguishable pure states can be mapped to each other: the pentagon is bit-symmetric. The dashed lines denote the level sets of an effect e which distinguishes ω and φ (and, accidentally, also ω' and φ'). For the square state space, there are two types of inequivalent logical bits: lines generated by adjacent pure states like ω, φ, and the square itself which is generated by diametral states like ω', φ'. For the pentagon – and any other bit-symmetric theory – all logical bits generated by pairs of perfectly distinguishable pure states are isometric (in this case, all pairs generate the full pentagon).

Quantum Example 3.20. *For any quantum system, two pure states $|\varphi\rangle$ and $|\omega\rangle$ are perfectly distinguishable if and only if $\langle\varphi|\omega\rangle = 0$. The logical bit,* face$(|\varphi\rangle, |\omega\rangle)$*, they generate contains all pure states of the form $\alpha|\varphi\rangle + \beta|\omega\rangle$ and their convex mixtures – that is, a full Bloch ball:*

$$\begin{aligned} \text{face}(|\varphi\rangle, |\omega\rangle) &= \text{conv}\left\{|\psi\rangle\langle\psi| \;\middle|\; |\psi\rangle = \alpha|\varphi\rangle + \beta|\omega\rangle,\ \alpha, \beta \in \mathbb{C},\ |\alpha|^2 + |\beta|^2 = 1\right\} \\ &= \left\{\hat{\rho} \in \mathcal{H}_{(2)} \mid \hat{\rho} \geq 0,\ \text{Tr}[\hat{\rho}] = 1\right\} \end{aligned} \tag{3.16}$$

In fact, in quantum theory, any *pair of pure states generate a Bloch ball [34].*

All quantum state spaces are bit-symmetric: every pair of orthogonal pure states can be mapped to every other by some unitary [34]. In fact, quantum state spaces are even more symmetric than this: analogous statements hold for triples, quadruples, etc., of orthogonal pure states. See Quantum Example 5.17 for more details and a generalization of this fact.

There is a close connection between bit symmetry and an important property of the circuit model for quantum computation. In this model we can without loss of generality start with an input 2-level system (qubit) in a particular state, as well as a number of other qubits which can without loss of generality be taken to be in the $|0\rangle$ state. Then we implement the circuit representing the computation we wish to carry out, and at the end we must measure a specific observable to determine the output of the particular computation. This last step can in principle be done without loss of generality by first reversibly transforming the – generally entangled – logical bit of interest into the first physical qubit, and then doing the desired measurement only on this. In fact this ability to transfer without destroying coherence is an important prerequisite for many algorithms [137] and is possible because quantum theory is bit-symmetric.

Classical probability theory (See Example 2.10) is bit-symmetric as well. The reversible transformations are the permutations of the d entries of the probability vectors, which can map every pair of pure states to every other. As another example, consider a regular polygon state space with n vertices. This state space is symmetric for all n, and bit-symmetric if and only if n is odd, as can be seen by inspection (see Figure 3.1 for more discussion of $n = 4$ and $n = 5$). In fact, in low dimensions bit-symmetric state spaces are rare. Using the classification of symmetric state spaces given in [110], it follows that the only bit-symmetric 2-dimensional (normalized) state spaces are the unit disc and the regular polygons with an odd number of vertices. In 3 dimensions, there is only the unit ball (i.e., the qubit) and the unique regular self-dual polytope, the tetrahedron (i.e., the classical 4-level system).

It is easy to see that for any theory the logical bit, face(φ, ω), generated by φ and ω cannot contain a third state ψ which is perfectly distinguishable from both φ and ω. If this were the case, then φ and ω would be in a supporting hyperplane (defined by $e(x) = 0$, where e is an effect distinguishing the pair φ, ω from ψ) of face(φ, ω), and therefore in a face of face(φ, ω). However, this does *not* imply that there cannot be *some* set of three perfectly distinguishable states in face(φ, ω). An example of this phenomenon is the triangular pillow theory from Example 4.55, where the face generated by the 'north' and 'south' pole states is the full pillow, and for which the three pure states generating the triangular base are perfectly distinguishable. For bit-symmetric theories, the situation is simpler.

Proposition 3.21. *Suppose $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$ is bit-symmetric, and let φ and*

ω be pure and perfectly distinguishable states. Then the maximum number of mutually perfectly distinguishable states in the logical bit, face(φ, ω)*, is two.*

Proof. Suppose that face(φ, ω) contains three perfectly distinguishable states ψ_1, ψ_2, ψ_3. Because $\Omega(A)$ is bit-symmetric and ψ_1, ψ_2 are perfectly distinguishable, there is a reversible transformation $T \in \mathcal{G}_A$ such that $T\psi_1 = \varphi$ and $T\psi_2 = \omega$. It is then clear that $T\psi_3$ is perfectly distinguishable from φ and ω, contradicting the fact that face(φ, ω) cannot contain a third state which is perfectly distinguishable from both φ and ω. □

3.3.1 Bit symmetry implies self-duality

In this section we present the main technical result of this chapter: bit symmetry implies self-duality.

Lemma 3.22. *If the state space $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$ is bit-symmetric, then it is symmetric. Further, there exists a unique state $\mu_A \in \Omega(A)$ with the property that $T\mu_A = \mu_A$ for all reversible transformations $T \in \mathcal{G}(A)$. We call μ_A the* maximally mixed state *on A.*

Proof. If $\omega \in \Omega(A)$ is any pure state, then there is always another pure state φ that is perfectly distinguishable from ω (unless the state space contains only a single point). Therefore bit symmetry implies symmetry. Now fix any pure state ω and define

$$\mu_A := \int_{\mathcal{G}_A} G\omega \, dG. \tag{3.17}$$

Since the group $\mathcal{G}_A$ of reversible transformations acts transitively on the pure states, this definition does not depend on the choice of the pure state ω. It is easy to check that $T\mu_A = \mu_A$ for all reversible transformations T. Now suppose that ν is another state which has the same invariance property as μ_A. We can decompose ν as a mixture of a finite number of pure states as $\nu = \sum_i p_i \phi_i$, where $\sum_i p_i = 1$. We then have

$$\nu = \int_{\mathcal{G}_A} G\nu \, dG = \sum_i p_i \int_{\mathcal{G}_A} G\phi_i dG = \sum_i p_i \mu_A = \mu_A.$$

Therefore μ_A is the unique state which is invariant with respect to all reversible transformations. □

Note that we can decompose the space A as the direct sum

$$A = \hat{A} \oplus \mathbb{R} \cdot \mu_A, \tag{3.18}$$

where $\hat{A} = \{x \in A \mid u_A(x) = 0\}$, and u_A is the order unit on $S_+(A)$. In particular, given a normalized state $\omega \in \Omega(A)$, define its *Bloch vector* as

$$\hat{\omega} := \omega - \mu_A. \tag{3.19}$$

The Bloch vector of any normalized state $\omega \in \Omega(A)$ is an element of the vector space $\hat{A}$.

Lemma 3.23. *Assume $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$ is bit-symmetric and satisfies the no-restriction hypothesis, i.e., $E_+(A) = S_+(A)^*$. Then there exists an inner product $(\cdot, \cdot)$ on $\hat{A}$ and a constant $c < 0$ such that the following hold for all $\omega, \varphi \in \Omega(A)$:*

(i) $c \leq (\hat{\omega}, \hat{\varphi}) \leq 1$.
(ii) *If ω is pure and φ is arbitrary, and $(\hat{\omega}, \hat{\varphi}) = c$, then ω and φ are perfectly distinguishable.*
(iii) *If ω and φ are arbitrary perfectly distinguishable states, then $(\hat{\omega}, \hat{\varphi}) = c$.*

Proof. Because reversible transformations preserve normalization, they leave the subspace $\hat{A}$ invariant. According to group representation theory [157], there is an inner product $(\cdot, \cdot)$ on $\hat{A}$ such that $(Tx, Ty) = (x, y)$ for all $T \in \mathcal{G}_A$ and $x, y \in \hat{A}$. By transitivity, the inner product $(\hat{\omega}, \hat{\omega})$ is independent of the pure state ω, and we can scale this product by an arbitrary positive factor such that $(\hat{\omega}, \hat{\omega}) = 1$ for all pure states ω.

Next define

$$c := \min_{\omega, \varphi \in \Omega(A)} (\hat{\omega}, \hat{\varphi}), \tag{3.20}$$

to be the minimal inner product between the Bloch vectors of any two states. Note that $c \leq (\hat{\mu}^A, \hat{\mu}^A) = 0$. Further, define a linear map $e_\omega : A \to \mathbb{R}$ for every pure state ω by linearly extending

$$e_\omega(\varphi) := \frac{(\hat{\omega}, \hat{\varphi}) - c}{1 - c}, \qquad \varphi \in \Omega(A). \tag{3.21}$$

Because $c \leq 0$, this is well-defined, and since $(\hat{\omega}, \hat{\varphi}) \geq c$, we have $e_\omega(\varphi) \geq 0$ for all $\varphi \in \Omega(A)$. Due to convexity of the induced norm $\|\hat{\omega}\| \equiv \sqrt{(\hat{\omega}, \hat{\omega})}$, all mixed states $\omega \in \Omega(A)$ satisfy $\|\hat{\omega}\| \leq 1$, with equality for the pure states. Therefore, the Cauchy-Schwarz inequality implies $(\hat{\omega}, \hat{\varphi}) \leq \|\hat{\omega}\| \cdot \|\hat{\varphi}\| \leq 1$, which further gives $e_\omega(\varphi) \leq 1$ for all $\varphi \in \Omega(A)$. In other words, for every pure state ω, the map e_ω is an effect.

Now suppose that $\omega \in \Omega(A)$ is pure and $\varphi \in \Omega(A)$ is arbitrary, and $(\hat{\omega}, \hat{\varphi}) = c$. Then $e_\omega(\varphi) = 0$ and $e_\omega(\omega) = 1$, hence φ and ω are perfectly distinguishable. This proves (ii). Moreover, if $c = 0$, we would have $(\hat{\omega}, \hat{\mu}^A) = 0 = c$, and so ω and μ^A would be perfectly distinguishable, which is impossible. Hence $c < 0$, proving (i).

Next take $\omega, \varphi \in \Omega(A)$ such that $(\hat{\omega}, \hat{\varphi}) = c$. We can decompose ω and φ into pure states ω_i and φ_j: $\omega = \sum_i \alpha_i \omega_i$, $\varphi = \sum_j \beta_j \varphi_j$, where $\alpha_i, \beta_j > 0$. Since $c = \sum_{ij} \alpha_i \beta_j (\hat{\omega}_i, \hat{\varphi}_j)$, and c is the minimal possible value, then by convexity every addend must have this value, so $(\hat{\omega}_i, \hat{\varphi}_j) = c$ for all i, j. Therefore ω_i and φ_j are pure and perfectly distinguishable. Now fix some pair i, j. If ω' and φ' are another pair of pure and perfectly distinguishable states, there is a reversible transformation T such that $T\omega_i = \omega'$ and $T\varphi_j = \varphi'$, and so $(\hat{\omega}', \hat{\varphi}') = (T\hat{\omega}_i, T\hat{\varphi}_j) = (\hat{\omega}_i, \hat{\varphi}_j) = c$. That is, every pair of perfectly distinguishable pure states have inner product c between their Bloch vectors. Now suppose that ω and φ are arbitrary perfectly distinguishable states. Decomposing them as above, it follows that every ω_i is perfectly distinguishable from every φ_j, hence $(\hat{\omega}, \hat{\varphi}) = \sum_{ij} \alpha_i \beta_j (\hat{\omega}_i, \hat{\varphi}_j) = c$. This proves statement (iii). □

Theorem 3.24. *Assume $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$ is bit-symmetric and satisfies the no-restriction hypothesis, i.e., $E_+(A) = S_+(A)^*$. Then there exists an inner product $\langle \cdot, \cdot \rangle$ on A such that $S_+(A)$ is self-dual, and the following hold for all $\omega, \varphi \in \Omega(A)$:*

(i) $0 \leq \langle \omega, \varphi \rangle \leq 1$.

(ii) *$\langle \cdot, \cdot \rangle$ is invariant under all reversible transformations.*

(iii) *If ω is pure, then $\langle \omega, \omega \rangle = 1$.*

(iv) *If ω and φ are perfectly distinguishable, then $\langle \omega, \varphi \rangle = 0$.*

Proof. Let $x, y \in A$, and use the decomposition from (3.18) to write $x = x_0 \mu^A + \hat{x}$ with $\hat{x} \in \hat{A}$ (and similarly for y), and define

$$\langle x, y \rangle := \lambda x_0 y_0 + (1 - \lambda)(\hat{x}, \hat{y}), \tag{3.22}$$

where $\lambda := -c/(1-c)$. Because $c < 0$, $\lambda \in (0, 1)$. It is easy to check that this is an inner product on A, and the statements (i), (ii) and (iv) follow from the analogous statements proved in Lemma 3.23 for the inner product $(\cdot, \cdot)$. Further, because $\langle u_A, \omega \rangle = 1$, we have $\omega_0 = 1$ in $\omega = \omega_0 \mu^A + \hat{\omega}$ for all pure states (and therefore all normalized states), and (iii) follows as well.

Take an extremal ray of $S_+(A)$ which is spanned by some pure state ω, and consider the functional e_ω defined in Lemma 3.23. For arbitrary $\varphi \in \Omega(A)$ we have $e_\omega(\varphi) = \langle \omega, \varphi \rangle$. By construction, $\langle \omega, \varphi \rangle \geq 0$ for all $\omega, \varphi \in \Omega(A)$, hence the corresponding functional e_ω is contained in $S_+(A)^*$. This shows that $S_+(A) \subseteq S_+(A)^*$.

In order to show the reverse inclusion, let e be any (mathematically valid) effect such that $\mathbb{R}_0^+ \cdot e$ is an exposed ray of $S_+(A)^*$ (Appendix A.21). That is, there is some $x \in A$ with the following property:

$$\forall\ f \in S_+(A)^*,\ f(x) = 0 \ \Rightarrow\ f = \lambda e \text{ for some } \lambda \geq 0. \tag{3.23}$$

The point x defines a supporting hyperplane of $S_+(A)^*$ (Appendix A.20), touching it in the ray generated by e. Thus, either $f(x) \geq 0$ for all $f \in S_+(A)^*$, or $f(x) \leq 0$ for all $f \in S_+(A)^*$. For the latter case, we simply redefine $x \mapsto (-x)$, so that $f(x) \geq 0$ for all $f \in S_+(A)^*$, or in other words, $x \in (S_+(A)^*)^* = S_+(A)$ (Appendix A.15). Since $x \neq 0$, we have $u(x) \neq 0$, and $\omega := x/u(x)$ defines a state $\omega \in \Omega(A)$ which depends on e, and will be mixed in general.

Setting $\lambda := \max_{\varphi \in \Omega(A)} e(\varphi) > 0$ and $f := e/\lambda$, we have $f(\omega) = 0$, and the set of states φ with $f(\varphi) = 1$ is a non-empty face of $\Omega(A)$. Letting ω' be any extremal point of this face, we can see that it is a pure state which is by construction perfectly distinguishable from ω. From Lemma 3.23 (iii) we have $(\hat{\omega}, \hat{\omega}') = c$, and so $e_{\omega'}(x) = u(x)e_{\omega'}(\omega) = 0$. Due to (3.23), it follows that there is some $\lambda \geq 0$ such that $e_{\omega'} = \lambda e$. We have thus shown that every ray-exposed effect of $E_+(A) = S_+(A)^*$ is of the form $\lambda' e_{\omega'}$ for some $\lambda' > 0$ and pure state ω'. According to Straszewicz' Theorem [175], the exposed rays are dense in the set of extremal rays, so every ray-extremal effect of $S_+(A)^*$ is of this form. Thus, under the identification (3.2) of functionals with vectors through the inner product, all extremal rays of $S_+(A)^*$ are contained in $S_+(A)$. Since they generate the full cone $S_+(A)^*$, we have $S_+(A)^* \subseteq S_+(A)$. In summary, we have $S_+(A) = S_+(A)^*$ under the inner product $\langle \cdot, \cdot \rangle$ – that is, A is self-dual. □

Note that we have shown that for bit-symmetric theories if ω and φ are pure, then $\langle \omega, \varphi \rangle = 0$ implies that they are perfectly distinguishable. However, we were not able to prove this implication if both are mixed.

Recalling our discussion in the introduction of this chapter of postulates which require certain states to be closely related to certain effects, it is interesting to compare these to what we have shown above. In particular, there is a subtle distinction between functionals which are extremal in the cone $E_+(A)$ (i.e., ray extremal), and functionals which are extremal in the set of valid effects $\mathsf{E}(A) = E_+(A) \bigcap (u_A - E_+(A))$ (which we will call $\mathsf{E}(A)$-extremal). A functional which is $\mathsf{E}(A)$-extremal is necessarily ray extremal, but the converse is not the case in general. For example, the pentagon model has effects of the form $\tilde{e} = u_A - e$ where e is $\mathsf{E}(A)$-extremal, such that $\tilde{e}$ is not $\mathsf{E}(A)$-extremal but is nevertheless ray extremal. Self-duality (and therefore bit-symmetry) only guarantees a bijection between pure states and ray extremal effects. It would be interesting to look for conditions stronger than bit-symmetry which guarantee a bijection between pure states and $\mathsf{E}(A)$-extremal effects.

Finally, in Section 3.2 we discussed state distinguishability in self-dual theories. It would interesting to further consider this question for bit-symmetric theories, and to see if there are any parallels to quantum theory.

3.4 Discussion

In this chapter we have studied a special kind of relationship, called self-duality, between states and effects, which is an important part of the formalism of quantum theory. By studying the problem of optimally distinguishing (or discriminating) pairs of states, we have shown how a mathematical property of the state space has consequences for the form and features of operationally relevant quantities. We have also shown that self-duality follows from the computational primitive of bit-symmetry, which is closely related to the possibility of powerful reversible computation.

One interesting question we have not addressed is what are the possible bit-symmetric state spaces in various dimensions. In two dimensions, it is not difficult to see that the only bit-symmetric state spaces are the circle, and the regular polygons with an odd number of vertices. In higher dimensions it is not clear what non-classical and non-spherical states spaces are bit-symmetric, or if there are any at all. Further, the obvious generalizations of bit-symmetry, namely, "n-symmetry" for $n \geq 3$, may yield other interesting properties and operational consequences.

Finally, a promising direction for future research is to study composite systems, where the subsystems and/or the whole system are bit-symmetric. In [104] Janotta et al. have shown that in theories where the local state spaces are self-dual, bi-partite correlations must satisfy Tsirelson's bound (at least for certain states), but not necessarily the stronger Uffink bound. A natural conjecture is that in theories where the local state spaces are bit-symmetric, bi-partite correlations satisfy both bounds, for all states.

Chapter 4

Interference in generalized probabilistic theories

The double-slit experiment, and more generally the concept of interference have played a central role in the development and interpretation of quantum theory. The form of interference that is manifested in the double-slit experiment is one of the most characteristically quantum phenomena, and is often considered to capture the essence of quantum mechanics. For example, in the introductory chapter on quantum theory in the *Feynman Lectures on Physics* [62], Richard Feynman describes the double-slit experiment with individual electrons as "...impossible, absolutely impossible, to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the only mystery." In fact, rather than attempting to explain how it works and dissolve the mystery, Feynman explicitly limits himself to stating what the effect is and showing how to use quantum theory to predict the results.

However, in spite of its pedagogical and conceptual importance, in the vast literature on interference in quantum theory the focus has largely been on describing, analyzing, or attempting to explain only *two slit* interference, with little attention paid to the possibility of new and interesting phenomena arising when more than two slits are involved. More generally, there has been relatively little analysis of what is essential to a general notion of an 'interference experiment', in particular, *probabilistic* interference (rather than classical wave, or wave-function interference), and how it relates to other properties of a probabilistic theory.

An exception to this is the pioneering work of Raphael Sorkin in [162], where a hierarchy of interference-type experiments involving multiple slits was introduced. This hierarchy is

described by a sequence of expressions I_k, for $k = 2, 3, ...$, where each I_k is defined in terms of the outcome probabilities of a k-slit experiment. If I_k is nonzero, then the experiment is said to exhibit *k-th order interference*. Sorkin discovered the remarkable property that in quantum theory only the lowest-order expression I_2 is non-zero.

Recently several experiments have been carried out testing for the absence of third-order interference [159, 158, 161]. However, in the absence of a theoretical framework broader than the quantum formalism, it is not clear precisely why quantum theory does not exhibit higher than second-order interference, or more generally, what characteristic property of a theory (besides directly testing whether the expression I_3 is zero) three-slit experiments are testing.

In this chapter we will adapt Sorkin's hierarchy of interference expressions – originally defined in a space-time histories and measures language – to the general probabilistic theories framework. First, this will help us understand the implications of three slit type experiments. Second, it will help us understand the probabilistic structure of theories with particular interference properties, as well as how interference is related to other nonclassical phenomena possible in such theories. Finally, in a later chapter we will also see how interference can be used in reconstructions of quantum theory.

We begin in Sections 4.1 and 4.2 by analyzing two- and three-slit interference in quantum theory from an operational point of view. Next, in Section 4.3 we briefly discuss an interesting relationship between three-slit interference in quantum theory and state tomography – asymptotically convergent statistical estimation of a preparation. Section 4.4 is devoted to introducing our abstraction of the slits in the quantum multiple slit experiments as *filters*, which are simply transformations with properties that make them especially suited for describing a multiple-slit experiment. Then in Section 4.5 we are finally able to define the notion of a *generalized interference experiment*.

Our first technical results, presented in Sections 4.6, 4.7, and 4.8, characterize experiments which exhibit k-th order interference as ones in which the states that pass the slits when all slits are open cannot be written as linear combinations of states that pass when one or more of the slits are closed. The additional components can be interpreted as higher-order analogues of the off-diagonal elements of a density matrix that are often called 'coherences', and which are responsible for interference in two-slit experiments. An important corollary of this is that the lack of k-th order interference is equivalent to the possibility of doing tomography via specific sets of 'k-1 slit filtering' experiments. The content of Sections 4.1–4.7 are partly based on the publication [169].

Next, in Section 4.9, we focus in particular on I_3 and obtain a sharper characterization of theories for which $\mathsf{I}_3 = 0$ by using an interesting decomposition of the state space of

the theory analogous to the Peirce decomposition for Jordan algebras (see Chapter IV of [61]). Finally, in Section 4.9.3 we prove a connection between bit symmetric theories (see Chapter 3) which possess certain natural transformations called *non-mixing filters* [7, 99], and those bit symmetric theories which display at most second-order interference. This allows us to also prove in Section 4.10 that the state spaces of finite-dimensional Jordan algebras exhibit at most second-order interference.

4.1 Two slit interference

What do we mean when we say some phenomenon exhibits interference, or more precisely, what are the essential features of an interference experiment? Generally, a discussion of a classical or quantum interference experiment consists of analyzing the experimental setup in terms of a set of waves or wave-functions which arise from distinct – but usually correlated – sources and a common endpoint. Classically, a difference between the intensity of the wave at some endpoint and the sum of the intensities of the set of waves which are used in the decomposition is taken to be the signature of interference between the sources. Quantum mechanically the waves are replaced with wave-functions and the intensity of the wave at some point is replaced with the modulus squared of the wave-function, namely the probability of detection.

Suppose however that we are only given the probabilities of the outcomes of a set of observations, and we don't have any physical or mathematical objects which add or subtract with which to analyze the experiment. Can we define a notion of interference which is based only on experimental statistics, and which can be applied to a general probabilistic theory?

To motivate our subsequent considerations and definitions, we will first study an idealized Stern-Gerlach experiment with spin-$\frac{1}{2}$ systems (for example silver atoms). Consider the setup shown in Figure 4.1, where we have a source $\mathcal{S}$ of independent and identically prepared spin-$\frac{1}{2}$ systems, with the spin degrees of freedom of each system described by a density matrix $\hat{\rho} \in \Omega(\mathcal{H}_{(2)})$ (recall the notation from Quantum Example 2.11). Each system is then sent through a modified Stern-Gerlach apparatus, often called a *Feynman filter* [62, 73], with axis aligned along the $\vec{b}$ direction.[1] After passing the Feynman filter, the systems are measured with a standard Stern-Gerlach apparatus aligned along the $\vec{d}$ axis, together with two detectors denoted by d_1 and d_2. We represent this measurement with the POVM $\mathcal{M}_{\vec{d}} = \{\hat{D}_1, \hat{D}_2\}$, where a positive outcome of the effect $\hat{D}_l$ is associated with

the detector d_l firing.

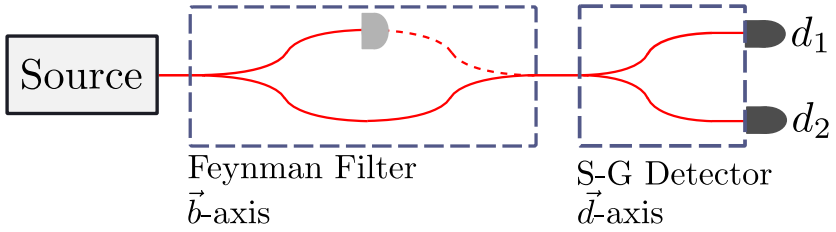

Figure 4.1: Schematic representation of a 'two-slit' experiment based on Stern-Gerlach devices. A system (with possible paths in red) emitted by the source is sent through an apparatus which acts as a filter for the spin degree of freedom. The systems that pass this filter are then subjected to a standard Stern-Gerlach measurement.

Given that we are using spin-$\frac{1}{2}$ systems, there are three possible (nontrivial) Feynman filters: one where we do not filter at all, one where we block the top path, and one where we block the bottom path. Let $\mathcal{P}_J$ denote the device constructed by *leaving open* the path(s) indexed by J, where $J \subseteq \{1,2\}$ and $J \neq \emptyset$. The transformation implemented by this device is given by $\hat{\rho} \mapsto \hat{P}_J \hat{\rho} \hat{P}_J$, where $\hat{P}_J$ is a projection operator on $\mathbb{C}^2$. Further, let e_J represent the experimental event "the system passed the filter $\mathcal{P}_J$", and let $\hat{E}_J$ be the effect associated with this event. In the idealized situation considered, we have $\hat{E}_J = \hat{P}_J$.

The probability that a system will pass the Feynman filter $\mathcal{P}_J$ given that the initial state is $\hat{\rho}$ is then given by

$$Pr(e_J | \mathcal{P}_J \;\&\; \mathcal{S}) = \mathrm{Tr}[\hat{P}_J \hat{\rho}]. \tag{4.1}$$

Further, the joint probability that a system passes the $\mathcal{P}_J$ filter *and then* the detector d_l fires is given by

$$Pr(d_l \;\&\; e_J | \mathcal{M}_{\vec{d}} \;\&\; \mathcal{P}_J \;\&\; \mathcal{S}) = Pr(d_l | \mathcal{M}_{\vec{d}} \;\&\; e_J \;\&\; \mathcal{S}) Pr(e_J | \mathcal{P}_J \;\&\; \mathcal{S}) = \mathrm{Tr}[\hat{D}_l \hat{P}_J \hat{\rho} \hat{P}_J], \tag{4.2}$$

where '&' represents time sequential conjunction – read as "and then" – and the time ordering in the conditionals is read from right to left. To simplify the notation, we will subsequently drop the $\mathcal{M}_{\vec{d}}$ from the conditional of the probabilities, it being implicit that the events d_l are conditioned on the final measurement $\mathcal{M}_{\vec{d}}$ being carried out.

[1] A Feynman filter consists of three Stern-Gerlach magnets in series. The magnets at each end are identical, while the middle one is twice as long and has reversed polarity. A beam of spin-s particles is first split into $2s+1$ spatially separated beams, which are then brought back together into one beam upon leaving the apparatus. A set of internal gates/detectors (one for each separated path) can be introduced in the middle magnet. This gives the possibility to filter the beam in many different ways by blocking a path.

Because the effects $\hat{E}_1$ and $\hat{E}_2$ form a decomposition of the effect $\hat{E}_{12} = \hat{E}_1 + \hat{E}_2 = \mathbb{I}_2$, we have the equality

$$Pr(e_2|\mathcal{P}_2 \;\&\; \mathcal{S}) + Pr(e_1|\mathcal{P}_1 \;\&\; \mathcal{S}) = Pr(e_{12}|\mathcal{P}_{12} \;\&\; \mathcal{S}). \tag{4.3}$$

This expression is in a sense trivial; all it states is that the probabilities of passing the various filters are additive. However, as Feynman points out his lectures [62], the probability of a system passing the filter $\mathcal{P}_1$ and then setting off detector d_l, plus the probability of a system passing filter $\mathcal{P}_2$ and then setting off the same detector d_l, is *not* equal to the probability of the system setting off d_l with no intermediate filtering, i.e.,

$$Pr(d_l \;\&\; e_2|\mathcal{P}_2 \;\&\; \mathcal{S}) + Pr(d_l \;\&\; e_1|\mathcal{P}_1 \;\&\; \mathcal{S}) \neq Pr(d_l \;\&\; e_{12}|\mathcal{P}_{12} \;\&\; \mathcal{S}). \tag{4.4}$$

The surprise or mystery of the quantum double-slit experiment is essentially the non-additivity of the probabilities in an expression like (4.4).[2] In classical (particle) theories it is generally assumed that experimentally inferring conditional probabilities of the form $Pr(d_l \;\&\; e_J|\mathcal{P}_J \;\&\; \mathcal{S})$ can in principle be done in a passive or non-disturbing way, or more precisely, in such a way that the filters $\mathcal{P}_J$ are not needed in the conditionals of the relevant probabilities. This is equivalent to assuming that a 'which-way' measurement at the slits has no effect on the state of the system, i.e., that we can effectively ignore the fact that a physical interaction has taken place there.

We take the lack of equality in an expression like (4.4) to be the operational meaning of a statement like "there is probabilistic interference between the top and bottom paths". More precisely,

Definition 4.1 (**Second-order interference**). *Given the preparation device $\mathcal{S}$, intermediate filtering devices $\{\mathcal{P}_{12}, \mathcal{P}_1, \mathcal{P}_2\}$, and final outcome d_l of some measurement $\mathcal{M}$, the* second order interference expression *(with respect to these devices) is given by*

$$I_2[d_l, \{\mathcal{P}_{12}, \mathcal{P}_1, \mathcal{P}_2\}, \mathcal{S}] = Pr(d_l \;\&\; e_{12}|\mathcal{P}_{12} \;\&\; \mathcal{S}) - Pr(d_l \;\&\; e_2|\mathcal{P}_2 \;\&\; \mathcal{S}) - Pr(d_l \;\&\; e_1|\mathcal{P}_1 \;\&\; \mathcal{S}). \tag{4.5}$$

Note that in the above definition we have use the term *filtering devices*, which we have not yet defined. We will define and discuss the notion of a filter in great detail in Section 4.4, but for now a filtering device can be understood as a device with similar properties as the devices implementing the slits in a quantum interference experiment, such as the Feynman filters discussed above.

[2] In the above expression, and in all of the following, it is convenient to not renormalize the state conditional on passing the slits, and to work with joint probabilities rather than conditional ones.

An important point to notice about the expression I_2 is that each of its three terms is a probability conditioned on a *distinct* set of open and blocked slits. In other words, the notion of interference between several alternatives presupposes a particular decomposition of some experimental setup into a set of separate but precisely related experiments. There is no a-priori reason for this type of expression to be identically zero. In the absence of physical input, probability theory does not constrain probabilities conditioned on different experimental situations [14, 105, 114, 112]. Nevertheless, it should not be surprising that a physical theory will in fact constrain probabilities pertaining to related experimental contexts. This gives us an interesting tool for studying the structure of a theory [151].

4.2 Three slit interference

Now that we have introduced the type of analysis we will be using, we generalize the above setup and replace the source of spin-$\frac{1}{2}$ systems with a spin-1 source, again denoted by $\mathcal{S}$. We also add another detector d_3 to the final Stern-Gerlach apparatus aligned along the $\vec{d}$ direction, and let the effect $\hat{D}_3$ be associated with the detector d_3 firing. There are now *seven* possible (non-trivial) Feynman filters, so the index set $J \subseteq \{1,2,3\}$. Let $\mathcal{P}_J$, $\hat{P}_J$, and e_J be defined as above, but with the new indexing set. With this setup in mind, we define:

Definition 4.2 (**Third-order interference**)**.** *Given a preparation device $\mathcal{S}$, a set of filters $\{\mathcal{P}_J\}_{J\subseteq\{1,2,3\}}$, and an outcome d_l of some measurement $\mathcal{M}$, we define the* third order interference *(with respect to these devices) as:*

$$\begin{aligned} \mathsf{I}_3[d_l, \{\mathcal{P}_J\}, \mathcal{S}] = Pr(d_l \;\&\; e_{123} | \mathcal{P}_{123} \;\&\; \mathcal{S}) \quad &- \sum_{1=i<j}^{3} Pr(d_l \;\&\; e_{ij} | \mathcal{P}_{ij} \;\&\; \mathcal{S}) \\ &+ \sum_{i=1}^{3} Pr(d_l \;\&\; e_i | \mathcal{P}_i \;\&\; \mathcal{S}). \end{aligned} \tag{4.6}$$

This type of expression was introduced by Raphael Sorkin in [162], who considered a set of three-slit experiments with electrons, and superimposed the seven interference patterns by using a plus sign when an odd number of the slits are open and a minus sign when an even number are open. It is interesting to note that this expression is closely related to the 'inclusion-exclusion principle', or 'sieve principle' of combinatorics and measure theory, which relates the sizes or measures of sets and their unions and intersections. This principle can be stated roughly as "the number of elements in the union of three sets is the sum of

the elements in each set respectively, plus the number of elements in the intersection of all three, minus the number of elements that are in the intersections of pairs of sets." Note that an analogous statement for the dimensions of three arbitrary subspaces – and their intersections and unions – of a real vector space fails.

As we have mentioned previously, quantum theory predicts no third-order interference:

Proposition 4.3. *Given a set of projection operators* $\{\hat{P}_J\}_{J\subseteq\{1,2,3\}}$ *acting on* $\mathbb{C}^d$*, which satisfy the relations* $\hat{P}_J\hat{P}_K = \hat{P}_K\hat{P}_J = \hat{P}_{J\cap K}$ *(and represent filtering devices* P_J *which act as* $P_J(\hat{\rho}) = \hat{P}_J\hat{\rho}\hat{P}_J$*), we have* $\mathsf{I}_3[d_l, \{\mathcal{P}_J\}, \mathcal{S}] = 0$ *for all* $\hat{\rho} \in \mathsf{S}(\mathcal{H}_{(d)})$ *(representing preparations* $\mathcal{S}$*), all final effects* $\hat{D}_l \in \mathsf{E}(\mathcal{H}_{(d)})$ *(representing events* d_l*).*

Proof. This is easy to see by working backwards: use (4.2) to expand out the joint probabilities in I_3, then use the linearity of the trace as well as the fact that the expression must hold *for all* initial states and final effects. The statement $\mathsf{I}_3[d_l, \{\mathcal{P}_J\}, \mathcal{S}] = 0$ is then seen to be equivalent to a constraint on the linear combination of projectors involved which is trivially satisfied due to the assumed relationships between them. □

4.3 Two-slit filtering tomography

Changing direction briefly, suppose we are given a device which outputs a set of identically and independently prepared d-level quantum systems (say spin-s atoms, with $d = 2s+1$), and we wish to infer a density matrix which describes the spin degrees of freedom of each prepared system. Further, suppose that to accomplish this task we are given a Stern-Gerlach apparatus which can be aligned along the z, x, and y axes *only*. Unless $d = 2$, these measurements are clearly not sufficient to estimate the density matrix for an arbitrary preparation.

However, if in addition to the above Stern-Gerlach apparatuses, we are given a Feynman filter with the possibility of blocking off all but pairs of paths i and j, such that $1 \leq i < j = 2, \ldots, d$, then we *can* in fact determine the state for any d. For each of the given filters $\mathcal{P}_{ij}$ take a sub-ensemble of the given systems and pass them through the filter, and then use the given Stern-Gerlach apparatuses aligned along the z, x, and y axes on sub-ensembles of the resulting filtered systems, and determine the frequencies of each outcome. The information gained from this filtering and measuring procedure will in fact be sufficient to infer a density matrix which describes each prepared system [73]. In other words, in order to specify a $d \times d$ density matrix, it is sufficient to do tomography on a specific set of two dimensional subspaces of filtered states. The reconstruction formula for the $d \times d$ density

matrix $\hat{\rho}$ in terms of the filtered states $\hat{\rho}_{ij} = \hat{P}_{ij}\hat{\rho}\hat{P}_{ij}$ is given by

$$\hat{\rho} = \sum_{1=i<j}^{d} \hat{\rho}_{ij} - (d-2)\sum_{k=1}^{d} \hat{\rho}_k, \tag{4.7}$$

where the $\hat{\rho}_i := P_i\hat{\rho}_{ij}P_i$ can easily be inferred once the $\hat{\rho}_{ij}$ are determined experimentally. Operationally this means that the outcome probabilities of all measurements on a d-level system can be predicted knowing only the outcome probabilities of all possible measurements on the above set of $\frac{d(d-1)}{2}$ two dimensional subspaces of filtered states. In this sense, a d-level quantum system can be thought of as an 'overlapping patchwork' of a set of $\frac{d(d-1)}{2}$, two-dimensional subspaces.

As we will see in the following, there is a very close relationship between the sufficiency of this kind of filtering tomography and the fact that quantum theory predicts the absence of third-order interference. In order to understand the structure of theories satisfying $\mathsf{I}_3 = 0$ and how tomography is related to this, we first need to abstract the essential elements of the above considerations into a setting more general than the quantum formalism.

4.4 Filters

In this section we will introduce and discuss a special class of transformations which will abstract and generalize the Feynman filters from the quantum experiments. There are many types of experimental devices which are generally thought of as filters: pinholes and slits, frequency filters, velocity filters, etc.. In spite of their often very different purposes and constructions, they generally have certain features in common: they preferentially select a type of system out of a larger system (or a certain range of values of some observable), and leave the selected system unchanged.

Filters are an essential part of the structure of quantum theory, and have played a large role in many studies and axiomatizations of the formalism. In [129, 130] Mielnik extensively analyzes filters, and connects these with the concept of "propositions" (sometimes called "yes-no measuring devices", or "ideal first-kind measurements") which have played an important role in the quantum logic tradition [32, 144]. He points out that the properties usually assumed of propositions are closely related to the properties of filters. For more on this connection see [30, 43, 88, 89]. For applications of concepts similar to filters see [143], as well as the assumptions used in the general coding theorem in [156]. Filters have also played an important role in various axiomatizations of quantum theory, such as in Guz

[90], Kummer [115, 116], Alfsen and Shultz [7], Araki [9], and more recently by Fivel [64] and Hardy [99]. In particular, Hardy takes as an axiom that filters act in a very natural way, which we will study in Sections 4.9.2 and 4.9.3, as well as in Chapter 5.

Before defining filters precisely, we make an assumption and remind the reader of a definition.

Assumption (**No-restriction hypothesis**). *Throughout this chapter we will assume the no-restriction hypothesis 2.6, i.e., $E_+(A) = S_+(A)^*$ for all abstract state spaces.*

Further, recall from Section 2.5 that a transformation T from an abstract state space $(A, S_+(A), u_A)$ to itself is a linear map which is positive, $T(S_+(A)) \subseteq S_+(A)$, and normalization non-increasing, $u_A(T_k\omega_A) \leq u_A(\omega_A)$.

Definition 4.4 (**Filters**). *Given an abstract state space $(A, S_+(A), u_A)$, a* filter*, P, is a transformation on $S_+(A)$ with the following properties:*

(i) *P is a* projection*: $PP = P$,*

(ii) *P is* complemented*: there exists at least one positive projection P', such that for all $\omega \in S_+(A)$, $P\omega = \omega$ if and only if $P'\omega = 0$, and $P'\omega = \omega$ if and only if $P\omega = 0$. If these conditions are satisfied we will say P and P' are* complementary *projections, and that P' is a* complement *of P, and vice versa.*

The first property abstracts the requirement that the state of a system which has been acted on by a filter will be unchanged if it passes through that type of filter again. The second property represents the requirement that for every filter there is another filter which acts as a 'negation' in the sense that the states which pass the filter P (respectively P') unchanged are exactly the same states which do not pass the filter P' (respectively P). Note that we do not assume that there is a unique complement to P in this definition.

In the following we will make extensive use of the notion of positive projections, and in order to deal with them efficiently and clearly we will need certain concepts which we now define and discuss.

Definition 4.5 (**Images and kernels**). *If P is a positive projection on a (closed and pointed) cone $S_+(A) \subset A$, define the* kernel*, $\ker(P)$, and the* positive kernel*, $\ker^+(P)$, of P by*

$$\ker(P) = \{x \in A \mid Px = 0\}, \quad \ker^+(P) = S_+(A) \cap \ker(P). \tag{4.8}$$

Further, define the image*, $\mathrm{im}(P)$, and the* positive image*, $\mathrm{im}^+(P)$, of P by*

$$\mathrm{im}(P) = \{x \in A \mid Px = x\}, \quad \mathrm{im}^+(P) = S_+(A) \cap \mathrm{im}(P). \tag{4.9}$$

Next recall the notion of a subspace A being *positively generated* by a pointed cone $S_+(A) \subset A$ (see Appendix A.10), in the sense that

$$A = S_+(A) - S_+(A) := \{x - y \mid x, y \in S_+(A)\}. \tag{4.10}$$

Lemma 4.6. *Let P be a positive projection on a (closed and pointed) cone $S_+(A) \subset A$. Then the image of P is positively generated, i.e.,* $\mathrm{im}(P) = \mathrm{im}^+(P) - \mathrm{im}^+(P)$.

Proof. From the property that A is positively generated by the cone $S_+(A)$, i.e., $A = S_+(A) - S_+(A)$, we have that any $x \in A$ can be decomposed as $x = y^+ - z^+$, where $y^+, z^+ \in S_+(A)$. Taking $x \in \mathrm{im}(P) \subset A$, we have $x = Px = Py^+ - Pz^+$, and using the positivity of P we have $Py^+, Pz^+ \in \mathrm{im}^+(P)$, which proves that $\mathrm{im}(P) = \mathrm{im}^+(P) - \mathrm{im}^+(P)$. □

The requirement that a filter P should have a complement P' with the given properties is equivalent to the following equalities:

$$\mathrm{im}^+(P) = \ker^+(P'), \quad \mathrm{im}^+(P') = \ker^+(P). \tag{4.11}$$

For any positive projection P the set $\ker^+(P)$ is an exposed face of $S_+(A)$, which implies that associated with any filter P and its complement P' there exists a pair of *complementary exposed faces* F, F' defined by $F = \mathrm{im}^+(P)$, and $F' = \mathrm{im}^+(P')$. Further, note that because a filter is a valid transformation of the state space, we also have $u_A(P\omega) \leq u_A(\omega)$, or equivalently, $u_A \circ P \leq u_A$. This means that every filter P induces an effect (see Definition 2.15) $f = u_A \circ P$, which evaluates to 1 on the face of normalized states $\Omega_F = \mathrm{im}^+(P) \cap \Omega(A)$, and 0 on the face $\Omega_{F'} = \ker^+(P) \cap \Omega(A)$. Finally, note that while the subspaces $\mathrm{im}(P)$ and $\ker(P)$ together span the space A, the subspaces $\mathrm{lin}(\mathrm{im}^+(P))$ and $\mathrm{lin}(\ker^+(P))$ in general do not span A.

In order to define the state spaces we will be working with in the remainder of this chapter, we need one more notion.

Definition 4.7 (**Dual projections**)**.** *If P is a positive projection on a (closed and pointed) cone $S_+(A) \subset A$, then the* dual projection*, $P^* : A^* \to A^*$ is defined by the action*

$$(P^*y)(x) := (y \circ P)(x) = y(Px), \quad \textit{where } x \in A,\ y \in A^*. \tag{4.12}$$

It is clear that P is positive on $S_+(A)$ if and only if P^* is positive on $S_+(A)^*$. Essentially, given a filter P acting on the state cone $S_+(A)$, the dual positive projection P^* acting on the effect cone $E_+(A)$ can be thought of as representing P in a 'generalized Heisenberg picture'.

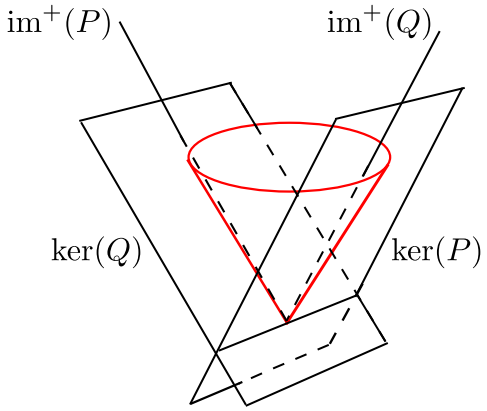

Figure 4.2: The state cone of a two-level real quantum system with boundary in red, along with a complementary pair of filters P, Q and their positive images and kernels. Note that $\mathrm{im}^+(P) = \ker^+(Q) \simeq \mathbb{R}_+$ as well as $\mathrm{im}^+(Q) = \ker^+(P) \simeq \mathbb{R}_+$, while $\ker(P) \simeq \ker(Q) \simeq \mathbb{R}^2$.

It is interesting to note that given a pair of complementary filters P, Q on a state space $(A, S_+(A), u_A)$, the dual positive projections P^*, Q^* need not be complementary, i.e., one or both of the relations $\mathrm{im}^+(P^*) = \ker^+(Q^*)$, $\mathrm{im}^+(Q^*) = \ker^+(P^*)$ may not hold! See Figure 7.3 of [7] for an example. This motivates the following class of models which we will use in the remainder of the chapter.

Assumption (**Projective state space**, [7] Chapter 8)**.** *A state space $(A, S_+(A), u_A)$ is said to be a* projective state space *if the following conditions are satisfied:*

(i) *For each exposed face F of $S_+(A)$ there exists a filter P_F such that $F = \mathrm{im}^+(P_F)$.*
(ii) *The dual projection P_F^* is a filter, i.e., if Q_F is a complement to P_F, then Q_F^* is a complement to P_F^*.*

We make this assumption for several reasons. First, as we will soon see, in order to formulate a general notion of an interference experiment we need to work with filters, and in particular filters with certain properties which hold for projective state spaces. Second, in order to say anything meaningful about the structure of the state space, we need our filters to have enough structure and to be closely related to the faces of the model. Further, assuming that the duals to filters are also filters is convenient because it gives us a useful mathematical framework to work with. In particular, in Chapter 7 of [7], Alfsen and Shultz extensively study projections satisfying condition (ii) above (see Appendix B.6), and further study projective state spaces in Chapter 8. One set of interesting results of these chapters are distinct conditions equivalent to condition (ii). See for example Appendices B.13, and

B.14. For more on these state spaces see the discussion after Proposition 7.49 of [7].

Quantum Example 4.8. *In quantum theory filters correspond to maps of the form* $\hat{B} \mapsto \hat{P}\hat{B}\hat{P}$*, where* $\hat{B} \in \mathcal{H}_{(d)}$*, and* $\hat{P}$ *is a projection operator on* $\mathbb{C}^d$*. These maps have many special properties:*

(a) *They are positive projections. In more detail, if* $\hat{\rho}$ *is a positive operator in* $S_+(\mathcal{H}_{(d)})$*, then* $\hat{P}\hat{\rho}\hat{P}$ *is positive, and clearly* $\hat{P}(\hat{P}\hat{\rho}\hat{P})\hat{P} = \hat{P}\hat{\rho}\hat{P}$*.*

(b) *They are* neutral*: if* $\mathrm{Tr}[\hat{P}\hat{B}\hat{P}] = \mathrm{Tr}[\hat{B}]$*, then* $\hat{P}\hat{B}\hat{P} = \hat{B}$ *(see [7] Proposition 1.41).*

(c) *They are* uniquely *complemented: for every map* $\hat{B} \mapsto \hat{P}\hat{B}\hat{P}$*, there exist a unique map* $\hat{B} \mapsto \hat{Q}\hat{B}\hat{Q}$*, with the properties that for all* $\hat{B} \in S_+(\mathcal{H}_{(d)})$*,* $\hat{P}\hat{B}\hat{P} = \hat{B}$ *if and only if* $\hat{Q}\hat{B}\hat{Q} = 0$*, and* $\hat{Q}\hat{B}\hat{Q} = \hat{B}$ *if and only if* $\hat{P}\hat{B}\hat{P} = 0$*. It should be clear that* $\hat{Q} = \mathbb{I}_d - \hat{P}$ *(see [7] Proposition 7.9 and Theorem 7.12).*

(d) *There is a unique correspondence (isomorphism) between the set of such maps and the set of faces of the cone of positive semi-definite operators. In particular, each face is the positive image of such a map (see [7] Proposition 5.10 and Theorem 5.32).*

(e) *They are* non-mixing *in the sense that if* $\hat{\rho}$ *is a pure state, then* $\hat{P}\hat{\rho}\hat{P} = \lambda\hat{\sigma}$ *with* $\hat{\sigma}$ *also a pure state, and* $0 \leq \lambda \leq 1$*. See [7] Proposition 9.3, as well as [99] and Section 4.9.2.*

(f) *By Theorems 5.11, 5.13 and 5.14 quantum state spaces satisfy the projective state space assumption.*

Before moving on, we define one particularly useful relation between filter.

Definition 4.9 (**Orthogonal filters, faces, and effects**)**.** *Let* $(A, S_+(A), u_A)$ *be a projective state space. A pair of filters* P_E, P_F *which satisfy*

$$P_E P_F = P_F P_E = 0_A, \tag{4.13}$$

where 0_A *is the map taking all vectors in* A *to the* 0 *vector, will be called* orthogonal*, as will the associated faces* E, F*, and the induced effects* $f = u_A \circ P_F, e = u_A \circ P_E$*.*

The relation $P_E P_F = P_F P_E = 0_A$ implies that the two associated faces are perfectly distinguishable in the sense that any pair of states $\omega \in E, \varphi \in F$ are perfectly distinguishable (Definition 2.12). In fact, the elements of these faces are distinguished by the effects $e = u_A \circ P_E$, and $f = u_A \circ P_F$, which are such that $e + f \leq u_A$, namely, they are part of some measurement. Note that if two filters are orthogonal they are not necessarily complementary, but if they are complementary then they are orthogonal.

Lemma 4.10. *Let* $(A, S_+(A), u_A)$ *be a projective state space. If* P, Q *are complementary filters then they are orthogonal, i.e.,* $PQ = QP = 0_A$*.*

Proof. P, Q being complementary is equivalent to the equalities $\mathrm{im}^+(P) = \ker^+(Q)$, and $\mathrm{im}^+(Q) = \ker^+(P)$. Using Lemma 4.6, we have $\mathrm{im}(P) = \mathrm{im}^+(P) - \mathrm{im}^+(P)$ and $\mathrm{im}(Q) = \mathrm{im}^+(Q) - \mathrm{im}^+(Q)$, which along with the complementarity of P, Q, imply the relations $\mathrm{im}(P) \subseteq \ker(Q)$ and $\mathrm{im}(Q) \subseteq \ker(P)$. These relations are equivalent to $PQ = QP = 0_A$, which proves the claim. □

4.4.1 Properties of projective state spaces

In this section we outline a few central results of Chapters 7 and 8 of Alfsen and Shultz [7] without proof. More details and other useful results can be found in Appendix B.

The set of exposed faces of a projective state space has a very elegant structure captured by the next two propositions. First recall the notion of a lattice from Appendix B.15, and that every exposed face F of a projective state space has a natural complement F' induced by the associated filter P_F, namely, $F' := \ker^+(P_F) = \mathrm{im}^+(P'_F)$. On a related note, we have previously mentioned that the definition of a filter does not assume a unique complementary filter. However, by Appendices B.13 and B.12, for projective state spaces, filters do indeed have unique complements!

Proposition 4.11 ([7] Proposition 8.1)**.** *Let $(A, S_+(A), u_A)$ be a projective state space. Then the set of exposed faces of $S_+(A)$, the set of filters, and the set of filter induced effects ($f = u_A \circ P_F$ where P_F is a filter), are isomorphic lattices with the partial order given by set inclusion $\subseteq$, and the least upper bound and greatest lower bound operations on pairs of faces F, G given by the following:*

$$F \vee G = (F' \cap G')', \quad F \wedge G = F \cap G. \tag{4.14}$$

In this context the greatest lower bound $F \wedge G$ has the interpretation of the largest face contained in both F and G, and the least upper bound $F \vee G$ has the interpretation of the smallest face containing both F and G. These operations induce lattice operations $(\subseteq, \vee, \wedge)$ on the set of filters by using the projective state space assumption in the obvious manner, which can in turn be used to define lattice operations on the set of filter induced effects.

Next recall the notion of an orthomodular lattice from Appendix B.16. These and related structures have played a large role in the quantum logic tradition [7, 32, 144, 82], and certain of their properties will also be very useful in this chapter.

Proposition 4.12 ([7] Theorem 8.10)**.** *Let $(A, S_+(A), u_A)$ be a projective state space. The lattice of filters (as well as the isomorphic lattices of exposed faces and filter induced effects) is orthomodular; that is, for each pair of filters P, Q, we have*

(i) $P'' = P$,
(ii) $\mathrm{im}(Q) \subseteq \mathrm{im}(P) \Rightarrow \mathrm{im}(P') \subseteq \mathrm{im}(Q')$,
(iii) $P \wedge P' = 0_A$ *and* $P \vee P' = \mathbb{I}_A$,
(iv) $\mathrm{im}(Q) \subseteq \mathrm{im}(P) \Rightarrow P = Q \vee (P \wedge Q')$.

Lemma 4.13 ([7] Lemma 7.42)**.** *Let $(A, S_+(A), u_A)$ be a projective state space. If P, Q are filters which satisfy $PQ = QP$, then $P'Q = QP'$, and $P'Q' = Q'P'$ as well.*

Theorem 4.14 ([7] Theorem 8.3)**.** *Let $(A, S_+(A), u_A)$ be a projective state space. If P, Q are filters which satisfy $PQ = QP$, then $PQ = P \wedge Q$.*

4.5 Generalized interference experiments

We now have all the tools to make the transition from the quantum interference experiments and expressions discussed above to a generalized notion of probabilistic interference. We simply take the quantum devices and mathematical objects representing them and replace these with preparations and operations of any other projective state space $(A, S_+(A), u_A)$. More precisely, we make the following replacements:

(i) Instead of an initial quantum state $\hat{\rho} \in S_+(\mathcal{H}_{(d)})$, take a state $\omega \in S_+(A)$,
(ii) Instead of a POVM $\{\hat{D}_l\} \subset \mathsf{E}(\mathcal{H}_{(d)})$, without loss of generality consider a single final outcome represented by an effect $q \in \mathsf{E}(A)$,
(iii) Instead of Feynman filters, take sets of transformations, $\{P_J\}$, which have the properties of filters on the state cone $S_+(A)$, and further necessary properties defined below.

We also have that a positive outcome for the effect defined as

$$e_J := u_A \circ P_J, \tag{4.15}$$

represents the event "the transformation P_J outputs a system". Further, the probability that a system passes the filter P_J and then the detector represented by q fires is given by

$$Pr(q \;\&\; e_J|\omega) = q(P_J\omega). \tag{4.16}$$

Next we define more precisely the generalization of the Feynman filters used in the quantum analysis. We will discuss the following definitions and results further at the end of Section 4.5.1.

Definition 4.15 (N**-slit mask**). *Let $(A, S_+(A), u_A)$ be a projective state space. Given an indexing set $\mathcal{I} = \{1, 2, ..., N\}$, we will say that $S_+(A)$* supports an N-slit mask *if there exists a set $\{P_i\}_{i\in\mathcal{I}}$ of N non-zero filters which satisfy the relations*

$$P_i P_j = P_j P_i = \delta_{ij} P_i, \quad \forall\ i, j \in \mathcal{I}. \tag{4.17}$$

In particular, this means that the faces defined by filters in an N-slit mask are all perfectly distinguishable as discussed after Definition 4.9. Recalling Propositions 4.11 and 4.12, we define the generalization of m-slit filters in the following way.

Definition 4.16 (m**-slit filter**). *Let $(A, S_+(A), u_A)$ be a projective state space. Given an N-slit mask $\{P_i\}$, for each subset $J \subset \mathcal{I}$, we define the* generalized m-slit filter generated by J *(where $m := |J|$) as*

$$P_J = \bigvee_{j\in J} P_j. \tag{4.18}$$

The following theorem relating different m-slit filters will be essential in the following.

Theorem 4.17. *Let $(A, S_+(A), u_A)$ be a projective state space. Given an N-slit mask $\{P_i\}_{i\in\mathcal{I}}$, then the filters $P_J = \bigvee_{j\in J} P_j$ satisfy the relations*

$$P_J P_K = P_K P_J = P_{J\cap K}, \tag{4.19}$$

for all $J, K \subseteq \mathcal{I}$.

Proof. From Definition 4.15 we have that the P_i are mutually orthogonal. Using this, Appendix B.19, and Appendix B.21, we have

$$u_A \circ P_K = u_A \circ \bigvee_{j\in J} P_j = u_A \circ \sum_{j\in J} P_j = \sum_{j\in J} p_j, \tag{4.20}$$

where $p_j = u_A \circ P_j$. Next, note that for all k, j, we have $p_j \circ P_k = u_A \circ (P_j P_k) = \delta_{jk} u_A \circ P_j$, as well as $p_j \circ P'_k = u_A \circ (P_j P'_k) = (1 - \delta_{jk}) u_A \circ P_j$, and so $p_j = p_j \circ (P_k + P'_k)$. Using Appendix B.19 again, this implies

$$u_A \circ P_K P_J = \sum_{j\in J} p_j \circ \left(\bigvee_{k\in K} P_k \right) = \sum_{j\in J} p_j \circ \left(\sum_{k\in K} P_k \right) = \sum_{j\in J\cap K} p_j. \tag{4.21}$$

Interchanging P_K and P_J, we also have $u_A \circ P_J P_K = \sum_{j\in J\cap K} p_j$. From Appendix B.20 we have that an effect p satisfies $p = \sum_{j\in J\cap K} p_j$ if and only if the associated filter satisfies

$P = \bigvee_{j \in J \cap K} P_j$. Together with the above equalities, this finally gives $P_J P_K = P_K P_J = P_{J \cap K}$. □

We are now ready to define our generalization of the set of multiple slit experiments with all possible combinations of open and closed slits:

Definition 4.18 (**Generalized N-slit system**). *Let $(A, S_+(A), u_A)$ be a projective state space which supports an N-slit mask $\{P_i\}_{i \in \mathcal{I}}$. Then define a* generalized N-slit system *as the set of filters $\{P_J\}_{J \subset \mathcal{I}}$, where the P_J are m-slit filters as defined above.*

4.5.1 Two, three, and N-slit interference expressions

An essential prerequisite for formulating a non-trivial two-slit interference experiment is that the projective state space $(A, S_+(A), u_A)$ supports at least one 2-slit mask $\{P_1, P_2\}$. The generalized two-slit filter generated by $\{P_1, P_2\}$ is then given by $P_{12} = P_1 \vee P_2$. Note that in general the filter $P_{12} \neq P_1 + P_2$! The second order interference expression (see Definition 4.1) with respect to the generalized slit system $\{P_{12}, P_1, P_2\}$, state $\omega \in \mathsf{S}(A)$, and final outcome $q \in \mathsf{E}(A)$ (which can be considered as an effect in some observation, or by itself), becomes:

$$\mathsf{I}_2[q, \{P_{12}, P_1, P_2\}, \omega] = q(P_{12}\omega - P_1\omega - P_2\omega). \tag{4.22}$$

Similarly, for a non-trivial third-order expression, the projective state space must support at least one 3-slit mask, $\{P_1, P_2, P_3\}$. Using such a mask, we then generate the set of all generalized two and three-slit filters, $\{P_J\}_{J \subset \{1,2,3\}}$. The third order interference expression (see Definition 4.2) with respect to the generalized slit system $\{P_J\}_{J \subseteq \{1,2,3\}}$, state $\omega \in \mathsf{S}(A)$, and final outcome $q \in \mathsf{E}(A)$, can then be written as:

$$\mathsf{I}_3[q, \{P_J\}_{J \subseteq \{1,2,3\}}, \omega] = q\left[(P_{123} - P_{12} - P_{13} - P_{23} + P_1 + P_2 + P_3)\omega\right]. \tag{4.23}$$

We can generalize the above and define an N-th order interference expression in the following way. Take a projective state space $(A, S_+(A), u_A)$ which supports an N-slit mask $\{P_i\}_{i \in \mathcal{I}}$. Then build a generalized slit system $\{P_J\}_{J \subseteq \mathcal{I}}$, and define the operator

$$P^{(n)} := \sum_{l=1}^{n-1} (-1)^{l-1} \sum_{|J| = n-l} P_J. \tag{4.24}$$

This object is the generalization of the signed sum of filters used in the second and third order expressions. For example, if $\mathcal{I} = \{1, 2\}$, then $P^{(2)} = P_1 + P_2$, and if $\mathcal{I} = \{1, 2, 3\}$, then $P^{(3)} = P_{12} + P_{13} + P_{23} - P_1 - P_2 - P_3$.

Definition 4.19 (**N-th order interference expression**, [162])**.** *Given a projective state space $(A, S_+(A), u_A)$ which supports an N-slit mask $\{P_i\}_{i\in\mathcal{I}}$, the* N-th order interference expression *with respect to the generalized slit system $\{P_J\}_{J\subseteq\mathcal{I}}$, initial state $\omega \in \mathsf{S}(A)$, and final outcome $q \in \mathsf{E}(A)$, is given by:*

$$I_N[q, \{P_J\}_{J\subseteq\mathcal{I}}, \omega] = q\left[(P_{\mathcal{I}} - P^{(N)})\omega\right]. \tag{4.25}$$

The reasons we have focused on m-slit filters as defined above, with the specified relations holding between them rather than more general transformations, should now be clearer. First, we want to allow for the possibility that the N-th order interference expression can be zero for all initial states and final effects. It is easy to find quantum or classical transformations (noisy filters for example) which by our definitions will give the appearance of second or higher order interference. This is clearly a kind of spurious or trivial interference and is simply a result of the noise. If we use idealized filters we will never have second-order interference in a finite dimensional classical probabilistic model, just as we will never have third-order interference in a quantum model. The structure of the state space is thus brought out much more strongly by using ideal filters with the properties we required above.

Second, the above requirements on the transformations P_i representing single slits from the mask generalizes the idea that systems which pass a particular single-slit filter should be perfectly distinguishable from systems which pass another single-slit filter. The condition (4.17) is sufficient for ensuring this, because if we take the effect representing "the transformation P_j outputs a system", $e_j = u_A \circ P_j$, and act on the re-normalized post-filter state $\omega_i = P_i\omega/u_A(P_i\omega)$ where ω is any normalized state in Ω_A, we find:

$$e_j(\omega_i) = u_A(P_jP_i\omega)/u_A(P_i\omega) = \delta_{ij}. \tag{4.26}$$

It can also be seen as a translation of Sorkin's requirement that the sets of histories that pass through distinct single-slits should be mutually disjoint.

Further, the definition of the m-slit filters P_J and the implied multiplicative properties expressed by (4.19) capture what is essential in the usual notion of an idealized multiple-slit experiment, and in particular, the operational meaning of leaving two or more slits open in the experiment. The fact that we have an orthomodular lattice of filters is essential for this property. This can also be taken as a justification of the projective state space assumption.

All of the above are only *sufficient* conditions on transformations in order to have a "well behaved" notion of interference. It may in fact be possible to weaken our assumptions, in

particular the projective state space assumption, and still say something interesting about interference phenomena in the larger class of theories. We will discuss this issue again briefly at the end of the chapter.

At a risk of redundancy, in the following three sections we present characterizations of the first two levels of the interference hierarchy, as well as a result for the general case. The results and proofs for the three cases are very similar in structure, but each is sufficiently important to merit its own separate analysis. As we will see, I_2 is fundamental to understanding interference in quantum theory, as well as the structure of the classical probabilistic model. I_3 is the first level which is zero for quantum theory, and it will make the general I_N case easier to present.

4.6 The structure of models with $\mathbf{I}_2 = 0$

Suppose we take a projective state space $(A, S_+(A), u_A)$, which supports a 2-slit mask $\{P_1, P_2\}$. The question we will address in the section is the following: what is the structure of $(A, S_+(A), u_A)$ if $\mathsf{I}_2[q, \{P_{12}, P_1, P_2\}, \omega] = 0$ for *all* initial states $\omega \in \mathsf{S}(A)$ and *all* final effects $q \in \mathsf{E}(A)$? To this end, define the operator

$$R_{12} := P_{12} - (P_1 + P_2) = P_{12} - P^{(2)}, \tag{4.27}$$

where recall from (4.24) that $P^{(2)} = P_1 + P_2$. An important object in the following is the intersection of the kernels of our single slit filters, so we define

$$N_{12} := \ker(P_2) \cap \ker(P_1). \tag{4.28}$$

The following result characterizes R_{12}.

Proposition 4.20. $\mathrm{im}(R_{12}) = N_{12} \cap \mathrm{im}(P_{12})$.

Proof. Let $x \in A$, and let $R_{12}x = x_{12} - x_1 - x_2 := y$, where $x_i := P_i x$ and $x_{12} := P_{12}x$. First, it is easy to see that R_{12} is a (not necessarily positive) projection. Letting $z \in \mathrm{im}(R_{12})$, and using Theorem 4.17 we have $P_{12}x = P_{12}R_{12}x = R_{12}x = x$, from which it follows that $\mathrm{im}(R_{12}) \subseteq \mathrm{im}(P_{12})$. It is also clear that $P_i y = 0$, so $\mathrm{im}(R_{12}) \subseteq \ker(P_i)$ for $i = 1, 2$. This in turn implies that $\mathrm{im}(R_{12}) \subseteq \ker(P_2) \cap \ker(P_1) \cap \mathrm{im}(P_{12}) = N_{12} \cap \mathrm{im}(P_{12})$. For the other direction, take $x \in N_{12} \cap \mathrm{im}(P_{12})$. Since $P_1 x = P_2 x = 0$ and $P_{12}x = x$, we have that $R_{12}(x) = x$, so $x \in \mathrm{im}(R_{12})$, which proves the claim. □

In order to make the main result of this section simpler to state, first recall the notion of a direct convex sum of convex sets from Appendix A.6. Next define the faces

$$F_i := \mathrm{im}^+(P_i), \quad F_{12} := \mathrm{im}^+(P_{12}), \tag{4.29}$$

and further let

$$\Omega_i := F_i \cap \Omega(A), \quad \Omega_{12} := F_{12} \cap \Omega(A), \tag{4.30}$$

where Ω_A is the compact convex set of normalized states. A useful result is that if Ω_1 and Ω_2 split Ω_{12}, then $\mathrm{lin}(F_{12}) = \mathrm{lin}(F_1) \oplus \mathrm{lin}(F_2)$, where $\oplus$ denotes the direct sum of subspaces (see Proposition 10.2 of [9]).

The following theorem gives a set of equivalent geometric conditions in terms of the faces F_1, F_2 and F_{12} such that the model $(A, S_+(A), u_A)$ displays no second-order interference.

Theorem 4.21. *Let $(A, S_+(A), u_A)$ be a projective state space which supports a 2-slit mask $\{P_1, P_2\}$. For the generalized 2-slit system $\{P_{12}, P_1, P_2\}$, the following are equivalent:*

(i) $I_2[q, \{P_{12}, P_1, P_2\}, \omega] = 0$ *for all* $\omega \in \mathsf{S}(A)$, $q \in \mathsf{E}(A)$,
(ii) $P_{12} = P_1 + P_2$,
(iii) $\mathrm{im}(P_{12}) = \mathrm{im}(P_1 + P_2) = \mathrm{im}(P_1) \oplus \mathrm{im}(P_2)$,
(iv) $F_{12} \subset \mathrm{lin}(F_1 \cup F_2)$,
(v) $\Omega_{12} = \Omega_1 \oplus_c \Omega_2$.

Proof. (i) $\Leftrightarrow$ (ii) is clear from the definitions.

(ii) $\Leftrightarrow$ (iii) It is easy to check that $R_{12}(P_1 + P_2) = (P_1 + P_2)R_{12} = 0_A$, and since $P_{12} = R_{12} + P^{(2)}$, the image of P_{12} can be written as $\mathrm{im}(P_{12}) = \mathrm{im}(R_{12}) \oplus \mathrm{im}(P_1 + P_2)$. Using the fact that P_1 and P_2 are projections which satisfy $P_1P_2 = P_2P_1 = 0_A$, and therefore $P_1 + P_2$ is also a positive projection with $\mathrm{im}(P_1 + P_2) = \mathrm{im}(P_1) \oplus \mathrm{im}(P_2)$, we have that $\mathrm{im}(P_{12}) = \mathrm{im}(P_1) \oplus \mathrm{im}(P_2)$ if and only if $\mathrm{im}(R_{12}) = \{0\}$.

(iii) $\Rightarrow$ (iv) $\mathrm{im}(P_1) \oplus \mathrm{im}(P_2) = \mathrm{lin}(F_1 \cup F_2)$ follows from the definitions, and the fact that the images of the P_i meet only at the zero vector. $F_{12} = \mathrm{im}^+(P_{12}) \subset \mathrm{im}(P_{12})$ is trivial.

(iv) $\Rightarrow$ (v) Let $x \in \Omega_{12}$ (i.e., $x \in F_{12}$ with $u_A(x) = 1$), and take

$$x = P_{12}x = P_1x + P_2x := x_1 + x_2,$$

where x_1 and x_2 are unique. This can be rewritten as $x = u_A(x_1)\tilde{x}_1 + u(x_2)\tilde{x}_2$, where $\tilde{x}_i := \frac{x_i}{u_A(x_i)} \in \Omega_i$, and $u_A(x_2) = 1 - u_A(x_1)$. Therefore $\Omega_{12} = \mathrm{conv}(\Omega_1 \cup \Omega_2)$. From the definitions of the P_i it is clear that the subspaces $\mathrm{lin}(F_i)$ are linearly independent, hence the Ω_i will be affinely independent, and we have $\Omega_{12} = \Omega_1 \oplus_c \Omega_2$.

(v) $\Rightarrow$ (iii) follows from the fact that if Ω_1 and Ω_2 split Ω_{12}, then $\mathrm{lin}(F_{12}) = \mathrm{lin}(F_1) \oplus \mathrm{lin}(F_2)$. □

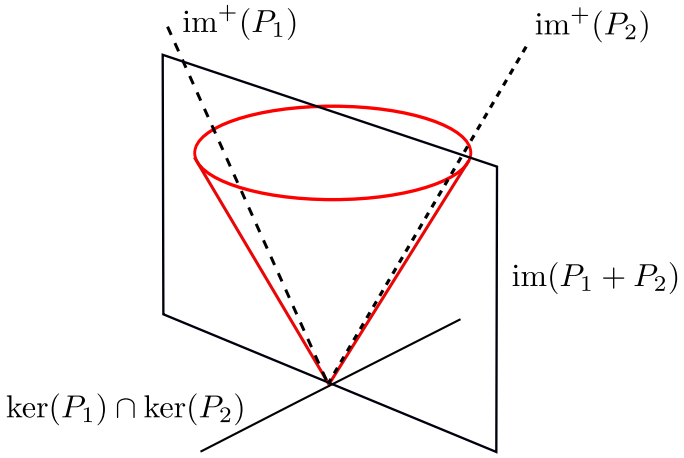

Figure 4.3: The state cone of a two-level real quantum system with boundary in red, along with a complementary pair of filters P_1, P_2 and various objects from Theorem 4.21. Initial states and final effects outside of the plane $\mathrm{im}(P_1+P_2)$ will exhibit second order interference with respect to the set of filters $\{P_1, P_2, P_{12}\}$.

In the simplest case where the projections P_i are both rank-one, then the faces F_i are rays of $S_+(A)$, and the faces $\Omega(A)_i$ are single pure states. We can therefore think of the set $\Omega_1 \oplus_c \Omega_2$ as consisting of the normalized states of a two-level "classical subsystem" of $\Omega(A)$, namely a line segment. If we further have that $S_+(A) = \mathrm{im}^+(P_{12})$, and $\mathsf{I}_2[q, \{P_{12}, P_1, P_2\}, \omega] = 0$ for all ω and q, it follows from condition (v) that $(A, S_+(A) = F_{12}, u_A)$ is exactly the state space of a two-level classical system. More generally, we can see that the absence of second-order interference with respect to a pair of filters means that the full set of normalized states is just the convex hull of the faces defined by these filters.

Quantum Example 4.22. *Take a two level* real *quantum system, for which the state cone* $S_+(\mathcal{H}_{(2)}(\mathbb{R}))$ *consists of real, positive semi-definite,* 2×2 *symmetric matrices. Further let* $\hat{P}_1, \hat{P}_2$ *be orthogonal rank-one projections on* $\mathbb{R}^2$*. As we have seen before, the filtering maps* $P_i(\hat{B}) = \hat{P}_i \hat{B} \hat{P}_i$ *are such that* $\hat{B} \neq P_1(\hat{B}) + P_2(\hat{B})$ *unless* $\hat{B}$ *is diagonal in the basis defined by* $\hat{P}_1, \hat{P}_2$*. In the rest of this example we take all matrices written in the basis defined by* $\hat{P}_1, \hat{P}_2$*. It is clear that* P_{12} *is the identity map on* $S_+(\mathcal{H}_{(2)}(\mathbb{R})) \subset \mathbb{R}^3$*, while* $\mathrm{im}(P_1+P_2) \simeq \mathbb{R}^2$ *is the set of diagonal matrices, and in particular* $\Omega_1 \oplus_c \Omega_2$ *is the set of diagonal density matrices. It is also not difficult to see that* $N_{12} = \ker(P_2) \cap \ker(P_1) \simeq \mathbb{R}$ *is exactly the set of symmetric matrices with* 0 *on the diagonals, and that states with a non-zero component in the subspace* N_{12} *will display second order interference. Given Theorem 4.21, it is then*

clear that real quantum theory for two-level systems exhibits second-order interference. See Figure 4.3 for a pictorial representation of these facts.

4.7 The structure of models with $\mathsf{I}_3 = 0$

In this section we will begin to study the structure of models in which there exists a generalized 3-slit system $\{P_J\}_{J\subseteq\{1,2,3\}}$ such that $\mathsf{I}_3[q, \{P_J\}, \omega] = 0$ for all initial states $\omega \in \mathsf{S}(A)$ and all final effects $q \in \mathsf{E}(A)$. To this end, take a projective state space $(A, S_+(A), u_A)$ which supports a 3-slit mask $\{P_i\}_{i=1}^3$. Take the generalized slit system $\{P_J\}_{J\subset\{1,2,3\}}$ generated by this mask and define the faces $F_J := \mathrm{im}^+(P_J)$ for all $J \subset \{1,2,3\}$. Further, recall the definition of the operator $P^{(3)} = P_{12} + P_{13} + P_{23} - P_1 - P_2 - P_3$. The following two technical Lemmas will be useful below.

Lemma 4.23. *$P^{(3)}$ is a (not necessarily positive) projection.*

Proof. Checking that $P^{(3)}P^{(3)} = P^{(3)}$ is a simple exercise in applying the definition of $P^{(3)}$ and then using the fact that $P_K P_J = P_J P_K = P_{J\cap K}$. □

Lemma 4.24. $\mathrm{lin}(F_{12} \cup F_{23} \cup F_{12}) = \mathrm{im}(P^{(3)})$.

Proof. That $\mathrm{im}(P^{(3)}) \subseteq \mathrm{lin}(F_{12}\cup F_{23}\cup F_{12})$, is immediate from the definition of $P^{(3)}$. We also have that $P^{(3)}P_i = P_iP^{(3)} = P_i$ and $P^{(3)}P_{ij} = P_{ij}P^{(3)} = P_{ij}$ for all $1 \leq i < j \leq 3$, so $P^{(3)}$ acts as the identity on $\mathrm{lin}(F_{12}\cup F_{23}\cup F_{12})$, and therefore $\mathrm{lin}(F_{12}\cup F_{23}\cup F_{12}) \subseteq \mathrm{im}(P^{(3)})$. □

Analogously to the two-slit case, define

$$R_{123} := P_{123} - [P_{12} + P_{13} + P_{23} - P_1 - P_2 - P_3] = P_{123} - P^{(3)}, \tag{4.31}$$

and the intersection of the kernels of the two-slit filters,

$$N_{123} := \ker(P_{23}) \cap \ker(P_{13}) \cap \ker(P_{12}). \tag{4.32}$$

Relating these objects, the analogue of Proposition 4.20 for the I_3 case is the following:

Proposition 4.25. $\mathrm{im}(R_{123}) = N_{123} \cap \mathrm{im}(P_{123})$.

Proof. Let $x \in A$, and take

$$R_{123}x = x_{123} - x_{12} - x_{13} - x_{23} + x_1 + x_2 + x_3 := y,$$

where $x_i \in \mathrm{im}(P_i)$, $x_{ij} \in \mathrm{im}(P_{ij})$, and $x_{123} \in \mathrm{im}(P_{123})$. First, it is clear that $\mathrm{im}(R_{123}) \subseteq \mathrm{im}(P_{123})$ and that $P_i y = P_{ij} y = 0$, so $\mathrm{im} R_{123} \subseteq \ker P_{ij} \subset \ker P_i$. This implies that $\mathrm{im} R_{123} \subseteq (\ker P_{23} \cap \ker P_{13} \cap \ker P_{12} \cap \mathrm{im} P_{123}) = N_{123} \cap \mathrm{im} P_{123}$. For the other direction, take $x \in N_{123} \cap \mathrm{im} P_{123}$. Since $P_i x = P_{ij} x = 0$ for all i, j, and $P_{123} x = x$, we have $R_{123} x = x$, so $x \in \mathrm{im} R_{123}$, and we are done. □

The following theorem gives a set of equivalent geometric conditions in terms of the faces F_{ij} and F_{123} such that the model $(A, S_+(A), u_A)$ exhibits no third-order interference.

Theorem 4.26. *Let $(A, S_+(A), u_A)$ be a projective state space, which supports at least one 3-slit mask $\{P_i\}_{i=1}^3$. Given the generalized slit system $\{P_J\}_{J \subset \{1,2,3\}}$ generated by this mask, the following are equivalent:*

(i) $\mathsf{I}_3[q, \{P_J\}, \omega] = 0$ *for all* $\omega \in \mathsf{S}(A)$ *and* $q \in \mathsf{E}(A)$,
(ii) $P_{123} = P^{(3)}$,
(iii) $F_{123} \subset \mathrm{lin}(F_{12} \cup F_{23} \cup F_{13})$.

Proof. (i) $\Leftrightarrow$ (ii) is clear from the definitions.

(ii) $\Leftrightarrow$ (iii) It is not difficult to see that $R_{123} P^{(3)} = P^{(3)} R_{123} = 0_A$, and using the fact that $P^{(3)}$ is a projection, we have $\mathrm{im}(P_{123}) = \mathrm{im}(R_{123}) \oplus \mathrm{im}(P^{(3)})$ and $\ker(P_{123}) = \ker(R_{123}) \cap \ker(P^{(3)})$. These equalities (along with $P^{(3)} R_{123} = 0_A$) imply that $P_{123} = P^{(3)}$ if and only if $R_{123} = 0_A$. Combining this with Lemma 4.23 gives $\mathrm{im}(P_{123}) = \mathrm{lin}(F_{12} \cup F_{23} \cup F_{12})$ if and only if $R_{123} = 0_A$. Since P_{123} is a positive projection, from Lemma 4.6 we have $\mathrm{im}(P_{123}) = \mathrm{im}^+(P_{123}) - \mathrm{im}^+(P_{123})$ and $F_{123} = \mathrm{im}^+(P_{123})$, so the statement $\mathrm{im}(P_{123}) = \mathrm{lin}(F_{12} \cup F_{23} \cup F_{12})$ is equivalent to $F_{123} \subset \mathrm{lin}(F_{12} \cup F_{23} \cup F_{12})$. □

The condition that $F_{123} \subset \mathrm{lin}(F_{12} \cup F_{23} \cup F_{12})$ expresses the property that states conditioned on all three slits being open (i.e., states of systems that have passed the three slits and are therefore in the face F_{123}) can be written as linear combinations of states which are conditioned only two of the slits being open. This may seem like a mysterious property at first sight, but it has an intuitive and operational interpretation we discuss in the next section.

Finally, notice that an easy corollary of Theorem 4.26 and Proposition 4.25 is that $\mathsf{I}_3[q, \{P_J\}, \omega] = 0$ for all $\omega \in \mathsf{S}(A)$ and $q \in \mathsf{E}(A)$ if and only if $N_{123} \cap \mathrm{im}(P_{123}) = \{0\}$. If $S_+(A) = F_{123}$, then the kernels making up N_{123} are certainly all contained in $\mathrm{im}(P_{123})$, so the relevant issue is whether they have a non-trivial intersection. We will study this in more detail in Section 4.9.

4.7.1 Third-order interference and tomography

We are now in a position to state the relationship between third-order interference and the two-slit filtering tomography discussed in Section 4.3. Suppose we are given a device that outputs a set of identically and independently prepared systems, each described the same state of a model $(A, S_+(A), u_A)$ (for simplicity, in the following we will assume that the filter P_{123} acts as the identity on the whole state space under consideration, i.e., $S_+(A) = F_{123}$) which satisfies the requirements of Theorem 4.26. Our task is to reconstruct the state ω which represents this device by measurements on the systems it outputs. To accomplish this task we are only allowed to use the following: (1) the three 'two-slit' filters P_{ij}, and (2) for each P_{ij}, a measurement device M_{ij} which is 'informationally complete'[3] for the systems which pass the filter P_{ij}. What we can do (for each of the given filters) is take a sub-ensemble of the systems produced by the source, pass them through P_{ij}, and then use the device M_{ij} on the resulting filtered ensemble to determine the state $\omega_{ij} = P_{ij}\omega \in F_{ij}$.

Quantum Example 4.27. *As we have already discussed in Section 4.3, the information gained from the above filtering and measuring procedure will in fact be sufficient to find a density matrix which describes the source [73]. In other words, in order to specify a $d \times d$ density matrix (where now $d^2 = rank(P_{123})$), it is sufficient to do tomography on the three subspaces F_{ij} of filtered states. In particular, if we temporarily assume that the three filters P_i are all rank-one projections, then F_{123} is the cone of un-normalized states of a three-level system, and the faces F_{ij} correspond to two-level subsystems of F_{123}. So in order to specify a 3×3 density matrix, it is sufficient to do tomography on three, two-level subspaces of filtered states.*

The sufficiency of this kind of tomography for the quantum model generalizes to all models (A, F_{123}, u_A) satisfying the conditions of Theorem 4.26 and for which $\mathsf{I}_3[q, \{P_J\}, \omega] = 0$ for all $\omega \in F_{123}$ and $q \in \mathsf{E}(A)$. In other words, if the model (A, F_{123}, u_A) exhibits no third-order interference (with respect to the set of experiments defined by the generalized slit system $\{P_J\}$), then the components of a state $\omega \in F_{123}$ can be reconstructed from the measurements M_{ij} on the faces F_{ij} of filtered states. This follows from condition (iii) of Theorem 4.26, together with the fact that the P_{ij} act as the identity on the states they transmit with probability one. In this scenario, the reconstruction formula for a state

[3] An informationally complete measurement is one whose outcome probabilities are sufficient to uniquely specify a state. In other words, the linear map induced by the set of effects from states to probability distributions is injective. All finite dimensional state spaces have informationally complete measurements [26]

$\omega \in F_{123}$ in terms of the filtered states $\omega_{ij} \in F_{ij}$ is given by

$$\omega = \omega_{12} + \omega_{13} + \omega_{23} - \omega_1 - \omega_2 - \omega_3, \tag{4.33}$$

where the $\omega_i = P_i(\omega_{ij}) \in F_{ij}$ can easily be inferred once the ω_{ij} are determined experimentally. Note that the projective state space assumption plays a hidden role here. If the filters P_{ij} did not act trivially on the faces F_{ij}, i.e., if $F_{ij} \neq \mathrm{im}^+(P_{ij})$, then characterizing the states after the three filters may not be sufficient to infer the initial state. Conversely, if for some model (A, F_{123}, u_A) which satisfies the conditions of Theorem 4.26, and for which the components of a state $\omega \in F_{123}$ can be reconstructed from some measurements M_{ij} on states filtered by the P_{ij}, then there will be no third-order interference (again, with respect to the experiments defined by the filters $\{P_J\}$).

On the other hand, if a model does exhibit third-order interference for some generalized slit system, then there are extra parameters which are needed to describe states filtered by P_{123} over and above all the parameters needed to describe states filtered by each of the P_{ij}. Operationally this means that there are measurements which can be performed on states $\omega \in F_{123}$, the outcome probabilities of which cannot be predicted knowing only the outcome probabilities of all possible measurements on the filtered states ω_{ij}. The additional parameters can be interpreted as higher-order analogues of the off-diagonal elements of a density matrix – often called 'coherences' – which are related to interference in two-slit experiments.

4.8 The structure of models with $\mathbf{I}_N = 0$

The above considerations and results for the second and third order interference expressions have a simple generalization for each level N of the interference expression hierarchy. Take a projective state space $(A, S_+(A), u_A)$, and take an N-slit mask $\{P_i\}_{i\in\mathcal{I}}$. Take the set of generalized slits $\{P_J\}_{J\subset\mathcal{I}}$ generated by this mask, and let $F_J := \mathrm{im}^+(P_J)$ be the faces defined by these slits. Further, recall the definition of the operator

$$P^{(N)} = \sum_{l=1}^{N-1} (-1)^{l-1} \sum_{|J|=N-l} P_J. \tag{4.34}$$

Lemma 4.23 generalizes to the following for arbitrary $N \geq 2$. The proof is long and not particularly instructive, so we leave it to Appendix C.

Lemma 4.28. *$P^{(N)}$ is a (not necessarily positive) projection with image*

$$\mathrm{im}\left(P^{(N)}\right) = \mathrm{lin}\Big(\bigcup_{|J|=N-1} F_J\Big). \tag{4.35}$$

Proof. See Appendix C. □

Analogously to the two-slit case, define

$$R_{\mathcal{I}} := P_{\mathcal{I}} - P^{(N)}, \tag{4.36}$$

and also define the intersection of the kernels of the N-1-slit filters,

$$N_{\mathcal{I}} := \bigcap_{|J|=N-1} \ker(P_J). \tag{4.37}$$

Relating these objects, the analogue of Proposition 4.20 and 4.25 is the following:

Proposition 4.29. $\mathrm{im}(R_{\mathcal{I}}) = N_{\mathcal{I}} \cap \mathrm{im}(P_{\mathcal{I}})$.

Proof. Let $x \in A$, and take $y := R_{\mathcal{I}}x$. First, it is easy to check that $\mathrm{im}(R_{\mathcal{I}}) \subseteq \mathrm{im}(P_{\mathcal{I}})$, and that $P_J y = 0$ for all $J \subset \mathcal{I}$ with $|J| \leq N-1$. Therefore, $\mathrm{im}(R_{\mathcal{I}}) \subseteq \ker(P_J)$ for $|J| \leq N-1$. This implies that

$$\mathrm{im}(R_{\mathcal{I}}) \subseteq \Big(\bigcap_{|J|=n-1} \ker(P_J)\Big) \cap \mathrm{im}(P_{\mathcal{I}}) = N_{\mathcal{I}} \cap \mathrm{im}(P_{\mathcal{I}}).$$

For the other direction, take $x \in N_{\mathcal{I}} \cap \mathrm{im} P_{\mathcal{I}}$. Since $P_J x = 0$ for all $J \subset \mathcal{I}$ with $|J| \leq N-1$, and $P_{\mathcal{I}}x = x$, we have $R_{\mathcal{I}}(x) = x$, so $x \in \mathrm{im}(R_{\mathcal{I}})$, and we are done. □

The following theorem gives a set of equivalent geometric conditions in terms of the faces faces $\{F_J\}_{|J|=N-1}$ and $F_{\mathcal{I}}$ such that the model $(A, S_+(A), u_A)$ exhibits no N-th order interference.

Theorem 4.30. *Let $(A, S_+(A), u_A)$ be a projective state space which supports at least one N-slit mask $\{P_i\}_{i\in\mathcal{I}}$. Given the generalized slit system $\{P_J\}_{J\subset\mathcal{I}}$ generated by this mask, the following are equivalent:*

(i) $I_N[q, \{P_J\}_{J\in\mathcal{I}}, \omega] = 0$ *for all* $\omega \in \mathsf{S}(A)$ *and* $q \in \mathsf{E}(A)$,
(ii) $P_{\mathcal{I}} = P^{(N)}$,
(iii) $F_{\mathcal{I}} \subset \mathrm{lin}\left(\bigcup_{|J|=N-1} F_J\right)$.

Proof. (i) $\Leftrightarrow$ (ii) is clear from the definitions.

(ii) $\Leftrightarrow$ (iii) It is not difficult to see that $R_{\mathcal{I}}P^{(N)} = P^{(N)}R_{\mathcal{I}} = 0_A$ (see Appendix C), so $\mathrm{im}(P_{\mathcal{I}}) = \mathrm{im}(R_{\mathcal{I}}) \oplus \mathrm{im}(P^{(N)})$. Using this equality, it is clear that $\mathrm{im}(P_{\mathcal{I}}) = \mathrm{im}(P^{(N)})$ if and only if $R_{\mathcal{I}} = 0_A$. Combining this with Lemma 4.28, gives $\mathrm{im}P_{\mathcal{I}} = \mathrm{lin}(\bigcup_{|J|=N-1} F_J)$. Using the facts that $\mathrm{im}(P_{\mathcal{I}})$ is positively generated, and $F_{\mathcal{I}} = \mathrm{im}^+(P_{\mathcal{I}})$, we see that $\mathrm{im}(P_{\mathcal{I}}) = \mathrm{lin}(\bigcup_{|J|=N-1} F_J)$ is equivalent to $F_{\mathcal{I}} \subset \mathrm{lin}(\bigcup_{|J|=N-1} F_J)$. □

In Sections 4.3 and 4.7.1 we discussed the relationship between third-order interference and two-slit filtering tomography. This relationship can be generalized to N-th order interference and N-1-slit filtering tomography. Rather than going into the details, we simply mention that this follows easily from the close analogy between Theorem 4.26 and Theorem 4.30, along with the discussion in Section 4.7.1.

4.9 A Peirce decomposition, $\mathsf{I}_3 = 0$, and non-mixing filters

In this section we will explore third-order interference in more detail. We will prove a decomposition of projective state spaces, which is analogous to the Peirce decomposition for Jordan algebras (see Chapter IV of [61]). We will also show that $\mathsf{I}_3 = 0$ is equivalent to a certain property of this decomposition. Finally, we will connect third order interference with the existence of an interesting type of filter, often called non-mixing, which takes pure states to multiples of pure states [7, 99].

We begin by introducing an operator built from the pair of complementary filters associated with some face. This operator plays an important role in the study of Jordan algebras (in fact, it is the Jordan multiplication operator in that setting!), and will eventually be used in the final step of our reconstruction of Jordan algebras from operational principles in Chapter 5.

Definition 4.31. *Let $(A, S_+(A), u_A)$ be a projective state space, and let P_F be the filter associated with a face $F \subseteq S_+(A)$. Define $T_F : A \to A$ by*

$$T_F = \frac{1}{2}(\mathbb{I}_A + P_F - P_{F'}), \tag{4.38}$$

where $\mathbb{I}_A$ is the identity operator on A.

We will now give a few important properties of these operators.

Proposition 4.32. *Let $(A, S_+(A), u_A)$ be a projective state space. Then the operators T_F satisfy the following properties:*

(i) $T_F P_F = P_F T_F = P_F$, *and* $T_F P_{F'} = P_{F'} T_F = 0_A$.
(ii) $P_F = 2T_F^2 - T_F$.
(iii) $2T_F^3 - 3T_F^2 + T_F = 0_A$.
(iv) *If E, F are complementary faces, i.e., they are orthogonal (see Definition 4.9) and satisfy the equality $u_A \circ (P_E + P_F) = e + f = u_A$, then $T_E + T_F = T_{E \vee F} = \mathbb{I}_A$.*
(v) *If E, F are orthogonal faces, then $T_E T_F = T_F T_E$, and further, $f \circ T_E = 0$ where $f = u_A \circ P_F$.*

Proof. (i) and (ii) are trivial to check.

(iii) Follows from (ii) by an application of T_F to both sides, and using (i).

(iv) Because E and F are complementary faces, and the lattice of faces is orthomodular (Theorem 4.12), we have that $P_E \vee P_F = P_{E \vee F} = \mathbb{I}_A$, from which it follows easily that $T_{E \vee F} = \mathbb{I}_A$. Further, the complementarity of E and F is equivalent to the equalities $P'_E = P_F$ and $P'_F = P_E$. Expanding out $T_E + T_F$ and using the above, it is easily seen that $T_E + T_F = \mathbb{I}_A$.

(v) Because E, F are orthogonal faces, it follows that $P_E P_F = P_F P_E$, and furthermore, from Lemma 4.13 the filters $P_F, P_E, P_{F'}, P_{E'}$ all commute pairwise. Next, because E, F are orthogonal, $f \leq e' = u_A - e$, which in turn implies that $P_F P_{E'} = P_{E'} P_F = P_F$, and then it is easy to see that $f \circ P_{E'} = f$, and $f \circ P_E = 0$, which gives the claim. □

In fact, (iii) above gives us the eigenvalues and spectral decomposition of T_F.

Theorem 4.33 (Peirce decomposition). *Let $(A, S_+(A), u_A)$ be a projective state space. Then the operators T_F have eigenvalues $\{0, \frac{1}{2}, 1\}$, and spectral decomposition*

$$T_F = \frac{1}{2} \cdot (\mathbb{I}_A - P_F - P_{F'}) + 1 \cdot P_F + 0 \cdot P_{F'}. \tag{4.39}$$

Further, the eigenspaces can be written as:

$$\begin{aligned} V_1(F) &= \{x \in A \mid P_F x = x\} = \mathrm{im}(P_F), \\ V_{1/2}(F) &= \{x \in A \mid P_F x = P_{F'} x = 0\}, \\ V_0(F) &= \{x \in A \mid P_{F'} x = x\} = \mathrm{im}(P_{F'}), \end{aligned} \tag{4.40}$$

with the projections onto these subspaces given by:

$$\begin{aligned} \textit{onto } V_1(F) &: \quad P_F, \\ \textit{onto } V_{1/2}(F) &: \quad \mathbb{I}_A - P_F - P_{F'}, \\ \textit{onto } V_0(F) &: \quad P_{F'}. \end{aligned} \tag{4.41}$$

Proof. That the eigenvalues are $\{0, \frac{1}{2}, 1\}$ follows from Lemma 4.32 (iii). The spectral decomposition is then clear, and so are the expression for the eigenspaces. □

Lemma 4.34. *Let $(A, S_+(A), u_A)$ be a projective state space. For every face F and associated filter P_F, we have the following:*

(i) $V_0(F) = V_1(F')$, $V_0(F') = V_1(F)$,
(ii) $V_{1/2}(F) = V_{1/2}(F') = \ker(P_{F'}) \cap \ker(P_F)$, *and*
(iii) $\ker(P_F) = V_1(F') \oplus V_{1/2}(F) = V_0(F) \oplus V_{1/2}(F)$.

Proof. (i) and (ii) are clear from the definitions and the Peirce decomposition.

(iii) First, it is clear that $V_{1/2}(F) \subset \ker(P_F)$. Taking $x \in V_0(F) = V_1(F')$, then $P_{F'}x = x$, and using Lemma 4.10, we have $P_F x = 0$, which proves $x \in \ker(P_F)$. This shows that $V_0(F) \oplus V_{1/2}(F) \subseteq \ker(P_F)$.

Next, using the Peirce decomposition $A = V_1(F) \oplus V_1(F') \oplus V_{1/2}(F)$, we can write $x \in A$ as $x = x_1^F + x_1^{F'} + x_{1/2}^F$. Assuming that $x \in \ker(P_F)$, we have $x_1^F = 0$, which proves $x \in V_1(F') \oplus V_{1/2}(F)$. Finally using $V_0(F) = V_1(F')$ and the inclusion above proves $\ker(P_F) = V_1(F') \oplus V_{1/2}(F) = V_0(F) \oplus V_{1/2}(F)$. □

It is interesting to note the relationship between $V_{1/2}(F) = \ker(P_{F'}) \cap \ker(P_F)$ in the above proof, and the subspace N_{12} introduced in the discussion of second order interference in Section 4.6. Next we define a special restriction of the subspace $V_{1/2}(F)$.

Definition 4.35. *Let $(A, S_+(A), u_A)$ be a projective state space, and let E, F be two orthogonal faces of $S_+(A)$. Define the restriction of $V_{1/2}(E)$ to the subspace defined by the face $E \vee F$ as*

$$V_{1/2}(E)|_{E \vee F} = V_{1/2}(E) \cap \mathrm{lin}(E \vee F). \tag{4.42}$$

Lemma 4.36. *Let $(A, S_+(A), u_A)$ be a projective state space. Given two orthogonal faces E, F, we have the following:*

(i) $V_0(E \vee F) = V_0(E) \cap V_0(F)$.
(ii) $V_1(E \vee F) \subseteq V_1(E) \oplus V_1(F) \oplus V_{1/2}(E) \cap V_{1/2}(F)$.

Proof. (i) From Lemma 4.34 (i) we can write the intersection

$$V_0(E) \cap V_0(F) = V_1(E') \cap V_1(F') = \mathrm{im}(P_{E'}) \cap \mathrm{im}(P_{F'}) = \mathrm{lin}(E') \cap \mathrm{lin}(F').$$

We also have from Proposition 4.11 that for all exposed faces E, F,

$$E' \cap F' = (E \vee F)' = \mathrm{im}^+(P_{(E \vee F)'}).$$

Next, using result B.18 we can write $\mathrm{lin}(E') \cap \mathrm{lin}(F') = \mathrm{lin}(E' \cap F')$, and combining this with $\mathrm{lin}(E' \cap F') = \mathrm{im}(P_{(E \vee F)'}) = V_0(E \vee F)$, proves the claim.

(ii) From Appendix B.17 we have that in the subspace $V_1(E \vee F) = \mathrm{lin}(E \vee F)$ the face $E \vee F$ is a projective state space in its own right. Therefore, we can take the Peirce decomposition of $\mathrm{lin}(E \vee F)$ with respect to E (or equivalently F), namely,

$$\mathrm{lin}(E \vee F) = V_1(E \vee F) = V_1(E)|_{E \vee F} \oplus V_0(E)|_{E \vee F} \oplus V_{1/2}(E)|_{E \vee F}.$$

Because E is a face of $E \vee F$, we have $V_1(E) \subset \mathrm{lin}(E \vee F)$, which implies $V_1(E)|_{E \vee F} = V_1(E)$. By the orthomodularity of the lattice of exposed faces (Proposition 4.12), we have that $F = (F \vee E) \cap E'$, and by taking linear spans in this expression, and using result B.17 once again, $V_1(F) = V_1(E') \cap \mathrm{lin}(E \vee F) = V_0(E)|_{E \vee F}$ follows. By a similar argument with E and F interchanged, we have $V_0(F)|_{E \vee F} = V_1(E)$. Using the Peirce decomposition in terms of F, namely,

$$\mathrm{lin}(E \vee F) = V_1(F)|_{E \vee F} \oplus V_0(F)|_{E \vee F} \oplus V_{1/2}(F)|_{E \vee F} = V_1(F) \oplus V_1(E) \oplus V_{1/2}(F)|_{E \vee F},$$

together with the above, shows that we can without loss of generality take $V_{1/2}(F)|_{E \vee F} = V_{1/2}(E)|_{E \vee F}$. This equality then gives the inclusion $V_{1/2}(E)|_{E \vee F} \subset V_{1/2}(E) \cap V_{1/2}(F)$, and the statement follows. □

One step in the above proof is a statement of a property which will be useful below, so we state it as a corollary:

Corollary 4.37. *Let $(A, S_+(A), u_A)$ be a projective state space. For all pairs of orthogonal faces E, F, we have the symmetry*

$$V_{1/2}(E)|_{E \vee F} = V_{1/2}(F)|_{E \vee F}. \tag{4.43}$$

Proof. Follows from the proof of Lemma 4.36 (ii). □

Lemma 4.38. *Let $(A, S_+(A), u_A)$ be a projective state space. Take two orthogonal faces, E, F, and set $G = (E \vee F)'$. Then*

$$V_{1/2}(E)|_{E \vee F} \cap V_{1/2}(E)|_{E \vee G} = \{0\}. \tag{4.44}$$

Proof. By definition

$$V_{1/2}(E)|_{E\vee F} \cap V_{1/2}(E)|_{E\vee G} = [V_{1/2}(E) \cap V_1(E \vee F)] \cap [V_{1/2}(E) \cap V_1(E \vee G)].$$

Next, note that $V_1(E\vee G)\cap V_1(E\vee F) = \mathrm{im}(P_{E\vee G})\cap \mathrm{im}(P_{E\vee F})$, and from the orthogonality of the faces E, F, G, and Theorem 4.17 these filters satisfy $P_{E\vee G}P_{E\vee F} = P_{E\vee F}P_{E\vee G} = P_E$. Using result B.18, we then have $\mathrm{im}(P_{E\vee G}) \cap \mathrm{im}(P_{E\vee F}) = \mathrm{im}(P_E) = V_1(E)$. Finally, by definition $V_1(E) \cap V_{1/2}(E) = \{0\}$, and the statement follows. □

4.9.1 The Peirce decomposition and interference

There is also an interesting connection between the T_F operators studied above and third order interference, which was first noticed by Gerd Niestegge in [138]. In this section we review and expand on this connection.

Let $(A, S_+(A), u_A)$ be a projective state space. Take two orthogonal faces (see Definition 4.9) F_1, F_2, and let $F_3 := (F_1\vee F_2)'$. The filters $\{P_1, P_2, P_3\}$ associated with these faces make up a 3-slit mask (see Definition 4.15), and can be used to form a generalized 3-slit system $\{P_J\} = \{P_1, P_2, P_3, P_{12}, P_{13}, P_{23}, P_{123}\}$. Note that P_{123} is defined as the filter onto the face $F_{123} = F_1\vee F_2\vee F_3 = F_1\vee F_2\vee(F_1\vee F_2)'$, and by orthomodularity, $F_1\vee F_2\vee(F_1\vee F_2)' = S_+(A)$, so $P_{123} = \mathbb{I}_A$.

Lemma 4.39 ([138] Lemma 8.2)**.** *Let $(A, S_+(A), u_A)$ be a projective state space. Take two orthogonal faces F_1, F_2, such that $F_3 := (F_1 \vee F_2)' \neq \{0\}$, and form the generalized 3-slit system $\{P_J\}$ defined by these faces. Then $I_3[q, \{P_J\}, \omega] = 0$ for all $q \in \mathsf{E}(A)$, $\omega \in \mathsf{S}(A)$, if and only if $T_{F_1\vee F_2} = T_{F_1} + T_{F_2}$.*

Proof. Writing out the T_F operators, and using Theorem 4.12 we have

$$\begin{aligned} T_{F_1\vee F_2} &= (\mathbb{I}_A + P_{12} - P_3)/2, \\ T_{F_1} &= (\mathbb{I}_A + P_1 - P_{23})/2, \\ T_{F_2} &= (\mathbb{I}_A + P_2 - P_{13})/2. \end{aligned}$$

Putting these together and simplifying, gives

$$T_{F_1} + T_{F_2} - T_{F_1\vee F_2} = (\mathbb{I}_A - P_{12} - P_{13} - P_{23} + P_1 + P_2 + P_3)/2 = (\mathbb{I}_A - P^{(3)})/2.$$

By Theorem 4.26, we then have that $T_{F_1\vee F_2} = T_{F_1} + T_{F_2}$ if and only if $I_3[q, \{P_J\}, \omega] = 0$ for all $q \in \mathsf{E}(A)$, $\omega \in \mathsf{S}(A)$. □

Next we give a simple result which shows that any apparent asymmetry between F_1, F_2 and $(F_1 \vee F_2)'$ above is in fact only apparent.

Lemma 4.40. *Let $(A, S_+(A), u_A)$ be a projective state space, and take two orthogonal faces F_1, F_2, and let $F_3 := (F_1 \vee F_2)'$. Then the following are equivalent:*

(i) $T_{F_1} + T_{F_2} = T_{F_1 \vee F_2}$.

(ii) $T_{F_3} + T_{F_2} = T_{F_3 \vee F_2}$.

(iii) $T_{F_1} + T_{F_3} = T_{F_1 \vee F_3}$.

Proof. Recall from Lemma 4.32 (iv) that if G, H are orthogonal faces such that $u_A \circ (P_G + P_H) = g + h = u_A$, or equivalently, $(G \vee H)' = \{0\}$, then $T_G + T_H = T_{G \vee H} = \mathbb{I}_A$. This implies that

$$\mathbb{I}_A = T_{F_1 \vee F_2 \vee F_3} = T_{F_1} + T_{F_{23}} = T_{F_2} + T_{F_{13}} = T_{F_3} + T_{F_{12}}.$$

Using these equalities it is then easy to see that if any of the statements (i), (ii), (iii) hold, then the other two hold as well. □

The next set of results pinpoint exactly where in the space A we do not have equality of the actions of $T_{F_1 \vee F_2}$ and $T_{F_1} + T_{F_2}$.

Lemma 4.41. *Let $(A, S_+(A), u_A)$ be a projective state space, and take two orthogonal faces E, F. Then the following are equivalent:*

(i) $T_E + T_F = T_{E \vee F}$,

(ii) $V_{1/2}(E) \cap V_{1/2}(F) \subseteq V_1(E \vee F)$,

Proof. (i) $\Rightarrow$ (ii) Assume that $T_E + T_F = T_{E \vee F}$, and take $x \in V_{1/2}(E) \cap V_{1/2}(F)$. This is equivalent to the two Equations $T_E x = x/2$ and $T_F x = x/2$. By assumption this implies that $T_{E \vee F} x = x$, which implies that $x \in V_1(E \vee F)$.

(ii) $\Rightarrow$ (i) Assume that $V_{1/2}(E) \cap V_{1/2}(F) \subseteq V_1(E \vee F)$. We will show that in each of the subspaces in the Peirce decomposition of A in terms of $E \vee F$, we have that the action of the operator $T_E + T_F$ is identical to the action of $T_{E \vee F}$. First take $x \in V_1(E \vee F)$. By Lemma 4.36 (ii), we know that

$$x \in V_1(E) \oplus V_1(F) \oplus V_{1/2}(E) \cap V_{1/2}(F),$$

so we can write $x = x_1^E + x_1^F + x_{1/2}^{E,F}$ (with the obvious interpretation of these components). It is then easy to see that $(T_E + T_F)x = T_{E \vee F} x = x$. Next, take $x \in V_0(E \vee F)$. By Lemma 4.36 (i), we have $x \in V_0(E) \cap V_0(F)$, so $T_E x = T_F x = 0$, and we have $(T_E + T_F)x = T_{E \vee F} x = 0$ on this subspace as well. Finally, take $x \in V_{1/2}(E \vee F)$. By assumption, if x were in $V_{1/2}(E) \cap V_{1/2}(F)$, then $T_{E \vee F} x = x$ would hold, but this would contradict $x \in V_{1/2}(E \vee F)$, so we cannot have both $x_{1/2}^E, x_{1/2}^F \neq 0$. Without loss of generality, take $x_{1/2}^E = 0$. We have three possible Peirce decompositions of x:

$$x = x_0^E + x_1^E, \quad x = x_0^F + x_1^F + x_{1/2}^F, \quad x = x_{1/2}^{E \vee F}.$$

However, since $x \in V_{1/2}(E \vee F)$, then $P_{E\vee F}x = 0$ implies that $x_1^E = x_1^F = 0$. Now notice that

$$(T_E + T_F)x = T_E(x_0^E) + T_F(x_0^F + x_{1/2}^F) = 0 \cdot x + 0 \cdot x + \frac{1}{2} \cdot x_{1/2}^F,$$

and $T_{E\vee F}x = \frac{1}{2} \cdot x_{1/2}^{E\vee F}$, where we have used each of the three decompositions in turn. Next we want to show that $x_0^F = 0$, i.e., $x_{1/2}^F = x_{1/2}^{E\vee F}$. Suppose that $x_0^F \neq 0$. Recalling that $V_0(F) = V_1(F') = \mathrm{im}(P_{F'})$, and using $F' = E \vee (E \vee F)'$, and a Peirce decomposition of $V_1(F')$ in terms of E and $(E \vee F)'$, from Lemma 4.36 (ii) we can write

$$x_0^F = x_1^E + x_1^{(E\vee F)'} + x_{1/2}^{E,(E\vee F)'}.$$

But $x \in V_{1/2}(E \vee F)$, so $x_1^E = x_1^{(E\vee F)'} = 0$, and by assumption $x_{1/2}^E = 0$, so $x_0^F = 0$. This gives $T_E + T_F = T_{E\vee F}$ on all of A. □

Before we can prove the main result of this section, we require one more technical result. Recall from our initial discussion of third-order interference the operator R_{123} defined in (4.31), namely,

$$R_{123} = P_{123} - [P_{12} + P_{13} + P_{23} - P_1 - P_2 - P_3] = P_{123} - P^{(3)}.$$

We also have Lemma 4.23 and Proposition 4.25 which in the current case with $P_{123} = \mathbb{I}_A$ state that

$$\ker(P^{(3)}) = \mathrm{im}(R_{123}) = \ker(P_{12}) \cap \ker(P_{13}) \cap \ker(P_{23}). \tag{4.45}$$

Defining the subspace

$$N(F_1, F_2, F_3) := V_{1/2}(F_1) \cap V_{1/2}(F_2) \cap V_{1/2}(F_3), \tag{4.46}$$

we have the following interesting equality:

Lemma 4.42. *Let $(A, S_+(A), u_A)$ be a projective state space, and take two orthogonal faces F_1, F_2, and let $F_3 := (F_1 \vee F_2)'$. Then*

$$N(F_1, F_2, F_3) = \ker(P_{12}) \cap \ker(P_{13}) \cap \ker(P_{23}) = \ker(P^{(3)}), \tag{4.47}$$

where the filters P_{ij} are part of the generalized 3-slit system $\{P_J\}$ and are defined by the faces $F_i \vee F_j$.

Proof. Suppose that $x \in V_{1/2}(F_1) \cap V_{1/2}(F_2) \cap V_{1/2}(F_3)$, which by the orthocomplement relations $F_1' = F_2 \vee F_3$, $F_2' = F_1 \vee F_3$, and $F_3' = F_1 \vee F_2$, along with the relations $V_{1/2}(F_i) = \ker(P_i) \cap \ker(P_i')$ is equivalent to the equalities

$$0 = P_1x = P_{23}x = P_2x = P_{13}x = P_3x = P_{12}x.$$

Clearly, this implies $x \in \ker(P_{12}) \cap \ker(P_{13}) \cap \ker(P_{23})$.

Next suppose that $x \in \ker(P_{12}) \cap \ker(P_{13}) \cap \ker(P_{23})$. From the definition of F_3, it is clear that $P_{123} = P_1 \vee P_2 \vee P_3 = \mathbb{I}_A$, and from Proposition 4.25 we then have $\mathrm{im}(R_{123}) = \mathrm{im}(\mathbb{I}_A - P^{(3)}) = \ker(P_{12}) \cap \ker(P_{13}) \cap \ker(P_{23})$. This equality along with our assumption imply that

$$P^{(3)}x = (P_{12} + P_{13} + P_{23} - P_1 - P_2 - P_3)x = -P_1x - P_2x - P_3x = 0.$$

Using the orthogonality of the P_i, we have $P_i(P_1 + P_2 + P_3) = P_i$, for $i = 1, 2, 3$, and using the above shows that $P_ix = 0$. This implies that $x \in V_{1/2}(F_1) \cap V_{1/2}(F_2) \cap V_{1/2}(F_3)$, which proves the claim. □

Proposition 4.43. *Let $(A, S_+(A), u_A)$ be a projective state space, and take two orthogonal faces F_1, F_2. Then the following are equivalent:*

(i) $T_{F_1} + T_{F_2} = T_{F_1 \vee F_2}$,
(ii) $V_{1/2}(F_1) \cap V_{1/2}(F_2) \cap V_{1/2}((F_1 \vee F_2)') = \{0\}$,
(iii) $V_1(F_1 \vee F_2) = V_1(F_1) \oplus V_1(F_2) \oplus V_{1/2}(F_1) \cap V_{1/2}(F_2)$

Proof. (i) ⇒ (ii) Follows from the statement (i) ⇒ (ii) of Lemma 4.41 and the fact that $V_1(F_1 \vee F_2) \cap V_{1/2}((F_1 \vee F_2)') = V_1(F_1 \vee F_2) \cap V_{1/2}(F_1 \vee F_2) = \{0\}$.

(ii) ⇒ (i) Letting $F_3 = (F_1 \vee F_2)'$, from Lemma 4.42 we have

$$\{0\} = V_{1/2}(F_1) \cap V_{1/2}(F_2) \cap V_{1/2}((F_1 \vee F_2)') = \ker(P_{12}) \cap \ker(P_{13}) \cap \ker(P_{23}).$$

This implies that $R_{123} = 0_A$, which from the proof of Theorem 4.26 (ii) ⇔ (iii) is equivalent to $I_3[q, \{P_J\}, \omega] = 0$ for all $q \in \mathsf{E}(A)$, $\omega \in \mathsf{S}(A)$. Finally, using Lemma 4.39 proves the claim.

(i) ⇔ (iii) First recall from Theorem 4.17 that $P_1P_{12} = P_{12}P_1 = P_1$, as well as $P_2P_{12} = P_{12}P_2 = P_2$. These equalities then imply the inclusions $V_1(F_1) \subseteq V_1(F_1 \vee F_2)$ and $V_1(F_2) \subseteq V_1(F_1 \vee F_2)$, The claim follows from Lemma 4.41 and part (ii) of Lemma 4.36. □

We are now ready to prove the main result of this section: a complete Peirce decomposition of the space A in terms of the V_1 and $V_{1/2}$ spaces associated with a triple of orthogonal faces.

Theorem 4.44. *Let $(A, S_+(A), u_A)$ be a projective state space, and suppose there exist two orthogonal faces E, F, such that $G := (E \vee F)' \neq \{0\}$. Then we have the full direct sum decomposition:*

$$A = V_1(E) \oplus V_1(F) \oplus V_1(G) \oplus V_{1/2}(F)|_{G \vee F} \oplus V_{1/2}(G)|_{E \vee G} \oplus V_{1/2}(E)|_{E \vee F} \oplus \ker(P^{(3)}), \quad (4.48)$$

where $P^{(3)} = P_{E\vee F} + P_{E\vee G} + P_{F\vee G} - P_E - P_F - P_G$ *is the third-order interference operator defined in* (4.24).

Proof. We start with the Peirce decomposition of A into eigenspaces of T_E:

$$A = V_1(E) \oplus V_1(F \vee G) \oplus V_{1/2}(E). \tag{4.49}$$

Peirce decomposing $V_1(F \vee G)$ gives:

$$V_1(F \vee G) = V_1(F) \oplus V_1(G) \oplus V_{1/2}(F)|_{F\vee G}. \tag{4.50}$$

Next, it is clear that $V_{1/2}(E)|_{E\vee G} \subset V_{1/2}(E)$, and $V_{1/2}(E)|_{E\vee F} \subset V_{1/2}(E)$, and from Lemma 4.38, $V_{1/2}(E)|_{E\vee F} \cap V_{1/2}(E)|_{E\vee G} = \{0\}$. These relations imply that we can write

$$V_{1/2}(E) = V_{1/2}(E)|_{E\vee G} \oplus V_{1/2}(E)|_{E\vee F} \oplus Z_1, \tag{4.51}$$

where Z_1 is defined as a complement of $V_{1/2}(E)|_{E\vee G} \oplus V_{1/2}(E)|_{E\vee F}$ in $V_{1/2}(E)$. Substituting (4.50) and (4.51) in (4.49) gives

$$A = V_1(E) \oplus V_1(F) \oplus V_1(G) \oplus V_{1/2}(F)|_{F\vee G} \oplus V_{1/2}(E)|_{E\vee G} \oplus V_{1/2}(E)|_{E\vee F} \oplus Z_1. \tag{4.52}$$

Using the same procedure, but starting with a Peirce decomposition relative to T_F, gives

$$A = V_1(E) \oplus V_1(F) \oplus V_1(G) \oplus V_{1/2}(E)|_{E\vee G} \oplus V_{1/2}(F)|_{F\vee G} \oplus V_{1/2}(F)|_{E\vee F} \oplus Z_2, \tag{4.53}$$

and similarly, by decomposing relative to T_G we find that

$$A = V_1(E) \oplus V_1(F) \oplus V_1(G) \oplus V_{1/2}(F)|_{E\vee F} \oplus V_{1/2}(G)|_{E\vee G} \oplus V_{1/2}(G)|_{F\vee G} \oplus Z_3, \tag{4.54}$$

where Z_2, Z_3 are defined analogously to Z_1 above. Equating (4.52) and (4.53), and using $V_{1/2}(E)|_{E\vee F} = V_{1/2}(F)|_{E\vee F}$ from Corollary 4.37, we see that the complements of Z_1 and Z_2 in these expressions are equal, so without loss of generality we can take $Z_1 = Z_2$. Operating similarly with (4.53) and (4.54) shows that without loss of generality we can also take $Z_2 = Z_3$. This shows that we have a direct sum decomposition $A = B \oplus Z$ where for convenience we have defined

$$B := V_1(E) \oplus V_1(F) \oplus V_1(G) \oplus V_{1/2}(F)|_{G\vee F} \oplus V_{1/2}(G)|_{E\vee G} \oplus V_{1/2}(E)|_{E\vee F}.$$

From Lemma 4.23 and Proposition 4.25, along with $P_{E\vee F\vee G} = \mathbb{I}_A$, we have $A = \mathrm{im}(P^{(3)}) \oplus \ker(P^{(3)})$, where

$$\ker(P^{(3)}) = \mathrm{im}(R_{E\vee F\vee G}) = \ker(P_{E\vee F}) \cap \ker(P_{E\vee G}) \cap \ker(P_{F\vee G}).$$

It remains to show that Z ($:= Z_1 = Z_2 = Z_3$) can be chosen equal to $\ker(P^{(3)})$. First, suppose $x \in B$, namely,

$$x = x_1^E + x_1^F + x_1^G + x_{1/2}^F|_{G\vee F} + x_{1/2}^G|_{E\vee G} + x_{1/2}^E|_{E\vee F},$$

with the obvious notation. Taking $P^{(3)}x$, and using the definitions of the various subspaces in the decomposition of B, along with Theorem 4.17, we find

$$\begin{aligned} P_{E\vee F}(x_1^E) &= P_{E\vee F}P_E x = P_E x = x_1^E, && (4.55)\\ P_{E\vee F}(x_1^F) &= P_{E\vee F}P_F x = P_F x = x_1^F, && (4.56)\\ P_{E\vee F}(x_1^G) &= P_{E\vee F}P_G = 0_A x = 0, && (4.57) \end{aligned}$$

as well as

$$\begin{aligned} P_{E\vee F}(x_{1/2}^F|_{G\vee F}) &= P_{E\vee F}P_{F\vee G}x_{1/2}^F = P_F x_{1/2}^F = 0, && (4.58)\\ P_{E\vee F}(x_{1/2}^G|_{E\vee G}) &= P_{E\vee F}P_{E\vee G}x_{1/2}^G = P_E x_{1/2}^G = 0, && (4.59)\\ P_{E\vee F}(x_{1/2}^E|_{E\vee F}) &= x_{1/2}^E|_{E\vee F}, && (4.60) \end{aligned}$$

and similar relations for $P_{E\vee G}$ and $P_{F\vee G}$ with the appropriate replacements of E, F, G. Putting all these equalities together gives $P^{(3)}x = x$, so $B \subseteq \mathrm{im}(P^{(3)})$. Next, recall from above that $Z_1 \subseteq V_{1/2}(E)$, $Z_2 \subseteq V_{1/2}(F)$, and $Z_3 \subseteq V_{1/2}(G)$, which from the definition of Z gives $Z \subseteq V_{1/2}(E) \cap V_{1/2}(F) \cap V_{1/2}(G)$. Using Lemma 4.42, we have

$$Z \subseteq \ker(P_{E\vee F}) \cap \ker(P_{E\vee G}) \cap \ker(P_{F\vee G}) = \ker(P^{(3)}).$$

Finally, because $A = B \oplus Z$, this last inclusion is equivalent to $\mathrm{im}(P^{(3)}) \subseteq B$, which proves the claim. □

4.9.2 The Peirce decomposition and non-mixing filters

In this section we will use the Peirce decomposition to gain a better understanding of an interesting type of filter, often called *non-mixing*, which takes pure states to multiples of pure states [7, 99, 87]. First recall the notion of an extremal of a cone from Appendices A.3 and A.12.

Definition 4.45 (**Non-mixing filters**). *A filter P on a state space $(A, S_+(A), u_A)$ is called* non-mixing *if and only if for all $\omega \in \mathrm{ext}(S_+(A))$, $P\omega \in \mathrm{ext}(S_+(A))$.*

Theorem 4.46. *Let $(A, S_+(A), u_A)$ be a projective state space, and take two orthogonal faces E, E' such that $E \vee E' = S_+(A)$. Then the filter P_E is non-mixing if and only if every $\alpha \in \mathrm{ext}(S_+(A))$ has a Peirce decomposition $\alpha = \alpha^E + \alpha^{E'} + \alpha^E_{1/2}$ relative to T_E with $\alpha^E \in \mathrm{ext}(E)$.*

Proof. From the Peirce decomposition relative to T_E, namely,

$$A = V_0(E) \oplus V_1(E) \oplus V_{1/2}(E) = \mathrm{im}(P_{E'}) \oplus \mathrm{im}(P_E) \oplus \ker(P_E) \cap \ker(P_{E'}),$$

and the positivity of P_E and $P_{E'}$, any $\alpha \in \mathrm{ext}(S_+(A))$ can be written as $\alpha = \alpha^{E'} + \alpha^E + \alpha^E_{1/2}$, where $\alpha^E \in E$, $\alpha^{E'} \in E'$ and $\alpha^E_{1/2} \in V_{1/2}(E)$. Noting that because $P_E P_{E'} = 0$, we have $P_E(\alpha) = \alpha^E$, it then follows that $\alpha^E \in \mathrm{ext}(E)$ if and only if P_E is non-mixing. □

Quantum Example 4.47. *Recall from Quantum Example 4.8 that filters are maps of the form $P(\hat{B}) = \hat{P}\hat{B}\hat{P}$, where $\hat{B} \in \mathcal{H}_{(d)}$, and $\hat{P}$ is a projection operator on $\mathbb{C}^d$. It is easy to see that all filters are non-mixing by taking a pure state $|\varphi\rangle = \sum_{i=1}^d \alpha_i |i\rangle$ (with $\sum_i |\alpha_i|^2 = 1$) written in a basis in which $\hat{P}$ is diagonal. Looking at the Peirce decomposition of $|\varphi\rangle$ relative to T_F, where F is the face defined by P, we have*

$$|\varphi\rangle\langle\varphi| = \hat{P}|\varphi\rangle\langle\varphi|\hat{P} + (\mathbb{I}_d - \hat{P})|\varphi\rangle\langle\varphi|(\mathbb{I}_d - \hat{P}) + (\mathbb{I}_d - \hat{P})|\varphi\rangle\langle\varphi|\hat{P} + \hat{P}|\varphi\rangle\langle\varphi|(\mathbb{I}_d - \hat{P}).$$

The last two terms correspond to the part of $|\varphi\rangle\langle\varphi|$ in the space $V_{1/2}(F)$. It is then clear that, for example, $\hat{P}|\varphi\rangle\langle\varphi|\hat{P}$ is simply the state $|\psi\rangle = \sum_{i\in\mathcal{I}} \alpha_i |i\rangle$, where $\mathcal{I}$ is the set of basis elements on which $\hat{P}$ is the identity. Up to normalization this is a pure state.

It is also interesting to picture the set of all convex combinations of the form $p\omega^E + (1-p)\omega^{E'}$ where $\omega^E \in \mathrm{ext}(\Omega(E))$ and $\omega^{E'} \in \Omega(E')$, in the real *quantum theory case. In particular, take $\Omega(E)$ a rank-two face of the normalized states, i.e., a disk, and $\Omega(E')$ a rank-one face of the normalized states, i.e, a single pure state perfectly distinguishable from the face $\Omega(E)$. Then the set of all convex combinations of the above form is the boundary of a part of a 3-dimensional Lorentz cone embedded in the set of normalized states of a three-level system, with its point at the pure state $\Omega(E')$. This boundary is made up of line segments each from $\Omega(E')$ to some pure state of $\Omega(E)$. The property that P_E is non-mixing is equivalent to every pure state of the embedding three-level system being 'above' one of these boundary line segments, rather than 'above' the interior of the set $\mathrm{conv}(\Omega(E), \Omega(E'))$. In other words, the image of any pure state, when projected with $P_E + P_{E'}$ onto the hyperplane ($\simeq \mathbb{R}^3$) containing $\mathrm{conv}(\Omega(E), \Omega(E'))$, is exactly on one of these boundary lines!*

It is remarkable that it is possible for any state space that for *all* pure states, their components in *every* Peirce decomposition are multiples of two further pure states in the appropriate faces, along with an extra component. From the above theorem, this is in fact equivalent to all filters being non-mixing.

4.9.3 Mixing and interference given bit symmetry

In this section we will show that in all theories which are bit-symmetric (Chapter 3) and projective, the requirement that all filters are non-mixing is equivalent to the absence of third-order interference.

Before we can prove this equivalence, there is one issue we must deal with: we must ensure that the various direct sum decompositions we have proved in Sections 4.9 and 4.9.1 are in fact orthogonal under the inner product provided by bit symmetry. In particular, we must ensure that the direct sum decompositions of the form $V_{1/2}(E) = V_{1/2}(E)|_{E\vee G} \oplus V_{1/2}(E)|_{E\vee F} \oplus Z_1$ from the proof of Theorem 4.44 are orthogonal.

First, recall from Theorem 3.24 that bit symmetry implies the existence of an inner product $\langle \cdot, \cdot \rangle$ on A such that:

(i) $S_+(A) = S_+(A)^*$,
(ii) $\langle \omega, \omega \rangle = 1, \ \forall\ \omega \in \text{ext}(\Omega(A))$, and
(iii) if $\omega, \varphi \in \Omega(A)$ are perfectly distinguishable, then $\langle \omega, \varphi \rangle = 0$.

Further recall from Appendix A.19 that the positive annihilator of a subset B of a cone $S_+(A)$ is defined by $B^\bullet = \{y \in S_+(A)^* \mid y(x) = 0, \forall\ x \in B\}$. For bit symmetric cones we can use the inner product to write this as

$$B^\perp = \{y \in S_+(A) \mid \langle y, x \rangle = 0, \forall\ x \in B\}. \tag{4.61}$$

We write $B^\perp$ for this positive annihilator to emphasize that we are working with an inner product. First we prove a useful lemma relating the positive annihilator of a face and its complementary face.

Lemma 4.48. *Assume $(A, S_+(A), u_A)$ is bit symmetric and projective. For all exposed faces $F \subset S_+(A)$ we have $F' = F^\perp$.*

Proof. Let F be an exposed face of $S_+(A)$, and take $\omega \in \Omega_{F'} = \ker^+(P_F) \cap \Omega(A)$ and $\varphi \in \Omega_F = \text{im}^+(P_F) \cap \Omega(A)$. The observation $\{p = u_A \circ P_F, p' = u_A - p\}$ is a discriminating observation (see Definition 2.12) for the states ω, φ, namely, $\langle p, \omega \rangle = 0$, and $\langle p, \varphi \rangle = 1$. For bit symmetric theories we also have that if ω, φ are perfectly distinguishable, then

$\langle \omega, \varphi \rangle = 0$, which implies that $F' \subseteq F^\perp$. Because both faces are exposed, we can use Theorem 4.12 (iv) and the isomorphism between the lattice of filters and the lattice of exposed faces to write

$$F^\perp = F' \vee (F^\perp \wedge F'').$$

However, from orthomodularity we also have $F'' = F$, and $F^\perp \wedge F = \{0\}$ because these faces are defined by inner product orthogonality, and only the 0 vector is orthogonal to itself. Using these facts in the above expression for $F^\perp$, we finally have $F^\perp = F' \vee \{0\} = F'$.

Proposition 4.49. *Assume $(A, S_+(A), u_A)$ is bit-symmetric and projective. Every filter P_F associated with some exposed face F is symmetric under the self-dualizing inner product on A, i.e., $P_F^* = P_F$.*

Proof. From Lemma 4.48 we have that $F' = F^\perp$, and because F is exposed, namely $F^{\perp\perp} = F$, we therefore have $(F')^\perp = F$. Using Appendix B.11, as well as the form of the positive annihilator for bit symmetric cones from (4.61), we find

$$(F')^\perp = (\ker^+(P_F))^\perp = \mathrm{im}^+(P_F^*). \tag{4.62}$$

Together with $(F')^\perp = F = \mathrm{im}^+(P_F)$, this implies $\mathrm{im}^+(P_F^*) = \mathrm{im}^+(P_F)$.

Next let $x, y \in A$, and take $\langle (\mathbb{I}_A - P_F)x, P_F y \rangle$. It is clear that $z := P_F y \in \mathrm{im}(P_F)$, and from above also $z \in \mathrm{im}(P_F^*)$. Then

$$\langle (\mathbb{I}_A - P_F)x, z \rangle = \langle x, (\mathbb{I}_A - P_F)^* z \rangle = \langle x, z - z \rangle = 0.$$

This is equivalent to the equality $\langle x, P_F y \rangle = \langle P_F x, P_F y \rangle$ for all $x, y \in A$. Exchanging the roles of x, y we also have $\langle P_F x, y \rangle = \langle y, P_F x \rangle = \langle P_F y, P_F x \rangle$, and using $\langle P_F x, P_F y \rangle = \langle P_F y, P_F x \rangle$, we find

$$\langle P_F x, y \rangle = \langle P_F x, P_F y \rangle = \langle x, P_F y \rangle,$$

which proves the claim. □

As an aside, recalling Definition 3.8, we have the following:

Corollary 4.50. *If $(A, S_+(A), u_A)$ is bit symmetric and projective, then every exposed face $F \subset S_+(A)$ is self-dual in its span, i.e., $S_+(A)$ is perfect.*

Proof. Follows from Proposition 4.49, and Proposition 3.9. □

We are now ready to prove the inner product orthogonality of the Peirce decompositions.

Proposition 4.51. *Assume $(A, S_+(A), u_A)$ is bit symmetric and projective. For any pair of exposed orthogonal faces E, F, with $G := (E \vee F)'$, we have:*

(i) $\langle V_{1/2}(E), V_1(E)\rangle = \langle V_{1/2}(E), V_0(E)\rangle = \langle V_0(E), V_1(E)\rangle = 0$.
(ii) $\langle V_{1/2}(E)|_{E\vee G}, V_{1/2}(E)|_{E\vee F}\rangle = 0$.

Proof. (i) From bit symmetry we have that the filters P_E, P'_E associated with the face E are symmetric under the inner product, and from the definition of these filters we have $P_E P'_E = P'_E P_E = 0$. The statement then follows from the definitions of the subspaces $V_0(E), V_1(E), V_{1/2}(E)$.

(ii) If $G = \{0\}$, then $V_{1/2}(E)|_{E\vee G} = \{0\}$ and the statement is trivial. Suppose that $G \neq \{0\}$. Take $x \in V_{1/2}(E)|_{E\vee G}$, which by the definition $V_{1/2}(E)|_{E\vee G} = V_{1/2}(E) \cap V_1(E\vee G)$ means that $P_E x = 0$, and $P_{E\vee G} x = x$, and further take $y \in V_{1/2}(E)|_{E\vee F}$, which means $P_E y = 0$, and $P_{E\vee F} y = y$. The inner product between x and y satisfies

$$\langle x, y\rangle = \langle P_{E\vee G} x, P_{E\vee F} y\rangle = \langle x, P_{E\vee G} P_{E\vee F} y\rangle,$$

which follows from the above, and the inner product orthogonality of the filters. We further have that $P_{E\vee G} P_{E\vee F} = P_{E\vee F} P_{E\vee G} = P_E$ from Theorem 4.17. So, $\langle x, y\rangle = \langle x, P_E y\rangle = 0$ follows from $P_E y = 0$. □

Theorem 4.52. *Assume $(A, S_+(A), u_A)$ is bit symmetric and projective. Suppose there exist two orthogonal faces E, F, such that $G := (E\vee F)' \neq \{0\}$. Then the filters P_E, P_F, P_G, $P_{E'}, P_{F'}, P_{G'}$ are non-mixing if and only if $N(E, F, G) = \{0\}$.*

Proof. Take a pure normalized state $\omega \in \mathrm{ext}(\Omega(A))$ and decompose it in terms of the full Peirce decomposition from Theorem 4.44:

$$\omega = \langle e, \omega\rangle \omega^E + \langle f, \omega\rangle \omega^F + \langle g, \omega\rangle \omega^G + \omega^{FG} + \omega^{EF} + \omega^{EG} + \omega^N, \tag{4.63}$$

with the obvious notation. To see that the coefficients in the first three terms are as stated, take say $P_E \omega =: \tilde{\omega}^E$, and note that $\langle u_A, P_E \omega\rangle = \langle u_A, \tilde{\omega}^E\rangle = \langle e, \omega\rangle$ where we have used $e = u_A \circ P_E = P_E u_A$ and the inner product. By defining the normalized state $\omega^E := \tilde{\omega}^E / \langle u_A, \tilde{\omega}^E\rangle$, we have $P_E \omega = \langle e, \omega\rangle \omega^E$. Using Proposition 4.51 and the above, it follows that

$$\begin{aligned} 1 = ||\omega||^2 = \langle e, \omega\rangle^2 ||\omega^E||^2 + \langle f, \omega\rangle^2 ||\omega^F||^2 + \langle g, \omega\rangle^2 ||\omega^G||^2 \\ + ||\omega^{FG}||^2 + ||\omega^{EF}||^2 + ||\omega^{EG}||^2 + ||\omega^N||^2. \end{aligned} \tag{4.64}$$

Next, taking the Peirce decomposition with respect to T_E, we also have

$$\omega = \langle e, \omega\rangle \omega^E + \langle e', \omega\rangle \omega^{E'} + \omega^E_{12}, \tag{4.65}$$

where $P_{E'} \omega = \langle e', \omega\rangle \omega^{E'}$. We also have similar decompositions with respect to T_F and T_G.

Now assume that the filters $P_E, P_F, P_G, P_{E'}, P_{F'}, P_{G'}$ are non-mixing. This implies that $\omega^E, \omega^{E'}$ are pure normalized states, and similarly for ω^F, ω^G as well as $\omega^{F'}, \omega^{G'}$ from $P_{G'}\omega = \langle g', \omega\rangle \omega^{G'}$, and $P_{F'}\omega = \langle f', \omega\rangle \omega^{F'}$. Further, we also have more detailed decompositions of these states from the Peirce decomposition $V_1(E') = V_1(F \vee G) = V_1(F) \oplus V_1(G) \oplus V_{1/2}(F)|_{F\vee G}$ (as in (4.50)), and analogous decompositions of $V_1(F')$ and $V_1(G')$:

$$\begin{aligned}
\omega^{E'} &= \frac{1}{\langle e', \omega\rangle}\left(\langle f, \omega\rangle \omega^F + \langle g, \omega\rangle \omega^G + \omega^{FG}\right), \\
\omega^{F'} &= \frac{1}{\langle f', \omega\rangle}\left(\langle e, \omega\rangle \omega^E + \langle g, \omega\rangle \omega^G + \omega^{EG}\right), \\
\omega^{G'} &= \frac{1}{\langle g', \omega\rangle}\left(\langle f, \omega\rangle \omega^F + \langle e, \omega\rangle \omega^E + \omega^{EF}\right).
\end{aligned} \tag{4.66}$$

In particular, these are normalized pure states, so we have

$$\begin{aligned}
1 = ||\omega^{E'}||^2 &= \frac{1}{\langle e', \omega\rangle^2}\left(\langle f, \omega\rangle^2 + \langle g, \omega\rangle^2 + ||\omega^{FG}||^2\right), \\
1 = ||\omega^{F'}||^2 &= \frac{1}{\langle f', \omega\rangle^2}\left(\langle e, \omega\rangle^2 + \langle g, \omega\rangle^2 + ||\omega^{EG}||^2\right), \\
1 = ||\omega^{G'}||^2 &= \frac{1}{\langle g', \omega\rangle^2}\left(\langle f, \omega\rangle^2 + \langle e, \omega\rangle^2 + ||\omega^{EF}||^2\right).
\end{aligned} \tag{4.67}$$

Further, from $e' = u_A - e = f + g$, it follows that $\langle e', \omega\rangle = \langle f, \omega\rangle + \langle g, \omega\rangle$, and therefore

$$\langle e', \omega\rangle^2 = \langle f, \omega\rangle^2 + \langle g, \omega\rangle^2 + 2\langle f, \omega\rangle\langle g, \omega\rangle, \tag{4.68}$$

with similar expressions for $\langle f', \omega\rangle^2$ and $\langle g', \omega\rangle^2$ as well. Equating $\langle e', \omega\rangle^2$ in 4.68 and in 4.67 (as well as the other two norms and squares) gives

$$\begin{aligned}
||\omega^{FG}||^2 &= 2\langle f, \omega\rangle\langle g, \omega\rangle, \\
||\omega^{EG}||^2 &= 2\langle e, \omega\rangle\langle g, \omega\rangle, \\
||\omega^{EF}||^2 &= 2\langle e, \omega\rangle\langle f, \omega\rangle.
\end{aligned} \tag{4.69}$$

The equality $\langle u_A, \omega\rangle = \langle e, \omega\rangle + \langle f, \omega\rangle + \langle g, \omega\rangle = 1$, further implies

$$1 = \langle e, \omega\rangle^2 + \langle f, \omega\rangle^2 + \langle g, \omega\rangle^2 + 2\langle e, \omega\rangle\langle f, \omega\rangle + 2\langle e, \omega\rangle\langle g, \omega\rangle + 2\langle f, \omega\rangle\langle g, \omega\rangle. \tag{4.70}$$

Equating this with 4.64, and using Equations 4.69, we see that $||\omega^N||^2 = 0$, so $\omega^N = 0$ for all $\omega \in \mathrm{ext}(\Omega(A))$.

Running the above argument backwards, assume that $||\omega^N||^2 > 0$ for some set of orthogonal faces E, F, G, and using the various Peirce decompositions and norm equalities above, we conclude that one of the states $\omega^E, \omega^F, \omega^G, \omega^{E'}, \omega^{F'}, \omega^{G'}$ must have norm strictly less than 1, and therefore must be mixed. □

It seems plausible that this relationship between the subspace $N(E, F, G)$ and the properties of the filters associated with the faces E, F, G is true in general, and not only for bit symmetric theories. Note however the essential use of the property that pure states have norm equal to one in the above proof. Therefore a proof of such a relationship would have to have a very different structure from the above.

Corollary 4.53. *Assume $(A, S_+(A), u_A)$ is bit symmetric and projective. Take two orthogonal faces F_1, F_2, such that $F_3 := (F_1 \vee F_2)' \neq \{0\}$, and form the generalized 3-slit system $\{P_J\}$ defined by these faces. If the filters P_J are all non-mixing, then $\mathsf{I}_3[q, \{P_J\}, \omega] = 0$ for all $\omega \in \mathsf{S}(A)$ and $q \in \mathsf{E}(A)$.*

Proof. By Theorem 4.52, $N(F_1, F_2, F_3) = \{0\}$, and by Lemma 4.42 this implies that $\ker(P^{(3)}) = \{0\}$, or equivalently $\mathrm{im}(P^{(3)}) = A$. Then Lemma 4.24, implies that $\mathrm{lin}(F_{12} \cup F_{23} \cup F_{12}) = \mathrm{im}(P^{(3)}) = A$, and Theorem 4.26 finally gives $\mathsf{I}_3[q, \{P_J\}, \omega] = 0$ for all $\omega \in \mathsf{S}(A)$ and $q \in \mathsf{E}(A)$. □

4.10 Third-order interference and other state spaces

As we have already stated, quantum theory, and therefore also classical probability theory satisfy $\mathsf{I}_3 = 0$. The following cones (as well as all direct convex sums of them) are also projective and possess many 3-slit masks:

- $d \times d$, positive semi-definite real symmetric matrices,
- $d \times d$, positive semi-definite quaternionic self-adjoint matrices,
- 3×3, positive semi-definite octonionic self-adjoint matrices.

These, along with the $d \times d$, positive semi-definite self-adjoint complex matrices and the so called Lorentz cones (for which the normalized states form a ball in d dimensions), are in fact the state spaces – made up of squares of elements – of the simple finite dimensional Jordan-Banach algebras (see Section 5.1 and in particular Definition 5.12 and Quantum Example 5.17 for more detail on these models). The Lorentz cones were excluded from the above list because they do not support any nontrivial 3-slit masks; they are all essentially two-level systems, so fail to have higher order interference by default.

In [138] Gerd Niestegge has proven that every state space associated with a finite dimensional Jordan-Banach algebra which supports a 3-slit mask displays no third-order interference. The proof is based on the fact that these state spaces possess a certain symmetry (see [7] Theorem 9.43), along with Lemma 4.39. Here we present a different proof based on the fact that these state spaces are bit symmetric, along with Corollary 4.53.

Theorem 4.54. *Let $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$ be a Jordan dynamical state space (Definition 5.12) which supports a 3-slit mask $\{P_i\}_{i=1}^3$ that generates the generalized slit system $\{P_J\}$. Then $\mathsf{I}_3[q, \{P_J\}, \omega] = 0$ for all $\omega \in \mathsf{S}(A)$ and $q \in \mathsf{E}(A)$.*

Proof. First, by Theorems 5.11 and 5.13, for each idempotent $c \in A$, the associated Jordan compression U_c (Definition 5.10) is a filter (Definition 4.4). Further, from Theorem 5.14, we have that filters on Jordan dynamical state spaces are in one-to-one correspondence with Jordan compressions, and from Theorem 5.15 we have that Jordan dynamical state space spaces are projective. Then from Theorem 5.16 the Jordan compressions are non-mixing. Finally, from Theorem 5.9 and Quantum Example 5.17 we have that Jordan dynamical state spaces are bit symmetric. The claim then follows from Corollary 4.53. □

It is also interesting to consider models which have some initial states and generalized slit systems for which $\mathsf{I}_3 \neq 0$. We now present a model which is projective and supports a 3-slit mask.

Example 4.55. *Take the triangle defined by the vertices*

$$\omega_1 = (0, 0, 0),\ \omega_2 = (1, 1, 0),\ \omega_3 = (1, -1, 0) \in \mathbb{R}^3, \tag{4.71}$$

and take each line segment of the form $[(0, 0, 0), (1, y_0, 0)]$, with $|y_0| \leq 1$, and expand it to the elliptical disk

$$D_{y_0} := \left\{ (x, y, t) \mid y = y_0 x,\ t^2 \leq x(1 - x)\sqrt{1 - y_0^2} \right\}. \tag{4.72}$$

Next, take the union of all the above elliptical disks, namely $\bigcup_{y_0} D_{y_0}$, and define the set of normalized states $\Omega_A := \bigcup_{y_0} D_{y_0}$ to be the resulting "triangular pillow" (see Figure 4.4 below). Then embed Ω_A in $\mathbb{R}^4$ in the hyperplane $H := \{(x, y, t, z) \mid z = 1\}$, and let the cone $S_+(A)$ of un-normalized states be defined as $S_+(A) := \bigcup_{\lambda \geq 0} \lambda \cdot \Omega_A$. Further, let the order unit for this model be $u_A := (0, 0, 0, 1)$. The resulting state space $(\mathbb{R}^4, S_+(A), u_A)$ projective (for a proof see Chapter 8 of [7]).

It is not difficult to see that Ω_A has two types of (nontrivial) faces. The three vertices ω_i, as well as all boundary points on the curved surface off the central triangle make up the

pure states of the model. The three edges of the triangle (which we denote by Ω_{ij}) are the other faces, and are analogous to two-level classical systems. We also have that for each face $\tilde{F}$ of Ω_A, there is a face F of the cone $S_+(A)$ such that $\tilde{F} = F \cap H$.

Now define three projections P_i ($i = 1, 2, 3$) onto the faces F_i (which are generated by the vertices ω_i), with positive kernels F_{jk} ($j, k \neq i$), and further take $P_{123} = \mathbb{I}_{\mathbb{R}^4}$. These four projections are in fact filters, and the P_i form a size three mask. Finally, the filters defined by $P_{jk} = P_j + P_k$ will be projections onto the faces F_{jk}. Since $\text{lin}\{\tilde{F}_{12}, \tilde{F}_{13}, \tilde{F}_{23}\}$ just gives the plane defined by $(x, y, t = 0, z = 1)$ (in which the central triangle is embedded), it is clear that $S_+(A) \nsubseteq \text{lin}\{F_{12}, F_{13}, F_{23}\}$; the linear span of these faces is missing the t dimension. Therefore, any normalized state which is not in the central triangle will exhibit third order interference with respect to the generalized slit system generated by $\{P_i\}_{i=1}^3$.

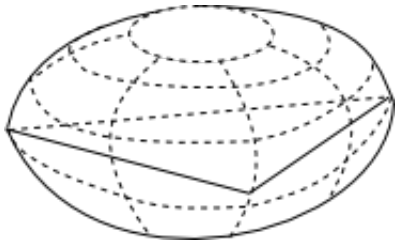

Figure 4.4: The triangular pillow state space discussed above, and in Chapter 8 of [7]. Pure states on the smooth top or bottom parts of the pillow will display third-order interference with respect to a 3-slit mask defined by the pure states of the central embedded triangle.

It is possible to construct many more similar examples simply by replacing the central triangle embedded in $\mathbb{R}^2$ with a direct sum of an n-ball and an m-ball embedded in $\mathbb{R}^{n+m}$ (see Chapter 8 of [7] for more detail). For example, for $n = m = 3$, the resulting state space can roughly be considered as a direct sum of two qubits (or a four-level system with a kind of super-selection rule) with an extra degree of freedom (t in the example above). The subset of states which are in the 'quantum' sector all have $t = 0$.

One important point about the above type of construction is that the pure states of the initial state space (the triangle of the triangular pillow) are of a different type than the pure states on the top and bottom of the smooth part of the pillow. In particular, these theories display a high degree of *asymmetry*. In particular, it is not difficult to convince oneself that the filters, P_{jk}, defined on the triangular pillow are in fact mixing, and further, the triagular pillow is not bit-symmetric.

4.11 Discussion

In this chapter we have studied a notion of probabilistic interference by adapting Raphael Sorkin's hierarchy of interference expressions to the operational probabilistic theories framework. Our first technical result was that the absence of third-order interference is equivalent to the possibility of reconstructing a state via specific sets of two-slit filtering experiments. This gives new insight into the structure of quantum theory and the implications of three-slit experiments: the presence of third-order interference in a set of experiments implies that more parameters are needed to describe a system than those specified by the quantum formalism. This suggests a novel way of testing for third-order interference: test whether the tomographic procedure outlined in Section 4.7.1 is in fact sufficient to fully characterize actual preparations.

We also discovered a connection between bit symmetric theories (see Chapter 3) which possess certain natural transformations called non-mixing filters, and those bit symmetric theories which display at most second-order interference. An interesting question is whether the assumption of bit-symmetry is essential here, or whether the existence of non-mixing filters is related to the absence of third-order interference in all theories.

One issue we have not discussed is the use of filters to represent the generalized slits, as well the role of the standing condition in our result. It may in fact be possible to drop the projective state space assumption. One promising weaker assumption is that made by Araki in [9].

More generally, it would be interesting to begin a study of how each interference expression is related to other nonclassical phenomena that generalized models exhibit, such as information processing properties, non-locality, symmetry properties, etc.

Chapter 5

Jordan algebras from operational principles

Current descriptions of quantum theory are simply statements of its mathematical formalism: a very specific and abstract set of propositions for how to represent experiments and make predictions. Throughout the history of physics there have been many theories or laws which were only understood, and their predictions made intuitive, when natural principles were found which imply the theory or law. Standard examples are Kepler's laws and their explanation through Newton's laws of motion and gravitation, and the Lorentz transformations and their explanation in Einstein's two relativity postulates.

Providing this form of explanation for the quantum formalism is important for several reasons. First, deeper and more reasonable principles can dissolve the mysteries of quantum phenomena and make them more intuitive. Second, it can be argued that this is essential to making progress on problems such as formulating a quantum theory of gravity, as well as for developing other potentially more accurate and more fundamental theories. In fact, defining consistent modifications of the standard formalism is a surprisingly difficult task (see for example [1, 2], as well as the references and discussion in [124]). More practically, this approach can shed light on what is responsible for the power of quantum information processing and cryptography. Finally, it can help the philosophically minded get a better night's sleep knowing the nature of reality underlying quantum phenomena.

The first question is then what kinds of principles should we expect for quantum theory, which applies to an incredibly broad range of physical systems and phenomena. Any potential principles should therefore also be natural in any situation in which quantum theory can be applied. Typically, an application of quantum theory involves finding the

right Hilbert space(s) for the phenomenon of interest, and correspondence rules between operators on this Hilbert space and physical quantities of interest [14]. This is a question of *which* complex Hilbert space. Here we are interested in the question of *why always* complex Hilbert spaces. This suggests that the principles should not make reference to particular physical quantities, but instead should be more abstract [70, 95, 83, 181].

The first attempt along these lines was made by Birkhoff and von Neumann [37]. Their approach was to take the collection of all yes-no statements (often called propositions) which can be made about a quantum system and to consider these as forming a new form of "logic", distinct from the classical or Boolean sort, in which the distributive law no longer holds. This work launched the "quantum logic" tradition and led to many axiomatizations of quantum theory, such as by Mackey [120], Piron [144], and Zierler [182]. See also [32, 50, 67, 171, 10] for more discussion of the quantum logic approach. It may be argued that this approach has not succeeded in providing simple and fundamental principles for the quantum formalism, partly because of the focus on infinite dimensional systems, as well as on the structure of single systems.

More recently, the rise of quantum information science has shown that the mathematical principles of quantum theory have many interesting operational or information theoretic consequences, such as the possibility of teleportation, the no cloning theorem, secure key distribution, new and powerful algorithms, and much more besides [137]. This has led to a renewed interest in the axiomatization problem, and in particular whether at least some of the principles can have an information-theoretic flavor [70]. In 2001 Lucien Hardy [95] made the first effort in this new direction, which was largely based on the requirements of continuous and reversible time evolution and the structure of composite systems. However, one of the axioms – which requires that states are specified by the smallest number of probabilities consistent with the other axioms (see **??** for an interesting attempt to construct a theory violating this) – is essentially a mathematical requirement and is not well motivated other than by an appeal to simplicity. The question of whether this could be replaced with something more compelling was left open until the success of Dakic and Brukner [51]. Their argument was subsequently sharpened further in another reconstruction by Masanes and Müller [124]. Another approach introduced in [45] by Chiribella, D'Ariano, and Perinotti, and completed in [46], has as its centerpiece the requirement that ignorance about a part of a system is always consistent with maximal knowledge of the whole system. Besides these, there have been many other partial or full reconstructions over the last decade [47, 53, 177, 82, 147, 80, 72, 64]. It is interesting to note that many of these new works have been within the general probabilistic theories framework discussed in Chapter 2, and have focused on finite-dimensional systems, which allows for a clean split between conceptual and technical issues, and have also emphasized the properties of

composite systems.

In this chapter we will focus in particular on reconstructing the algebraic structures of quantum theory for *individual* systems. While not recovering the exact (complex) structure of standard quantum theory, we will show that its closest mathematical cousins, namely the finite-dimensional formally real Jordan-algebras, can be derived from three simple principles having an informational flavor:

(1) a generalized spectral decomposition,
(2) a high degree of symmetry, and
(3) a generalization of the von Neumann-Luders projection postulate.

In Section 5.1 we begin by discussing the "algebraic approach" to reconstructing the formalism of quantum theory, and then give a short introduction to the theory Jordan algebras. Our precise principles are then presented and discussed in Section 5.2. We begin the reconstruction in Section 5.3 by deriving many interesting consequences only of Principles 1 and 2. Principle 3 is then included in the mix in Section 5.4, where we also complete the reconstruction. Finally, we present two principles which are equivalent to our third in Section 5.5, and conclude with a brief discussion in Section 5.6.

5.1 Jordan Algebras

Jordan algebras have played an important role in the development of quantum theory as well as various axiomatizations. In fact, the second oldest strategy for attacking the question "why quantum theory?" is to focus on the algebraic structure of the theory and ask "why should the set of observables of a system form the self-adjoint part of a C^*-algebra, or any algebra at all?".[1]

One of the most important questions in this approach is how to operationally motivate or define the product and sum of two observables which cannot be simultaneously measured. The earliest attempt to address this question was made by Pascual Jordan [106] in 1932. By focusing on sums and squares of observables, Jordan was led to define a commutative product, with the resulting algebra now called a formally real Jordan algebra. However, this commutative product still relies on the operationally dubious notion of a sum of

[1]It is interesting to note that the algebraic formulation of quantum theory [6, 7, 91] is in a sense more general than the Hilbert space formulation, as it includes classical mechanics as a special case, as well as quantum theory with super-selection rule. Perhaps most importantly however, a purely algebraic formulation (independent of Hilbert space representations) is especially relevant in quantum field theories where there are issues related to the existence of unitarily inequivalent representations [91].

observables, and perhaps more importantly, such a product is generally non-associative and non-distributive with respect to the sum, so a further 'distributivity axiom' is needed to recover the usual quantum formalism. For similar approaches and discussion of these issues see [152, 153, 166, 168].

A closely related approach, and the one we will take, is to begin within the general probability theories framework, namely to start with non-algebraic structures, and to look for postulates which allow for the definition of a product of observables which has the right properties to give the self-adjoint part of a C^*-algebra (or more generally, a formally real Jordan algebra). For example, see Gunson [87], Araki [9], Friedman and Russo [69], Niestegge [138], and most importantly for our development, Chapter 9 of Alfsen and Shultz [7].

Finally, we should point out the possibility of using the powerful Koecher-Vinberg Theorem [173, 29, 4], which states that the set of all formally real Jordan algebras is exactly the set of homogeneous[2] and self-dual cones. Given this result, the remaining task is then to operationally motivate homogeneity and self-duality (see Chapter 3). For more on the Koecher-Vinberg Theorem and this approach see [102, 28, 87, 115, 116], and for more recent discussion see [11, 180].

We will now give a short introduction to the theory Jordan-Banach algebras, focused largely on those elements essential to understanding the subsequent reconstruction. For more details, see Chapters 1-5 of [7], Chapters II-V of [61], as well as [11, 127]. Most of the following material can be skipped on first reading.

Definition 5.1 (**Jordan algebra**). *A finite-dimensional* Jordan algebra *over* $\mathbb{R}$ *is a real finite-dimensional vector space* A *equipped with a commutative bilinear product* $\circ$ *that satisfies the (Jordan) identity*

$$(a^2 \circ b) \circ a = a^2 \circ (b \circ a), \quad \forall\ a, b \in A, \tag{5.1}$$

where $a^2 := a \circ a$.

Note that the Jordan identity is a special case of the associative law; Jordan algebras are not associative in general.

[2] A homogeneous cone is one for which every pair of interior points can be mapped into each other by some invertible, positive linear map on the cone. Another way of phrasing this is that there is a distinguished state of the cone, which can be called 'the state of maximal ignorance', and that any other state can reached from this distinguished state by an appropriate 'observable' [11]. In particular, the cones of quantum states, $S_+(\mathcal{H}_{(d)})$ are homogeneous: for any two strictly positive operators $\hat{\rho}, \hat{\sigma} \in \text{int}S_+(\mathcal{H}_{(d)})$, there exists an invertible linear map on $\mathcal{H}_{(d)}$ (generally normalization increasing) which maps $\hat{\rho}$ to $\hat{\sigma}$.

A finite-dimensional Jordan algebra A over $\mathbb{R}$ is said to be *formally real* if $a^2 + b^2 = 0$ implies that $a = b = 0$ for all $a, b \in A$. Every finite-dimensional formally real Jordan algebra has a multiplicative identity (is *unital*), namely, there is an element $\mathbf{1} \in A$ such that $\mathbf{1} \circ a = a \circ \mathbf{1} = a$ for all $a \in A$ [11, 127].

In 1934, Jordan, von Neumann, and Wigner [107] classified all the finite-dimensional formally real Jordan algebras. The first step in this classification was to show that any such algebra is a direct sum of *simple* algebras, namely, algebras A which have as the only ideals $\{0\}$ and A itself, where an ideal is a vector subspace $B \subset A$ such that $b \in B$ implies $a \circ b \in B$ for all $a \in A$.

Theorem 5.2 ([107])**.** *Every finite-dimensional formally real simple Jordan algebra is isomorphic to one of the following:*

- *The algebra $\mathcal{H}_{(d)}(\mathbb{R})$ of $d \times d$ self-adjoint real matrices.*
- *The algebra $\mathcal{H}_{(d)}(\mathbb{C})$ of $d \times d$ self-adjoint complex matrices.*
- *The algebra $\mathcal{H}_{(d)}(\mathbb{H})$ of $d \times d$ self-adjoint quaternionic matrices.*
- *The algebra $\mathcal{H}_{(3)}(\mathbb{O})$ of 3×3 self-adjoint octonionic matrices.*
- *The 'spin factor' algebras $\mathbb{R}^d \oplus \mathbb{R}$ equipped with the product*

$$(a, t) \circ (b, u) = (tb + ua, \langle a, b \rangle + tu), \tag{5.2}$$

  where $\langle \cdot, \cdot \rangle$ is the standard inner product on $\mathbb{R}^d$.

The first four of the above algebras are equipped with the well known product $a \circ b = \frac{1}{2}(ab + ba)$. Further, a square matrix T is said to be *self-adjoint* if $T_{ji} = (T_{ij})^*$, where $()^*$ is the conjugation of the underlying field, $\mathbb{R}, \mathbb{C}, \mathbb{H}, \mathbb{O}$. The 3×3 self-adjoint octonionic matrices form a 27-dimensional formally real Jordan algebra called the *exceptional Jordan algebra*, because these matrices cannot be seen as linear operators acting on a quaternionic Hilbert space.

Definition 5.3 (**Idempotents and Jordan frames**, [61] Chapter III)**.** *Let A be a finite-dimensional Jordan algebra over $\mathbb{R}$ with multiplicative identity $\mathbf{1}$. An* idempotent *(sometimes a* projection*) is an element c of A satisfying $c^2 = c$. An idempotent c is* primitive *if it is non-zero and cannot be written as a sum of two non-zero idempotents. Two idempotents c, d are* orthogonal *if $c \circ d = 0$. A set of idempotents $c_1, \ldots, c_n$ is a* complete system of orthogonal idempotents *if*

$$c_j \circ c_k = 0, \quad j \neq k, \tag{5.3}$$

$$\sum_{j=1}^{n} c_j = \mathbf{1}. \tag{5.4}$$

Finally, a complete system of orthogonal idempotents $c_1, \ldots, c_n$ is called a Jordan frame *if each c_j is primitive.*

Theorem 5.4 (**Spectral theorem**, [61] Theorem III.1.1)**.** *Let A be a finite-dimensional Jordan algebra over $\mathbb{R}$. For every $a \in A$ there exist unique real numbers $\lambda_1, \ldots, \lambda_n$, all distinct, and a unique complete system of orthogonal idempotents $c_1, \ldots, c_n$ such that*

$$a = \sum_{j=1}^{n} \lambda_j c_j. \tag{5.5}$$

The numbers λ_j are said to be the eigenvalues *of a, and $\sum_{j=1}^{n} \lambda_j c_j$ the* spectral decomposition *of a.*

Note that the spectral decomposition of an element x can be refined to a Jordan frame (see [61] Theorem III.1.2). If it is so refined, the spectral decomposition is no longer unique in general, and eigenvalues can repeat.

Jordan algebras are also naturally equipped with a bilinear *trace* form. Defining (without loss of generality) the left multiplication operator as

$$L_a : A \to A, \quad b \mapsto L_a(b) := a \circ b, \tag{5.6}$$

we then have

$$tr : A \to \mathbb{R}, \quad a \mapsto tr(a) := \mathrm{Tr}(L_a), \tag{5.7}$$

where $\mathrm{Tr}(L_a)$ is the usual trace of operators on real vector spaces. If $a = \sum_{j=1}^{n} \lambda_j c_j$ is the spectral decomposition of a, then it is not difficult to see that $tr(a) = \sum_{j=1}^{n} \lambda_j$. The trace form induces a symmetric non-degenerate bilinear form $\langle \cdot, \cdot \rangle$, namely,

$$\langle \cdot, \cdot \rangle : A \times A \to \mathbb{R}, \quad (a, b) \mapsto \langle a, b \rangle := tr(a \circ b). \tag{5.8}$$

A finite-dimensional Jordan algebra A over $\mathbb{R}$ is said to be *Euclidean* if this induced bilinear form is positive definite, i.e., is an inner product on A. A finite-dimensional Jordan algebra A over $\mathbb{R}$ is Euclidean if and only if it is formally real (see [61] Proposition VIII.4.2).

If A is a finite-dimensional formally real Jordan algebra over $\mathbb{R}$, any linear functional $f : A \to \mathbb{R}$ can be written as

$$f(a) = tr(\rho_f \circ a) \tag{5.9}$$

for a unique element $\rho_f \in A$. The linear functional f is nonnegative if and only if ρ_f has non-negative eigenvalues. Conversely, every element $b \in A$ gives a linear functional by this formula [11].

Definition 5.5 (**Jordan-Banach algebra**, [7] Definition 1.5)**.** *A finite-dimensional* Jordan-Banach algebra *is a Jordan algebra A over $\mathbb{R}$ with multiplicative identity element $\mathbf{1}$, and a norm satisfying the following requirements for all $a, b \in A$:*

$$\begin{aligned} ||a \circ b|| &\leq ||a||\,||b||, && (5.10)\\ ||a^2|| &= ||a||^2, && (5.11)\\ ||a^2|| &\leq ||a^2 + b^2||. && (5.12) \end{aligned}$$

If A is Euclidean (equivalently if A is formally real), then the inner product induced by the trace functional can be used to define a norm (in the usual fashion) which makes A into a Jordan-Banach algebra. Further, an algebra is a finite-dimensional formally real Jordan algebra if and only if it is a finite-dimensional Jordan-Banach algebra (see [94] Corollary 3.1.7 and Corollary 3.3.8).

Definition 5.6 (**Cone of squares**, [61] Chapter III.2)**.** *Let A be a finite-dimensional Jordan-Banach algebra over $\mathbb{R}$. The* cone of squares *of A is defined by*

$$Q = \{a^2 \mid a \in A\}, \tag{5.13}$$

and its closed dual Q^* is defined by

$$Q^* = \{b \in A \mid \forall\ a \in A,\ \langle b, a^2 \rangle \geq 0\}. \tag{5.14}$$

It is clear that the idempotents of A are contained in the cone of squares Q, and in fact all extremal rays (Appendix A.12) of Q are generated by the primitive idempotents (see [61] Proposition IV.3.2). It is also interesting to note that every element of Q has non-negative eigenvalues. Recalling the notion of a self-dual cone from Definition 3.3, we have the following important result:

Theorem 5.7 ([61] Theorem III.2.1)**.** *Let A be a finite-dimensional Jordan-Banach algebra over $\mathbb{R}$. Then Q is self-dual under the inner product induced by the trace form, and furthermore,*

$$Q^* = Q = \{a \in A \mid L_a \text{ is positive semi-definite}\}. \tag{5.15}$$

Definition 5.8 (**Jordan automorphism**)**.** *Let A be a finite-dimensional Jordan algebra over $\mathbb{R}$. A* Jordan automorphism *of A is an invertible linear transformation T such that*

$$T(a \circ b) = T(a) \circ T(b), \tag{5.16}$$

for all $a, b \in A$.

The set of all Jordan automorphisms of A is a closed subgroup of $\mathrm{GL}(A)$, and therefore a Lie group. It is also interesting to note that Jordan automorphisms are also order automorphisms of the positive cone (i.e., bijective linear maps T with $T(Q) = Q$ and $T^{-1}(Q) = Q$) and preserve the multiplicative identity. In fact, cone automorphisms of Jordan-Banach algebras which preserve the multiplicative identity are Jordan automorphisms ([7] Theorem 2.80).

Theorem 5.9 ([61] Theorem IV.2.5)**.** *Let A be a finite-dimensional simple Jordan-Banach algebra over $\mathbb{R}$. If $c_1, \ldots, c_n$ and $d_1, \ldots, d_n$ are two Jordan frames, then there exists a Jordan automorphism T such that*

$$Tc_j = d_j, \quad 1 \leq j \leq n. \tag{5.17}$$

Further, the identity component of the group of Jordan automorphisms of A is transitive on the set of primitive idempotents and also on the set of Jordan frames.

Definition 5.10 (**Jordan compression**, [7] Definition 1.36)**.** *Let A be a finite-dimensional simple Jordan-Banach algebra over $\mathbb{R}$, and $c \in A$ an idempotent. Then define the* Jordan compression $U_c : A \to A$ *associated with c by*

$$U_c a = 2c \circ (c \circ a) - c \circ a. \tag{5.18}$$

Theorem 5.11 ([7] Proposition 1.38)**.** *Let c be an idempotent in a Jordan-Banach algebra A. Then the Jordan compression U_c satisfies $||U_c|| \leq 1$, and*

$$U_c^2 = U_c \quad \textit{and} \quad U_c U_{\mathbf{1}-c} = 0. \tag{5.19}$$

If $a \in A$ is positive, then

$$U_c a = 0 \quad \textit{iff} \quad U_{\mathbf{1}-c} a = a. \tag{5.20}$$

For every finite-dimensional simple Jordan-Banach algebra over $\mathbb{R}$, there exists a natural dynamical state space (Definition 2.18) defined in the following way:

Definition 5.12 (**Jordan dynamical state space**)**.** *Given a finite-dimensional simple Jordan-Banach algebra over $\mathbb{R}$, let the dynamical state space $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$ consist of the following elements:*

(i) *Let the vector space A be identical to the real vector space underlying the Jordan-Banach algebra, and let the cone of states $S_+(A)$ be the cone of squares Q of the algebra.*

(ii) *Let the cone of physical effects $E_+(A)$ be the full dual cone Q^*. By Theorem 5.7, we have that $E_+(A) = S_+(A)$.*

(iii) *Let the order unit u_A be the trace form, equivalently the functional defined by the multiplicative identity $\mathbf{1}$: for $a \in A$, let $u_A(a) := tr(\mathbf{1} \circ a)$. In particular, the set of normalized states is defined as $\Omega(A) = \{\rho \in Q \mid tr(\rho) = 1\}$.*

(iv) *Let the group of reversible transformations $\mathcal{G}(A)$ be the group of Jordan automorphisms.*

Next recall the notion of a positive projection on a state space, and also the notion of a bicomplemented normalized projection from Appendices B.6 and B.7. Note that by Theorems 5.11 and 5.13, for every idempotent c in a Jordan-Banach algebra A, the Jordan compressions U_c and $U_{\mathbf{1}-c}$ are bicomplementary normalized positive projections on the Jordan dynamical state spaces associated with A.

Theorem 5.13 ([7] Proposition 1.41)**.** *Let c be an idempotent in a Jordan-Banach algebra A, and let σ be a normalized state on A. Then $||U_c|| \leq 1$, and*

$$\sigma(c) = 1 \quad \Leftrightarrow \quad U_c\sigma = \sigma, \tag{5.21}$$

$$U_c\sigma = 0 \quad \Leftrightarrow \quad U_{\mathbf{1}-c}\sigma = \sigma, \tag{5.22}$$

$$||U_c\sigma|| = 1 \quad \Leftrightarrow \quad U_c\sigma = \sigma. \tag{5.23}$$

Theorem 5.14 ([7] Proposition 2.83)**.** *Let A be a Jordan-Banach algebra and let $P : A \to A$ be a normalized positive projection. There exists an idempotent $c \in A$ such that $P = U_c$ if and only if P is bicomplemented; in this case c is unique: $c = U_c\mathbf{1}$, and the complement $P' = U_{\mathbf{1}-c}$ is also unique.*

Theorem 5.15 ([7] Proposition 5.32)**.** *Let A be a Jordan-Banach algebra. Then every face of the resulting Jordan dynamical state space is exposed and projective (Appendix B.10).*

Theorem 5.16 ([7] Proposition 5.49)**.** *Let c be an idempotent in a Jordan-Banach algebra A, and σ a pure state of the associated Jordan dynamical state space. Then $U_c\sigma$ is a multiple of a pure state.*

5.2 The principles

Before we state the first principle, note that in the following sections we will generally work with a dynamical state space denoted by $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$ but drop explicit mentions of it in definitions and results. Further, recall from Definition 2.12 that a set of n normalized states $\omega_1, \ldots, \omega_n \in \Omega(A)$, are perfectly distinguishable if there is an n-outcome observation $\{e_j\}_{j=1}^n$ with $e_j(\omega_i) = \delta_{ij}$, for all $i, j = 1, \ldots, n$.

Principle 1. *Every normalized state is a mixture of a set of pure and perfectly distinguishable states.*

In more detail, given any normalized state $\omega \in \Omega(A)$, Principle 1 states that there exists a set of pure and perfectly distinguishable states $\omega_1, \ldots, \omega_n$, such that

$$\omega = \sum_{i=1}^{n} p_i \omega_i, \quad p_i > 0, \quad \sum_{i}^{n} p_i = 1. \tag{5.24}$$

In writing such a decomposition we are not assuming that the state ω is actually prepared by a device which on each run outputs a system in one of the states ω_i according to the distribution p_i, but only that ω can be represented or simulated with such a device. Also, we do not assume uniqueness of the decomposition, only existence. Further, if every normalized state is a mixture of a set of pure and perfectly distinguishable states, then in fact every state $\omega \in S_+(A)$ can also be simulated in this way, but with the coefficients no longer forming a probability distribution, i.e., $0 \leq p_i \not\leq 1$ and $\sum_i^n p_i \neq 1$ in general. Finally, it is implicit in Principle 1 that the discriminating effects for every set of pure and perfectly distinguishable states of the theory are physical effects.

Next, recall from Definition 2.17 that a transformation $T \in \mathsf{T}(A, A)$ is reversible if T is invertible and T^{-1} is a valid transformation as well.

Principle 2. *Every set of pure and perfectly distinguishable states can be mapped to any other set of pure and perfectly distinguishable states of the same size by some reversible transformation.*

In more detail, Principle 2 demands that for every integer $n \geq 1$, if $\omega_1, \ldots, \omega_n$ are perfectly distinguishable pure states, and so are $\omega'_1, \ldots, \omega'_n$, then there is a reversible transformation $T \in \mathcal{G}_A$ such that $T\omega_i = \omega'_i$ for all $i = 1, \ldots, n$. Recalling from Definition 3.19 the notion of bit symmetry, it is clear that if a theory satisfies Principle 2, then it is bit symmetric. In fact Principle 2 can be thought of as demanding n-symmetry for all n. In more physical terms this means that any state of any system with n distinguishable states can be transferred to any other system with n distinguishable states by a suitable reversible interaction. Principles 1 and 2 together can be understood as demanding that every state of every system can be prepared by first mixing a set of perfectly distinguishable pure states, and then reversibly transforming the resulting mixed state into the state we wish to prepare.

For the third principle recall the notion of a filter from Definition 4.4, and that a non-mixing filter is a filter which takes extremal rays of the state cone to extremal rays (Definition 4.45). For more discussion of filters and non-mixing filters see Sections 4.4 and 4.9.2.

Principle 3. *All filters are non-mixing.*

One way to understand this principle is to suppose that each filter is capable (at least in principle) of being constructed from a reversible transformation, and some selection process. In particular, we can imagine that the reversible process splits an ensemble of incoming systems into two sub-ensembles, in such a way that it is possible to reconstruct the original ensemble by applying the inverse of the reversible process. If this is were the case, then the resulting filter would be non-mixing. In the recent reconstruction of (complex) quantum theory [99], Hardy takes as an principle that filters have a property called *non-flattening*, which implies that they are non-mixing. More generally, filters have played an important role in various other axiomatizations of quantum theory, such as in Guz [90], Kummer [115, 116], Alfsen and Shultz [7], and in Araki [9] and Gunson [87] who also use essentially the same principle as ours.

Quantum Example 5.17. *In order to understand the relationship between Jordan-Banach algebras and our principles, first recall the notion of a Jordan frame (Definition 5.3), the notion of a Jordan automorphism (Definition 5.8), and the notion of a Jordan dynamical state space (Definition 5.12). Note the following facts relating the Jordan state space* $(A, S_+(A), E_+(A), u_A, \mathcal{G}(A))$ *with the underlying Jordan-Banach algebra* A*:*

(i) *Primitive idempotents* c *of the Jordan-Banach algebra are extremal elements of the positive cone* $S_+(A)$*, and have* $tr(c) = 1$*, i.e., they are pure states. Further, the primitive idempotents* $c_1, \ldots, c_n$ *making up any Jordan frame are perfectly distinguishable, with the discriminating observation* $\{e_j\}_{j=1}^n$ *identical to the Jordan frame (under the identification of non-negative functionals with states through the trace form), i.e.,* $e_j(c_i) = tr(c_j \circ c_i) = \delta_{ij}$*.*

(ii) *By the Spectral Theorem 5.4, every normalized state* $\rho \in \Omega(A)$ *has a decomposition* $\rho = \sum_{i=1}^n \lambda_i c_i$*, where* $\lambda_i \geq 0$*,* $\sum_{i=1}^n \lambda_i = 1$*, and the* c_i *are primitive idempotents. This implies that Jordan dynamical state spaces satisfy Principle 1.*

(iii) *By Theorem 5.9 the identity component of the group of Jordan automorphisms is transitive on the set of Jordan frames. Further, Jordan automorphisms and their inverses preserve the cone of states* $S_+(A)$ *which means they are reversible transformations in the sense of Definition 2.17. Next, because Jordan automorphisms preserve the multiplicative identity of the algebra, they also preserve the normalization of states. This implies that Jordan dynamical state spaces satisfy Principle 2.*

(iv) *By Theorems 5.11 and 5.13, for idempotent* $c \in A$*, the associated Jordan compression* U_c *(Definition 5.10) is a filter (Definition 4.4). Further, from Theorem 5.14, we have that filters on Jordan dynamical state spaces are in one-to-one correspondence with Jordan compressions, and from Theorem 5.16 the Jordan compressions are non-*

mixing. This implies that Jordan dynamical state spaces satisfy Principle 3.

There is a close connection between Principle 2 and certain properties of the circuit model for quantum computation. In this model we generally start with an input n-level system in a particular state, as well as a number of other n-level systems which can without loss of generality be taken to be in the $|0\rangle$ state. Then we implement the circuit representing the computation we wish to carry out, and at the end we must measure a specific observable to determine the output of the particular computation. This last step can be done without loss of generality by first reversibly transforming the – generally entangled – logical n-level system of interest into the first physical n-level system, and then doing the desired measurement on this. This transfer is possible because complex quantum theory satisfies Principle 2.

5.3 Consequences of Principles 1 and 2

In this section we will investigate the consequences of Principles 1 and 2 alone. Rather than assuming the no-restriction hypothesis (Definition 2.6), our first result is that it follows from Principle 1 alone.

Theorem 5.18. *Assume Principle 1. Then all mathematically well defined effects are physical effects, i.e., $E_+(A) = S_+(A)^*$.*

Proof. Take a functional $e \in S_+(A)^*$ which generates an exposed ray (Appendix A.21) of the cone. We will show that e is a measurable effect, namely that $e \in E_+(A)$, and because exposed rays generate the full cone $S_+(A)^*$ via convex combinations and closure, it will follow that all effects are measurable.

It follows from Appendix A.24 that if e is exposed then there is some $x \in A$ with $f(x) \geq 0$ for all $f \in S_+(A)^*$, and

$$f(x) = 0, \text{ and } f \in S_+(A)^* \;\Leftrightarrow\; f = \lambda e \text{ for some } \lambda \geq 0.$$

That is, the hyperplane $\{f \mid f(x) = 0\}$ supports $S_+(A)^*$ only in the exposed ray generated by e. Therefore,

$$x \in (S_+(A)^*)^* = S_+(A),$$

and further $\omega := x/u(x)$ is a state in $\Omega(A)$. Since $e \neq 0$ is a linear functional with $e(\omega) = 0$ and $e(\varphi) \geq 0$ for all $\varphi \in \Omega(A)$, the state ω is on the boundary of $\Omega(A)$. Invoking Principle

1, ω can be decomposed by a set of pure and perfectly distinguishable states $\omega_1, \ldots, \omega_n$, so,

$$\omega \in \operatorname{conv}(\omega_1, \ldots, \omega_n) \subseteq \Omega(A).$$

Therefore, ω must be in the boundary of $\operatorname{conv}(\omega_1, \ldots, \omega_n)$. This convex hull is a classical simplex and it is clear that all states ω in the boundary have one of the discriminating effects e_i with $e_i(\omega) = 0$. By definition, e_i is an allowed effect. Since $e_i(x) = u_A(x)e_i(\omega) = 0$, it follows that $e = \lambda e_i$ for some $\lambda \geq 0$ is an allowed effect. □

As mentioned above, Principle 2 implies that the state space is bit symmetric, which we know from Theorem 3.24 implies that the state cone is self-dual. This fact will be essential in our development, so we restate this theorem for convenience.

Proposition 5.19. *Assume Principles 1 and 2. Then $\Omega(A)$ is bit symmetric, and therefore there exists an inner product $\langle \cdot, \cdot \rangle$ on A such that the state cone $S_+(A)$ is self-dual and the following hold for all $\omega, \varphi \in \Omega(A)$:*

(i) $0 \leq \langle \omega, \varphi \rangle \leq 1$,
(ii) $\langle \cdot, \cdot \rangle$ *is invariant under all reversible transformations,*
(iii) $\langle \omega, \omega \rangle = 1$ *for all pure ω,*
(iv) $\langle \omega, \varphi \rangle = 0$ *if ω and φ are perfectly distinguishable.*

Proof. This follows from Theorem 5.18 and Theorem 3.24. □

Given that Principles 1 and 2 imply that $S_+(A)$ is self-dual, in the remainder of this chapter we will use a self-dualizing inner product to write functional evaluation $e(\omega)$ as $\langle e, \omega \rangle$. We will further identity A^* with A, and treat objects like u_A as elements of A when convenient.

In order to simplify terminology, in the following we will call a set of pure and perfectly distinguishable states a *frame*.

Definition 5.20 (**Frame**)**.** *A* frame of size n *for a state space $(A, S_+(A), E_+(A), u_A)$ is a set of n pure and perfectly distinguishable states in $\Omega(A)$. A frame is* maximal *if n is the maximal value among all frames in $\Omega(A)$ (this size will be denoted N_A).*

In the next few results we will explore the connection between frames and faces of the set of normalized states.

Lemma 5.21. *Assume Principles 1 and 2. Then every frame for $\Omega(A)$ can be extended to a maximal frame for $\Omega(A)$.*

Proof. Let $\varphi_1, \ldots, \varphi_{N_A}$ be a maximal frame for $\Omega(A)$, and let $\omega_1, \ldots, \omega_k$, with $k < N_A$, be another frame. Principle 2 implies that there exists a reversible transformation T such that $T\varphi_i = \omega_i$ for all $i = 1, \ldots, k$. The states $T\varphi_{k+1}, \ldots, T\varphi_{N_A}$ clearly extend the frame $\omega_1, \ldots, \omega_k$. □

Next recall from Appendix A.8 that for a subset M of a convex set C, the face generated by M is the intersection of all faces of C containing M.

Theorem 5.22. *Assume Principles 1 and 2. Then every face of $\Omega(A)$ is generated by a frame for $\Omega(A)$. Further, if two different frames generate the same face, they must be the same size.*

Proof. Let F be a face of $\Omega(A)$, and let ω be in the relative interior of F. It is clear that ω generates F. By Principle 1 the state ω is a mixture of the states in a frame $\omega_1, \ldots, \omega_n$, namely, $\omega = \sum_{i=1}^n p_i \omega_i$, with $p_i > 0$, and $\sum_i^n p_i = 1$. This implies that each $\omega_i \in F$, and the frame $\omega_1, \ldots, \omega_n$ generates F. Suppose there exists another frame $\varphi_1, \ldots, \varphi_m$ which generates F as well, with $m < n$. By Principle 2, there exists a reversible transformation T with $T\varphi_i = \omega_i$ for $i = 1, \ldots, m$. Because the frame $\varphi_1, \ldots, \varphi_m$ generates F, the states $T\varphi_1, \ldots, T\varphi_m = \omega_1, \ldots, \omega_m$ generate the face $T(F)$. Further, because the frame $\omega_1, \ldots, \omega_n$ generates F, we must then have $T(F) \subseteq F$. Due to the reversibility of T, this is only possible if $T(F) = F$, so the frame $\omega_1, \ldots, \omega_m$ generates F. Denote the effects discriminating between the states ω_i by e_j, i.e., $e_j(\omega_i) = \delta_{ij}$, for $i, j = 1, \ldots, n$. Then the set

$$G = \{\omega \in \Omega(A) \mid \langle e_n, \omega \rangle = 0\}$$

is a face of $\Omega(A)$, and further $\omega_1, \ldots, \omega_m \in G$. In particular, the face which this set of states generate must be a subset of this, namely $F \subseteq G$. On the other hand, $\omega_n \notin G$, so $\omega_n \notin F$. This contradicts the assumption that F is the face generated by $\omega_1, \ldots, \omega_n$, which implies that $m = n$. □

The above theorem shows that each face of a state space which satisfies Principles 1 and 2 has a well defined size or information carrying capacity, which we formalize as the *rank* of the face.

Definition 5.23 (**Rank of a face**). *Assume Principles 1 and 2, and let F be a face of $\Omega(A)$. The size of any frame which generates F is called the* rank *of F, and is denoted $|F|$.*

Lemma 5.24. *Assume Principles 1 and 2, and let F be a face of $\Omega(A)$. If $\omega_1, \ldots, \omega_{|F|}$ is a frame with all $\omega_i \in F$, then it generates F.*

Proof. Let G be the face generated by the frame $\omega_1, \ldots, \omega_{|F|}$. From Theorem 5.22 we have that F is generated by a frame, so let $\varphi_1, \ldots, \varphi_{|F|}$ be any such frame. Since $\omega_1, \ldots, \omega_{|F|} \in F$,

we have $G \subseteq F$. By Principle 2, there is a reversible transformation T such that $T\omega_i = \varphi_i$ for all $i = 1, \ldots, |F|$, and therefore $T(G) = F$. This is only possible if $G = F$. □

Proposition 5.25. *Assume Principles 1 and 2. Then there is a unique state $\mu_A \in \Omega(A)$ with the property that $T\mu_A = \mu_A$ for all reversible transformations $T \in \mathcal{G}(A)$. Further, μ_A has the following properties:*

(i) *If $\omega_1, \ldots, \omega_{N_A}$ is any maximal frame for $\Omega(A)$, then $\mu_A = \frac{1}{N_A} \sum_{i=1}^{N_A} \omega_i$.*

(ii) *$\mu_A = u_A / N_A$, and therefore $u_A = \sum_{i=1}^{N_A} \omega_i$.*

Proof. Take a pure state $\omega \in \Omega(A)$, and define

$$\mu_A := \int_{\mathcal{G}_A} T\omega \, dT. \tag{5.25}$$

Because all pure states are frames of size 1, by Principle 2 the group $\mathcal{G}_A$ of reversible transformations acts transitively on the pure states, which proves that this definition does not depend on the choice of the pure state ω.

It is easy to check that $T\mu_A = \mu_A$ for all reversible transformations T. Now suppose that χ is a state distinct from μ_A, which also satisfies $T\chi = \chi$. By Principle 1 and Lemma 5.21 there exists a maximal frame $\omega_1, \ldots, \omega_{N_A}$ for $\Omega(A)$ such that $\chi = \sum_{i=1}^{N_A} \lambda_i \omega_i$, with $\lambda_i \geq 0$ and $\sum_i^{N_A} \lambda_i = 1$. Therefore,

$$\chi = T\chi = \int_{\mathcal{G}_A} T\chi \, dT = \sum_{i=1}^{N_A} \lambda_i \int_{\mathcal{G}_A} T\omega_i \, dT = \sum_{i=1}^{N_A} \lambda_i \mu_A = \mu_A.$$

This proves that μ_A is the unique state which is invariant with respect to all reversible transformations.

From Principle 2 we have that for every permutation π on the indexing set $\{1, \ldots, N_A\}$, there is a reversible transformation T_π such that $T_\pi \omega_i = \omega_{\pi(i)}$. Using the maximal frame from above, and properties (iii) and (iv) from Proposition 5.19, we have

$$\lambda_j = \sum_{i=1}^{N_A} \lambda_i \langle \omega_j, \omega_i \rangle = \langle \omega_j, \mu_A \rangle = \langle \omega_j, T_\pi \mu_A \rangle = \sum_{i=1}^{N_A} \lambda_i \langle \omega_j, \omega_{\pi(i)} \rangle = \lambda_{\pi^{-1}(j)}.$$

By choosing different permutations, we can see that all λ_j must be identical, so $\lambda_j = 1/N_A$. This proves the claimed representation of μ_A in terms of the average of a maximal frame, for a particular maximal frame. However, this maximal frame can be reversibly transformed to any other, from which the claimed identity holds for all maximal frames.

Finally, it is clear that all reversible transformations preserve the order unit u_A. Therefore, by self-duality $u_A/\langle u_A, u_A\rangle$ is a normalized state which is preserved by all reversible transformations, so by the above, it must be μ_A. The proportionality constant follows from the following equalities:

$$\langle \mu_A, \omega_j\rangle = \frac{1}{N_A} = \frac{\langle u_A, \omega_j\rangle}{\langle u_A, u_A\rangle} = \frac{1}{\langle u_A, u_A\rangle}. \tag{5.26}$$

□

Next, recall from Definition 3.6 that for every face F of a cone $S_+(A)$ there exists a not necessarily positive projection, P_F, such that $\mathrm{im}(P_F) = \mathrm{lin}(F)$, and that for self-dual cones these projections can be chosen to be symmetric (i.e., $P^* = P$) under the self-dualizing inner product and have further useful properties (see Lemma 3.7).

Definition 5.26 (**Projective unit**). *Let $S_+(A)$ be a self-dual cone, and F a face of $S_+(A)$. For each symmetric (not necessarily positive on $S_+(A)$) projection P_F onto* $\mathrm{lin}(F)$*, define the* projective unit *u_F associated with F by*

$$u_F = u_A \circ P_F = P_F^*(u_A). \tag{5.27}$$

Note that because there is a one-to-one correspondence between faces of $\Omega(A)$ and faces of $S_+(A)$, we can also associate a projective unit u_F with a face $F \subset \Omega(A)$ by simply using the projection onto the corresponding face of $S_+(A)$. Further, notice that projective units have the same form as the effects induced by filters from Section 4.4 (given the fact that the P_F are symmetric projections under the inner product, which is equivalent to $P_F = P_F^*$, and therefore $u_F = u_A \circ P_F = P_F(u_A)$). However, because the projections P_F are not necessarily positive on the cone, projective units are not necessarily valid effects. The next proposition shows that in fact they are in the context of Principles 1 and 2.

Proposition 5.27. *Assume Principles 1 and 2. For every face F of the set of normalized states $\Omega(A)$, the projective unit associated with F satisfies:*

(i) *$u_F \in \mathsf{E}(A)$, i.e., u_F is a valid effect.*
(ii) *If $\omega_1, \ldots, \omega_{|F|}$ is any frame contained in F, then $u_F = \sum_{i=1}^{|F|} \omega_i$.*
(iii) *$\langle u_F, \omega\rangle = 1$ for all $\omega \in F$.*

Proof. By Lemma 5.24 the frame $\omega_1, \ldots, \omega_{|F|}$ generates F, and by Lemma 5.21 this can be extended to a maximal frame $\omega_1, \ldots, \omega_{N_A}$ for $\Omega(A)$. If $i > |F|$, then it follows from Proposition 5.19 (iv) that $\langle \omega_i, \omega_j\rangle = 0$ for all $j = 1, \ldots, |F|$. Because the frame $\omega_1, \ldots, \omega_{|F|}$ generates F, we must also have $\langle \omega_i, \omega\rangle = 0$ for all $\omega \in F$. This implies that

$\omega_i \in (\mathrm{im}(P_F))^\perp = \ker(P_F)$, so $P_F\omega_i = 0$ for all $i > |F|$. Using Proposition 5.25, we have that $u_A = \sum_{i=1}^{N_A} \omega_i$, and applying P_F to both sides, we find

$$u_F = P_F u_A = \sum_{i=1}^{|F|} \omega_i,$$

which proves (ii). Comparing the form of u_F to that of u_A, it is clear that $0 \leq u_F \leq u_A$ (see Appendix A.16), and so $u_F \in \mathsf{E}(A)$, which proves (i). From the equalities

$$\langle u_F, \omega_j \rangle = \langle P_F u_A, \omega_j \rangle = \langle u_A, P_F \omega_j \rangle = 1,$$

for $j = 1, \ldots, |F|$, and the fact that the ω_j generate F, we must also have $\langle u_F, \omega \rangle = 1$ for all $\omega \in F \cap \Omega(A)$, which proves (iii). □

Lemma 5.28. *Assume Principles 1 and 2, and suppose F and G are faces of the set of normalized states $\Omega(A)$ such that $F \subseteq G$. Then $|F| \leq |G|$. Further, if $F \subsetneq G$, then $|F| < |G|$.*

Proof. Suppose that $|F| > |G|$. Let $\varphi_1, \ldots, \varphi_{|G|}$ be a frame for G, and extend it to a maximal frame $\varphi_1, \ldots, \varphi_{N_A}$ for $\Omega(A)$. From Proposition 5.27 (i) we have that $u_F \leq u_A$, and using Proposition 5.19 (iii) and (iv), we have

$$\langle u_F, u_G \rangle \leq \langle u_A, u_G \rangle = \sum_{i=1}^{N_A} \sum_{j=1}^{|G|} \langle \varphi_i, \varphi_j \rangle = |G|. \tag{5.28}$$

Next, let $\omega_1, \ldots, \omega_{|F|}$ be a frame for F. Since $\omega_j \in F$, we have $\omega_j \in G$, and from Proposition 5.27 (iii) we have $u_G(\omega_j) = 1$ for all j. Therefore,

$$\langle u_F, u_G \rangle = \langle u_G, u_F \rangle = \sum_{j=1}^{|F|} \langle u_G, \omega_j \rangle = |F| > |G|,$$

which contradicts (5.28), thus proving that $|F| \leq |G|$.

Next suppose that F is a proper face of G; we will show that $|F| \neq |G|$. Suppose $|F| = |G| =: k$, and let $\omega_1, \ldots, \omega_k$ be a frame for F, and $\varphi_1, \ldots, \varphi_k$ a frame for G. By Principle 2, there exists a reversible transformation T such that $T\omega_i = \varphi_i$ for all i. Thus F is reversibly mapped onto G, which is impossible if F is properly contained in G. This proves $|F| \neq |G|$, and together with $|F| \leq |G|$, this proves that $|F| < |G|$. □

Proposition 5.29. *Assume Principles 1 and 2. Then every face F of the set of normalized states $\Omega(A)$ can be written as*

$$F = \{\omega \in \Omega(A) \mid \langle u_F, \omega\rangle = 1\}. \tag{5.29}$$

In particular, all faces are exposed.

Proof. Let F be a face of $\Omega(A)$, and let $G := \{\omega \in \Omega(A) \mid \langle u_F, \omega\rangle = 1\}$. It is clear that G is a face, and $F \subseteq G$ from Proposition 5.27. Assuming $F \neq G$, Lemma 5.28 implies that $|F| < |G|$. Next let $\omega_1, \ldots, \omega_{|F|}$ be a frame for F, and $\varphi_1, \ldots, \varphi_{|G|}$ a frame for G. Since $\varphi_j \in G$, we have $\langle u_F, \varphi_j\rangle = 1$ for all j. Using the decomposition from Proposition 5.27 (ii) for u_F and u_G we have

$$|G| = \sum_{j=1}^{|G|} \langle u_F, \varphi_j\rangle = \langle u_F, u_G\rangle = \langle u_G, u_F\rangle = \sum_{i=1}^{|F|} \underbrace{\langle u_G, \omega_i\rangle}_{\leq 1} \leq |F|,$$

which contradicts the above. This proves that $F = G$. It is then clear from the given form for F that it is exposed by the effect u_F. □

Lemma 5.30. *Assume Principles 1 and 2. Every face F of the cone $S_+(A)$ is generated by the associated projective unit u_F, or equivalently, the face $F \cap \Omega(A)$ is generated by the normalized state $u_F/|F|$. Further,*

$$\langle u_F, u_F\rangle = |F| = \langle u_A, u_F\rangle. \tag{5.30}$$

Proof. Let $\omega_1, \ldots, \omega_{|F|}$ a frame for F. Using Proposition 5.27 (ii) we have that $u_F = \sum_{i=1}^{|F|} \omega_i$, and it follows easily that $\langle u_F, u_F\rangle = |F|$. Further, $\langle u_A, u_F\rangle = |F|$ follows from completing this frame to a maximal frame for $\Omega(A)$ and repeating the calculation. It also follows that $u_F/|F|$ is a normalized state in F, and since the ω_i generate F, $u_F/|F|$ is in the relative interior of F. Therefore, $u_F/|F|$ generates $F \cap \Omega(A)$, or equivalently, u_F generates F as a face of $S_+(A)$. □

Proposition 5.31. *Assume Principles 1 and 2. Let F be a face of the cone $S_+(A)$, and take an effect $f \in F \cap \mathsf{E}(A)$. Then $f \leq u_F$.*

Proof. Without loss of generality, suppose that f is in the relative interior of F. From self-duality we have that $\varphi_f := f/\langle u_A, f\rangle$ is a normalized state, so by Principle 1 there exists a frame $\omega_1, \ldots, \omega_k$ such that

$$\varphi_f = \sum_{i=1}^{k} \beta_i \omega_i, \qquad \sum_{i=1}^{k} \beta_i = 1, \qquad \beta_i \geq 0.$$

Since φ_f is in the relative interior of the face $F \cap \Omega(A)$, the frame $\omega_1, \ldots, \omega_k$ generates $F \cap \Omega(A)$. It follows that all $\beta_i\, 0$, and from Theorem 5.22 and Lemma 5.24 that $k = |F|$, and $\omega_1, \ldots, \omega_k$ is a maximal frame for F. We then have that $f = \sum_{i=1}^{|F|} \alpha_i \omega_i$ with $\alpha_i = u_A(f)\beta_i > 0$. By assumption $f \leq u_A$, so we have for all $j = 1, \ldots, |F|$,

$$1 = \langle u_A, \omega_j \rangle \geq \langle f, \omega_j \rangle = \alpha_j.$$

Therefore, from the above and Proposition 5.27 (ii), we have $f \leq \sum_{i=1}^{|F|} \omega_i = u_F$. □

Next, recall from Lemma 3.7 that for every face F of a self-dual cone $S_+(A)$ there exists an orthogonal exposed face $F^\perp$ and an orthogonal (not necessarily positive) projection $P_{F^\perp}$ such that

$$F^\perp = \{\omega \in S_+(A) \mid P_F \omega = 0\} = \{\omega \in S_+(A) \mid P_{F^\perp}\omega = \omega\}. \tag{5.31}$$

Theorem 5.32. *Assume Principles 1 and 2, and let F, G be faces of the set of normalized states $\Omega(A)$ with $F \subset G$. Then every frame which generates F can be extended to a frame which generates G.*

Proof. First, from Proposition 5.27, $u_F \in F \subset G$, and using Proposition 5.31, we have $u_F \leq u_G$, or equivalently, $\Delta := u_G - u_F \geq 0$. By assumption $F \neq G$, so in fact $\Delta \neq 0$. From Lemma 3.7, and Proposition 5.29 we have that

$$G = \{\omega \in \Omega(A) \mid \langle u_G, \omega \rangle = 1\} = \{\omega \in \Omega(A) \mid \langle u_{G^\perp}, \omega \rangle = 0\}.$$

Because the faces G and $G^\perp$ are orthogonal under the inner product, and $u_G \in G$, $u_{G^\perp} \in G^\perp$, we also have

$$\langle u_{G^\perp}, \Delta \rangle = \langle u_{G^\perp}, u_G \rangle - \langle u_{G^\perp}, u_F \rangle = 0 - \langle P_{G^\perp} u_A, u_F \rangle = -\langle u_A, P_{G^\perp} u_F \rangle = 0.$$

Using Lemma 3.7 again, the above equalities prove that $\Delta \in G$. By Principle 1 there exits a frame $\omega_1, \ldots, \omega_m$ for the state $\tilde{\Delta} := \Delta / \langle u_A, \Delta \rangle$, where we can write

$$\Delta = \sum_{i=1}^{m} \alpha_i \omega_i, \qquad \alpha_i > 0.$$

Note also that Lemma 5.30 and Lemma 5.28 imply that $\langle u_A, \Delta \rangle = \langle u_A, u_G - u_F \rangle = |G| - |F| \geq 1$. Because the ω_i are in the face generated by Δ, and $\Delta \in G$, the ω_i must be elements of G. Further, from $u_G - \Delta = u_F \geq 0$, it follows that $\Delta \leq u_G$. This implies that

$$1 = \langle u_G, \omega_j \rangle \geq \langle \Delta, \omega_j \rangle = \rangle \sum_{i=1}^{m} \alpha_i \langle \omega_i, \omega_j \rangle = \alpha_j,$$

that is, all $\alpha_j \leq 1$.

Next we will show that the $\omega_1, \ldots, \omega_m$ can be appended to any maximal frame for F to give a maximal frame for G. First we will show that they are all orthogonal to F. Letting $\omega \in F$, we have

$$\langle \Delta, \omega \rangle = \langle u_G, \omega \rangle - \langle u_F, \omega \rangle = 1 - 1 = 0 = \sum_{i=1}^{m} \alpha_i \langle \omega_i, \omega \rangle.$$

Because $\alpha_i > 0$, this is only possible if all $\langle \omega_i, \omega \rangle = 0$. It remains to show that m is the right size to give a generating frame for G, namely, that $|F| + m = |G|$. First, note that if $\varphi_1, \ldots \varphi_n$ is a frame such that all $\varphi_i \in G$, then by considering the face generated by this frame and applying Lemma 5.28 it follows that $n \leq |G|$. In particular, this implies that $|F| + m \leq |G|$. Next, note that

$$\begin{aligned} m &= \sum_{i=1}^{m} 1 \geq \sum_{i=1}^{m} \alpha_i = \sum_{i=1}^{m} \alpha_i \langle u_G, \omega_i \rangle = \langle u_G, \Delta \rangle = \langle u_G, u_G \rangle - \langle u_G, u_F \rangle \geq |G| - \langle u_A, u_F \rangle \\ &\geq |G| - |F|. \end{aligned}$$

Combining with the above gives $m = |G| - |F|$. □

Next recall the notion of a lattice (Appendix B.15), and in particular the notion of an orthomodular lattice (Appendix B.16). We will use the last result to show that set of faces is an orthomodular lattice, which will then allow us to prove that the symmetric projections onto the faces of $S_+(A)$ are in fact positive.

Proposition 5.33. *The set of faces of an arbitrary cone $S_+(A)$ (or equivalently $\Omega(A)$) is a lattice with operations given by the inclusion relation $\subseteq$ and by the following Equations for each pair of faces F, G:*

$$F \wedge G = F \cap G, \quad F \vee G = \text{face}(F \cup G). \tag{5.32}$$

Proof. It is clear that the inclusion relation $\subseteq$ on faces gives a partial order, and further that under the operations $\wedge$ and $\vee$ the set of faces of $S_+(A)$ is a lattice in the sense of Appendix B.15. □

Lemma 5.34. *Assume $S_+(A)$ is self-dual. Then for every pair of faces $F, G \subseteq S_+(A)$ we have*

$$F^\perp \wedge G^\perp = (F \vee G)^\perp. \tag{5.33}$$

Proof. First, using the inclusions $F \subseteq F \vee G$, $G \subseteq F \vee G$, and the inclusion reversal property of the orthogonal complement, we have $(F \vee G)^\perp \subseteq F^\perp$ and $(F \vee G)^\perp \subseteq G^\perp$. These last inclusions are clearly equivalent to $(F \vee G)^\perp \subseteq F^\perp \wedge G^\perp$.

For the reverse inclusion, suppose that $\varphi \in F^\perp \wedge G^\perp$. In particular, this means that $\langle \varphi, F \rangle = \langle \varphi, G \rangle = 0$, which implies that $\langle \varphi, \mathrm{conv}(F \cup G) \rangle = 0$. From the definition $F \vee G = \mathrm{face}(F \cup G)$ it is clear that $\mathrm{face}(\mathrm{conv}(F \cup G)) = F \vee G$, and by Appendix A.22, we then have that $\mathrm{conv}(F \cup G) \cap \mathrm{relint}(F \vee G) \neq \emptyset$. In particular, this implies that φ evaluates to 0 on a point in the relative interior of $F \vee G$, which is equivalent to $\varphi \in (F \vee G)^\perp$. □

Proposition 5.35. *Assume Principles 1 and 2. Then the set of faces of $S_+(A)$ (or equivalently $\Omega(A)$) is an orthocomplemented lattice with the orthocomplement given by $F \to F^\perp$.*

Proof. We will show that this lattice satisfies the first three criteria from Appendix B.16, i.e., it is an orthocomplemented lattice. If F is an arbitrary face of $S_+(A)$, it is clear that $F^\perp$ is also a face. From Proposition 5.29 we have that F is exposed, and therefore from Appendix A.24 (iv) we have $F^{\perp\perp} = F$. Now let G be another face of $S_+(A)$ such that $F \subseteq G$. It is clear that taking orthogonal complements is inclusion reversing, so we have $G^\perp \subseteq F^\perp$. Further, $F \wedge F^\perp = \{0\}$ follows from the fact that the only vector $v \in A$ which satisfies $\langle v, v \rangle = 0$ is $v = 0$. Finally, in order to show that $F \vee F^\perp = S_+(A)$, we use Lemma 5.34, along with the fact that for every face H of $S_+(A)$ we have $H = H^{\perp\perp}$ as shown above. Putting these together, we have

$$\{0\} = F^\perp \wedge F = F^\perp \wedge F^{\perp\perp} = (F \vee F^\perp)^\perp.$$

Using $\{0\}^\perp = S_+(A)$ along with the above proves $F \vee F^\perp = S_+(A)$. □

Proposition 5.36. *Assume Principles 1 and 2. Then the lattice of faces of $S_+(A)$ (or equivalently $\Omega(A)$) is orthomodular.*

Proof. From Appendix B.16 and Proposition 5.35, it is sufficient to show that for every pair of faces $F, G \subseteq S_+(A)$, the inclusion $F \subseteq G$ implies $G = (G \wedge F^\perp) \vee F$. To this end, let $H := (G \wedge F^\perp) \vee F$. By Theorem 5.32, there exists a frame $\omega_1, \ldots, \omega_{|G|}$ which generates G, such that the sub-frame $\omega_1, \ldots, \omega_{|F|}$ generates F. If $1 \leq i \leq |F|$, then $\omega_i \in F$, and therefore $\omega_i \in H$. Further, if $|F| < i \leq |G|$, then $\omega_i \in G$ as well as $\omega_i \in F^\perp$, which implies that $\omega_i \in G \wedge F^\perp$, and therefore $\omega_i \in H$. Because $\omega_1, \ldots, \omega_{|G|}$ generates G, this shows that $G \subseteq H$. By definition H is the smallest face containing both F and $G \wedge F^\perp$, so $G \subsetneq H$ is not possible. This shows that $G = H$, thus proving the claim. □

Next recall from Definition 3.8 that for a self-dual cone $S_+(A)$, the dual of a face $F \subseteq S_+(A)$ in its own span is given by

$$F^\wedge = \{y \in \mathrm{lin}(F) \mid \langle x, y \rangle \geq 0, \ \forall\, x \in F\}. \tag{5.34}$$

Lemma 5.37 ([15] Theorem 1)**.** *If $S_+(A)$ is self-dual and its lattice of faces is orthomodular, then every face of $S_+(A)$ is self-dual in its span.*

Proof. Let F be a face of $S_+(A)$ and let $\omega \in F$ be in the relative interior of F. Because all elements of F are non-negative on F, if follows that $F \subseteq F^\wedge$. Suppose there exists a $y \neq 0$ such that $y \in F^\wedge \setminus F$. Because $\omega \in \text{relint}(F)$, and $y \notin F$ it is clear that the line segment $\{t\omega + (1-t)y \mid 0 \leq t \leq 1\}$ must intersect the boundary of F at some point which we will denote by z. It is clear there there exists a $\beta \in (0,1)$ such that $z = \beta\omega + (1-\beta)y$. Because z is on the boundary of F, $\text{face}(z)$ is contained in the boundary as well, and by Appendix A.22 we have $\text{face}(z) \subsetneq F$. From the orthomodularity of the lattice of faces of $S_+(A)$ we have $\text{face}(z) \vee (\text{face}(z)^\perp \wedge F) = F$, so there exists a $u \in \text{face}(z)^\perp \cap F$, with $u \neq 0$. Therefore,

$$0 = \langle u, z\rangle = \beta\langle u, \omega\rangle + (1-\beta)\langle u, y\rangle \geq \beta\langle u, \omega\rangle > 0,$$

where the last inequality follows from $\omega \in \text{relint}(F)$. This is a contradiction, which proves $F = F^\wedge$. □

Next we use an important result from [102], which we have already stated as Proposition 3.9, to prove that the symmetric projections onto the faces of $S_+(A)$ are in fact positive.

Theorem 5.38. *Assume Principles 1 and 2. Then for every face F of $S_+(A)$, the associated symmetric projection P_F is positive on $S_+(A)$, i.e., $P_F(S_+(A)) \subseteq S_+(A)$.*

Proof. By Proposition 5.19 and Proposition 5.36, $S_+(A)$ is self-dual and its lattice of faces is orthomodular. Then by Lemma 5.37, every face of $S_+(A)$ is self-dual its span. The claim then follows from the second part of Proposition 3.9. □

Recalling the notion of a compression from Appendix B.9, the final step of this section is to show that the positive symmetric projections onto the faces of $S_+(A)$ are in fact compressions.

Theorem 5.39. *Assume Principles 1 and 2. Then for every face F of $S_+(A)$, the associated symmetric projection P_F is a compression. In particular, it satisfies the following:*

(i) $P_F(S_+(A)) \subseteq S_+(A)$, *($P_F$ is positive),*
(ii) $\text{im}^+(P_F) = \ker^+(P_{F^\perp})$, *and* $\ker^+(P_F) = \text{im}^+(P_{F^\perp})$, *($P_F$, $P_{F^\perp}$ are complementary),*
(iii) $u_A \circ P_F \leq u_A$, *(P_F is normalized).*

Proof. (i) follows from Theorem 5.38.

(ii) follows from the fact that F is exposed and from Lemma 3.7 applied to F as well as $F^\perp$.

(iii) Follows from Proposition 5.27 (i), along the fact that $P_F^* = P_F$, which implies that $e \circ P_F = P_F e$ for all $e \in E_+(A)$.

Finally, using the fact that P_F and $P_{F^\perp}$ are symmetric projections together with (ii) above, shows that P_F^* and $P_{F^\perp}^*$ are also complementary and therefore bicomplementary (see Appendix B.6), which proves that P_F is a compression. □

Corollary 5.40. *Assume Principles 1 and 2. Then every face F of $S_+(A)$ is such that $F = S_+(A) \cap \mathrm{im}(P_F)$ for a compression P_F, namely $S_+(A)$ is a projective state space (see Appendix B.10).*

Proof. Follows from Proposition 5.29 along with Theorem 5.39. □

5.4 Principle 3 and the last steps of the reconstruction

In this section we will finish our reconstruction of the finite-dimensional Jordan-Banach algebras by using Principle 3. As we will see below, the extremal rays of the state cone and their associated rank-1 symmetric projections and projective units will be especially important, so we define the following:

Definition 5.41 (**Atom**). *Assume Principles 1 and 2. Given a face F of $S_+(A)$ which is an extremal ray, along with the associated symmetric projection P_F, the projective unit $u_F = P_F u_A$ will be called an* atom.

Note that we have a one-to-one correspondence between atoms and pure states, namely every atom u_F is also a pure state, and for every pure state ω there exists a rank-1 symmetric projection onto the ray generated by ω. Further, we also have a one-to-one correspondence between faces F of $S_+(A)$, projective units u_F and symmetric projections P_F. Given that the symmetric projections P_F are in fact compressions, we can use results developed by Alfsen and Shultz to our advantage. We now review two such results which will be essential in the last steps of our reconstruction.

Lemma 5.42 ([7] Lemma 8.32). *Assume Principles 1 and 2, and let F be a face of $S_+(A)$ with associated symmetric projection P_F. For any $\omega \in S_+(A)$, we have*

$$\mathrm{face}(P_F\omega) = (\mathrm{face}(\omega) \vee F^\perp) \wedge F. \tag{5.35}$$

Proof. From Corollary 5.40 we have that $S_+(A)$ is a projective state space (see Appendix B.10). The claim then follows from the fact that $S_+(A)$ is self-dual and we have a one-to-one correspondence between faces of $S_+(A)$ and projective units, along with [7] Lemma 8.32. □

Proposition 5.43 ([7] Proposition 9.7)**.** *Assume Principles 1 and 2, and let F be a face of $S_+(A)$ with associated orthogonal projection P_F. Then P_F is non-mixing if and only if for every $\omega \in \text{ext}(S_+(A))$, we have that $(\text{face}(\omega) \vee F^\perp) \wedge F$ is an extremal ray of $S_+(A)$ or is equal to $\{0\}$.*

Proof. From Lemma 5.42, we have that $\text{face}(P_F\omega) = (\text{face}(\omega) \vee F^\perp) \wedge F$. The claim follows from noting that for $\omega \in \text{ext}(S_+(A))$, $\text{face}(P_F\omega)$ is an extremal ray of $S_+(A)$ if and only if P_F is non-mixing. □

Our next step is to define a product on A. We will do this using an operator closely related to that defined in Section 4.9. Note that by Proposition 5.27, for every face F the associated projective unit $u_F = P_F u_A$ generates F, and because a symmetric projection P_F is uniquely defined by its image $\text{lin}(F)$, then P_F is uniquely defined by u_F.

Definition 5.44 ([7] 9.29)**.** *Assume Principles 1 and 2. For each atom $p = Pu_A$ (with P an symmetric projection onto a face of $S_+(A)$), and each $b \in A$, define*

$$p * b = \frac{1}{2}(\mathbb{I}_A + P - P^\perp)b. \tag{5.36}$$

Proposition 5.45 ([7] Lemma 9.29)**.** *Assume Principles 1, 2, and 3. Let p, q be atoms of $S_+(A)$. Then we have:*

(i) $p * q = q * p$.
(ii) *If* $\langle p, q\rangle = 0$, *and* $b \in A$, *then* $p * (q * b) = q * (p * b)$.
(iii) *If* $\langle p, q\rangle = 0$, *then* $p * q = 0$.
(iv) $p * p = p$.

Proof. Let P, R be the compressions such that $\text{im}^+(P) = \text{face}(p)$, and $\text{im}^+(R) = \text{face}(p) \vee \text{face}(q)$. First, on the subspace $\text{im}(R)$, the complement of P is exactly the restriction of $P^\perp$. This follows from Appendix B.17, along with the fact that Principles 1 and 2 imply the state space is projective. Therefore, $p * q$ can be evaluated in $\text{im}(R)$ rather than A, so without loss of generality we can assume that $\text{face}(p) \vee \text{face}(q) = S_+(A)$. If $p = q$, then (i) is clear, so we will also assume that $p \neq q$.

From Principle 3 it follows that $P^\perp$ is non-mixing, and using Proposition 5.43, as well as $p \neq q$, we have that for the pure state q, the face

$$(\text{face}(q) \vee \text{face}(p)) \wedge \text{face}(p^\perp) = S_+(A) \wedge \text{face}(p^\perp) = \text{face}(p^\perp),$$

is an extremal ray of $S_+(A)$. In particular, since $\text{face}(p^\perp) \subset S_+(A)$, the above implies that $\text{face}(p^\perp)$ is an extremal ray, and so $P^\perp$ is rank-1. By orthomodularity we have that $\text{face}(p) \vee \text{face}(p^\perp) = S_+(A)$, which together with the above implies that $|S_+(A)| = |\text{face}(p) \vee \text{face}(q)| = 2$.

Next, because P is rank-1, $\text{im}(P)$ consists of all real multiples of p, so for each $\omega \in \text{ext}(S_+(A))$,

$$P\omega = \lambda p,$$

for some $\lambda \geq 0$. Taking the inner product of both sides with p, we have

$$\langle p, \omega \rangle = \langle p, P\omega \rangle = \lambda \langle p, p \rangle = \lambda,$$

where we have used that $Pp = p$, and $\langle p, p \rangle = 1$ which follows from Proposition 5.19 (iii). Because $p^\perp = u_A - p$ also has corresponding compression $P^\perp$ with rank=1, we also have

$$P^\perp \omega = \langle p^\perp, \omega \rangle p^\perp.$$

Using the definition of the $*$ product and the above, we have

$$\begin{aligned} p * q &= \frac{1}{2}(q + \langle p, q \rangle p - \langle p^\perp, q \rangle p^\perp) \\ &= \frac{1}{2}(q + \langle p, q \rangle p - \langle p^\perp, q \rangle (u_A - p)) \\ &= \frac{1}{2}(q + \langle u_A, q \rangle p - \langle p^\perp, q \rangle u_A). \end{aligned} \tag{5.37}$$

Using $\langle u_A, q \rangle = 1$, the above simplifies to

$$p * q = \frac{1}{2}\left(q + p - (1 - \langle p, q \rangle) u_A\right),$$

which is clearly symmetric in p and q, so (i) follows.

(ii), (iii) Now assume $\langle p, q \rangle = 0$, and let P, Q be the symmetric projections corresponding to p, q. It follows that $p \in (\text{im}(Q))^\perp$ and $q \in (\text{im}(P))^\perp$, which imply $Pq = 0$ and $Qp = 0$. Taking an arbitrary pure state ω, as above we have $P\omega = \langle p, \omega \rangle p \leq p$, from

which it follows that $0 \leq QP\omega \leq Qp = 0$. By a similar calculation with P and Q interchanged, we have $PQ\omega = 0$. As the pure states span the full vector space A, this shows that $PQ = QP = 0$. Then by Lemma 4.13, $P, Q, P^\perp, Q^\perp$ all commute pairwise, so

$$\begin{aligned} p * (q * b) &= \frac{1}{4}\left(\mathbb{I}_A + P - P^\perp\right)\left(\mathbb{I}_A + Q - Q^\perp\right) b = \frac{1}{4}\left(\mathbb{I}_A + Q - Q^\perp\right)\left(\mathbb{I}_A + P - P^\perp\right) b \\ &= q * (p * b). \end{aligned}$$

Finally, $\langle p, q\rangle = 0$ also implies $P^\perp q = q$, and (iii) follows.

(iv) Clear from the definition of the $*$ product. □

Next we use the frame decomposition for states to give a *generalized frame decomposition* for every element of A in terms of a frame. We will then use this decomposition and the $*$ product to define a product on all of A.

Lemma 5.46. *Assume Principles 1 and 2. Then every $a \in A$ has a decomposition of the form*

$$a = \sum_{i=1}^{n} \alpha_i \omega_i, \tag{5.38}$$

where $\omega_1, \ldots, \omega_n$ is a frame, and $\alpha_i \in \mathbb{R} \setminus \{0\}$. In general this decomposition is not unique.

Proof. First, note that if $a \in A$ is an atom, then it is identical to its decomposition, and this is unique. For general $a \in A$, because $S_+(A)$ is self-dual, it follows from Proposition 3.5 that there exist a unique decomposition of a such that

$$a = a^+ - a^-, \quad a^\pm \in S_+(A), \quad \text{and} \quad \langle a^+, a^-\rangle = 0.$$

Because every normalized state is a mixture of the elements of a frame, and every state in $S_+(A)$ is a multiple of a normalized state, then in fact every state in $S_+(A)$ can also be decomposed as a mixture of the elements of a frame, but with the coefficients no longer forming a probability distribution, i.e., $0 < p_i \not\leq 1$ and $\sum_i p_i \neq 1$ in general. Therefore there exist frame decompositions for a^+ and a^-, namely,

$$a^+ = \sum_{i=1}^{m^+} \alpha_i^+ \omega_i^+, \quad \text{and} \quad a^- = \sum_{i=1}^{m^-} \alpha_i^- \omega_i^-,$$

where the only constraints are $\alpha_i^+, \alpha_i^- > 0$. Since $\langle a^+, a^-\rangle = 0$, the states ω_i^+ are orthogonal to the states ω_i^-. Therefore, $\omega_1^+, \ldots, \omega_{m^+}^+, \omega_1^-, \ldots, \omega_{m^-}^-$ is a frame. To simplify notation we will write

$$a = \sum_{i=1}^{m^+} \alpha_i^+ \omega_i^+ - \sum_{i=1}^{m^-} \alpha_i^- \omega_i^- = \sum_{i=1}^{m^+ + m^-} \alpha_i \omega_i, \tag{5.39}$$

where $\alpha_i = \alpha_i^+$, $\omega_i = \omega_i^+$ for $1 \leq i \leq m^+$, and $\alpha_i = -\alpha_{i-m^+}^-$, $\omega_i = \omega_{i-m^+}^-$ for $m^+ + 1 \leq i \leq m^+ + m^-$. □

Definition 5.47. *Assume Principles 1, 2, and 3. For each $a, b \in A$, fix once and for all generalized frame decompositions $a = \sum_i \alpha_i \omega_i$ and $b = \sum_j \beta_j \varphi_j$, and define*

$$a \circ b = \sum_{i,j} \alpha_i \beta_j \omega_i * \varphi_j. \tag{5.40}$$

As we have noted above, the generalized frame decomposition of an element is not unique in general. We will show below that the above product $a \circ b$ does not depend on which decomposition is chosen for a and b. Further, recalling Proposition 5.19 and the comments following it, in the remainder we will explicitly take $u_A \in A$. The proof of the following theorem is inspired by [5] Proposition 6.11.

Theorem 5.48. *Assume Principles 1,2, and 3. Then under the product $a \circ b$, A is a Jordan algebra with multiplicative identity u_A such that $p \circ q = p * q$ for all atoms p, q.*

Proof. Let $a, b \in A$, and take generalized frame decompositions $a = \sum_i \alpha_i \omega_i$, and $b = \sum_j \beta_j \varphi_j$ from Lemma 5.46. As each ω_i is a pure state, and therefore an atom, we can see from the definition of the $*$ product that

$$\omega_i \circ b = \sum_j \beta_j \omega_i * \varphi_j = \omega_i * \left(\sum_j \beta_j \varphi_j \right) = \omega_i * b. \tag{5.41}$$

Further, from its definition it is clear that $*$ is linear in the right element, and because we can absorb scalar multiples of an element into its generalized frame decomposition, a similar calculation to the above shows that for all $\lambda \in \mathbb{R}$

$$\omega_i \circ (\lambda b) = \lambda \omega_i \circ b.$$

Using the commutativity of $*$, we further have

$$\omega_i \circ b = \sum_j \beta_j \omega_i * \varphi_j = \sum_j \beta_j \varphi_j * \omega_i = b \circ \omega_i. \tag{5.42}$$

Multiplying the expressions $\omega_i \circ b$ by α_i, summing over i, and using (5.42), the linearity of $*$ in the right element, as well as its commutativity, gives

$$\begin{aligned} \sum_i \alpha_i \left(\omega_i \circ b \right) &= \sum_i \alpha_i \left(b \circ \omega_i \right) = \sum_i \alpha_i \Big(\sum_j \beta_j \varphi_j * \omega_i \Big) \\ &= \sum_{i,j} \beta_j \alpha_i \varphi_j * \omega_i = \sum_{i,j} \alpha_i \beta_j \omega_i * \varphi_j \\ &= b \circ a = a \circ b. \end{aligned} \tag{5.43}$$

Next, take a third element $c \in A$, and using (5.41), (5.42), and the linearity of $*$ in the right element, we have

$$\begin{aligned} \omega_i \circ (b+c) &= \omega_i * (b+c) = \omega_i * b + \omega_i * c \\ &= \omega_i \circ b + \omega_i \circ c, \\ &= (b+c) \circ \omega_i. \end{aligned} \tag{5.44}$$

Putting (5.43), and (5.44) together gives

$$a \circ (b+c) = \sum_i \alpha_i \left(\omega_i \circ (b+c)\right) = \sum_i \alpha_i \left(\omega_i \circ b + \omega_i \circ c\right) \tag{5.45}$$

$$= a \circ b + a \circ c, \tag{5.46}$$

which proves that $\circ$ is commutative and bilinear. This also proves that $a \circ b$ is independent of the decompositions chosen for a and b.

Next we will verify that the Jordan identity $a^2 \circ (b \circ a) = (a^2 \circ b) \circ a$ holds for all $a, b \in A$. Using Proposition 5.45 (iii) and (iv), we have

$$a^2 = a \circ a = \sum_{i,j} \alpha_i \alpha_j \omega_i * \omega_j = \sum_i \alpha_i^2 \omega_i.$$

We further have

$$a^2 \circ (b \circ a) = \sum_{i,j,k} \alpha_i^2 \beta_j \alpha_k \omega_i \circ (\varphi_j \circ \omega_k),$$

while,

$$(a^2 \circ b) \circ a = \sum_{i,j,k} \alpha_i^2 \beta_j \alpha_k (\omega_i \circ \varphi_j) \circ \omega_k.$$

Using the definition of $\circ$, and commutativity we further have

$$\omega_i \circ (\varphi_j \circ \omega_k) = \omega_i * (\omega_k * \varphi_j), \tag{5.47}$$

while,

$$(\omega_i \circ \varphi_j) \circ \omega_k = \omega_k * (\omega_i * \varphi_j). \tag{5.48}$$

Using Proposition 5.45 (ii) on (5.47) and (5.48) proves the Jordan identity.

Finally, let $p = P u_A$ be an atom, and note that

$$p \circ u_A = \frac{1}{2} \left(\mathbb{I}_A + P - P^\perp\right) u_A = p,$$

where we have used $u_A = p + p^\perp = Pu_A + P^\perp u_A$. Using commutativity and bilinearity of $\circ$ proves that u_A is the multiplicative identity on A. □

Our final step is to show that there exists a norm on A such that the requirements from Definition 5.5 are satisfied. To this end, we now the define the norm of an element $a \in A^*$.

Definition 5.49 (**Order unit norm**, [7] 1.12)**.** *Let $(A, S_+(A), E_+(A), u_A)$ be an abstract state space. Then the order unit norm of an element $a \in A^*$ is defined as*

$$||a|| = \inf\{\lambda > 0 \mid -\lambda u_A \leq a \leq \lambda u_A\}. \tag{5.49}$$

The following characterizations of this norm will be useful below.

Lemma 5.50 ([6] Lemma 1.18)**.** *Let $(A, S_+(A), E_+(A), u_A)$ be an abstract state space with the set of normalized states $\Omega(A)$. Then for $a \in A^*$ we have*

$$||a|| = \sup\{|a(\omega)| \mid \omega \in \Omega(A)\}. \tag{5.50}$$

Lemma 5.51. *Assume Principles 1,2, and 3. Then for $a \in A^*$ with generalized frame decomposition $a = \sum_i \alpha_i \omega_i$ we have*

$$||a|| = \max_i\{|\alpha_i|\}. \tag{5.51}$$

Proof. Because $S_+(A)$ is self-dual, the distinction between A^* and A is irrelevant, and as usual we use a self-dualizing inner product for functional evaluation. Then by Lemma 5.50, it is sufficient to find a normalized state ω such that $|\langle a, \omega\rangle|$ is maximized. Let $a = \sum_i \alpha_i \omega_i$ be the generalized frame decomposition from Lemma 5.46, where $\omega_1, \ldots, \omega_n$ is a frame, and $\alpha_i \neq 0$. We then have the following inequalities:

$$\begin{aligned} |\langle a, \omega\rangle| &= \left|\sum_i \alpha_i \langle \omega_i, \omega\rangle\right| \leq \sum_i |\alpha_i| \langle \omega_i, \omega\rangle \leq \max_i\{|\alpha_i|\} \sum_i \langle \omega_i, \omega\rangle \leq \max_i\{|\alpha_i|\} \langle u_A, \omega\rangle \\ &\leq \max_i\{|\alpha_i|\}, \end{aligned} \tag{5.52}$$

where we have used Proposition 5.25 (ii), and the fact that ω is normalized. It is then clear that taking $\omega = \omega_j$ where $j = \arg\max_i\{|\alpha_i|\}$ achieves the upper bound in (5.52), which proves that $||a|| = \max_i\{|\alpha_i|\}$. □

Proposition 5.52. *Assume Principles 1,2, and 3. Then the product $a \circ b$ together with the order unit norm satisfy the following requirements for all $a, b \in A$:*

(i) $||a \circ b|| \leq ||a||\,||b||$,
(ii) $||a^2|| = ||a||^2$,

(iii) $||a^2|| \leq ||a^2 + b^2||$.

Proof. (i) First, take $a \in A$ such that $-u_A \leq a \leq u_A$, and use the decomposition from Lemma 5.46 to write $a = \sum_i \alpha_i \omega_i$. Using Proposition 5.25 (ii) we can decompose u_A in terms of the same frame as a, and applying the projection P_i associated with the face generated by ω_i to the expression $-u_A \leq a \leq u_A$, we find

$$-\omega_i \leq P_i a = \alpha_i \omega_i \leq \omega_i.$$

This implies that $-1 \leq \alpha_i \leq 1$ for all i. Using Proposition 5.45 (iii) and (iv), it is then easy to see that $a^2 \in [0, u_A]$.

([6] Lemma 1.79) Suppose that $a, b \in A$ have order unit norms $||a|| \leq 1$, and $||b|| \leq 1$. Using the triangle inequality we have $||\frac{1}{2}(a+b)|| \leq 1$ and $||\frac{1}{2}(a-b)|| \leq 1$. From the definition of the order unit norm we have that for all $c \in A^*$, the inequalities $-||c||u_A \leq c \leq ||c||u_A$ are satisfied, and using the above implication $-u_A \leq a \leq u_A \Rightarrow 0 \leq a^2 \leq u_A$, gives

$$0 \leq \frac{1}{4}(a+b)^2 \leq u_A, \quad 0 \leq \frac{1}{4}(a-b)^2 \leq u_A. \tag{5.53}$$

Multiplying the second set of inequalities of (5.53) by -1 and then adding the first inequality gives

$$-1 \leq \frac{1}{4}(a+b)^2 - \frac{1}{4}(a-b)^2 \leq u_A.$$

It is easy to see that $\frac{1}{4}(a+b)^2 - \frac{1}{4}(a-b)^2 = a \circ b$, and so $||a \circ b|| \leq 1$. Therefore $||a \circ b|| \leq ||a||\,||b||$ for all $a, b \in A$ follows by scaling with some $\lambda \in \mathbb{R}$.

(ii) Take an arbitrary $a \in A$, and using its generalized frame decomposition, we have that $a^2 = \sum_i \alpha_i^2 \omega_i$. From Lemma 5.51 we then have $||a||^2 = (\max_i\{|\alpha_i|\})^2$, and $||a^2|| = \max_i\{|\alpha_i|^2\}$, which proves that $||a^2|| = ||a||^2$.

(iii) It follows from $a^2 = \sum_i \alpha_i^2 \omega_i$ and the closure of $S_+(A)$ under positive multiplication and addition that $a^2 \in S_+(A)$. Further, for all $a, b \in A$ we have $a^2 \leq a^2 + b^2$. Then $a^2 + b^2 \leq \lambda u_A$, implies $a^2 \leq \lambda u_A$, and from the definition of the order unit norm (5.49), it follows that $||a^2|| \leq ||a^2 + b^2||$. □

Theorem 5.53. *Assume Principles 1,2, and 3. Then under the product $a \circ b$ and the order unit norm, A is a Jordan-Banach algebra with multiplicative identity u_A.*

Proof. Follows from Theorem 5.48 and Proposition 5.52. □

5.5 Alternatives to Principle 3

In this section we introduce Principles 3′ and 3″, and show that in the context of Principles 1 and 2, Principles 3, 3′, and 3″ are in fact equivalent.

First recall the notion of a lattice (Appendix B.15), and that for an arbitrary abstract state space the set of faces of $S_+(A)$ (or equivalently $\Omega(A)$) is a lattice with operations given the inclusion relation $\subseteq$ and $F \wedge G = F \cap G$, and $F \vee G = \text{face}(F \cup G)$ (Proposition 5.33).

Definition 5.54 (**Covering property**). *Let $(A, S_+(A), E_+(A), u_A)$ be an arbitrary abstract state space, and let F, G be faces of $S_+(A)$ (or equivalently $\Omega(A)$). We say G* covers *F if $F \subsetneq G$ and there is no face B satisfying $F \subsetneq B \subsetneq G$. We say $S_+(A)$ has the* covering property *if for all faces F and all extremal rays ω of $S_+(A)$ (or equivalently pure states of $\Omega(A)$),*

$$\textit{either} \quad F \vee \omega = F \quad \textit{or} \quad F \vee \omega \textit{ covers } F. \tag{5.54}$$

This property has been very important in the quantum logic literature, and has often been taken as an axiom (usually in a slightly different form) [32]. The close relationship between this property and non-mixing filters seen in Lemma 5.42 and Proposition 5.43 was first noticed by Guz [88, 89], and further used by Alfsen and Shultz in their reconstruction of Jordan-Banach algebras in Chapter 9 of [7].

Principle 3′. *The covering property holds.*

In the context of Principles 1 and 2, we can give a more operational interpretation to this principle. Recall from Theorem 5.22 that each face F has a well defined size, namely, the number of states in any frame which generates F. Then the covering property states that the smallest face G, which contains both F and any given pure state ω, is either F itself (if $\omega \in F$), or $|G| = |F|+1$, i.e., G must have exactly $|F|+1$ perfectly distinguishable pure states. In general if $\omega \notin F$, then by Lemma 5.28, G must have at least $|F|+1$ perfectly distinguishable pure states, so the covering property requires that G contains no more than this.

Next recall the notion of third-order interference from Section 4.5.1, and in particular, that a state space $(A, S_+(A), u_A)$ (which satisfies the no-restriction hypothesis 2.6) exhibits no third order interference if $\mathsf{I}_3[q, \{P_J\}_{J\subseteq\{1,2,3\}}, \omega] = 0$ (Equation (4.23)) for all generalized slit systems $\{P_J\}_{J\subseteq\{1,2,3\}}$ (Definition 4.18), states $\omega \in \mathsf{S}(A)$, and final outcomes $q \in \mathsf{E}(A)$.

Principle 3″. *There is no third-order interference.*

Theorem 5.55. *Assume Principles 1 and 2, and that $S_+(A)$ has at least 3 perfectly distinguishable states. Then Principles 3, 3′, and 3″ are equivalent. In particular, Theorem 5.53 holds with Principle 3 replaced by Principle 3′, or 3″.*

Proof. It is clear that Principles 1 and 2 imply that $\Omega(A)$ is bit symmetric, and from Theorem 5.39 and Corollary 5.40 we have that the symmetric projections associated with faces of $S_+(A)$ are compressions, and this state space is projective. Because $S_+(A)$ has at least 3 perfectly distinguishable states, there exist orthogonal faces F_1, F_2 such that $F_3 := (F_1 \vee F_2)' \neq \{0\}$, namely we can take F_1, F_2 to be the faces generated by any two perfectly distinguishable states. More generally, take any orthogonal faces F_1, F_2 such that $F_3 := (F_1 \vee F_2)' \neq \{0\}$. The filters $\{P_1, P_2, P_3\}$ associated with these faces make up a 3-slit mask (Definition 4.15), and can be used to form a generalized 3-slit system $\{P_J\} = \{P_1, P_2, P_3, P_{12}, P_{13}, P_{23}, P_{123}\}$. Note that in this case we have $S_+(A) = F_1 \vee F_3 \vee F_3$, and so the filter P_{123} associated with the face $F_1 \vee F_3 \vee F_3$ is identical to $\mathbb{I}_A$.

$3' \Leftrightarrow 3''$ By Theorem 4.52, the filters in the 3-slit system $\{P_J\}$ are non-mixing if and only if $N(F_1, F_2, F_3) = \{0\}$. Lemma 4.42 then implies that $N(F_1, F_2, F_3) = \{0\}$ if and only if $\ker(P^{(3)}) = \{0\}$, or equivalently $\mathrm{im}(P^{(3)}) = A$. Theorem 4.26 together with $P_{123} = \mathbb{I}_A$ finally give $\mathsf{I}_3[q, \{P_J\}, \omega] = 0$ for all $\omega \in \mathsf{S}(A)$ and $q \in \mathsf{E}(A)$ if and only if the filters P_J are non-mixing, which proves that Principle 3 is equivalent to Principle 3″.

$3 \Leftrightarrow 3'$ By Proposition 5.43 we have that if F is a face of $S_+(A)$ with associated filter P_F, then P_F is non-mixing if and only if for every $\omega \in \mathrm{ext}(S_+(A))$, we have that $(\mathrm{face}(\omega) \vee F) \wedge F^\perp$ is an extremal ray of $S_+(A)$ or $\{0\}$. The statement that $(\mathrm{face}(\omega) \vee F) \wedge F^\perp = \{0\}$, is equivalent to $\mathrm{face}(\omega) \vee F \subseteq F^{\perp\perp} = F$, which is further equivalent to $\mathrm{face}(\omega) \vee F = F$.

Next, note that because $F \subseteq \mathrm{face}(\omega) \vee F$, and the lattice of faces is orthomodular (Proposition 5.36), we have $\mathrm{face}(\omega) \vee F = F \vee \big((\mathrm{face}(\omega) \vee F) \wedge F^\perp\big)$. Further, from Theorem 5.32, any frame for F can be extended to a frame for $\mathrm{face}(\omega) \vee F$, and any frame for $(\mathrm{face}(\omega) \vee F) \wedge F^\perp$ can also be extended to a frame for $\mathrm{face}(\omega) \vee F$. Because F is orthogonal to $(\mathrm{face}(\omega) \vee F) \wedge F^\perp$, this implies that $|\mathrm{face}(\omega) \vee F| = |F| + |(\mathrm{face}(\omega) \vee F) \wedge F^\perp|$.

Now suppose that $(\mathrm{face}(\omega) \vee F) \wedge F^\perp$ is an extremal ray of $S_+(A)$. The above then implies that $|\mathrm{face}(\omega) \vee F| = |F| + 1$, which by Lemma 5.28, implies that $\mathrm{face}(\omega) \vee F$ cover F. Finally, suppose that $\mathrm{face}(\omega) \vee F$ covers F. Using Lemma 5.28 again, this implies that $|\mathrm{face}(\omega) \vee F| = |F| + 1$, or $|(\mathrm{face}(\omega) \vee F) \wedge F^\perp| = 1$. This is equivalent to $(\mathrm{face}(\omega) \vee F) \wedge F^\perp$ being an extremal ray of $S_+(A)$. This proves that Principle 3 is equivalent to Principle 3′. □

Notice the restriction in the above theorem to state spaces which have at least three perfectly distinguishable states. This is necessary so that non-trivial three-slit experiments

can be formulated, namely that there exist two orthogonal faces F_1, F_2 such that $F_3 := (F_1 \vee F_2)' \neq \{0\}$. An interesting question then is what happens in the case where the state space has only two perfectly distinguishable states.

Proposition 5.56. *Assume Principles 1 and 2, and suppose that $S_+(A)$ has exactly two perfectly distinguishable states. Then under the product $a \circ b$ and the order unit norm, A is a Jordan-Banach algebra isomorphic to one of the spin factors $\mathbb{R}^d \oplus \mathbb{R}$. In particular, all filters are non-mixing, the product reduces to $(a,t) \circ (b,u) = (tb + ua, \langle a,b\rangle + tu)$, where $\langle \cdot,\cdot\rangle$ is the standard inner product on $\mathbb{R}^d$, and the set of normalized states is a ball in d dimensions.*

Proof. First note that if p, q are distinct atoms of $S_+(A)$, then the requirement that $S_+(A)$ has exactly 2 perfectly distinguishable states, along with Lemma 5.28 imply that that $\text{face}(p) \vee \text{face}(q) = S_+(A)$. Further note that for every atom p (equivalently, every pure state), we have $|\text{face}(p)| = |\text{face}(p)^\perp| = 1$, and therefore $p^\perp$ is also an atom, and the symmetric projection associated with $\text{face}(p)^\perp$ is rank-1. This implies that for every atom p, the associated symmetric projections $P, P^\perp$ are non-mixing as their positive images are faces of $S_+(A)$. Then following the steps of the proof of Proposition 5.45 (without the use of Principle 3 now), we conclude that the $*$ product satisfies the properties given in Proposition 5.45 without the assumption of Principle 3.

Defining the product $a \circ b$ exactly as in Definition 5.47 (without the use of Principle 3 now), it is easy to see that Theorem 5.48 and Theorem 5.52 follow as well without the use of Principle 3. This proves that under the product $a \circ b$ and the order unit norm, A is a Jordan-Banach algebra with multiplicative identity u_A.

Recalling the notion of a Jordan dynamical state space from Definition 5.12, and the discussion in Quantum example 5.17, it is not difficult to see that Theorem 5.2 implies the state space $S_+(A)$ is one of the spin factor algebras $\mathbb{R}^d \oplus \mathbb{R}$, where the product reduces to $(a,t) \circ (b,u) = (tb + ua, \langle a,b\rangle + tu)$, and $\langle \cdot,\cdot\rangle$ is the standard inner product on $\mathbb{R}^d$. Finally, using the definition of the cone of squares (Definition 5.6) and the normalization condition (Definition 5.12 (iii)), it is not difficult to see that $\Omega(A)$ is a ball in d dimensions. □

5.6 Discussion

In this chapter we have shown that the finite-dimensional formally real Jordan-algebraic theories can be reconstructed from three simple principles having an informational flavor. In particular, we have seen that many interesting consequences follow only from Principles

1 and 2. However, we could not complete the derivation without introducing a third principle requiring a special property for filters. An interesting question then is whether the third principle is essential, or whether it can be weakened or dropped altogether. Given the fact that we have not fully explored the group theoretic consequences of Principle 1, along with its apparent power when used together with Principle 2, we might conjecture that these alone are sufficient.

Conjecture 1. *Principles 1 and 2 are sufficient to reconstruct the finite-dimensional formally real Jordan-algebras.*

The other interesting possibility is that given Principles 1 and 2, Principle 3 is actually necessary for the reconstruction. In particular, this would mean that there are state spaces which satisfy Principles 1 and 2, but which are not Jordan-algebraic!

Conjecture 2. *There exist state spaces which are not Jordan-algebraic but which satisfy Principles 1 and 2.*

Evidence for this conjecture is the previous use of a principle like our third by Hardy [99], Araki [9], and Gunson [87].

It is also interesting to note that we have not discussed composite systems at all the reconstruction. In particular, recent reconstructions of complex quantum theory have used in a deep way the principle that any state of a composite system is determined just from correlations between measurements on subsystems.[3]One direction for further research is whether it is possible to drop Principle 3 and add local tomography in order to derive complex quantum theory. Another direction is to study whether it is possible to define "well behaved" composites of Jordan-algebraic systems, with the composite also a Jordan-algebraic system, but without local tomography.

Conjecture 3. *Composites of Jordan-algebraic systems which are Jordan-algebraic but not locally-tomographic can be defined, and are "well behaved".*

In particular, such composites would bypass recent results [123, 56] that any locally tomographic theory admitting continuous reversible interaction, where the local state spaces are Euclidean balls, must be standard complex quantum theory.

[3]This is often called local tomography [95], or local distinguishability [46], or tomographic locality [99].

Chapter 6

Equilibration of measurement statistics under complex dynamics

A major open problem in fundamental physics is that the processes by which the characteristic features of quantum statistical mechanics emerge from an exact treatment of the underlying quantum system are not fully understood. In particular, it is not obvious how (or even whether) a pure state evolving according to some quantum dynamics approaches some form of equilibrium. Recently however, there has been a resurgence of interest in this problem [49, 79, 146, 148, 155, 150, 172], and progress in providing rigorous justification for the existence and time scale of equilibration of small subsystems of a larger system [117, 154, 40, 125, 118, 78, 75, 149]. These results are based on the generic nature of random quantum states, and in particular their entanglement properties.

In this chapter we demonstrates an alternative, complementary mechanism for micro-canonical equilibration of isolated quantum systems which are at all times in a pure state, without invoking decoherence or coarse-graining of observables. Our central assumptions are:

(1) "sufficiently complex" dynamics in the sense that if the Hamiltonian is represented in the basis of an observable of interest, its eigenstates can be accurately modeled by those typical of a unitary chosen from the Haar measure [93, 128].

(2) an information theoretic restriction on the resources required to distinguish the dynamical quantum state from the relevant micro-canonical state.

The form of equilibration we will study is an *information theoretic* one in the sense that it is based on well-motivated operational constraints on the difficulty of predicting the detailed evolution of the system and the infeasibility of collecting an astronomically large amount of

measurement data. We will show that such equilibration is mathematically generic in the sense that it holds for almost all Hamiltonian systems, where "almost all" is defined with respect to the invariant (Haar) measure and the associated Gaussian Unitary Ensemble (GUE) [93, 128]. Similar techniques and arguments have recently also been used to show finite time equilibration for subsystems of a larger system [125, 101, 40].

The approach of studying an ensemble of random dynamical systems is motivated by the successful history in the fields of quantum chaos and nuclear physics of demonstrating that many properties of *individual* systems can be predicted simply from the average of the property over an appropriate *ensemble* of systems [86, 60, 59, 58, 93, 164, 165]. In particular, the well-known "random matrix theory conjecture" (also known as the BGS conjecture) states that certain important features of models whose classical counterparts are chaotic, such as the correlations between eigenvalues or between eigenvector components, can be modeled statistically by the those of a random matrix drawn from an appropriate ensemble sharing the same symmetries [38, 36, 86]. For example, GUE matrices are typically useful for modeling complex quantum systems that do not obey time–reversal symmetry [93].

In Section 6.1 we introduce our basic setup and discuss key concepts. Then in Section 6.2 we precisely define the notion of information theoretic equilibration and give a sufficient condition for it to hold. In Section 6.3 we introduce the Gaussian Unitary Ensemble, and state our main result, with the proof contained in Appendix D. Finally, in Section 6.4 we discuss the insights gained from our approach, and the possibility of information theoretic equilibration for more general systems. In particular, we identify a weaker condition on the system eigenvectors which we expect to be satisfied by physically realistic models.

6.1 General equilibration

Consider a dynamically isolated quantum system described by a Hilbert space $\mathcal{H} = \mathbb{C}^D$, and which is in an initial state ρ_0. Further, let H be a fixed Hamiltonian describing the reversible evolution of the system, $\rho(t) = U(t)\rho_0 U(t)^\dagger$, where $U(t) = e^{-itH}$ (taking $\hbar = 1$ for convenience).

Generally a system said to *equilibrate* if its state $\rho(t)$ evolves towards some particular mixed state ρ_{eq} and remains close to it (usually with respect to the trace distance [137, 146, 154]) for almost all times. If there is such an equilibrium state, it is not difficult to see that it must be the *infinite time average state* (sometimes called the generalized Gibbs state [150])

$$\rho_\infty := \mathbb{E}_\infty\big\{\rho(t)\big\} := \lim_{\tau\to\infty} \frac{1}{\tau}\int_0^\tau \rho(t)dt. \tag{6.1}$$

Note that this limit always exists, but in general it depends on the initial phases of the state when written in the energy eigenbasis. The assumption of non-degenerate energy gaps ensures independence of the initial phases [103]. In general the infinite time average state can differ from the *micro-canonical state* which is defined as

$$\varrho_{mc} := \mathbf{1}/D, \tag{6.2}$$

or any thermal state for that matter [154, 172]. This form of equilibration has only been rigorously proven for the states of sufficiently small subsystems of the full system, and depends on the generic nature of random quantum states and their entanglement properties [117, 154, 155, 40, 125, 118, 78, 75, 149].

Another approach is to take an observable $A = \sum_{mn} A_{mn}|e_m\rangle\langle e_n|$ (or more generally a given set of observables), and ask whether its expectation value

$$\mathrm{Tr}[A\rho(t)] = \sum_{mn} \rho_{mn} e^{i(E_n - E_m)t} A_{nm}, \tag{6.3}$$

equilibrates in the sense that it evolves towards its infinite time average and remains close to it for almost all times. In particular, closeness here is quantified by the infinite time variance

$$\mathbb{V}_\infty\big\{\,\mathrm{Tr}[A\rho(t)]\big\} := \lim_{\tau\to\infty} \frac{1}{\tau}\int_0^\tau \big[\mathrm{Tr}[A\rho(t)] - \mathbb{E}_\infty\big\{\,\mathrm{Tr}[A\rho(t)]\big\}\big]^2\, dt, \tag{6.4}$$

where

$$\mathbb{E}_\infty\big\{\,\mathrm{Tr}[A\rho(t)]\big\} = \mathrm{Tr}[A\rho_\infty] = \sum_m \rho_{mm} A_{mm}. \tag{6.5}$$

is the infinite time average of the expectation value of A (for non-degenerate energy gaps). This form of equilibration has been studied recently in [148, 155], under assumptions on the observable and the initial state, as well as in the older works [142, 58, 164, 165], which are closer in spirit to our approach.

More generally however, finite dimensional closed quantum systems cannot equilibrate either in the sense of the state approaching its infinite time average, or the expectation value of an observable approaching its infinite time average, because every finite dimensional unitary evolution is quasi-periodic [172]. In the next section we will instead consider a mathematically weaker but operationally well-motivated form of equilibration for the statistics of global observables.

6.2 Information theoretic equilibration

Consider a maximally fine-grained (non-degenerate) global observable A acting on $\mathcal{H}$, and let $A = \sum_{k=1}^{D} a_k \hat{P}_k$ be its spectral decomposition, where the projectors $\hat{P}_k$ are rank-one. Our central quantity of interest will be the probability of outcome k of the observable A, namely

$$\Pr\left(k|\rho(t)\right) := \mathrm{Tr}\left[\hat{P}_k U(t)\rho_0 U(t)^\dagger\right]. \tag{6.6}$$

Because we are interested in the dynamical properties of a probability distribution, we will use an information theoretic criterion for equilibration based on the ability to distinguish between data consistent with the precise quantum distribution $\Pr\left(k|\rho(t)\right)$, or the uniform or *micro-canonical distribution*

$$\mathrm{Pr_u(k)} := \frac{1}{\mathrm{D}}, \quad \text{for all k} = 1, \ldots, \mathrm{D}. \tag{6.7}$$

More precisely, as our measure of how distinguishable these distributions are, we will use the probability of successfully guessing which distribution a given set of data was sampled from. See Section 3.2, and [34, 137, 154] for more detail.

Definition 6.1 (**Information theoretic equilibration**). *A state $\rho(t)$ undergoing unitary dynamics $U(t)$ is* information theoretically equilibrated *with respect to a non-degenerate global observable A, after a time t_{eq}, if for almost all times $t > t_{\mathrm{eq}}$ the outcome distribution $\Pr\left(k|\rho(t)\right)$ cannot be distinguished from the uniform distribution $\mathrm{Pr_u(k)}$ with high probability using at most $O(\mathrm{polylog}(D))$ samples from $\Pr\left(k|\rho(t)\right)$ and at most $O(\mathrm{polylog}(D))$ computational time.*

This notion captures the idea that although the exact quantum distribution for the system is *in principle* distinguishable from the uniform distribution, the number of observations required scales polynomially (rather than poly–logarithmically) with the Hilbert space dimension. The Hilbert space dimension of a macroscopic body is typically astronomically large (for a many-body system D scales exponentially with the number of subsystems), which justifies the requirement that only a poly–logarithmic number of samples are allowed. The restriction on the number of computational steps is important because it forbids explicitly solving for $\rho(t)$, which is physically unrealistic (even in the presence of quantum computers) and could be used to distinguish $\rho(t)$ from ϱ_{mc}. See the discussion following Assumption 7 for more on this point.

Next, define the *outcome variance* of a probability distribution $\mathrm{P(k)}$ over D outcomes as

$$\mathbb{V}_k\{\mathrm{P(k)}\} := \mathrm{D}^{-1} \sum_{\mathrm{k}} \left[\mathrm{P(k)} - \mathrm{D}^{-1}\right]^2. \tag{6.8}$$

The following result, which will be essential in the next section, gives a sufficient condition for distinguishing an *unknown* distribution P(k) from the uniform distribution $\mathrm{Pr_u(k)}$ given $O(\mathrm{polylog}(D))$ samples from P(k).

Theorem 6.2 (proved by Nathan Wiebe)**.** *Take an unknown probability distribution* P(k) *over D outcomes, which has outcome variance* $\mathbb{V}_k\{P(k)\} \in O(D^{-2})$*. Then at least* $O(D^{1/4})$ *samples from the distribution* P(k) *are needed in order to distinguish it from* $\mathrm{Pr_u(k)}$ *with high probability.*

Proof. First, notice that because we have little knowledge about the form of the distribution P(k), the labels of the outcomes are of no significance. This means that we can only use the frequency of coincidences in outcomes in order to distinguish P(k) from the uniform distribution.

It is straightforward to see that $O(\sqrt{D})$ samples are needed in order to obtain a pair of identical outcomes drawn from the uniform distribution $\mathrm{Pr_u(k)}$. We will compare this with a distribution with $\mathbb{V}_k\{\mathrm{P(k)}\} = O(D^{-2})$ which has the highest probability of coincidence for some outcomes. Given the implication

$$\mathbb{V}_k\{\mathrm{P(k)}\} = O(D^{-2}) \;\Rightarrow\; \mathrm{P(k)} \leq \mathrm{O(D^{-1/2})},\ \forall\ \mathrm{k}, \tag{6.9}$$

we find that the distribution Q(k) which maximizes the probability of coincidence is, for some constant a, given by

$$\mathrm{Q(k)} = \begin{cases} \sqrt{\frac{a}{D}} & :\ k \leq \sqrt{\frac{D}{a}} \\ 0 & :\ \text{else} \end{cases}. \tag{6.10}$$

Given the distribution Q, the probability that there are no coincidental outcomes after N samples is given by:

$$\frac{\left(\sqrt{\frac{D}{a}}\right)!}{\left(\sqrt{\frac{D}{a}} - N + 1\right)! \left(\sqrt{\frac{D}{a}}\right)^N} = 1 - \frac{N(N-1)}{2\sqrt{\frac{D}{a}}} + O(N/D). \tag{6.11}$$

This implies that a distribution P(k) with $\mathbb{V}_k\{\mathrm{P(k)}\} = O(D^{-2})$ cannot be distinguished from the uniform distribution with high probability unless $N \geq O(D^{1/4})$, as claimed. □

Now that we have an operationally well motivated notion of equilibration based on the difficulty of predicting the detailed evolution of the system and the infeasibility of collecting an astronomically large amount of measurement data, the question is which quantum systems equilibrate.

6.3 The Gaussian Unitary Ensemble

In this section we show that information theoretic equilibration is mathematically generic in the sense that it holds for almost all Hamiltonian systems, where "almost all" is defined with respect to the natural, invariant (Haar) measure and the associated Gaussian Unitary Ensemble. To this end, consider the following class of dynamical models:

The Hamiltonian H is fixed but drawn uniformly at random from the Gaussian Unitary Ensemble (GUE) of Hermitian matrices.

The approach of studying an ensemble of random dynamical systems is motivated by the successful history in the fields of quantum chaos and nuclear physics of demonstrating that many properties of *individual* systems can be predicted simply from the average of the property over an appropriate *ensemble* of systems [86, 60, 59, 58, 93, 164, 165]. In particular, the well-known "random matrix theory conjecture" (also known as the BGS conjecture) states that certain important features of models whose classical counterparts are chaotic, such as the correlations between eigenvalues and eigenvector-components, can be modeled statistically by the those of a random matrix drawn from an appropriate ensemble sharing the same symmetries [38, 36, 86]. For example, GUE matrices are typically useful for modeling complex quantum systems that do not obey time–reversal symmetry [93]. The success of this approach is based on two independent elements:

(i) the particular physical system of interest has sufficient complexity to generate enough pseudo-randomness for the property of interest, and can thus be well modeled by a typical matrix drawn from an ensemble of random matrices, and
(ii) the distribution of the property is tightly peaked around the ensemble mean (this is often called concentration of measure [146, 117, 133], or typicality [78]).

In this section we will address the second element by showing that information theoretic equilibration occurs not only for the ensemble average of H but also holds with high probability for individual members of the ensemble. The first element will be addressed in the following section.

The Gaussian Unitary Ensemble is the unique probability distribution $P(H)$ over Hermitian matrices H which satisfies the following two requirements [93, 128]:

(1) invariance of the distribution $P(H)$ under any unitary transformation $H \to UHU^\dagger$,
(2) the joint distribution over all elements of H factorizes into a product of distributions each over an individual element of H.

The second requirement can be interpreted as a maximum ignorance property given the Hermitian and unitary invariance requirements [12]. The elements of H are such that the

diagonal elements h_{aa} are real valued random variables with distribution $\mathcal{N}(0, \sigma^2)$, and the off-diagonal elements h_{ab} have real and imaginary parts which are independent and identically distributed random variables with distribution $\mathcal{N}(0, \frac{1}{2}\sigma^2)$. The variance σ^2 is a free parameter, which is closely related to the expected maximum eigenvalue, as well as the average level spacing.

There are several important drawbacks to using GUE which deserve mention before moving on. First, the average level density, which takes the form of a semi-circle, is not a good model for real Hamiltonian systems, even chaotic ones. In particular, for most natural systems, such as those with only two particle interactions, the norm of the Hamiltonian scales polynomially with the number of particles [117]. However, the expected norm of a GUE Hamiltonian scales polynomially in the Hilbert space dimension D [93]. There are also other features of a typical spectrum which are unrealistic, such as the correlations between distant energies [93, 86, 128]. As will become clear in the calculations in Appendix D.2, these features have a large impact on what might be called the equilibration time for these dynamical systems, which is an important element in the next theorem. For this reason we do not state an explicit equilibration time in our results. This further allows us to simply the calculations by following the standard practice [125, 128] of taking the variance

$$\sigma^2 = \frac{1}{2}. \tag{6.12}$$

A final important fact is that the joint distribution over the elements of H factorizes into a product of a joint distribution over eigenvectors, and a joint distribution over the energy eigenvalues of H (see [128] Theorem 3.3.1, [92] Theorem 1.4, or [93] Chapter 4). Letting C be the unitary which transforms the eigenbasis of H to that of A, and working in the eigenbasis of A, we can write

$$U(t) = CF(t)C^\dagger, \tag{6.13}$$

where $F(t) := \text{diag}[e^{-itE_a}]$, and $\{E_a\}_{a=1}^D$ are the energy eigenvalues of H. The joint probability distribution over eigenvectors of H is the same as that over the change of basis matrix C, namely the Haar measure on the unitary group $\mathbb{U}(D)$ (see [93] Chapter 4, or [128]). In other words, the eigenvectors of a typical GUE Hamiltonian H (making up the columns of C) are columns of a typical Haar random unitary. We are now ready to state the main technical result of this chapter.

Theorem 6.3. *Take a non-degenerate observable A acting on $\mathcal{H} = \mathbb{C}^D$, an initial pure state $\rho_0 = |x\rangle\langle x|$ which is an eigenstate of A, and a unitary $U(t) = e^{-itH}$ where H is drawn uniformly at random from GUE, with $\sigma^2 = 1/2$. Then there exists a finite time $t_{\text{eq}}(D)$,*

such that for all $t > t_{\text{eq}}(D)$ *we have that the mean, variance, and fourth moment of the measurement outcome probabilities* $\Pr(k|x,t)$ *over the GUE are given by:*

$$(i) \quad \mathbb{E}_{\text{GUE}}\{\Pr(k|x,t)\} = \frac{\delta_{xk}+1}{D+1} + O(D^{-2}), \tag{6.14}$$

$$(ii) \quad \mathbb{V}_{\text{GUE}}\{\Pr(k|x,t)\} = O(D^{-2}), \tag{6.15}$$

$$(iii) \quad \mathbb{E}_{\text{GUE}}\{\Pr(k|x,t)^4\} = O(D^{-4}). \tag{6.16}$$

Proof. First, because the probability distribution over eigenvectors and eigenvalues factorizes, we can take separate expectations over these, namely, over the matrices C and F. To this end, we write $\Pr(k|x,t)$ more explicitly as

$$\Pr(k|x,C,F(t)) := \text{Tr}[|k\rangle\langle k|CF(t)C^\dagger|x\rangle\langle x|CF(t)^\dagger C^\dagger]. \tag{6.17}$$

In the following we will write the expectation over the Haar measure on the unitary group $\mathbb{U}(D)$ of change of basis matrices C, as $\mathbb{E}_C\{\cdot\}$, and $\mathbb{E}_{spec}\{\cdot\}$ for the expectation over the GUE distribution of eigenvalues of H which are contained in F.

Our strategy for proving this result consists of two steps:
(1) First derive expressions for the Haar expectations of $\Pr(k|x,t)$ and the second and fourth moments. See Lemmas D.1, D.2, and D.3.
(2) Find the expectations of these expressions over the GUE eigenvalue distribution, and show that there exists a finite time $t_{\text{eq}}(D)$ such that for all $t > t_{eq}(D)$, expressions (i)-(iii) hold. See Appendix D.2 for these expectations. In particular, see Appendix D.2.2 for the existence of the equilibration time, Appendix D.2.3 for the proof of expressions (i) and (ii), and Appendix D.2.5 for the proof of expression (iii). □

This result implies that $\Pr(k|x,t)$ is concentrated around $1/D$ for almost every Hamiltonian chosen from the GUE for $t \geq t_{\text{eq}}(D)$. In particular, by Chebyshev's inequality we have that in the limit of large D, with high probability over H drawn from GUE, the corresponding probabilities $\Pr(k|x,t)$ are at most a constant multiple of $1/D$:

$$\Pr_{H\sim\mu_D}\left(\left|\Pr(k|x,t) - \frac{\delta_{xk}+1}{D+1}\right| \geq \epsilon\right) \leq \frac{2\delta_{xk}+1}{\epsilon^2 D^2}, \tag{6.18}$$

where μ_D represents the measure for the GUE. In other words, the probabilities $\Pr(k|x,t)$ for an individual system behave like the ensemble average.

Before stating the central result of this chapter, we must make one further plausible assumption:

Assumption 7. *For a typical GUE Hamiltonian H, it is not possible to simulate the unitary $U(t) = e^{-itH}$, either classically or quantum mechanically in less than $O(\text{polylog}(D))$ computational time. In the following we will say such a unitary is* unknown.

This is assumption is motivated by the following considerations:

(1) The evolution of generic many–dimensional quantum systems is believed to be difficult to simulate classically. If this were not the case, then quantum computers would be no more powerful than classical computers. As such, even completely specifying the Hamiltonian parameters for a many–body system is insufficient for efficient simulation.

(2) There are no known efficient quantum simulation algorithms for non-sparse Hamiltonians [35, 44], and further, "almost all" unitary evolutions are exponentially hard to simulate, even on a quantum computer [137].

In particular, if $U(t)$ is unknown for a typical GUE Hamiltonian, then the probability distribution $\Pr(k|x,t)$ is unknown as well.

Theorem 6.4. *Take a non-degenerate observable A acting on $\mathcal{H} = \mathbb{C}^D$, an initial pure state $\rho_0 = |x\rangle\langle x|$ which is an eigenstate of A, and an unknown unitary $U(t) = e^{-itH}$ where H is drawn uniformly at random from GUE with $\sigma^2 = 1/2$. Given Assumption 7, then with high probability over GUE, for almost all times $t > t_{\text{eq}}$, the outcome distribution $\Pr(k|x,t)$ cannot be distinguished from the uniform distribution $\Pr_{\text{u}}(\text{k})$ with less than $O(D^{1/4})$ samples from $\Pr(k|x,t)$ and less than $O(\text{polylog}(D))$ computational time, i.e., the system is information theoretically equilibrated.*

Proof. First, we will show that $\mathbb{V}_{\text{GUE}}\{\mathbb{V}_k\{\Pr(k|x,t)\}\} = O(D^{-4})$. Expanding out the variance $\mathbb{V}_k\{\Pr(k|x,t)\}$ as well as the GUE variance in terms of expectations, and using Theorem 6.3, we find:

$$\mathbb{V}_{\text{GUE}}\left\{\mathbb{V}_k\{\Pr(k|x,t)\}\right\} = D^{-2}\sum_{j,k}\mathbb{E}_{\text{GUE}}\left\{\Pr(k|x,t)^2\Pr(j|x,t)^2\right\} + O(D^{-4}). \tag{6.19}$$

Then using the Cauchy-Schwartz inequality for expectations, we have

$$\begin{aligned}\sum_{j,k}\mathbb{E}_{\text{GUE}}\left\{\Pr(k|x,t)^2\Pr(j|x,t)^2\right\} &\leq \sum_{j,k}\sqrt{\mathbb{E}_{\text{GUE}}\left\{\Pr(k|x,t)^4\right\}\mathbb{E}_{\text{GUE}}\left\{\Pr(j|x,t)^4\right\}} \\ &= \left(\sum_{j=1}^{D}\sqrt{\mathbb{E}_{\text{GUE}}\left\{\Pr(k|x,t)^4\right\}}\right)^2 = O(D^{-2}),\end{aligned} \tag{6.20}$$

where we have also used Theorem 6.3. This proves that $\mathbb{V}_{\mathrm{GUE}}\{\mathbb{V}_k\{\Pr(k|x,t)\}\} = O(D^{-4})$.

Chebyshev's inequality then implies (in the same fashion as in expression (6.18)) that $\mathbb{V}_k\{\Pr(k|x,t)\} \in O(D^{-2})$ with high probability over GUE, for $t \geq t_{\mathrm{eq}}(D)$. Finally, by using Assumption 7 and Theorem 6.2, we have that $\Pr(k|x,t)$ cannot be distinguished from the uniform distribution $P_{\mathrm{u}}(k)$ with less than $O(D^{1/4})$ samples from $\Pr(k|x,t)$ and less than $O(\mathrm{polylog}(D))$ computational time, i.e., the system is information theoretically equilibrated. □

This result states the vast majority of *individual* systems (which are at all times in a pure state) with dynamics controlled by individual Hamiltonians chosen according to the GUE, information theoretically equilibrate in finite time. It is important to stress that this result holds for a *single* realization of a system, and that the averaging used in Theorem 6.3 is only done in order to show that the typical system behaves like the ensemble average.

6.4 Going beyond GUE

We have proved analytically that information theoretic equilibration holds generically for closed Hamiltonian systems that satisfy a condition of sufficient complexity. Remarkably, for each of these systems information theoretic equilibration is observed to hold even for pure quantum states, without requiring any form of decoherence or restricting to coarse-grained measurements. Our key insight is that, although the dynamical pure states exhibit coherent fluctuations away from true micro-canonical equilibrium, these fluctuations remain small and appear pseudo-random, and hence their detection requires extraordinary resources, such as collecting $O(D^{1/4})$ measurement outcomes from repetitions of the exact experimental conditions, or pre–computation of the dynamical state in a D-dimensional Hilbert space, or performing joint (entangling) measurements on identical copies of the system.

However, assuming GUE Hamiltonians or more simply just Haar-randomness of the eigenstates seems unnecessarily strong for Theorem 6.2 to hold. First, recall that the GUE average spectrum and the long range correlations between eigenvalues are not physically realistic. Therefore we would like to replace terms of the form $\mathbb{E}_{spec}\{|\mu(t)|^2\}$ with expectations over more realistic ensembles, or simply $|\mu(t)|^2$ for *individual* systems. Recall that $|\mu(t)|^2$ appears for example in $\mathbb{E}_C\{\Pr(k|x,C,F(t))\}$ in Lemma D.1, and so controls the equilibration time (given the assumption of Haar random eigenstates). Second, a Haar-random unitary exhibits the maximal possible independence of its matrix elements. In particular their independence is constrained only by normalization and orthogonality (the

rows, or equivalently columns, must consist of D orthonormal pure states). For a physical system this is clearly not realistic. For example, a physical system might have of only two-body local interactions, which would create correlations between different eigenvectors, and between elements of each eigenvector. Therefore we would like to replace the expectations $\mathbb{E}_C\{\cdot\}$ over Haar measure with more realistic ensembles, or study the properties of analogous expressions for individual systems.

To this end, recall from Appendix D.1 that the measurement outcome probabilities can be written as

$$\Pr\left(k|x,C,F(t)\right) = \mathrm{Tr}\left[|k\rangle\langle k|CF(t)C^\dagger|x\rangle\langle x|CF(t)^\dagger C^\dagger\right], \tag{6.21}$$

where C is the unitary which transforms the eigenbasis of H to that of the observable A. Information theoretic equilibration simply requires that $\mathbb{V}_k\{\Pr\left(k|x,C,F(t)\right)\} \in O(D^{-2})$, which in turn requires that we know certain properties of $\Pr\left(k|x,C,F(t)\right)^2$. Recall from Lemma D.2 that $\Pr\left(k|x,C,F(t)\right)$ and $\Pr\left(k|x,C,F(t)\right)^2$ can be concisely represented by

$$\Pr\left(k|x,C,F(t)\right) = \langle L_2|C^{\otimes 2}\otimes\bar{C}^{\otimes 2}|R_2(t)\rangle, \tag{6.22}$$

$$\Pr\left(k|x,C,F(t)\right)^2 = \langle L_4|C^{\otimes 4}\otimes\bar{C}^{\otimes 4}|R_4(t)\rangle, \tag{6.23}$$

where $\langle L_2| := \langle k,x,x,k|$, $\langle L_4| := \langle k,x,k,x,x,k,x,k|$, and

$$|R_2(t)\rangle := \sum_{b,b'} e^{it(E_b-E_{b'})}|b,b',b,b'\rangle, \tag{6.24}$$

$$|R_4(t)\rangle := \sum_{b,b',d,d'} e^{it(E_b-E_{b'}+E_d-E_{d'})}|b,b',d,d',b,b',d,d'\rangle. \tag{6.25}$$

We refer to a term of the form $\langle L_2|C^{\otimes 2}\otimes\bar{C}^{\otimes 2}|b,b',b,b'\rangle$ in the sum in (6.22) as a $(2,2)$–term because there are two basis change matrices acting on each factor space. The analogous terms in (6.23) will be called $(4,4)$–terms.

Recent studies of subsystem equilibration [40, 125] show that the scalings given in Equations (6.14), (6.15) and (6.16) are satisfied provided that the matrix elements of C, the eigenvector components, satisfy a unitary k-design condition [52] for some finite k. In particular, if the ensemble average of the $(2,2)$ terms is $O(D^{-2})$, and for the $(4,4)$ terms $O(D^{-4})$, and the variance of the $(4,4)$ terms $O(D^{-8})$, then by Chebyshev's inequality the variance of $\Pr\left(k|x,C,F(t)\right)$ will be $O(D^{-2})$ for almost all Hamiltonians chosen from the matrix ensemble. Hence, if these underlying scalings are obeyed by the eigenvector components of any ensemble of Hamiltonians then information theoretic equilibration is a direct consequence according to Theorem 6.2.

For example, consider the ensemble consisting of 2–local Hamiltonians, i.e., Hamiltonians composed at most 2-body interactions between qudits. These Hamiltonians are ubiquitous in condensed matter theory and quantum computing. We anticipate that such Hamiltonians have sufficient complexity to induce equilibration because the interactions that comprise them generate a universal gate for quantum computing. Another interesting question is whether an individual physical system, rather than those drawn at random from some mathematical ensemble, will exhibit information theoretic equilbration as D increases. This will require studying the properties of the above $(2, 2)$ and $(4, 4)$ terms for the individual system.

Chapter 7

Future Directions

In this dissertation we have tried to shed a little light on quantum theory, or at least on its formal mathematical structure, from an operational point of view. After over 100 years of development and spectacular success, quantum theory remains essentially a mysterious theory. Every current description of quantum theory is a statement of its abstract mathematical formalism, without any understanding of the nature or origin of these principles or the underlying reality. The natural question is then: how can we understand this formalism better?

There are several possible approaches to this type of question. One is based on the idea that understanding of some physical theory comes from placing it in a relevant context of other theories, and by studying the various basic parts or elements of the theory independently, so that they can be generalized, modified, contrasted, and related to each other without the whole package necessarily coming along for the ride. A second approach is based the idea that we understand some phenomenon or physical theory when we can take it apart and put it back together conceptually, or when we can derive or reconstruct its abstract structure from simpler, or more natural, or intuitive principles. A third approach is simply to apply the formalism of the theory in a rigorous fashion and try to find interesting consequences or to answer interesting questions.

In the spirit of the first approach, in Chapter Three we studied an interesting correspondence between the states and observables of quantum theory, which underpins the well known state transition probability rule. In particular, we showed that this correspondence follows naturally from a simple symmetry principle, called bit-symmetry, which is closely related to reversible time evolution and the possibility of powerful reversible computation. Recent work on operational probabilistic theories has largely focused on a weaker form

of symmetry, so a very interesting direction for future work is to understand what other consequences bit-symmetry, as well as generalizations of it, have for the geometric and information processing properties of a theory.

Again in the spirit of the first approach, in Chapter Four we studied the concept of interference, which has played a central role in the development and interpretation of quantum theory. We proved several results characterizing theories which display various forms of interference, and in particular, theories which share the interference properties of quantum theory. These results give new insight into the structure of quantum theory and the implications of interference experiments, and also suggests a novel way of testing the interference properties of quantum theory. We also showed that there is an interesting connection between the existence of certain forms of interference and whether the state update rules of the theory are similar to those of quantum theory. More generally, it would be interesting to study how various forms of interference are related to other nonclassical phenomena, such as information processing properties, non-locality, symmetry properties, etc.

In Chapter Five we took the second approach and focused on reconstructing the algebraic structures of quantum theory for individual systems. We showed that the closest mathematical cousins to standard quantum theory can be characterized by three simple principles which have an informational flavor. It is interesting to note how powerful the first two principles seem, which suggests the possibility the third principle can be weakened or dropped altogether. Alternatively, if the third principle is necessary for the reconstruction, it would mean that there are unknown theories with posses the simplicity and symmetry properties demanded by the first two principles, but which are not Jordan-algebraic. Another interesting direction for future research is whether it is possible to drop the third principle and add a commonly used principle for the behavior of composite systems (local tomography) in order to derive the standard complex quantum theory.

Finally, in Chapter Six, we focused on quantum statistical mechanics and the problem of understanding how its characteristic features can be derived from an exact treatment of the underlying quantum system. Based on an assumption of sufficiently complex but reversible dynamics, and a restriction on measurement resources, we proved that a natural information theoretic form of equilibration is mathematically generic in the sense that it holds for almost all Hamiltonian systems drawn from the Gaussian Unitary Ensemble. Remarkably, for each of these systems information theoretic equilibration occurs even for pure quantum states, without requiring any form of decoherence or coarse-grained measurements. Our key insight is that although the dynamical pure states exhibit coherent fluctuations away from micro-canonical equilibrium, these fluctuations remain small and appear pseudo-random, and hence their detection requires extraordinary measurement or

computational resources. We also noted that assuming GUE Hamiltonians seems unnecessarily strong, and suggested less restrictive assumptions on the Hamiltonian eigenvectors under which information theoretic equilibration is expected to hold. This further holds the promise of experimental tests of information theoretic equilibration.

While the topics and results presented may be of intrinsic interest to those who wish to better understand the mathematical structure of finite dimensional quantum theory, they can also be seen as stepping stones to a better understanding of quantum theory, or to developing other more accurate and more fundamental theories. In the absence of some fundamental picture of the world, the approaches we have taken allow us to proceed in a careful and conceptual fashion. Important ideas can be formulated in operational terms, without committing to a particular mathematical framework beyond what is needed to describe experiments and to make predictions. This methodology helps us listen more carefully to what quantum theory is telling us, rather than imposing our preconceptions or prejudices on it.

In particular, deeper and more reasonable principles for quantum theory, such as those presented in Chapter Five, can dissolve the mysteries of quantum phenomena and make them more intuitive. Further, the results on Chapters Three and Five not only shed light on apparently coincidental features of quantum theory, but can also serve as pedagogical tools. For example, when asked about the meaning or origin of the state transition probability rule, the teacher of quantum theory can point to the bit-symmetry principle as an explanation. More generally, the three principles we have presented can be used to teach and motivate the mathematical structure of quantum theory. More practically, the approaches taken in this dissertation also hold the promise of explaining what is responsible for the power of quantum information processing and cryptography. In particular, the results in Chapters Three and Five can be used as first steps in this important project.

A more philosophical and open ended question which we have not addressed is what is the most fruitful way to think of the quantum formalism – given an operational or information theoretic point of view on it – and the reconstruction program. In particular, what do the recent reconstructions tell us about the formalism, and which set of principles is the most natural or will be most fruitful? One possibility is that the recent reconstructions are simple and transparent characterization of the formalism of quantum theory, and the principles used are similar to lego blocks or subroutines in an algorithm in the sense that they can be used or swapped out appropriately, depending on one's intuition and need [177]. Another point of view is that reconstructions of quantum theory are similar to derivations or characterizations of various classical probability distributions, like the Gaussian or Poisson distributions; the goal is simply to understand in what situations it is appropriate to use these distributions, or where to expect to see them in nature. Yet another possibility is

that principles should be understood as "laws of thought", and the quantum formalism as a way to reason under uncertainty [147]. A final possibility is that we should only expect some of the principles to be information theoretic or probabilistic in nature, and that what is left over will represent the true essence of quantum theory or the underlying reality [70].

APPENDICES

Appendix A

Convexity

In this appendix we define and discuss various well known notions from the theory of convex geometry in finite dimensional real vector spaces. We focus on aspects of this theory which are most relevant for our purposes. The various definitions are standard, but the particulars used are largely based on [7, 15, 39, 175].

Unless otherwise stated, all sets and vectors are assumed to be contained in a real, finite-dimensional vector space A.

A.1. *A vector $x = p_1x_1 + \ldots + p_kx_k \in A$, where $p_i \geq 0$ for $i = 1, \ldots, k$, and $\sum_{i=1}^{k} p_i = 1$, is called a **convex combination** or **mixture** of the vectors $x_1, \ldots, x_k$.*

A.2. *A set $C \subseteq A$ is **convex** if $x_1, x_2 \in C$, implies that $px_1 + (1-p)x_2 \in C$, for all $0 \leq p \leq 1$.*

The condition defining a convex set has the simple geometric interpretation that the entire line segment between any two points in C lies in C. It is easy to see that any convex combination of elements in a convex set C is contained in C. Further, the intersection of two convex sets is also a convex set [175].

A.3. *A point x in a convex set $C \subseteq A$ is an **extreme point** of C if $x = px_1 + (1-p)x_2$ with $x_1, x_2 \in C$ and $0 < p < 1$, implies that $x_1 = x_2 = x$. The set of extreme points of a convex set C will be denoted by $\mathrm{ext}(C)$.*

Essentially, the extreme points of a convex set are the elements which cannot be decomposed as (non-trivial) convex combinations of other elements of the set.

A.4. *The* ***convex hull****,* $\mathrm{conv}(B)$*, of a set* $B \subseteq A$ *is the set of all convex combinations of points in* B*, namely*

$$\mathrm{conv}(B) = \left\{ \sum_i p_i x_i \mid x_i \in B,\ 0 \leq p_i,\ \sum_i p_i = 1 \right\}. \tag{A.1}$$

Further, the ***affine hull****,* $\mathrm{aff}(B)$*, of* B *is the set of all affine combinations of points in* B*, namely*

$$\mathrm{aff}(B) = \left\{ \sum_i p_i x_i \mid x_i \in B,\ p_i \in \mathbb{R},\ \sum_i p_i = 1 \right\}. \tag{A.2}$$

Finally, the ***linear hull****,* $\mathrm{lin}(B)$*, of* B *is the set of all linear combinations of points in* B*, namely*

$$\mathrm{lin}(B) = \left\{ \sum_i p_i x_i \mid x_i \in B,\ p_i \in \mathbb{R} \right\}. \tag{A.3}$$

The convex hull of a set B is the smallest convex set which contains B. An important result – known as Minkowski's theorem – is that any closed and bounded (i.e., contained in some ball of finite radius) convex subset C of a finite dimensional real vector space A is the convex hull of its extreme points, namely $C = \mathrm{conv}(\mathrm{ext}(C))$ [34, 175]. Further, if $A \simeq \mathbb{R}^n$, the number of extreme points needed in any convex decomposition of a point $x \in \mathrm{conv}(C)$ is at most $n+1$. This is known as Caratheodory's theorem [34].

A.5. *The* ***interior****,* $\mathrm{int}(D)$*, of a set* $D \subseteq A$ *is defined as*

$$\mathrm{int}(D) = \{x \in D \mid \exists\ r > 0,\ \textit{with}\ B(x,r) \subseteq D\}, \tag{A.4}$$

where $B(x,r) = \{y \in A \mid ||y - x|| < r\}$ *is the open ball of radius* r *centered at* x *(here* $||\cdot||$ *is any norm on* A*).*
The ***closure****,* $\mathrm{cl}(D)$*, of* D *is defined as*

$$\mathrm{cl}(D) = \{y \in A \mid \forall\ r > 0,\ B(x,r) \cap D \neq \emptyset\}. \tag{A.5}$$

We will say D *is* ***closed*** *if* $\mathrm{cl}(D) = D$*.*
The ***boundary****,* $\mathrm{bd}(D)$*, of* D *is defined as* $cl(D) \setminus \mathrm{int}(D)$*.*
The ***relative interior*** $\mathrm{relint}(D)$ *of* D *is defined as*

$$\mathrm{relint}(D) = \{x \in D \mid \exists\ r > 0,\ \textit{with}\ B(x,r) \cap \mathrm{aff}(D) \subseteq D\}. \tag{A.6}$$

The ***relative boundary*** $\mathrm{relbd}(D)$ *of* D *is defined as* $cl(D) \setminus \mathrm{relint}(D)$*.*

Note that the extreme points of a convex set are a subset of the boundary points, but not all boundary points are extreme points. See Figure 2.2.

A.6. *A convex set $C \subseteq A$ is a* ***direct convex sum*** *of two convex subsets $C_1, C_2 \subset C$, if any $k \in C$ has a* unique *decomposition of the form $k = \lambda k_1 + (1-\lambda)k_2$, with $\lambda \in [0,1]$, and $k_i \in C_i$. If the above is satisfied, we write $C = C_1 \oplus_c C_2$, and C_1 and C_2 are said to* ***split*** *C.*

A.7. *A nonempty convex subset F of a convex set $C \subseteq A$ is a* ***face*** *of C if for all $0 < \lambda < 1$, and $k_1, k_2 \in C$, the inclusion $\lambda k_1 + (1-\lambda)k_2 \in F$ implies $k_1, k_2 \in F$. F is a* ***proper face*** *of C if F is a face and $F \neq C$.*

The basic idea behind the notion of a face of C is that it is a convex set which is closed under convex decompositions in terms of elements of C. The extreme points of C are faces.

A.8. *Let D be a subset of a convex set $C \subseteq A$. The* ***face generated by*** *D is defined to be the intersection of all faces of C containing D, namely,*

$$\text{face}(D) = \bigcap \{F \mid F \text{ is a face of } C, \text{ and } D \subset F\}. \tag{A.7}$$

It is not difficult to see that $D \subset \text{face}(D)$, and that $D = \text{face}(D)$ if and only if D is a face.

A.9. *A face F of a convex set $C \subseteq A$ is called a* ***maximal face*** *if $F \neq C$, and if $F \subseteq G \subseteq C$ for a face G, then $G = F$ or $G = C$.*

A.10. *A subset A_+ of a vector space A is called a* ***cone*** *if the following conditions are satisfied:*

(i) $A_+ + A_+ \subset A_+$ *(A_+ is closed under addition)*
(ii) $\lambda \cdot A_+ \subset A_+$ *for all $\lambda \geq 0$ (A_+ is closed under non-negative multiplication),*
(iii) $A_+ - A_+ = A$ *(A is positively generated [7]),*

A.11. *A cone $A_+ \subset A$ is called a* ***pointed cone*** *(sometimes called a proper cone) if $A_+ \cap (-A_+) = \{0\}$.*

A.12. *Given a point x in a cone $A_+ \subset A$, the line segment*

$$\mathbb{R}_+ \cdot x := \{y \in A_+ \mid y = \lambda x, \lambda \geq 0\}, \tag{A.8}$$

is an ***extremal ray*** *of A_+ if and only if x is an extreme point of A_+, i.e., the elements of $\mathbb{R}_+ \cdot x$ can only be decomposed as convex combinations of other elements of $\mathbb{R}_+ \cdot x$. We will write $\text{ext}(A_+)$ for the set of elements of A_+ which are contained in extremal rays.*

A.13. *Given a vector space A, the* ***dual space*** A^* *is defined as the set of all linear functionals $f : A \to \mathbb{R}$.*

A.14. *Given a cone $A_+ \subset A$, the* ***dual cone*** A_+^* *is defined as the set of all linear functionals on A which are non-negative on all of A_+, namely*

$$A_+^* = \{f \in A^* \mid f(x) \geq 0,\ \forall x\ \in A_+\}\,. \tag{A.9}$$

Next is a well known result, which we state without proof.

A.15. *Let A be a finite dimensional real vector space, and $A_+ \subset A$ a cone. The bi-dual space of A is isomorphic to A, namely, $(A^*)^* \simeq A$, and given the association $(A^*)^* = A$, the bi-dual cone $(A_+^*)^*$ is identical to A_+, namely $(A_+^*)^* = A_+$.*

A.16. *Given a cone $A_+ \subset A$, define a* ***partial order*** $\leq$ *on A induced by A_+ by the following relation:*

$$x \leq y \quad \text{if and only if} \quad y - x \in A_+. \tag{A.10}$$

As any other partial order, this is *reflexive* ($x \leq x$), *anti-symmetric* (if $x \leq y$ and $y \leq x$ then $x = y$), and *transitive* (if $x \leq y$ and $y \leq z$ then $x \leq z$).

A.17. *Given two elements x, y in a cone $A_+ \subset A$, with $x \leq y$, the* ***order interval*** *from x to y is defined as the set*

$$[x, y] = \{z \in A_+ \mid x \leq z \leq y\}\,. \tag{A.11}$$

It is not difficult to see that $[x, y] = (x + A_+) \bigcap (y - A_+)$, i.e., the cone A_+ translated such that $0 \to x$ intersected with the cone $-A_+$ translated such that $0 \to y$.

A.18. *Given a cone $A_+ \subset A$, an* ***order unit****, u_A, for A_+ is an element of A^* which is strictly positive on non-zero elements of A_+, namely $u_A(A_+ \setminus \{0\}) > 0$.*

It is easy to see that u_A is an order unit for A_+ only if it is in the interior of A_+^*.

A.19. *Given a cone $A_+ \subset A$, and a subset $B \subseteq A_+$, let the* ***annihilator*** *of B be defined as*

$$B^\circ = \{y \in A^* \mid y(x) = 0,\ \forall\ x \in B\}. \tag{A.12}$$

Further, let the ***positive annihilator*** *of B be defined as*

$$B^\bullet = B^\circ \cap A_+^* = \{y \in A_+^* \mid y(x) = 0,\ \forall\ x \in B\}. \tag{A.13}$$

A.20. *Given a closed convex set $C \subset A$, an affine subspace $H \subset A$ is called a* ***supporting hyperplane*** *of C if $H \cap C \neq \emptyset$, and C is entirely contained in one of the two closed half-spaces determined by H.*

It is easy to see that if an affine subspace H supports a closed convex C at the set $F := H \cap C$, then F is a face of C.

A.21. *A face F of a convex set $C \subset A$ is* ***exposed*** *if and only if for some $f \in A^*$, $F = f^\circ \cap C$.*

Next we give an important set of properties of faces and exposed faces of cones. These are well known, and simple to prove.

A.22 ([175] Corollary 2.6.11)**.** *Let C and D be convex sets in A such that $C \subseteq \text{relbd}(D)$. Then $\text{face}(C) \subsetneq D$.*

A.23. *The following relations between faces of convex sets hold:*

(a) *If F is a face of a convex set G, and G is a face of a convex set K, then F is a face of K.*
(b) *If F is a face of a convex set K, and G is a face of K such that $F \subset G$, then F is a face of G.*

Finally, we prove a set of basic results on exposed faces and positive annihilators.

A.24. *Given a cone $A_+ \subset A$, and a subset $M \subset A_+$, we have the following:*

(i) $M^{\bullet\bullet\bullet} = M^\bullet$,
(ii) *$M^\bullet$ is an exposed face of A_+^*,*
(iii) *if M is a face, then it is an exposed face if and only if $M = S^\bullet$ for some subset $S \subseteq A_+^*$.*
(iv) *if M is a face, then it is an exposed face if and only if $M = M^{\bullet\bullet}$.*
(v) *$M^{\bullet\bullet}$ is the smallest exposed face of A_+ containing M.*

Proof. (i) First, for any $M_1, M_2 \subseteq A_+$, we have the implication $M_1 \subseteq M_2 \Rightarrow M_1^\bullet \supseteq M_2^\bullet$, and we also have $M \subseteq M^{\bullet\bullet}$ in general. Substituting $M = C^\bullet$ in the last inclusion, we get $C^\bullet \subseteq C^{\bullet\bullet\bullet}$. Next, substituting C for M in the above inclusion, and applying the previous implication, we reverse the containment to get $C^\bullet \supseteq C^{\bullet\bullet\bullet}$, and we are done.

(ii) That $M^\bullet$ is a face should be clear. Further, we can write $M^\bullet = (A_+)^* \cap \bigcap_{x \in M} x^\circ$. In finite dimension, an intersection of a family of hyperplanes will equal the intersection of a finite subfamily. It is sufficient then to take this finite subfamily to be a spanning set

for M. Now, take an $x \in \text{relint}(M) \neq \emptyset$, and note that the set x° is the intersection of the kernels of all elements in a convex decomposition of x, so from above it is sufficient to take an x in $\text{int}(M)$, and we have $B^\bullet = (A_+)^* \cap x^\circ$, and this x exposes $B^\bullet$.

(iii) From (ii) we have that if $M = S^\bullet$, then M is an exposed face. Now suppose that M is an exposed face, namely there exists a $y \in A^*$ such that $y(M) = 0$ and and $y(x) > 0$ for all $x \in A_+ \setminus M$, so $y \in (A_+)^*$, and we can simply take $S = \{x\}$.

(iv) From (ii), if $M = M^{\bullet\bullet}$, then M is exposed. Now suppose that M is an exposed face. From (iii) we have that $M = S^\bullet$ for some subset $S \subseteq (A_+)^*$, which implies that $M^{\bullet\bullet} = S^{\bullet\bullet\bullet} = S^\bullet = M$ (where we have used (i)).

(v) Suppose that there exists an exposed face F containing M, and strictly contained in $M^{\bullet\bullet}$, namely $B \subseteq F \subsetneq M^{\bullet\bullet}$. Taking the positive annihilator of this and using (i), we get $M^\bullet \supseteq F^\bullet \supsetneq M^{\bullet\bullet\bullet} = M^\bullet$, which implies that $M^\bullet = F^\bullet$, as well as $M^{\bullet\bullet} = F^{\bullet\bullet} = F$ since F is an exposed face. $\square$

Appendix B

Alfsen and Shultz formalism

In this appendix we review the framework and several important results of Chapters 7 and 8 of Alfsen and Shultz [7].

Note on conventions: In Alfsen and Shultz the vector space A refers to the dual of the vector space in which the 'state' cone (in their notation $V_+ \subset V$) is embedded, and A_+ is the positive cone in this space. They generally use P for projections on this space, and P^* the dual projection on the space V. The following definitions and properties are not sensitive to which space we start with, but in order to be consistent with our notation and notion of a filter (Definition 4.4), we maintain the usage of A as the finite dimensional vector space containing the state cone, and projections P will be taken to act on A.

B.1 Basic Definitions

B.1. *If P is a positive projection on a cone $S_+(A) \subset A$ (i.e., $P : A \to A$ is such that $P^2 = P$, and $P(S_+(A)) \subset S_+(A)$), define the* ***kernel***, $\ker(P)$, *and the* ***positive kernel***, $\ker^+(P)$, *of P by*

$$\ker(P) = \{x \in A \mid Px = 0\}, \quad \ker^+(P) = S_+(A) \cap \ker(P). \tag{B.1}$$

Further, define the ***image***, $\mathrm{im}(P)$, *and the* ***positive image***, $\mathrm{im}^+(P)$, *of P by*

$$\mathrm{im}(P) = \{x \in A \mid Px = x\}, \quad \mathrm{im}^+(P) = S_+(A) \cap \mathrm{im}(P). \tag{B.2}$$

Because A is positively generated by $S_+(A)$, i.e., $A = S_+(A) - S_+(A)$, if P is a positive projection, then $\ker^+(P)$ is a face of $S_+(A)$, and $\mathrm{im}(P)$ is a positively generated linear subspace of A (see Lemma 4.6).

B.2. *If P is a positive projection on a cone $S_+(A) \subset A$, then the **dual projection**, P^* acting on A^* is defined by*

$$(P^*y)(x) = (y \circ P)(x) = y(Px), \quad \text{where } x \in A,\ y \in A^*. \tag{B.3}$$

B.3 ([7] Definition 7.1)**.** *Given a subset F of a convex set $C \subset A$, then the **tangent space** of C at F, denoted by $\mathrm{Tan}(F)$ is defined as the intersection of all supporting hyperplanes of C which contain F.*

B.4 ([7] Definition 7.5)**.** *A positive projection P on a cone $S_+(A) \subset A$ is **smooth** if $\ker(P)$ is equal to the tangent space of $S_+(A)$ at $\ker^+(P)$, i.e., if*

$$\mathrm{Tan}(\ker^+(P)) = \ker(P). \tag{B.4}$$

B.5. *Two positive projections P and Q on a cone $S_+(A) \subset A$ are **complementary** if*

$$\ker^+(Q) = \mathrm{im}^+(P), \quad \text{and} \quad \ker^+(P) = \mathrm{im}^+(Q). \tag{B.5}$$

B.6. *A positive projection P on a cone $S_+(A) \subset A$ is said to be **bicomplemented** if there exits a positive projection Q on A such that P, Q are complementary and P^*, Q^* are also complementary.*

B.7. *Given a state space $(A, S_+(A), u_A)$, a positive projection P on $S_+(A) \subset A$ is said to be **normalized** if and only if $P^*u_A \leq u_A$.*

B.8 ([7] Definition 7.19)**.** *A normalized positive projection P on state space $(A, S_+(A), u_A)$ is **neutral** if for all $\omega \in S_+(A)$, we have the implication*

$$u_A(P\omega) = u_A(\omega) \quad \Rightarrow \quad P\omega = \omega. \tag{B.6}$$

B.9 ([7] Definition 7.22)**.** *A **compression** is a bicomplemented, normalized, positive projection.*

In the language of this Appendix, the projective state space Assumption 4.4 can be phrased as follows:

B.10 (**Projective state space**, [7] Chapter 8)**.** *A state space $(A, S_+(A), u_A)$ is said to be a* projective state space *(alternatively, satisfies the Alfsen and Shultz standing hypothesis) if each exposed face F of $S_+(A)$ is such that $F = S_+(A) \cap \mathrm{im}(P_F)$ for a compression P_F.*

B.2 Useful Results

In this section we review several important results used throughout.

B.11 ([7] Proposition 7.7)**.** *If P is a positive projections on a cone $S_+(A) \subset A$, then*

$$(\mathrm{im}^+(P))^\bullet = \ker^+(P^*). \tag{B.7}$$

B.12 ([7] Theorem 7.9)**.** *If P is a positive projection on a cone $S_+(A) \subset A$, then if P is complemented, the dual projection P^* is smooth. Further, if P has a smooth complement Q, then Q is the unique complement of P.*

B.13 ([7] Theorem 7.10)**.** *If P and Q are positive projections on a cone $S_+(A) \subset A$, then the following are equivalent:*

(i) *P, Q are complementary smooth projections,*
(ii) *P^*, Q^* are complementary smooth projections,*
(iii) *P, Q are bicomplementary projections.*

B.14 ([7] Proposition 7.21)**.** *Let P and Q be two normalized positive projections on a cone $S_+(A) \subset A$, then P^* and Q^* are bicomplementary if and only if they are complementary and P, Q are both neutral.*

B.15. *A **lattice** is a set $\mathcal{L}$ with a partial order, $\leq$, (a reflexive, antisymmetric, and transitive binary relation on $\mathcal{L}$), such that*

(i) *for any $l, m \in \mathcal{L}$ there exists a* least upper bound *$l \vee m$,*
(ii) *for any $l, m \in \mathcal{L}$ there exists a* greatest lower bound *$l \wedge m$.*

B.16. *A lattice $\mathcal{L}$ with least element 0 and greatest element 1 is **orthocomplemented** if there is a map $l \mapsto l'$, called the **orthocomplementation**, that satisfies*

(i) *$l'' = l$,*
(ii) *$l \leq m$ implies $l' \geq m'$,*
(iii) *$l \vee l' = 1$ and $l \wedge l' = 0$.*

*Further, the lattice is called **orthomodular** if in addition to (i)-(iii) it satisfies the following implication:*

$$\text{if} \quad l \leq m, \quad \text{then} \quad m = l \vee (m \wedge l'). \tag{B.8}$$

B.17 ([7] Theorem 8.17)**.** *Let $(A, S_+(A), u_A)$ be a projective state space, and let P_0 be a compression with induced effect $p_0 = P_0^* u_A$, and associated exposed face $F_0 = \mathrm{im}^+(P_0)$. Then the projected state space $(P_0(A), P_0(S_+(A)), P_0^*(E_+(A)), p_0)$ will also be a projective*

state space. For this state space the compressions are the compressions P on $S_+(A)$ such that $\mathrm{im}(P) \subseteq \mathrm{im}(P_0)$ *(restricted to $P_0(A)$), the induced effects are those effects p on $S_+(A)$ such that $p \leq p_0$, and the associated faces are those faces F of $S_+(A)$ such that $F \subset F_0$. For this state space the complementary compressions of P is P' (restricted to $P_0(A)$).*

B.18. *Let $(A, S_+(A), u_A)$ be a projective state space, and let E and F be exposed faces of $S_+(A)$, with associated compressions P_E, P_F such that $P_E P_F = P_F P_E$. Then*

$$\mathrm{lin}(E \wedge F) = \mathrm{lin}(E) \cap \mathrm{lin}(F). \tag{B.9}$$

Proof. Since $E \wedge F \equiv E \cap F \subseteq \mathrm{lin}(E) \cap \mathrm{lin}(F)$, it is clear that $\mathrm{lin}(E \wedge F) \subseteq \mathrm{lin}(E) \cap \mathrm{lin}(F)$. For the other direction, consider the associated compressions P_E with $\mathrm{im}_+(P_E) = E$, and P_F with $\mathrm{im}_+(P_F) = F$. By Theorem 4.14, if P_E and P_F commute, then $P_E P_F = P_F P_E = P_E \wedge P_F$. Consider $x \in \mathrm{im}(P_E) \cap \mathrm{im}(P_F)$, or equivalently, $P_E x = x$ and $P_F x = x$. Therefore $P_E P_F x = P_F P_E x = (P_E \wedge P_F)x = x$, so $x \in \mathrm{im}(P_E \wedge P_F)$. Next from the isomorphism between compressions and faces (Proposition 4.11), and from the fact that images of compressions are positively generated, we have $\mathrm{im}(P_E \wedge P_F) = \mathrm{lin}(E \wedge F)$, which proves the claim. □

B.19 ([7] Proposition 8.7)**.** *Let $(A, S_+(A), u_A)$ be a projective state space. Further let $P_1, \ldots, P_n$ be compressions which satisfy $P_i P_j = P_j P_i = \delta_{ij} P_i$, and $a \in A^*$ be such that $a = a \circ (P_i + P_i')$ for all $i = 1, \ldots, n$. Then $a = a \circ \bigvee_i P_i + a \circ (\bigvee_i P_i)'$, and*

$$a \circ \bigvee_i P_i = a \circ \sum_i P_i. \tag{B.10}$$

B.20 ([7] Proposition 8.8)**.** *Let $(A, S_+(A), u_A)$ be a projective state space, and let $P_1, \ldots, P_n$ be compressions with induced effects $p_i = u_A \circ P_i$. Then $P_i P_j = P_j P_i = 0$ for $i \neq j$ if and only if $\bigvee_i p_i = \sum_i p_i$. Further, $\bigvee_i P_i = P$ for a compression P if and only if $\sum_i p_i = p$ for $p = u_A \circ P$.*

B.21. *Let $(A, S_+(A), u_A)$ be a projective state space. For all compressions P on $S_+(A)$ we have $u_A = u_A \circ (P + P')$. Further, if the compressions P, Q satisfy $PQ = QP = 0$, then $p = u_A \circ P = p \circ (P + Q)$.*

Proof. First, because P is normalization non-increasing, we have $u_A \circ P \leq u_A$, which implies that $u_A - u_A \circ P \in \ker^+(P^*) = \mathrm{im}^+(P^{*'})$. Then $(u_A - u_A \circ P) \circ P' = u_A - u_A \circ P$, and because $PP' = 0$, we have $u_A - u_A \circ P = u_A \circ P'$.

Next, notice that $p \circ Q = u_A \circ PQ = 0 \leq p$. This implies that $p - p \circ Q \in \ker^+(Q^*) = \mathrm{im}^+(Q^{*'})$, and by a similar argument to the above, it follows that $p = p \circ (Q + Q')$. □

Appendix C

Proof of Lemma 4.28

In this Appendix we prove Lemma 4.28, which states that $P^{(N)}$ is a (not necessarily positive) projection with image $\mathrm{im}\left(P^{(N)}\right) = \mathrm{lin}\left(\bigcup_{|J|=N-1} F_J\right)$.

Proof. That $\mathrm{im}\left(P^{(N)}\right) \subseteq \mathrm{lin}\left(\bigcup_{|J|=N-1} F_J\right)$ is immediate from the fact that $J \subseteq K$ implies $\mathrm{im}(P_J) \subseteq \mathrm{im}(P_K)$ and the definition of $P^{(N)}$ as a linear combination of P_J's with $|J| \leq N-1$. For the rest of the statement, it suffices to show that $P^{(N)}P_K = P_K$ for all K with $|K| \leq N-1$. Using the fact that $P_J P_K = P_{J\cap K}$, we have

$$P^{(N)}P_K = \sum_{l=1}^{N-1}(-1)^{l-1}\sum_{|J|=N-l} P_{J\cap K}. \tag{C.1}$$

This is a signed sum of projectors P_L onto subsets of K. We will compute the total weight, a signed integer which we'll call $\mathrm{wt}_K(L)$, with which each projector P_L occurs this sum.

P_L is given by an alternating signed sum of terms, the l-th of which is the number of ways of obtaining $L = K \cap R$ where R is an $(N-l)$-subset of I. The upper limit on l in this sum can be $N - |L|$, as this gives us $|L|$-subsets and smaller subsets R cannot have $K \cap R = L$. At fixed l, there are $N - |K|$ elements not already in K (and hence L), which can be used as the $N - l - |L|$ elements of the $(N-l)$-set R that are not already in L. Hence we get:

$$\mathrm{wt}_K(L) = \sum_{l=|K|-|L|}^{N-|L|}(-1)^{l+1}\binom{N-|K|}{N-l-|L|}, \tag{C.2}$$

The lower limit, $l \geq |K| - |L|$, enforces $N - l - |L| \leq N - |K|$. With $m := N - |L|$, and $r := |K| - |L|$, we have:

$$\mathrm{wt}_K(L) = \sum_{l=r}^{m} (-1)^{l+1} \binom{m-r}{m-l} \tag{C.3}$$

Changing summation variables to $k = l - r$,

$$\mathrm{wt}_K(L) = \sum_{k=0}^{m-r} (-1)^{k+r-1} \binom{m-r}{m-r-k} \equiv (-1)^{r-1} \sum_{k=0}^{m-r} (-1)^k \binom{m-r}{k} \tag{C.4}$$

Next, an elementary identity for binomial coefficients,

$$\sum_{k=0}^{p} (-1)^k \binom{p}{k} = 0, \tag{C.5}$$

with $p = m - r$, shows that (C.4) is zero. The lower bound of $|K| - |L|$ on l in the sum is correct whenever $|K| - |L| \geq 1$. When $|K| = |L|$, which is possible only when $L = K$, the constraint $l \geq 1$ is the relevant one, and we lose the $k = 0$ term in (C.4). Since the $k = 0$ term is -1, the remaining terms add to 1 by (C.5). Using $P^{(N)} = \sum_{L \subseteq K} \mathrm{wt}_K(L) P_L$, $\mathrm{wt}_K(K) = 1$, and $\mathrm{wt}_K(L) = 0$ for all $L \subset K$, we have that $P^{(N)} P_K = P_K$.

Finally, (C.1) and the definition of $P^{(N)}$ gives $(P^{(N)})^2 = P^{(N)}$, so $P^{(N)}$ is a projector. Since any vector $v \in \mathrm{lin}(\bigcup_{|J|=N-1} \mathrm{im}(P_J))$ can be written as $v = \sum_{|J|=N-1} P_J v_J$, (C.1) gives $P^{(N)} v = v$ for all such v. Therefore the image of $P^{(N)}$ is precisely $\mathrm{lin}(\bigcup_{|J|=N-1} \mathrm{im}(P_J))$ as claimed. □

Appendix D

Equilibration

D.1 Expectation over eigenvectors

In this section we carry out the first step to proving Theorem 6.3. Recall that C is the unitary which transforms the eigenbasis of the GUE Hamiltonian H to that of the observable A, and that we can write

$$U(t) = CF(t)C^\dagger, \tag{D.1}$$

where $F(t) = \text{diag}[e^{-itE_a}]$, and $\{E_a\}_{a=1}^D$ are the energy eigenvalues of H. Further, we can write $\Pr(k|x,t)$ more explicitly as

$$\Pr(k|x, C, F(t)) = \text{Tr}\left[|k\rangle\langle k|CF(t)C^\dagger|x\rangle\langle x|CF(t)^\dagger C^\dagger\right]. \tag{D.2}$$

Finally, recall that the joint probability distribution over eigenvectors of H is the same as that over the change of basis matrix C, namely the Haar measure on the unitary group $\mathbb{U}(D)$ (see [93] Chapter 4, or [128]). In other words, the eigenvectors of H (making up the columns of C) are columns of a typical Haar random unitary.

Lemma D.1. *Take a non-degenerate observable A acting on $\mathcal{H} = \mathbb{C}^D$, an initial pure state $\rho_0 = |x\rangle\langle x|$ which is an eigenstate of A, and a unitary $U(t) = e^{-itH}$ where H is drawn uniformly at random from GUE. Then the expectation of the measurement outcome probabilities over the Haar measure on the unitary group $\mathbb{U}(D)$ of change of basis matrices*

C is given by:

$$\begin{aligned}\mathbb{E}_C\left\{\Pr\left(k|x,C,F(t)\right)\right\} &= \int_{\mathbb{U}(D)} \mathrm{Tr}\left[|k\rangle\langle k|CF(t)C^\dagger|x\rangle\langle x|CF(t)^\dagger C^\dagger\right]\mu_H(dC)\\ &= \frac{D-\frac{1}{D}|\mu(t)|^2+\delta_{xk}\left(|\mu(t)|^2-1\right)}{D^2-1},\end{aligned} \tag{D.3}$$

where we have defined $\mu(t) := \mathrm{Tr}[U(t)] = \mathrm{Tr}[F(t)]$ (this is often called the spectral form factor [93, 128]).

Proof. The ensemble average $\mathbb{E}_C\left\{\Pr\left(k|x,C,F(t)\right)\right\}$ can be calculated in a similar but simpler fashion as the ensemble variance in the next lemma, so we leave the proof of the above as an exercise. □

Lemma D.2. *Take a non-degenerate observable A acting on $\mathcal{H} = \mathbb{C}^D$, an initial pure state $\rho_0 = |x\rangle\langle x|$ which is an eigenstate of A, and a unitary $U(t) = e^{-itH}$ where H is drawn uniformly at random from GUE. Then the variance of the measurement outcome probabilities over the Haar measure on the unitary group $\mathbb{U}(D)$ of change of basis matrices C is given by:*

$$\begin{aligned}\mathbb{V}_C\Big\{\Pr\left(k|x,C,F(t)\right)\Big\} &= \frac{1}{D^4}\Big\{D^2-2D+4+(7-2D)\delta_{xk}+|\mu(t)|^2(2D\delta_{xk}-10\delta_{xk}-2)\\ &\quad +\delta_{xk}|\mu(2t)|^2+2\delta_{xk}Re[\mu(t)^2\mu(-2t)]\Big\}+O(D^{-5}).\end{aligned} \tag{D.4}$$

Proof. First, note that the squares of the outcome probabilities $\Pr\left(k|x,C,F(t)\right)$ can be written in the form:

$$\begin{aligned}&\mathrm{Tr}\left[|k\rangle\langle k|CF(t)C^\dagger|x\rangle\langle x|CF(t)^\dagger C^\dagger\right]^2 \qquad\qquad (D.5)\\ =\ &\sum_{s,s'}\langle s|k\rangle\langle k|CF(t)C^\dagger|x\rangle\langle x|CF(t)^\dagger C^\dagger|s\rangle\langle s'|k\rangle\langle k|CF(t)C^\dagger|x\rangle\langle x|CF(t)^\dagger C^\dagger|s'\rangle.\end{aligned}$$

Expanding each $F(t) = \sum_b e^{-itE_b}|b\rangle\langle b|$, and using $\langle i|\bar{C}^T|j\rangle = \langle j|\bar{C}|i\rangle$, the above expression becomes

$$\begin{aligned}(D.5) &= \sum_{b,b',d,d'} e^{it(E_b-E_{b'}+E_d-E_{d'})}\langle k,x,k,x,x,k,x,k|C^{\otimes 4}\otimes\bar{C}^{\otimes 4}|b,b',d,d',b,b',d,d'\rangle\\ &= \langle L_4|C^{\otimes 4}\otimes\bar{C}^{\otimes 4}|R_4(t)\rangle,\end{aligned} \tag{D.6}$$

where we have defined

$$\langle L_4| := \langle k,x,k,x,x,k,x,k|, \tag{D.7}$$

$$|R_4(t)\rangle := \sum_{b,b',d,d'} e^{it(E_b-E_{b'}+E_d-E_{d'})}|b,b',d,d',b,b',d,d'\rangle, \tag{D.8}$$

The expectation of the expression $C^{\otimes 4}\otimes \bar{C}^{\otimes 4}$ over Haar measure can be written as the projector onto the subspace spanned by the vectors

$$|\Phi_\pi\rangle = V_\pi \otimes \mathbf{1}|\phi\rangle_{1,5}|\phi\rangle_{2,6}|\phi\rangle_{3,7}|\phi\rangle_{4,8}, \tag{D.9}$$

where $|\phi\rangle_{ij} = \sum_{a=1}^{D}|a\rangle_i|a\rangle_j$, and the index π runs over the 4! permutations of the elements $\{1,2,3,4\}$, and V_π is the unitary permutting the first four factor spaces according to π. From [40] Fact 22 and Lemma 23, this projector is given by

$$\mathbb{E}_C\{C^{\otimes 4}\otimes\bar{C}^{\otimes 4}\} = \sum_{\pi,\sigma}|\Phi_\pi\rangle(M^{-1})_{\pi,\sigma}\langle\Phi_\sigma|, \tag{D.10}$$

where the matrix M has components $M_{\pi,\sigma} = \langle V_\pi|V_\sigma\rangle = \mathrm{Tr}\,[V_{\pi^{-1}}V_\sigma] = d^{l(\pi^{-1}\sigma)}$, and $l(\sigma)$ is number of cycles in the cycle decomposition of the permutation $\pi^{-1}\sigma$. The inverse of M can be found in [40] Theorem 20, or simply by using a symbolic mathematics package.

Putting this all together, we have

$$\mathbb{E}_C\Big\{\Pr\left(k|x,C,F(t)\right)^2\Big\} = \sum_{\pi,\sigma}\langle L_4|\Phi_\pi\rangle(M^{-1})_{\pi,\sigma}\langle\Phi_\sigma|R_4(t)\rangle, \tag{D.11}$$

where

$$\langle L_4|\Phi_\pi\rangle = \langle k,x,k,x|V_\pi|x,k,x,k\rangle \tag{D.12}$$

$$\langle\Phi_\sigma|R_4(t)\rangle = \sum_{b,b',d,d'} e^{it(E_b-E_{b'}+E_d-E_{d'})}\langle b,b',d,d'|V_\sigma|b,b',d,d'\rangle. \tag{D.13}$$

Recalling that $\mu(t) = \mathrm{Tr}[F(t)] = \sum_b e^{itE_b}$, the inner products $\langle\Phi_\sigma|R_4(t)\rangle$ can be found explicitly to be:

$$\langle\Phi_{(1,2,3,4)}|R_4(t)\rangle = |\mu(t)|^4, \tag{D.14}$$

$$\langle\Phi_{(1,2,4,3)}|R_4(t)\rangle = \langle\Phi_{(1,3,2,4)}|R_4(t)\rangle = \langle\Phi_{(2,1,3,4)}|R_4(t)\rangle = \langle\Phi_{(4,2,3,1)}|R_4(t)\rangle = d|\mu(t)|^2, \tag{D.15}$$

$$\langle\Phi_{(1,3,4,2)}|R_4(t)\rangle = \langle\Phi_{(1,4,2,3)}|R_4(t)\rangle = \langle\Phi_{(2,3,1,4)}|R_4(t)\rangle = \langle\Phi_{(2,4,3,1)}|R_4(t)\rangle = \tag{D.16}$$

$$\langle\Phi_{(3,1,2,4)}|R_4(t)\rangle = \langle\Phi_{(3,2,4,1)}|R_4(t)\rangle = \langle\Phi_{(4,1,3,2)}|R_4(t)\rangle = \langle\Phi_{(4,2,1,3)}|R_4(t)\rangle = |\mu(t)|^2,$$

$$\langle \Phi_{(3,4,1,2)} | R_4(t) \rangle = |\mu(2t)|^2, \tag{D.17}$$

$$\langle \Phi_{(1,4,3,2)} | R_4(t) \rangle = \mu(t)^2 \bar{\mu}(2t), \tag{D.18}$$

$$\langle \Phi_{(3,2,1,4)} | R_4(t) \rangle = \mu(2t) \bar{\mu}(t)^2, \tag{D.19}$$

$$\langle \Phi_{(2,1,4,3)} | R_4(t) \rangle = \langle \Phi_{(4,3,2,1)} | R_4(t) \rangle = d^2, \tag{D.20}$$

$$\begin{aligned} \langle \Phi_{(2,3,4,1)} | R_4(t) \rangle &= \langle \Phi_{(3,4,2,1)} | R_4(t) \rangle = \langle \Phi_{(2,4,1,3)} | R_4(t) \rangle = \\ \langle \Phi_{(3,1,4,2)} | R_4(t) \rangle &= \langle \Phi_{(4,1,2,3)} | R_4(t) \rangle = \langle \Phi_{(4,3,1,2)} | R_4(t) \rangle = d, \end{aligned} \tag{D.21}$$

Further it is easy to see that

$$\langle L_4 | \Phi_{(4,3,2,1)} \rangle = \langle L_4 | \Phi_{(4,1,2,3)} \rangle = \langle L_4 | \Phi_{(2,3,4,1)} \rangle = \langle L_4 | \Phi_{(2,1,4,3)} \rangle = 1. \tag{D.22}$$

and that for all other π, $\langle L_4 | \Phi_\pi \rangle = \delta_{ik}$.

Next, taking the sum with M^{-1} as in Equation (D.11), we have

$$\begin{aligned} \mathbb{E}_C \left\{ \Pr\left(k|x, C, F(t)\right)^2 \right\} = \frac{1}{\alpha} \Big\{ & |\mu(t)|^4 \left((D^2 - D - 2)\delta_{xk} + 2 \right) \\ & + |\mu(t)|^2 \left(-4D^2 - 12D - 8 + (4D^3 + 8D^2 + 4D + 8)\delta_{xk} \right) \\ & + (D^2 - D - 2)\delta_{xk} \left(|\mu(2t)|^2 + \mu(2t)^* \mu(t)^2 + \mu(2t) \left(\mu(t)^*\right)^2 \right) \\ & + 2D^4 + 8D^3 + 6D^2 - (4D^3 + 12D^2)\delta_{xk} \\ & + 2|\mu(2t)|^2 + 2\mu(2t)^* \mu(t)^2 + 2\mu(2t) \left(\mu(t)^*\right)^2 \Big\}, \end{aligned} \tag{D.23}$$

where $\alpha = D^2(D-1)(D+1)(D+2)(D+3)$. Finally, recalling $\mathbb{E}_C \left\{ \Pr\left(k|x, C, F(t)\right) \right\}$ from Lemma D.1, we take $\mathbb{E}_C \{ \Pr\left(k|x, C, F(t)\right)^2 \} - \mathbb{E}_C \{ \Pr\left(k|x, C, F(t)\right) \}^2$, and then approximate to order $O(D^{-5})$ to find:

$$\begin{aligned} \mathbb{V}_C \{ \Pr\left(k|x, C, F(t)\right) \} = \frac{1}{D^4} \Big\{ & D^2 - 2D + 4 + (7 - 2D)\delta_{xk} + |\mu(t)|^2 (2D\delta_{xk} - 10\delta_{xk} - 2) \\ & + \delta_{xk} |\mu(2t)|^2 + 2\delta_{xk} Re[\mu(t)^2 \mu(-2t)] \Big\} + O(D^{-5}). \end{aligned} \tag{D.24}$$

□

Lemma D.3. *Take a non-degenerate observable A acting on $\mathcal{H} = \mathbb{C}^D$, an initial pure state $\rho_0 = |x\rangle\langle x|$ which is an eigenstate of A, and a unitary $U(t) = e^{-itH}$ where H is drawn uniformly at random from GUE. Then the fourth moment of the measurement outcome*

probabilities over the Haar measure on the unitary group $\mathbb{U}(D)$ of change of basis matrices C is given by:

$$\mathbb{E}_C \left\{ \Pr\left(k|x, C, F(t)\right)^4 \right\} \leq \frac{1}{D^8} \sum_{\pi} \langle \Phi_\pi | R_8(t) \rangle + \sum_{\pi \neq \sigma} B(D^{-9})_{\pi,\sigma} \langle \Phi_\sigma | R_8(t) \rangle, \tag{D.25}$$

where $B(D^{-9})$ is a matrix with each component being $\leq (D^{-9})$.

Proof. First, note that analogously to the second power, the fourth power of the outcome probabilities can be written in the form:

$$\mathrm{Tr}\left[|k\rangle\langle k| CF(t)C^\dagger |x\rangle\langle x| CF(t)^\dagger C^\dagger \right]^4 = \langle L_8 | C^{\otimes 8} \otimes \bar{C}^{\otimes 8} | R_8(t) \rangle, \tag{D.26}$$

where $\langle L_8|$ and $|R_8(t)\rangle$ are defined analogously to (D.7), but with twice the number of tensor factors. Further, the average of the expression $C^{\otimes 8} \otimes \bar{C}^{\otimes 8}$ over Haar measure can be written as the projector onto the subspace spanned by the vectors

$$|\Phi_\pi\rangle = V_\pi \otimes \mathbf{1} |\phi\rangle_{1,5} |\phi\rangle_{2,6}, \ldots, |\phi\rangle_{8,16}, \tag{D.27}$$

where $|\phi\rangle_{ij} = \sum_{a=1}^{D} |a\rangle_i |a\rangle_j$, and the index π runs over the 8! permutations of the elements $\{1, 2, \ldots, 8\}$, and V_π is the unitary permutting the first eight factor spaces according to π. Similarly to the proof of Lemma D.2, this projector is given by

$$\mathbb{E}_C \{ C^{\otimes 8} \otimes \bar{C}^{\otimes 8} \} = \sum_{\pi,\sigma} |\Phi_\pi\rangle (M^{-1})_{\pi,\sigma} \langle \Phi_\sigma |. \tag{D.28}$$

We will determine the order of the expression

$$\mathbb{E}_C \Big\{ \Pr\left(k|x, C, F(t)\right)^4 \Big\} = \sum_{\pi,\sigma} \langle L_8 | \Phi_\pi \rangle (M^{-1})_{\pi,\sigma} \langle \Phi_\sigma | R_8(t) \rangle, \tag{D.29}$$

rather than its precise form, by finding an approximation for M^{-1} good up to some order in D. Recall that the matrix M has components $M_{\pi,\sigma} = \langle V_\pi | V_\sigma \rangle = Tr[V_{\pi^{-1}} V_\sigma] = D^{l(\pi^{-1}\sigma)}$, where $l(\sigma)$ is number of cycles in the cycle decomposition of the permutation $\pi^{-1}\sigma$. It is clear that the diagonal components of M are all equal to $M_{\pi,\pi} = Tr[V_{\pi^{-1}} V_\pi] = D^{l(\mathbf{1})} = D^8$, where $\mathbf{1}$ is the identity permutation. All other components of M are strictly $< D^8$, as the identity permutation is the only one with 8 cycles.

Define

$$A := \mathbb{I} - \frac{M}{D^8}. \tag{D.30}$$

From the above observations we have that A has all zeros on the diagonal, and all off diagonal terms are $< 1/D$. It is clear that

$$(\mathbb{I} - A)\left(\sum_{k=0}^{N} A^k\right) = \left(\sum_{k=0}^{N} A^k\right)(\mathbb{I} - A) = \mathbb{I} - A^{N+1}, \tag{D.31}$$

and that this product converges to $\mathbb{I}$ in the limit $N \to \infty$ if and only if $||A||_{\text{op}} < 1$. Now, because A is of fixed size $8! \times 8!$ and all its elements are $< 1/D$, we have $||A||_{\text{op}} < 1$ for D sufficiently large. Therefore,

$$(\mathbb{I} - A)^{-1} = \mathbb{I} + A + O(D^{-2}), \tag{D.32}$$

where by $O(D^{-2})$ we mean that the inverse is $\mathbb{I} + A$ up to a matrix with each of its components $\leq D^{-2}$. This implies that

$$M^{-1} = \frac{\mathbb{I}}{D^8} + \frac{A}{D^8} + O(D^{-10}). \tag{D.33}$$

Again, because each element of A is $< 1/D$, we can simply take

$$M^{-1} = \frac{\mathbb{I}}{D^8} + O(D^{-9}). \tag{D.34}$$

Going back to the sum $\sum_{\pi,\sigma}\langle L_8|\Phi_\pi\rangle(M^{-1})_{\pi,\sigma}\langle\Phi_\sigma|R_8(t)\rangle$, and recalling the analogues of the vectors $\langle L_8|\Phi_\pi\rangle$ and $\langle\Phi_\sigma|R_8(t)\rangle$ from the variance calculation, we see that all components of $\langle L_8|\Phi_\pi\rangle$ are equal to 1 or δ_{xk} as well.

Therefore, we can split the sum into a diagonal and non-diagonal part as

$$\begin{aligned}\mathbb{E}_C\Big\{\Pr\left(k|x,C,F(t)\right)^4\Big\} &= \frac{1}{D^8}\sum_{\pi}\langle L_8|\Phi_\pi\rangle\langle\Phi_\pi|R_8(t)\rangle + \sum_{\pi\neq\sigma}\langle L_8|\Phi_\pi\rangle O(D^{-9})_{\pi,\sigma}\langle\Phi_\sigma|R_8(t)\rangle \\ &\leq \frac{1}{D^8}\sum_{\pi}\langle\Phi_\pi|R_8(t)\rangle + \sum_{\pi\neq\sigma} O(D^{-9})_{\pi,\sigma}\langle\Phi_\sigma|R_8(t)\rangle, \end{aligned} \tag{D.35}$$

where we have used that $\langle L_8|\Phi_\pi\rangle \leq 1$ for all π. □

D.2 Expectations over GUE eigenvalues

In this section we focus on the spectrum dependent functions appearing in Lemma D.2. In particular, we must find the GUE expectation of these functions. First however, a few more details on this ensemble.

The elements of a GUE matrix H are such that the diagonal elements h_{aa} are real valued random variables with distribution $\mathcal{N}(0,\sigma^2)$, and the off-diagonal elements h_{ab} have real and imaginary parts that are independent and identically distributed random variables with distribution $\mathcal{N}(0,\frac{1}{2}\sigma^2)$. More explicitly, for any $a \neq b$, we have probability densities (w.r.t. to Lebesgue measure on $\mathbb{R}$)

$$\begin{aligned} P(h_{aa}) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{h_{aa}^2}{2\sigma^2}}, \\ P(h_{ab}) &= \frac{1}{\pi\sigma^2} e^{-\frac{|h_{ab}|^2}{\sigma^2}}. \end{aligned} \tag{D.36}$$

The joint probability distribution over H is then given by

$$P(H) = 2^{-D/2}(\pi\sigma^2)^{-D^2/2} \exp\Big[-\frac{1}{2\sigma^2}\operatorname{Tr}[H^2]\Big]. \tag{D.37}$$

The joint distribution over the un-ordered energy eigenvalues $\{E_a\}_{a=1}^D$ of H (see [128] Theorem 3.3.1, [92] Theorem 1.4, or [93] Chapter 4), is given by:

$$P(\{E_a\}^D) = \frac{1}{C_D}\prod_{1=a<b}^{D}(E_a - E_b)^2 \exp\Big[-\frac{1}{2\sigma^2}\sum_{a=1}^{D}E_a^2\Big], \tag{D.38}$$

where C_D is the normalization constant and is given by

$$C_D = (2\pi)^{D/2}\sigma^{D^2}\prod_{j=1}^{D} j!. \tag{D.39}$$

Next we define the the m-point correlation function ([128] 6.1.1)

$$R_m(E_1,\ldots,E_m) := \frac{D!}{(N-m)!}\int_{-\infty}^{\infty} P(\{E_a\}^D)dx_{m+1},\ldots,dx_D, \tag{D.40}$$

which is the probability density of finding a level (regardless of labeling) around each of the energies $E_1,\ldots,E_m$, with the positions of the remaining levels being unobserved. Further, let

$$K_D(E_1,E_2) := \sum_{k=0}^{D-1}\phi_k(E_1)\phi_k(E_2), \tag{D.41}$$

where the $\phi_k(x)$ are orthogonal monic polynomials of order k. We will give these functions in more detail below (see [128] section 6.2).

By [128] Theorem 5.1.4 (see also sections 5.3 and 6.2), integration of $P(\{E_a\}^D)$ over $D-m$ variables gives the m-point correlation function:

$$R_m(E_1,\ldots,E_m) = \det[K_D(E_i,E_j)]_{i,j=1,\ldots,m}. \tag{D.42}$$

For example, for $m=1$ we have

$$R_1(E) = K_D(E,E) = \sum_{k=0}^{D-1} \phi_k^2(E), \tag{D.43}$$

and for $m=2$,

$$\begin{aligned} R_2(E_1,E_2) &= K_D(E_1,E_1)K_D(E_2,E_2) - K_D(E_1,E_2)K_D(E_2,E_1) \\ &= \left(\sum_{k=0}^{D-1}\phi_k^2(E_1)\right)^2 - \left(\sum_{k=0}^{D-1}\phi_k(E_1)\phi_k(E_2)\right)^2. \end{aligned} \tag{D.44}$$

D.2.1 Calculation of $\mathbb{E}_{spec}\{|\mu(t)|^2\}$

In this section we calculate the expression $\Gamma(t,D) := \mathbb{E}_{spec}\{|\mu(t)|^2\} = \mathbb{E}_{spec}\left\{\left|Tr\left[e^{-itH}\right]\right|^2\right\}$. This expectation is needed for the second step of finding the full GUE variance of the outcome probabilities, as the expression $|\mu(t)|^2$ appears in the eigenvector expectation calculated in the first step in Lemma D.2. As discussed before the statement of Theorem 6.3, we take $\sigma^2 = 1/2$.

Expanding out the trace and grouping terms with the same index, we have

$$\begin{aligned} \Gamma(t,D) &= \mathbb{E}_{GUE}\left\{\left|Tr\left[e^{-itH}\right]\right|^2\right\} \\ &= \int_{-\infty}^{\infty}\left|\sum_{l=1}^{D} e^{itE_l}\right|^2 P(\{E_a\}^D)\prod_{j=1}^{D} dE_j \\ &= \int_{-\infty}^{\infty}\left(\sum_{l=1}^{D} e^{it(E_l-E_l)} + \sum_{l\neq m}^{D} e^{it(E_l-E_m)}\right) P(\{E_a\}^D)\prod_{j=1}^{D} dE_j. \end{aligned} \tag{D.45}$$

This can further be simplified to

$$\Gamma(t,D) = D\int_{-\infty}^{\infty} P(\{E_a\}^D)\prod_{j=1}^{D} dE_j + \int_{-\infty}^{\infty}\sum_{l\neq m}^{D} e^{it(E_l-E_m)} P(\{E_a\}^D)\prod_{j=1}^{D} dE_j. \tag{D.46}$$

The joint distribution over the un-ordered energy eigenvalues, $P(\{E_a\}^D)$ as given in Equation (D.38) is clearly invariant under permutations (i.e., relabeling) of the energies. Using this fact in the second term of (D.46), we can integrate out all but two energies to find the 2-point correlation function:

$$\begin{aligned}\Gamma(t,D) &= D + D(D-1)\int_{-\infty}^{\infty} e^{it(E_2-E_1)} P(E_1,\ldots,E_D)\prod_{j=1}^{D} dE_j \\ &= D + \int_{-\infty}^{\infty} e^{it(E_2-E_1)} R_2(E_1,E_2) dE_1 dE_2. \end{aligned} \tag{D.47}$$

Using Equation (D.44) for the 2-point correlation function, and switching to variables x, y, we have

$$\begin{aligned}\Gamma(t,D) &= D + \int_{-\infty}^{\infty} e^{it(x-y)} \left[\left(\sum_{k=0}^{D-1} \phi_k^2(x)\right)\left(\sum_{k=0}^{D-1} \phi_k^2(y)\right) - \left(\sum_{k=0}^{D-1} \phi_k(x)\phi_k(y)\right)^2 \right] dxdy \\ &= D + \left| \int_{-\infty}^{\infty} e^{itx} \sum_{k=0}^{D-1} \phi_k^2(x) dx \right|^2 - \sum_{k,l=0}^{D-1} \left| \int_{-\infty}^{\infty} e^{itx} \phi_k(x)\phi_l(x) dx \right|^2 \end{aligned} \tag{D.48}$$

It is convenient to take the $\phi_j(x)$ as the harmonic oscillator wave-functions:

$$\phi_j(x) = \frac{1}{\sqrt{2^j j! \sqrt{\pi}}} e^{-x^2/2} H_j(x). \tag{D.49}$$

where $H_j(x)$ are the Hermite polynomials, which can be written as:

$$H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j} e^{-x^2}. \tag{D.50}$$

Next, we derive simpler expressions for these integrals by first taking the $D \to \infty$ limit. Recalling the definition of the function $K_D(E_1, E_2)$ from (D.41), and comparing to the integrand of (D.48), we see that we must understand the $D \to \infty$ of $K_D(E_1, E_2)$. It is well known that the function $K_D(E, E)$ is related to the average level density $\rho_D(E) = \frac{1}{D}\sum_j \delta(E - E_j)$, by $K_D(E, E) = D\rho_D(E)$, and that its large D limit follows Wigner's semi-circle law [92, 128] (with $\sigma^2 = 1/2$):

$$\lim_{D\to\infty} \left(\sum_{k=0}^{D-1} \phi_k^2(E)\right) = \frac{1}{\pi}\sqrt{2D - E^2}\,\Theta(\sqrt{2D} - |E|), \tag{D.51}$$

where $\Theta(x)$ is the Heavyside step function. This form is a specific property of the Gaussian matrix ensembles [93].

Further, in taking the limit of the function $K_D(E_1, E_2)$ we must ensure that we distinguish between local fluctuations in the sequence of levels, and the global energy dependence of the average density. Generally this is done by rescaling (often called 'unfolding' - see [86] section III.A.1) the energy level density as well as the energies by the local mean spacing [93, 128]. Following the procedure from [128] Appendices 10, and 11, we have:

$$\begin{aligned}\Gamma(t,D) &= D + \frac{1}{2\pi^2}\left|\int_{-\sqrt{2D}}^{\sqrt{2D}} e^{itx}\sqrt{2D-x^2}dx\right|^2 \\ &\quad -\frac{2D}{\pi^2}\int_{-\infty}^{\infty} e^{it(x-y)}\left(\frac{\sin\left[\sqrt{2D}(x-y)\right]}{\sqrt{2D}(x-y)}\right)^2 dxdy \\ &= D + 2D\frac{J_1\left(\sqrt{2D}t\right)^2}{t^2} - (\sqrt{2D}-t/2)\Theta(2\sqrt{2D}-t), \end{aligned} \tag{D.52}$$

where $J_1(x)$ is the first Bessel function of the first kind.

D.2.2 Existence of equilibration time

In this section we discuss the equilibration time. First note that the full expectation over eigenvectors and eigenvalues of the outcome probabilities is given by

$$\mathbb{E}_{spec,C}\left\{\Pr\left(k|x,C,F(t)\right)\right\} = \frac{D - \frac{1}{D}\mathbb{E}_{spec}\{|\mu(t)|^2\} + \delta_{xk}\left(\mathbb{E}_{spec}\{|\mu(t)|^2\}-1\right)}{D^2-1}, \tag{D.53}$$

where we have used Lemma D.1 for $\mathbb{E}_C\left\{\Pr\left(k|x,C,F(t)\right)\right\}$. Notice that if $\mathbb{E}_{spec}\{|\mu(t)|^2\} = D + O(1)$, then

$$\mathbb{E}_{spec,C}\left\{\Pr\left(k|x,C,F(t)\right)\right\} = \frac{\delta_{xk}+1}{D+1} + O(D^{-2}). \tag{D.54}$$

In particular, if D is large, then the GUE expectation of the probability distribution from (D.54) is essentially the uniform distribution. We therefore take the *equilibration time* t_{eq} to be defined by the condition

$$t_{eq} := \{T \mid \mathbb{E}_{spec}\{|\mu(t)|^2\} = D + O(1), \forall\ t > T\}. \tag{D.55}$$

We will show in the following that such a t_{eq} exists and is finite, and then for all $t > t_{eq}$, Equations (6.14), and (6.15) hold. Further, in Appendix D.2.5 we will show that Equation 6.16 holds as well.

Recall from Appendix D.2.1 that

$$\mathbb{E}_{spec}\{|\mu(t)|^2\} = D + 2D\frac{J_1\left(\sqrt{2D}t\right)^2}{t^2} - (\sqrt{2D} - t/2)\Theta(\sqrt{8D} - t). \tag{D.56}$$

It is clear that the last term is monotonically decreasing, and is equal to 0 at the finite time $\sqrt{8D}$. As for the second term, note that for $x \gg 3/4$ we can approximate $J_1(x) \simeq \sqrt{2/\pi x}\cos[x - 3\pi/4]$. It the follows that

$$\lim_{t\to\infty} 2D\frac{J_1\left(\sqrt{2D}t\right)^2}{t^2} \to 0, \tag{D.57}$$

and therefore there exists a finite time t_{eq} such that $\mathbb{E}_{spec}\{|\mu(t)|^2\} = D + O(1), \forall\ t > t_{eq}$.

D.2.3 GUE expectation of first and second power

In this section we use the various expectations over the spectrum to prove Equations (6.14), and (6.15).

First, from the previous section we have that there exists a finite time t_{eq} such that $\mathbb{E}_{spec}\{|\mu(t)|^2\} = D + O(1), \forall\ t > t_{eq}$. Using Lemma D.1, and plugging in the above spectral expectation gives

$$\mathbb{E}_{spec,C}\left\{\Pr\left(k|x, C, F(t)\right)\right\} = \frac{\delta_{xk} + 1}{D + 1} + O(D^{-2}), \tag{D.58}$$

which proves that for all $t > t_{eq}$, Equation 6.14 holds.

Next, we prove that for all $t > t_{eq}$, Equation 6.15 holds. Using Lemma D.2, we have

$$\begin{aligned}
&\mathbb{E}_{spec}\Big\{\mathbb{V}_C\{\Pr\left(k|x, C, F(t)\right)\}\Big\} \\
&= \frac{1}{D^4}\Big\{D^2 - 2D + 4 + (7 - 2D)\delta_{xk} + \mathbb{E}_{spec}\left\{|\mu(t)|^2\right\}(2D\delta_{xk} - 10\delta_{xk} - 2) \\
&+\delta_{xk}\mathbb{E}_{spec}\left\{|\mu(2t)|^2\right\} + 2\delta_{xk}\mathbb{E}_{spec}\left\{Re[\mu(t)^2\mu(-2t)]\right\}\Big\} + O(D^{-5}).
\end{aligned} \tag{D.59}$$

In the next section we will prove that for $t > t_{eq}$, $\mathbb{E}_{spec}\left\{Re[\mu(t)^2\mu(-2t)]\right\} = O(D)$. Using this along with $\mathbb{E}_{spec}\{|\mu(t)|^2\} = D + O(1)$ in the above, we find

$$\mathbb{E}_{spec}\Big\{\mathbb{V}_C\{\Pr\left(k|x, C, F(t)\right)\}\Big\} = O(D^{-2}). \tag{D.60}$$

D.2.4 Bounding $\mathbb{E}_{spec}\{\mu(t)^2\mu(-2t)\}$

We will now bound the expression

$$\Delta(t,D) := \mathbb{E}_{spec}\{\mu(t)^2\mu(-2t)\} = \mathbb{E}_{spec}\left\{\mathrm{Tr}\left[e^{-itH}\right]^2 \mathrm{Tr}\left[e^{2itH}\right]\right\}. \tag{D.61}$$

This expectation is needed for the second step of finding the full GUE variance of the outcome probabilities, as the expression $\mu(t)^2\mu(-2t)$ appears in the eigenvector expectation calculated in the first step in Lemma D.2. As discussed before the statement of Theorem 6.3, we take $\sigma^2 = 1/2$.

Expanding out the expectation, we have:

$$\Delta(t,D) = \int_{-\infty}^{\infty} \sum_{i,j,k} e^{it(E_i+E_j-2E_k)} P(\{E_a\}^D) \prod_{l=1}^{D} dE_l. \tag{D.62}$$

We can expand the triple sum in the integrand into three parts depending on which indices are equal or not equal. Using the permutation symmetry of the distribution in the same fashion as in Appendix D.2.1, and integrating out all but three variables, the integrands become:

(i) $i \neq j \neq k$: $\quad R_3(E_1,E_2,E_3)e^{it(E_1+E_2-2E_3)}$
(ii) $i = j \neq k, i = k \neq j, j = k \neq i$: $\quad R_2(E_1,E_2)e^{2it(E_1-E_2)} + 2R_2(E_1,E_2)e^{it(E_1-E_2)}$
(iii) $i = j = k$: $\quad D$

where R_3 and R_2 have been defined in (D.42). We have already seen what terms with $R_2(E_1,E_2)$ look like in Appendix D.2.1, so we focus first on the R_3 term. By [128] Theorem 5.1.4 (and sections 5.3 and 6.2), integration of $P(\{E_a\}^D)_{GUE}$ over $D-3$ variables gives the 3-point correlation function:

$$\begin{aligned} R_3(E_1,E_2,E_3) &= det[K_D(E_i,E_j)]_{i,j=1,2,3} \\ &= K_D(E_1,E_1)K_D(E_2,E_2)K_D(E_3,E_3) - K_D(E_1,E_2)K_D(E_2,E_1)K_D(E_3,E_3) \\ &- K_D(E_1,E_1)K_D(E_2,E_3)K_D(E_3,E_2) - K_D(E_1,E_3)K_D(E_3,E_1)K_D(E_2,E_2) \\ &+ K_D(E_1,E_3)K_D(E_3,E_2)K_D(E_2,E_1) + K_D(E_1,E_2)K_D(E_2,E_3)K_D(E_3,E_1) \end{aligned} \tag{D.63}$$

Plugging into the integral and adding the R_2 terms, we find:

$$\begin{aligned}\Delta(t,D) = D \;&+\; \left(\int_{-\infty}^{\infty} K_D(E,E)e^{itE}\right)^2 \left(\int_{-\infty}^{\infty} K_D(E,E)e^{-2itE}\right) \\ &-\; \left(\int_{-\infty}^{\infty} K_D(E_1,E_2)^2 e^{it(E_1+E_2)}\right)\left(\int_{-\infty}^{\infty} K_D(E,E)e^{-2itE}\right) \\ &-\; 2\left(\int_{-\infty}^{\infty} K_D(E,E)e^{itE}\right)\left(\int_{-\infty}^{\infty} K_D(E_2,E_3)^2 e^{it(E_2-2E_3)}\right) \\ &+\; 2\int_{-\infty}^{\infty} K_D(E_1,E_2)K_D(E_2,E_3)K_D(E_3,E_1)e^{it(E_1+E_2-2E_3)} \\ &+\; \left|\int_{-\infty}^{\infty} K_D(E,E)e^{2itE}\right|^2 - \left(\int_{-\infty}^{\infty} K_D(E_1,E_2)^2 e^{2it(E_1-E_2)}\right) \\ &+\; 2\left|\int_{-\infty}^{\infty} K_D(E,E)e^{itE}\right|^2 - 2\left(\int_{-\infty}^{\infty} K_D(E_1,E_2)^2 e^{it(E_1-E_2)}\right).\end{aligned} \tag{D.64}$$

Rather than calculating the integrals of all these terms, we will instead bound the magnitude of each one, in a similar fashion to the calculation in [125]. First, take the fourth term, which is of the general form

$$\int_{-\infty}^{\infty} K_D(E_1,E_2)K_D(E_2,E_3)K_D(E_3,E_1)e^{i(t_1E_1+t_2E_2+t_3E_3)}dE_1dE_2dE_3, \tag{D.65}$$

for some combination of t_i. Next note that this expression is just

$$\mathrm{Tr}[Pe^{it_1E_1}Pe^{it_2E_2}Pe^{it_3E_3}], \tag{D.66}$$

where $P = \sum_{k=0}^{D-1} |\phi_k\rangle\langle\phi_k|$ is the projector onto the D-dimensional lower-energy subspace spanned by the harmonic oscillator wave-functions defined in Equation (D.49), and X is the position operator. This follows from the definition of $K_D(x,y)$ in (D.41). Next we use the Cauchy-Schwartz inequality twice on (D.66):

$$\begin{aligned}\left|\mathrm{Tr}[(Pe^{it_1E_1}Pe^{it_2E_2})Pe^{it_3E_3}]\right| &\leq \sqrt{\mathrm{Tr}[Pe^{-it_1E_1}Pe^{it_1E_1}]Tr[P]} \\ &\leq \sqrt{\sqrt{\mathrm{Tr}[P]\,\mathrm{Tr}[P]}\,\mathrm{Tr}[P]} \leq D.\end{aligned} \tag{D.67}$$

A similar argument also shows that terms with integrands of the form $K_D(E_1,E_2)^2$ are

also bounded by D. Then

$$\begin{aligned}|\Delta(t,D)| &\leq 6D + \left(\int_{-\infty}^{\infty} K_D(E,E)e^{itE}\right)^2 \left(\int_{-\infty}^{\infty} K_D(E,E)e^{-2itE}\right) \\ &+ D\left(\int_{-\infty}^{\infty} K_D(E,E)e^{-2itE}\right) + 2D\left(\int_{-\infty}^{\infty} K_D(E,E)e^{itE}\right) \\ &+ \left|\int_{-\infty}^{\infty} K_D(E,E)e^{2itE}\right|^2 + 2\left|\int_{-\infty}^{\infty} K_D(E,E)e^{itE}\right|^2. \end{aligned} \tag{D.68}$$

Next recall that

$$\int_{-\infty}^{\infty} K_D(E,E)e^{itE} = \sqrt{2D}\frac{J_1(\sqrt{2D}\ t)}{t}, \tag{D.69}$$

which approaches 0 as $t \to \infty$. Further, the equilibration condition (D.55) requires that $\sqrt{2D}J_1(\sqrt{2D}\ t)/t = O(1)$ for $t > t_{\text{eq}}$. Certainly, if this is satisfied, then for all $t > t_{\text{eq}}$, we have

$$|\Delta(t,D)| \leq O(D). \tag{D.70}$$

D.2.5 GUE expectation of fourth power

In this section we show that for $t > t_{\text{eq}}(D)$ we have that $\mathbb{E}_{spec,C}\left\{\Pr\left(k|x,C,F(t)\right)^4\right\} = O(D^{-4})$. Recall from Lemma D.3 that

$$\mathbb{E}_C\left\{\Pr\left(k|x,C,F(t)\right)^4\right\} \leq \frac{1}{D^8}\sum_{\pi}\langle\Phi_\pi|R_8(t)\rangle + \sum_{\pi\neq\sigma} B(D^{-9})_{\pi,\sigma}\langle\Phi_\sigma|R_8(t)\rangle. \tag{D.71}$$

It is then sufficient to show that for $t > t_{\text{eq}}(D)$, we have $\mathbb{E}_{spec}\{\langle\Phi_\sigma|R_8(t)\rangle\} \leq O(D^4)$, for all permutations σ.

Calculating some explicit examples of

$$\begin{aligned}&\langle\Phi_\sigma|R_8(t)\rangle \\ &= \sum_{\substack{a,a',b,b',\\ c,c',d,d'}} e^{it(E_a - E_{a'} + E_b - E_{b'} + E_c - E_{c'} + E_d - E_{d'})}\langle a,a',b,b',c,c',d,d'|V_\sigma|a,a',b,b',c,c',d,d'\rangle,\end{aligned} \tag{D.72}$$

we see that these are of the general form

$$D^a\mu(f_1 t)\mu(g_1 t)^*, \ldots, \mu(f_4 t)\mu(g_4 t)^*, \tag{D.73}$$

where $f_j, g_j \in \{0, 1, 2, 3, 4\}$, and

$$a + \sum_{j=1}^{4} f_j + \sum_{j=1}^{4} g_j = 8, \tag{D.74}$$

and if $f_j = 0$ then the corresponding $\mu(f_j t)$ does not appear in the product (and similarly for g_j). For example, there is a $|\mu(t)|^4$ term arising from $\sigma = \mathbf{1}$, and a D^4 term arising from $\sigma = (12)(34)(56)(78)$. The expectations $\mathbb{E}_{spec}\{D^a \mu(f_1 t)\mu(g_1 t)^*, \ldots \mu(f_4 t)\mu(g_4 t)^*\}$ can then be bounded in a similar fashion to Appendix D.2.4. In particular, we can expand the sums in the products of the terms $\mu(f_j t) = \mathrm{Tr}[e^{if_j tH}]$ into various parts depending on which indices are equal or not equal, just as was done for expression (D.62). These will then give a constant D^b factor, and various combinations of integrals of m-point correlation functions, similarly to Equation (D.64). Each of the integrals with 2-point or higher order correlation functions can be bounded using the Cauchy-Schwartz inequality by D, just as was done for Equation (D.65). We will then be left with various powers of integrals of the form

$$\int_{-\infty}^{\infty} K_D(E, E) e^{itE} = \sqrt{2D} \frac{J_1(\sqrt{2D}\ t)}{t}, \tag{D.75}$$

which as we have seen approach 0 as $t \to \infty$, and so are irrelevant for the $t > t_{eq}$ regime. The only remaining question then is power of the constant D^b factor for each term. It is not difficult to see that a power of D arises from each pairing $\mu(f_j t)\mu(g_k t)^*$ with $f_j = g_k$. For example, the term $|\mu(t)|^4$ gives a contribution of D^2. From the constraint (D.74), it is not difficult to see that D^4 is the highest power of D which can arise. This shows that for $t > t_{eq}$

$$\mathbb{E}_{spec,C}\Big\{ Pr(k|i, C, F(t))^4 \Big\} = O(D^{-4}). \tag{D.76}$$

References

[1] S. Aaronson. Is quantum mechanics an island in theoryspace? arXiv:0401062.

[2] D. S. Abrams and S. Lloyd. Nonlinear quantum mechanics implies polynomial-time solution for NP-complete and ♯P problems. *Phys. Rev. Lett.*, 81:3992, 1998. arXiv:9801041.

[3] S. Abramsky and B. Coecke. A categorical semantics of quantum protocols. *Proceedings of the 19th Annual IEEE Symposium on Logic in Computer*, pages 415–425, 2004.

[4] Sh.A. Ajupov, B. Iochum, and N.Dz. Yadzgorov. Symmetry versus facial homogeneity for self-dual cones. *Lin. Alg. and its Applications*, 142:83–89, 1990.

[5] E. Alfsen and F. Shultz. State spaces of jordan algebras. *Acta. Math.*, 140:155–190, 1978.

[6] E. Alfsen and F. Shultz. *State Spaces of Operator Algebras*. Birkhauser, 2001.

[7] E. Alfsen and F. Shultz. *Geometry of State Spaces of Operator Algebras*. Birkhauser, 2003.

[8] C.D. Aliprantis and R. Tourky. *Cones and Duality*. American Mathematical Society, 2007.

[9] H. Araki. On a characterization of the state space of quantum mechanics. *Commun. Math. Phys.*, 75:1–24, 1980.

[10] D.J. Moore B. Coecke and A. Wilce. Test spaces. In D.J. Moore B. Coecke and A. Wilce, editors, *Current Research in Operational Quantum Logic: Algebras, Categories, Languages*, pages 1–36. Kluwer Academic Publishers, 2000. arXiv:0008019.

[11] J. Baez. Division algebras and quantum theory. 2011. arXiv:1101.5690.

[12] R. Balian. Random matrices and information theory. *Il Nuovo Cimento B (1965-1970)*, 57:183–193, 1968.

[13] L. E. Ballentine. Probability in quantum mechanics. In D. M. Greenberger, editor, *New Techniques and Ideas in Quantum Mechanics*, pages 382–392, New York, 1986. Annals of the New York Academy of Sciences, volume 480, New York Academy of Sciences.

[14] L. E. Ballentine. Quantum mechanics: A modern development. Singapore, 1998. World Scientific.

[15] G. Barker. Perfect cones. *Lin. Alg. and its Applications*, 22:211–221, 1978.

[16] G.P. Barker. Faces and duality in convex cones. *Linear and Multilinear Algebra*, 6(3):161–169, 1978.

[17] G.P. Barker. Theory of cones. *Linear Algebra and its Applications*, 39:263–291, 1981.

[18] G.P. Barker and Foran J. Self-dual cones in euclidean spaces. *Linear Algebra and its Applications*, 13(12):147–155, 1976.

[19] G.P. Barker, M. Laidacker, and G. Poole. Projectionally exposed cones. *SIAM J. Algebraic Discrete Methods*, 8(1):100–105, 1987.

[20] H. Barnum, J. Barrett, L. 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. arXiv:0909.5075.

[21] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Cloning and broadcasting in generalized probabilistic models. 2006. arXiv:0611295.

[22] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. A generalized no-broadcasting theorem. *Phys. Rev. Lett.*, 99:240501, 2007. arXiv:0707.0620.

[23] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Teleportation in general probabilistic theories. 2008. arXiv:0805.3553.

[24] H. Barnum, S. Beigi, S. Boixo, M.B. Elliott, and S. Wehner. Local quantum measurement and no-signaling imply quantum correlations. *Phys. Rev. Lett.*, 104:140401, 2010. arXiv:0910.3952.

[25] H. Barnum, C.P. Gaebler, and A. Wilce. Ensemble Steering, Weak Self-Duality, and the Structure of Probabilistic Theories. 2010. arXiv:0912.5532.

[26] J. Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75:032304, 2007. arXiv:0508211.

[27] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Non-local correlations as an information theoretic resource. *Phys. Rev. A*, 71:022101, 2005.

[28] J. Bellissard and B. Iochum. Homogeneous self dual cones versus jordan algebras: The theory revisited. *Annales de l'institut Fourier*, 28(1):27–67, 1978.

[29] J. Bellissard, B. Iochum, and R. Lima. Homogeneous and facially homogeneous self-dual cones. *Lin. Alg. and its Applications*, 19:1–16, 1978.

[30] E. Beltrametti and G. Cassinelli. On state transformations induced by yes-no experiments, in the context of quantum logic. *Journal of Philosophical Logic*, 6:369–379, 1977.

[31] E. G. Beltrametti and S. Bugajski. Effect algebras and statistical physical theories. *J. Math. Phys.*, 38:3020–3030, 1997.

[32] E.G. Beltrametti and J. Cassinelli. *The Logic of Quantum Mechanics*. Addison-Wesley, Reading, 1981.

[33] I. Bengtsson, S. Weis, and K. Życzkowski. Geometry of the set of mixed quantum states: An apophatic approach. 2011. arXiv:1112.2347.

[34] I. Bengtsson and K. Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, Cambridge, 2011.

[35] D.W. Berry and A. M. Childs. Black-box Hamiltonian simulation and unitary implementation. *Quantum Information and Computation*, 12:29, 2012.

[36] M.V. Berry. Quantizing a classically ergodic system: Sinai's billiard and the KKR method. *Annals of Physics*, 131(1):163–216, 1981.

[37] G. Birkhoff and J. Von Neumann. The logic of quantum mechanics. *The Annals of Mathematics*, 37(4):823–843, 1936.

[38] O. Bohigas, M. J. Giannoni, and C. Schmit. Characterization of chaotic quantum spectra and universality of level fluctuation laws. *Phys. Rev. Lett.*, 52:1–4, Jan 1984.

[39] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[40] F. Brandao, P. Cwiklinski, M. Horodecki, P. Horodecki, J. Korbicz, and M. Mozrzymas. Convergence to equilibrium under a random hamiltonian. 2011. arXiv:1108.2985.

[41] G. Brassard, H. Buhrman, N. Linden, A. A. Methot, A. Tapp, and F. P. Unger. Limit On Nonlocality In Any World In Which Communication Complexity Is Not Trivial. *Physical Review Letters*, 96, 2006.

[42] P. Busch. Quantum states and generalized observables: a simple proof of Gleasons theorem. 1999. arXiv:9909073.

[43] G. Cassinelli and E. G. Beltrametti. Ideal, first-kind measurements in a proposition-state structure. *Communications in Mathematical Physics*, 40:7–13, 1975.

[44] A. M. Childs. On the Relationship Between Continuous- and Discrete-Time Quantum Walk. *Communications in Mathematical Physics*, 294(2):581–603, 2010.

[45] G. Chiribella, G.M. D'Ariano, and P. Perinotti. Probabilistic theories with purification. *Physical Review A*, 81(6):062348, 2010. arXiv:0908.1583.

[46] G. Chiribella, G.M. D'Ariano, and P. Perinotti. Informational derivation of quantum theory. *Phys. Rev. A*, 84:012311, 2011. arXiv:1011.6451.

[47] R. Clifton, J. Bub, and H. Halvorson. Characterizing quantum theory in terms of information-theoretic constraints. *Foundations of Physics*, 33(11):1561–1591, 2003.

[48] B. Coecke. Quantum picturalism. arXiv:0402130.

[49] M. Cramer and J. Eisert. A quantum central limit theorem for non-equilibrium systems: exact local relaxation of correlated states. *New Journal of Physics*, 12(5):055020, 2010. arXiv:0911.2475.

[50] Finkelstein D. The logic of quantum theory. *Trans. N.Y. Acad. Sci.*, 25(2):621–637, 1962–1963.

[51] B. Dakic and C. Brukner. Quantum theory and beyond: is entanglement special? 2009. arXiv:0911.0695.

[52] C. Dankert, R. Cleve, J. Emerson, and E. Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A*, 80:012304, 2009. arXiv:0606161.

[53] G. M. D'Ariano. Probabilistic theories: what is special about Quantum Mechanics? 2008. arXiv:0807.4383.

[54] E.B. Davies. Example related to the foundations of quantum theory. *Journal of Mathematical Physics*, 13(1):39–41, 1972.

[55] E.B. Davies and J.T. Lewis. An operational approach to quantum probability. *Commun. Math. Phys.*, 17:239–260, 1970.

[56] G. de la Torre, L. Masanes, A.J. Short, and M. Müller. Deriving quantum theory from its local structure and reversibility. 2011. arXiv:1110.5482.

[57] H. De Raedt, K. Michielsen, and K. Hess. Analysis of multipath interference in three-slit experiments. *Phys. Rev. A*, 85:012101, 2012.

[58] J.M. Deutsch. Quantum statistical mechanics in a closed system. *Phys. Rev. A*, 43:2046–2049, 1991.

[59] J. Emerson and L. E. Ballentine. Quantum-classical correspondence for the equilibrium distributions of two interacting spins. *Phys. Rev. E*, 64:026217, 2001.

[60] J. Emerson and L.E. Ballentine. Characteristics of quantum-classical correspondence for two interacting spins. *Phys. Rev. A*, 63:052103, 2001.

[61] J. Faraut and A. Koranyi. *Analysis on Symmetric cones*. Oxford University Press, 1994.

[62] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics*. Addison-Wesley, Reading, MA, 1965.

[63] D. I. Fivel. How interference effects in mixtures determine the rules of quantum mechanics. *Physical Review A*, 50(3):27, 1994. arXiv:9405042.

[64] D.I. Fivel. Derivation of the rules of quantum mechanics from information-theoretic axioms. 2010. arXiv:1010.5300.

[65] D. J. Foulis and C. H. Randall. An approach to empirical logic. *American Math. Monthly*, 77:363–374, 1970.

[66] D. J. Foulis and C. H. Randall. Operational statistics I, II. *American Math. Monthly*, 11, 14:1667–1675, 1472–1480, 1972, 1973.

[67] D. J. Foulis and C. H. Randall. What are quantum logics, and what ought they to be? In E. Betrametti and B. van Fraassen, editors, *Current Issues in Quantum Logic*. Plenum, 1981.

[68] D.J. Foulis. Mathematical metascience. *J. Natural Geometry*, 13:1–50, 1998.

[69] Y. Friedman and B. Russo. Classification of atomic facially symmetric spaces. *Canad. J. Math.*, 45:33–87, 1993.

[70] C.A. Fuchs. Quantum mechanics as quantum information (and only a little more). 2002. arXiv:0205039.

[71] C.A. Fuchs. QBism, the perimeter of quantum Bayesianism. 2010. arXiv:1003.5209.

[72] C.A. Fuchs and R. Schack. A quantum-bayesian route to quantum-statespace. *Foundations of Physics*, 41(3):1–12, 2010. arXiv:0912.4252.

[73] W. Gale, E. Guth, and G. T. Trammell. Determination of the quantum state by measurements. *Phys. Rev.*, 165(5):1434–1436, 1968.

[74] I. Garcia-Mata, C. Pineda, and D. A. Wisniacki. Non-markovian quantum dynamics and classical chaos. 2012. arXiv:1204.3614.

[75] J. Gemmer, M. Michel, and G. Mahler. *Quantum Thermodynamics: Emergence of Thermodynamic Behavior within Composite Quantum Systems*. Springer: Berlin, 2004.

[76] D. Gillies. *Philosophical theories of probability*. Routledge, 2002.

[77] A.M. Gleason. Measures on the closed subspaces of a Hilbert space. *J. Math. Mech.*, 6:885, 1957.

[78] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghì. Canonical typicality. *Phys. Rev. Lett.*, 96:050403, 2006.

[79] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghì. Long-time behavior of macroscopic quantum systems: Commentary accompanying the english translation of John von Neumann's 1929 article on the quantum ergodic theorem. 2010. arXiv:1003.2129.

[80] P. Goyal. Information-geometric reconstruction of quantum theory. *Physical Review A*, 78(5):052120, 2008.

[81] P. Goyal, H. Knuth, and J. Skilling. Origin of complex quantum amplitudes and Feynman's rules. *Physical Review A*, 81(2):022109, 2010.

[82] A. Grinbaum. *The Significance of Information in Quantum Theory*. PhD thesis, Ecole Polytechnique, Paris, 2004. arXiv:0410071.

[83] A. Grinbaum. On the notion of reconstruction of quantum theory. 2005. arXiv:0509104.

[84] D. Gross, M. Müller, R. Colbeck, and Dahlsten O. C. O. All reversible dynamics in maximally non-local theories are trivial. *Phys. Rev. Lett.*, 104:080402, 2010. arXiv:0910.1840.

[85] S. Gudder. Convex structures and effect algebras. *Int. J. Theor. Phys.*, 28(12):3179–3187, 1999.

[86] T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller. Random-matrix theories in quantum physics: common concepts. *Physics Reports*, 299(4–6):189–425, 1998.

[87] J. Gunson. On the algebraic structure of quantum mechanics. *Comm. Math. Phys.*, 6(4):262–285, 1967.

[88] W. Guz. Filter theory and covering law. *Annales de l'institut Henri Poincaré (A) Physique théorique*, 29:357–378, 1978.

[89] W. Guz. Pure operations and the covering law. *Reports on Mathematical Physics*, 16(1):125–141, 1979.

[90] W. Guz. Conditional probability in quantum axiomatics. *Annales de l'institut Henri Poincaré (A) Physique théorique*, 33(1):63–119, 1980.

[91] R. Haag. *Local Quantum Physics*. Springer, 1996.

[92] U. Haagerup and S. Thorbjørnsen. Random matrices with complex gaussian entries. *EXPO. MATH*, 21:293–337, 1998.

[93] F. Haake. *Quantum Signatures of Chaos* 2nd *Edition*. Springer-Verlag New York, 2004.

[94] H. Hanche-Olsen and E. Stormer. *Jordan Operator Algebras*. Pitman Publ., Boston-London-Melbourne, 1984.

[95] L. Hardy. Quantum theory from five reasonable axioms. 2001. arXiv:0101012.

[96] L. Hardy. Probability theories with dynamic causal structure: A new framework for quantum gravity. 2005. arXiv:0509120.

[97] L. Hardy. Foliable operational structures for general probabilistic theories. 2009. arXiv:0912.4740.

[98] L. Hardy. A formalism-local framework for general probabilistic theories including quantum theory. 2010. arXiv:1005.5164.

[99] L. Hardy. Reformulating and reconstructing quantum theory. 2011. arXiv:1104.2066.

[100] A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*. North-Holland, New York, 1980.

[101] M. Žnidarič, C. Pineda, and I. García-Mata. Non-Markovian behavior of small and large complex quantum systems. *Phys. Rev. Lett.*, 107:080404, 2011. arXiv:1104.5263.

[102] B. Iochum. *Cones Autopolaires et Algebres de Jordan*. Lecture Notes in Mathematics, Vol. 1049. Springer, 1984.

[103] R. Jancel. *Foundations of Classical and Quantum Statistical Mechanics*. Pergamon Press, 1959.

[104] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner. Limits on nonlocal correlations from the structure of the local state space. *New J. Phys.*, 13:063024, 2011. arXiv:1012.1215.

[105] E.T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[106] P. Jordan. Ueber verallgemeinerungsmöglichkeiten des formalismus der quantenmechanik. *Nachr. Akad. Wiss. Göttingen Math. Phys. Kl.*, 41:209–217, 1933.

[107] P. Jordan, J. von Neumann, and E. Wigner. On an algebraic generalization of the quantum mechanical formalism. *Ann. Math.*, 35:29–64, 1934.

[108] V.T. Khudalov. Description for all regular cones in Hilbert space. *Siberian Mathematial Journal*, 42:205–208, 2001.

[109] G. Kimura, T. Miyadera, and H Imai. Optimal observables for minimum-error state discrimination in general probabilistic theories. *J. Math. Phys.*, 51:093505, 2010.

[110] G. Kimura and K. Nuida. On affine maps on non-compact convex sets and some characterizations of finite-dimensional solid ellipsoids. 2011. arXiv:1012.5350.

[111] G. Kimura, Miyadera T., and H. Imai. Optimal state discrimination in general probabilistic theories. *Phys. Rev. A*, 79:062306, 2009. arXiv:0808.3844.

[112] K.A. Kirkpatric. "Quantal" behavior in classical probability. 2001. arXiv:0106072.

[113] M. Koecher. Die geodätischen von positivitätsbereichen. *Math. Annalen*, 135:192–202, 1958.

[114] B.O. Koopman. Quantum theory and the foundations of probability. In L.A. MacColl, editor, *Proceedings of Symposia in Applied Mathematics, Volume VII: Applied Probability*, pages 97–102, New York, 1957. McGraw-Hill.

[115] H. Kummer. A constructive approach to the foundations of quantum mechanics. *Foundations of Physics*, 17:1–62, 1987.

[116] H. Kummer. The foundation of quantum theory and noncommutative spectral theory, I,II. *Foundations of Physics*, 21:1021–1069,1183–1236, 1991.

[117] N. Linden, S. Popescu, A. J. Short, and A. Winter. Quantum mechanical evolution towards thermal equilibrium. *Phys. Rev. E*, 79:061103, 2009.

[118] S. Lloyd. *Black Holes, Demons and the Loss of Coherence: How complex systems get information, and what they do with it.* PhD thesis, Rockefeller University, New York, 1988.

[119] G. Ludwig. *An Axiomatic Basis of Quantum Mechanics*, volume I, II. Springer–Verlag, 1985, 1987.

[120] G. Mackey. *Mathematical Foundations of Quantum Mechanics.* The Mathematical Physics Monograph Series. W. A. Benjamin, Inc., New York, 1963.

[121] P. G. L. Mana. *Studies in plausibility theory, with applications to physics.* PhD thesis, Kungliga Tekniska Hogskolan, Stockolm, 2007. http://www.tp.umu.se/ mana/.

[122] P.G.L. Mana. Why can states and measurement outcomes be represented as vectors? 2003. arXiv:0305117.

[123] L. Masanes, M. Müller, M.D. Perez-Garcia, and R. Augusiak. Entangling dynamics beyond quantum theory. 2011. arXiv:1111.4060.

[124] L. Masanes and M. P. Müller. A derivation of quantum theory from physical requirements. *New Journal of Physics*, 13(6):063001, 2011. arXiv:1004.1483.

[125] L. Masanes, A. J. Roncaglia, and A. Acin. The complexity of energy eigenstates as a mechanism for equilibration. 2011. arXiv:1108.0374.

[126] Ll. Masanes, A. Acin, and N. Gisin. General properties of nonsignaling theories. *Phys. Rev. A*, 73:012112, 2006. arXiv:0508016.

[127] K. McCrimmon. *A Taste of Jordan Algebras*. Springer, Berlin, 2004.

[128] M. L. Mehta. *Random Matrices; 3rd ed.* Elsevier, San Diego, CA, 2004.

[129] B. Mielnik. Geometry of quantum states. *Comm. Math. Phys.*, 9:55–80, 1968.

[130] B. Mielnik. Theory of filters. *Comm. Math. Phys.*, 15(1):1–46, 1969.

[131] Bogdan Mielnik. Convex geometry: a travel to the limits of our knowledge. 2012. arXiv:1202.2164.

[132] J.J. Moreau. Décomposition orthogonale d'un espace Hilbertien selon deux cônes mutuellement polaires. *C. R. Acad. Sci.*, 255:238–240, 1962.

[133] M. Müller, D. Gross, and J. Eisert. Concentration of measure for quantum states with a fixed expectation value. *Commun. Math. Phys.*, 303(3):785–824, 2006. arXiv:1003.4982.

[134] M. P. Müller and C. Ududec. The structure of reversible computation determines the self-duality of quantum theory. *Phys. Rev. Lett.*, 108:130401, 2012. arXiv:1110.3516.

[135] M.P. Müller, O.C.O. Dahlsten, and V. Vedral. Unifying typical entanglement and coin tossing: on randomization in probabilistic theories. 2011. arXiv:1107.6029.

[136] M. Navascues and H. Wunderlich. A glance beyond the quantum model. *Proc. Roy. Soc. Lond. A*, 466:881–890, 2009. arXiv:0907.0372.

[137] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 1st edition, 2004.

[138] G. Niestegge. Sorkin's third-order interference term in quantum logics with unique conditional probabilities. 2009. arXiv:0912.0203.

[139] J. Oppenheim and S. Wehner. The uncertainty principle determines the nonlocality of quantum mechanics. *Science*, 330(6007):1072–1074, 2010.

[140] Barker G. P. The lattice of faces of a finite dimensional cone. *Linear Algebra and its Applications*, 7:71, 1973.

[141] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, 2009.

[142] A. Peres. Ergodicity and mixing in quantum theory. I. *Phys. Rev. A*, 30:504–508, 1984.

[143] C. Pfister. One simple postulate implies that every polytopic state space is classical. 2012. arXiv:1203.5622.

[144] C. Piron. Axiomatique quantique. *Helv. Phys. Acta*, 37:439–468, 1964.

[145] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Found. Phys.*, 24:379, 1994.

[146] S. Popescu, A. J. Short, and A. Winter. Entanglement and the Foundations of Statistical Mechanics. *Nature Physics*, 2:754–758, 2006. arXiv:0511225.

[147] J. Rau. Consistent reasoning about a continuum of hypotheses on the basis of finite evidence. 2007. arXiv:0706.2274.

[148] P. Reimann. Canonical thermalization. *New Journal of Physics*, 12(5):055027, 2010.

[149] A. Riera, C. Gogolin, and J. Eisert. Thermalization in nature and on a quantum computer. arXiv:1102.2389 .

[150] M. Rigol, V. Dunjko, and M. Olshanii. Thermalization and its mechanism for generic isolated quantum systems. *Nature*, 452:854–858, 2008.

[151] R. Schack and C.A. Fuchs. Quantum-Bayesian coherence. 2009. arXiv:0906.2187.

[152] I.E. Segal. Postulates for general quantum mechanics. *Ann. Math.*, 48:930, 1947.

[153] S. Sherman. On Segal's postulates for general quantum mechanics. *Ann. Math.*, 64:593, 1956.

[154] A. J. Short. Equilibration of quantum systems and subsystems. *New Journal of Physics*, 13(5):053009, 2011. arXiv:1012.4622.

[155] A. J. Short and Farrelly T. C. Quantum equilibration in finite time. *New Journal of Physics*, 14:013063, 2012. arXiv:1110.5759.

[156] A. J. Short and S. Wehner. Entropy in general physical theories. *New Journal of Physics*, 12:033023, 2010.

[157] B. Simon. *Representations of Finite and Compact Groups*. American Mathematical Society, 1996.

[158] U. Sinha, C. Couteau, T. Jennewein, R. Laflamme, and G. Weihs. Ruling out multi-order interference in quantum mechanics. *Science*, 329(5990):418–421, 2010.

[159] U. Sinha, C. Couteau, Z. Medendorp, I. Sollner, R. Laflamme, R. Sorkin, and G Weihs. Testing born's rule in quantum mechanics with a triple slit experiment. In L. et al. Accardi, editor, *Foundations of Probability and Physics - 5, Vaxjo, August, 2008*, pages 200–207, Melville, NY, 2009. American Institute of Physics, Ser. Conference Proceedings, vol. 1101.

[160] J. A. Smolin. Can quantum cryptography imply quantum mechanics? 2003. arXiv:0310067.

[161] I. Söllner, B. Gschösser, P. Mai, B. Pressl, Z. Vörös, and G. Weihs. Testing borns rule in quantum mechanics for three mutually exclusive events. *Foundations of Physics*, pages 1–10, 2011.

[162] R.D. Sorkin. Quantum mechanics as quantum measure theory. *Mod. Phys. Lett.*, A9:3119–3128, 1994. arXiv:9401003.

[163] R.W. Spekkens. In defense of the epistemic view of quantum states: a toy theory. *Physical Review A*, 75:032110, 2007. arXiv:0401052.

[164] M. Srednicki. Chaos and quantum thermalization. *Phys. Rev. E*, 50:888–901, 1994.

[165] M. Srednicki. Does quantum chaos explain quantum statistical mechanics? 1994. arXiv:9410046.

[166] F. Strocchi. *An introduction to the mathematical structure of quantum mechanics: A short course for mathematicians*. World Scientific, 2005.

[167] C.H. Sung and B.S. Tam. A study of projectionally exposed cones. *Linear Algebra and its Applications*, 139(0):225 – 252, 1990.

[168] W. Thirring. *Quantum mathematical physics*. Springer, 2004.

[169] C. Ududec, H. Barnum, and J. Emerson. Three slit experiments and the structure of quantum theory. *Foundations of Physics*, 41:396–405, 2011. arXiv:0909.4787.

[170] W. van Dam. Implausible Consequences of Superstrong Nonlocality. 2005. arXiv:0501159.

[171] V.S. Varadarajan. *Geometry of Quantum Theory*. Van Nostrand, vol. 1 (1968), vol. 2 (1970).

[172] C. L. Venuti and P. Zanardi. Unitary equilibrations: Probability distribution of the Loschmidt echo. *Phys. Rev. A*, 81:022113, 2010.

[173] E.B. Vinberg. Homogeneous cones. *Soviet Math. Dokl.*, 2:1416–1619, 1961.

[174] J. von Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, 1955.

[175] R. Webster. *Convexity*. Oxford University Press, 1994.

[176] S. Weis. Quantum convex support. 2011. arXiv:1101.3098.

[177] A. Wilce. Four and a half axioms for finite dimensional quantum mechanics. 2009. arXiv:0912.5530.

[178] A. Wilce. Test spaces. In K. Engesser, D. Gabbay, and D. Lehmann, editors, *Handbook of Quantum Logic*, pages 443–549. Elsevier, 2009.

[179] A. Wilce. Symmetry and composition in probabilistic theories. *Electron. Notes Theor. Comput. Sci.*, 270(2):191–207, 2011.

[180] A. Wilce. Symmetry, self-duality and the jordan structure of quantum mechanics. 2011. arXiv:1110.6607.

[181] Alexander Wilce. Formalism and interpretation in quantum theory. *Foundations of Physics*, 40(4):434–462, 2010.

[182] N. Zierler. Axioms for non-relativistic quantum mechanics. *Pacific Journal of Mathematics*, 11:1151–1169, 1961.

[183] K. Życzkowski. Quartic quantum theory: an extension of the standard quantum mechanics. *J. Phys. A*, 41(35):355302, 2008. arXiv:0804.1247.