Negative Quasi-Probability in the Context of Quantum Computation

by

Victor Veitch

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Mathematics
in
Applied Mathematics

Waterloo, Ontario, Canada, 2013

DECLARATION

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.

ABSTRACT

This thesis deals with the question of what resources are necessary and sufficient for quantum computational speedup. In particular, we study what resources are required to promote fault tolerant stabilizer computation to universal quantum computation. In this context we discover a remarkable connection between the possibility of quantum computational speedup and negativity in the discrete Wigner function, which is a particular distinguished quasi-probability representation for quantum theory. This connection allows us to establish a number of important results related to magic state computation, an important model for fault tolerant quantum computation using stabilizer operations supplemented by the ability to prepare noisy non-stabilizer ancilla states. In particular, we resolve in the negative the open problem of whether every non-stabilizer resource suffices to promote computation with stabilizer operations to universal quantum computation.

Moreover, by casting magic state computation as resource theory we are able to quantify how useful ancilla resource states are for quantum computation, which allows us to give bounds on the required resources. In this context we discover that the sum of the negative entries of the discrete Wigner representation of a state is a measure of its usefulness for quantum computation. This gives a precise, quantitative meaning to the negativity of a quasi-probability representation, thereby resolving the 80 year debate as to whether this quantity is a meaningful indicator of quantum behaviour.

We believe that the techniques we develop here will be widely applicable in quantum theory, particularly in the context of resource theories.

Has a bit the quantum nature?
This is the most serious question of all.
If you say yes or no,
You lose your own quantum nature.

— Qumon, zen poet.

ACKNOWLEDGMENTS

Firstly, I thank my supervisor Joseph Emerson for his support and encouragement throughout my time as a graduate student.

Next, I thank my friends and colleagues at the Institute for Quantum Computation for innumerable interesting conversations, good times and the occassional bout of productivity. These include Maris Ozols, Christopher Ferrie, Nathan Wiebe, Thom Bohdanowicz, Yuval Sanders, Dmitry Serbin and far too many others to list.

Third, I thank my partner Yomna Nasser for her love and support during this part of my life.

Finally, I thank my parents and sister for their continued love and support.

Dedicated to the loving memory of Tom (Teunis) Denbok.

1923 – 2012

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

While it is widely believed that quantum computers can solve certain problems with exponentially fewer resources than their classical counterparts, the scope of the physical resources of the underlying quantum systems that enable universal quantum computation is not well understood. For example, for the standard circuit model of quantum computation, Vidal has shown that high-entanglement is necessary for an exponential speed-up [75]; however, it is also known that access to high-entanglement is not sufficient [26]. Moreover, in alternative models of quantum computation such as DQC1 [48], algorithms that may be performed on highly-mixed input states appear to be more powerful than classical computation even though there appears to be a negligible amount of entanglement in the underlying quantum system [15]. This suggests that large amounts of entanglement, purity or even coherence may not be necessary resources for quantum-computational speed-up. One of the central open problems of quantum information is to understand which sets of quantum resources are jointly necessary and sufficient to enable an exponential speed-up over classical computation. Any solution to this important problem may point to more practical experimental means of achieving the benefits of quantum computation.

The question of whether a restricted subset of quantum theory is still sufficient for a given task is meaningful when there is a specific context that divides the full set of possible quantum operations into two classes: the restricted subset of operations that are accessible or easy to implement and the remainder that are not. In such a context it is then natural to consider the difficult operations as resources and ask how much, if any, of these resources are required. For example, a common paradigm in quantum communication is that of two or more spatially separated parties for which local quantum operations and classical communication define a restricted set of operations that are accessible or "free resources", whereas joint quantum operations are not free; in this context entanglement is the natural resource for quantum *communication*. Astonishingly, there is not yet a corresponding resource theory for the task of quantum *computation*.

The major obstacle to physical realizations of quantum computation is that real world devices suffer noise when they execute quantum algorithms. Fault tolerant quantum computation offers a partial resolution to this problem by allowing the effective error rates on logically encoded computations to be reduced below the error rate of the physical (unencoded) computation. These schemes work by encoding a large block of physical qudits into a single logical qudit that is robust against *isolated* errors on the physical qudits. However, it is possible for logical operations to spread errors between the physical qudits with the con-

sequence that the quantum computer will be overwhelmed by errors. Logical unitary gates that do not spread errors within a code block are called transversal; these play a critical role in fault-tolerant quantum computation. Recent theoretical work has shown that a set of quantum gates which is both universal and transversal does not exist[17, 80, 13]. Thus any scheme for fault tolerant quantum computation using transversal gates divides the quantum operations into two classes: those with a fault-tolerant implementation – these are the "free resources" – and the remainder – these are not free but are required to achieve universality. For a fixed fault tolerant scheme the critical question is: what are necessary and sufficient physical resources to promote fault-tolerant computation to universal quantum computation?

The known fault tolerant schemes with the best performance are built around the well-known stabilizer formalism [29], in which a distinguished set of preparations, measurements, and unitary transformations (the "stabilizer operations") have a fault tolerant implementation. Stabilizer operations also arise naturally in some physical systems with topological order[47, 23]. As described above, the transversal set of stabilizer operations do not give a universal gate set and must be supplemented with some additional (non-stabilizer) resource. A celebrated scheme for overcoming this limitation is the magic state model of quantum computation devised by Bravyi and Kitaev [6] where the additional resource is a set of ancilla systems prepared in a some (generally noisy) non-stabilizer quantum state. Hence in this important paradigm the question of which physical resources are required for universal fault-tolerant quantum computation reduces to the following: which non-stabilizer states are necessary and sufficient to promote stabilizer computation to universal quantum computation?

In this thesis we identify a non-trivial closed, convex subset of the space of quantum states which is incapable of producing universal fault-tolerant quantum computation. In particular, this convex subset strictly contains the convex hull of stabilizer states, and thereby proves that there exists a class of *bound magic states*, i.e. states that can not be prepared from convex combinations of stabilizer states and yet are not useful for quantum computation. Thus the proof of the existence of bound magic states resolves in the negative the open problem raised by Bravyi and Kitaev [6] of whether *all* non-stabilizer states promote stabilizer computation to universal quantum computation. Furthermore, we will see that there is an efficient simulation algorithm for the subset of quantum theory that consists of operations from the stabilizer formalism acting on inputs from the non-universal region, which includes mixed states both inside and outside the convex hull of stabilizer states. This simulation scheme is an extension of the Gottesman-Knill theorem [26, 1] to a broader class of input states and should be of independent interest.

The results we have just described deal with the binary question of whether a given non-stabilizer state is at all useful for quantum computation. We further extend this to quantify the degree to which a given resource state is useful for promoting stabilizer computation to universal quantum computation. In magic state computation the required non-stabilizer unitary gates are implemented by using stabilizer operations to consume non-stabilizer resource states; this is closely analogous to how entangled states can be consumed using local operations to implement non-local operations. In order to avoid introducing errors

into the computation the non-stabilizer resource states that are consumed must be very pure. However, the only pure states with fault tolerant preperation are stabilizer states; the available resource states will generally be highly mixed. For this reason a critical step in magic state computation is the distillation of a large number of noisy resource states ρ_{res} into a small number of very pure non-stabilizer states σ_{target} that will be consumed to implement non-stabilizer gates. In this context the natural measure of how useful a resource state ρ_{res} is for quantum computation is the number of copies required to produce the target state σ_{target}. To quantify how useful a state is for quantum computation the question we must address is: assuming it is possible to use stabilizer operations to distill a target state σ_{target} from a resource state ρ_{res}, how efficiently can it be done? That is, how many copies of ρ_{res} are required to produce m copies of σ_{target}?

Finding distillation protocols to minimize the amount of resources required is an extremely important problem. Currently stabilizer codes provide the best hope for practical quantum computation, but the physical resource requirement for known distillation protocols is enormous. For example, reference [22] analyzes the requirements for using Shor's algorithm to factor a 2000 bit number using physical qubits with realistic error rates[1]. A surface code construction is used to achieve fault tolerance, from which it is found that roughly a billion physical qubits are required. About 94% of these physical qubits are used in the distillation of the ancilla states required to perform the non-stabilizer gates. More efficient distillation protocols are critical for the realization of quantum computation, and there has been a recent flurry of effort on this front eg. [16, 7, 22, 45, 54, 11]. Unfortunately, although these innovations offer improvement over the original magic state distillation protocols, the physical requirements remain extravagant. Moreover, it is unclear whether these protocols are near optimal or if dramatic improvements might still be made. The current work addresses this problem by developing a coherent theory for the treatment of resources for stabilizer computation.

The key insight is the identification of magic[2] as a resource that can not be created by stabilizer operations. This is analogous to the role of entanglement as a resource that can not be created by local operations. We discover two quantitative measures of the magic of a quantum state. These measures are *magic monotones*, functions mapping quantum states to real numbers that are non-increasing under stabilizer operations, ie. $\mathcal{M}(\Lambda(\rho)) \leq \mathcal{M}(\rho)$ if Λ is a stabilizer operation. As an important application these measures allow us to upper bound the efficiency of magic state distillation. For example, suppose the target state is five times as magical as the resource state according to some measure, then we can immediately infer that at least five resource states will be required for each copy of the target state. More broadly, these measures provide an important tool for development of improved magic state distillation protocols and the study of stabilizer computation.

The theoretical method used to prove many of these results is to construct a classical hidden variable model for the subtheory of quantum theory that consists of the stabilizer formalism and then determine the scope of additional quantum resources that are also de-

1 Physical qubit error rate 0.1%, ancilla preparation error rate 0.5% , 10ons for measurement

2 This somewhat whimsical name stems from two sources. First, the use of the magic moniker in the original Bravyi and Kitaev paper to describe states that are, apparently magically, both distillable and useful for state injection. Second, the long held desire of the present author to refer to himself as a mathemagician.

scribed by this model. In fact, this local hidden variable model is given by the non-negative elements of a distinguished quasi-probability representation. A quasi-probability representation is a way of representing quantum theory as a classical probabilistic theory where the quasi-probability distributions are allowed to take on negative values; such a representation thus affords a classical model for the elements of quantum theory with positive representation. Perhaps unsurprisingly, it has been shown that the full quantum theory can not be represented with positive elements in any such representation[18, 19, 20, 65]. However, one might expect that a subtheory of quantum theory that is inadequate for quantum computational speed-up might be represented non-negatively, i.e. as a true classical theory, in some natural choice of quasi-probability representation. For the context described above, we seek a quasi-probability representation reflecting our natural operational restriction: we require that stabilizer states and projective measurements onto stabilizer states have non-negative representation and that unitary stabilizer operations (i.e., Clifford transformations) correspond to stochastic processes. Conveniently, for quantum systems with odd Hilbert space dimension such a representation is already known to exist: this is the discrete Wigner function first defined by Wooters[79] and connected to the stabilizer formalism by Gross [34, 35]. In such a representation it is natural to examine whether the resources required for quantum speed-up correspond to those that are represented with negative probabilities.

The results covered by this thesis may now be stated more carefully:

Classically efficient simulation of positive Wigner functions: The set of fault tolerant quantum logic gates in the stabilizer formalism are known as the Clifford gates. The first contribution is an explicit simulation protocol for quantum circuits composed of Clifford gates acting on input states with positive discrete Wigner representation. We also allow arbitrary product measurements with positive discrete Wigner representation. This simulation is efficient (linear) in the number of input registers to the quantum circuit. This simulation scheme is an extension of the celebrated Gottesman-Knill theorem.

Negativity is necessary for magic state distillation: This simulation protocol implies that states outside the stabilizer formalism with positive discrete Wigner function (bound magic states) are not useful for magic state distillation. We give a direct proof of this fact exploiting only the observation that negative discrete Wigner representation can not be created by stabilizer operations. This proof has a more general range of applicability than the efficient simulation scheme and also makes clear the conceptual importance of negative quasi-probability as a *resource* for stabilizer computation.

Geometry of positive Wigner functions: The set of quantum states with positive discrete Wigner function strictly contains the set of (convex combinations of) stabilizer states. To prove this fact we fully describe the geometry of the region of quantum state space with positive discrete Wigner representation. Concretely, the facets of the classical probability simplex defining the discrete Wigner function are also facets of the polytope with the (pure) stabilizer states as its vertices. Since there are many more facets of the stabilizer polytope than of the simplex this suffices to establish the existence of non-stabilizer states with positive representation.

The Relative Entropy of Magic: The relative entropy distance between two states is $S(\rho\|\sigma) = \mathrm{Tr}(\rho\log\rho) - \mathrm{Tr}(\rho\log\sigma)$. We use this to define the *relative entropy of magic*

$\mathrm{r}_{\mathcal{M}}(\rho) \equiv \min_{\sigma \in \mathrm{STAB}(\mathcal{H}_d)} S(\rho \| \sigma)$, the minimum relative entropy distance between the resource state and any (convex combination of) stabilizer states. Roughly speaking, this is a measure of how distinguishable ρ is from any (convex combination of) stabilizer states. We establish that the relative entropy of magic is a magic monotone. This monotone is most interesting in the asymptotic regime, where it's asymptotic variant is $\mathrm{r}^{\infty}_{\mathcal{M}}(\rho) = \lim_{n\to\infty} \mathrm{r}_{\mathcal{M}}(\rho^{\otimes n})/n$. Using generic resource theory results[40] we show that if it is possible, using stabilizer operations, to reversibly interconvert between magic states ρ and σ with vanishingly small error in the asymptotic limit then the rate at which this can be done is given by $\mathrm{r}^{\infty}_{\mathcal{M}}(\rho)/\mathrm{r}^{\infty}_{\mathcal{M}}(\sigma)$.

Mana: The preceding results dealing with the discrete Wigner function are purely binary: if an ancilla preparation is positively represented then it can not promote stabilizer computation to universal computation, but if it is even very slightly negatively represented then none of the preceding work applies. It is natural to wonder whether the *amount* of negativity is meaningful measure of how useful a state is for promoting stabilizer operations to full quantum power. We show that this is indeed the case by proving that the sum of the negative entries of the Wigner representation of a state ρ is a magic monotone, the *sum negativity* $\mathrm{sn}(\rho)$. We will find it is actually more convenient to work with the *mana* $\mathscr{M}(\rho) \equiv \log(2\mathrm{sn}(\rho) + 1)$ because this measure is additive in the sense that $\mathscr{M}(\rho \otimes \sigma) = \mathscr{M}(\rho) + \mathscr{M}(\sigma)$. As a particular application we find explicit bounds on the efficiency of magic state distillation: to distill m copies of a target state σ from n copies of a resource state ρ at least $n \geq m\frac{\mathscr{M}(\sigma)}{\mathscr{M}(\rho)}$ copies are required on average.

The first three results first appeared in [71], a collaboration with Christopher Ferrie, David Gross and Joseph Emerson. The final two results will appear in [72], a collaboration with Ali Hamed, Daniel Gottesman and Joseph Emerson.

It is important to bear in mind that the results based on the discrete Wigner function only apply to systems of qudits with odd Hilbert space dimension as the discrete Wigner function is only defined for such systems. It remains to be established whether this distinction between odd and even Hilbert space dimension is merely mathematical frippery (as in the case of error correction, which requires a similar distinction between bits and dits) or if it reveals something deep about the quantum formalism.

Conceptually, the main contributions of this thesis are:

1. Casting quantum computation as a resource theory, allowing precise, quantitative statements to be made about how useful any particular quantum resource is for computational speedup and,

2. Showing that the negativity of the discrete Wigner function has a well defined, concrete meaning as a measure of how useful a quantum state is for promoting stabilizer computation to universal quantum computation.

It is natural to expect that these insights can be applied to the study of quantum phenomena beyond computation (eg. to the study of quantum communication). One of the main goals of this thesis is to give a comprehensive explanation of the intuition and reasoning that lead to the results in order to clarify how they might be exported to the broader study of quantum theory.

2

BACKGROUND

There are three main topics we must introduce to explain the contents of this thesis: the stabilizer formalism, quasi-probability representations of quantum theory and the discrete Wigner function. Our aim is here both to introduce the mathematical formalism of each of these tools and explain their conceptual significance. In particular we aim to make it clear how all of these are related to quantum computation and to each other. As such the presentation we give of these topics is not canonical. The stabilizer formalism is usually introduced in the context of fault tolerant quantum computation, quasi-probability representations are usually introduced in the context of quantum foundations and discrete Wigner functions are usually introduced by analogy to the famous infinite dimensional counterpart. Our perspective does not align with any of these, so even experts in these subjects may find some new insights in this chapter.

One of the main goals of this thesis is to give future workers a solid grounding for extending the results contained herein. A consequence is that we go into considerably more detail than the minimum required to articulate the results. Indeed, this is the main contribution of the thesis beyond what is already contained in the associated papers[71, 73, 72].

A number of the more advanced topics in the background chapters, particularly relating to the stabilizer formalism, make use of the representation theory of finite groups. Hopefully the conceptual points are still clear to readers who are not familiar with the mathematics. In any case this material is not required to understand the remainder of the thesis so the reader who is intimidated by the mathematics is free to skip it.

2.1 THE STABILIZER FORMALISM

Known schemes for fault tolerant quantum computation allow for only a limited set of operations to be implemented directly on the encoded quantum information. For most known fault tolerance schemes this restricted set is the stabilizer operations consisting of preparation and measurement in the computational basis and a restricted set of unitary operations. We will now review the important parts of its structure for systems of power of prime dimension. For an overview of the stabilizer formalism in the context of fault tolerance see [26, 30]. For an overview of the phase space techniques for the stabilizer formalism see [34, 36].

2.1.1 *The Basics*

We begin by defining the generalized Pauli operators for prime dimension and we will build up the formalism from these. We will denote Hilbert space of dimension d by $\mathcal{H}_d$ and the standard (computational) basis by $\{|j\rangle\}_{j=0\ldots d-1}$. Let p be a prime number and define the boost and shift operators $X, Z \in L(\mathcal{H}_p)$:

$$\begin{aligned} X|j\rangle &= |j+1 \mod p\rangle \\ Z|j\rangle &= \omega^j |j\rangle, \ \omega = \exp\left(\frac{2\pi\imath}{p}\right). \end{aligned}$$

From these we can define the Heisenberg-Weyl (generalized Pauli) operators in prime dimension:

$$T_{(a_1,a_2)} = \begin{cases} Z^{a_1}X^{a_2} & p = 2 \\ \omega^{-2^{-1}a_1a_2}Z^{a_1}X^{a_2} & p \neq 2 \end{cases} \tag{2.1.1}$$

where $a_1, a_2 \in \mathbb{Z}_p$, $\mathbb{Z}_p$ are the integers modulo p and $2^{-1} = \frac{p+1}{2} \mod p$.[1] For a system with composite Hilbert space $\mathcal{H}_{p_A} \otimes \mathcal{H}_{p_B} \otimes \cdots \otimes \mathcal{H}_{p_V}$ the Heisenberg-Weyl operators are the tensor product of the subsystem Heisenberg-Weyl operators:

$$T_{(a_1,a_2)\oplus(b_1,b_2)\cdots\oplus(v_1,v_2)} \equiv T_{(a_1,a_2)} \otimes T_{(b_1,b_2)} \cdots \otimes T_{(v_1,v_2)}.$$

The vector $(a_1, a_2) \oplus (b_1, b_2) \cdots \oplus (v_1, v_2) = (a_1, a_2, b_1, b_2, \ldots, v_1, v_2)$ is an element of $(\mathbb{Z}_{p_A} \times \mathbb{Z}_{p_A}) \times (\mathbb{Z}_{p_B} \times \mathbb{Z}_{p_B}) \cdots \times (\mathbb{Z}_{p_V} \times \mathbb{Z}_{p_V})$.

On $\mathcal{H}_d$ the Clifford operators, $\mathcal{C}_d$, are the set of unitary operators that, up to phases, map Heisenberg-Weyl operators to Heisenberg-Weyl operators under conjugation:

$$U \in \mathcal{C}_d \iff \forall \boldsymbol{u} \exists \phi, \boldsymbol{u}' : \ UT_{\boldsymbol{u}}U^\dagger = \exp(i\phi)\, T_{\boldsymbol{u}'}.$$

These operators form a group, which in the context of quantum information is known as the Clifford group.

The pure stabilizer states for dimension d are defined as

$$\{S_i\} = \left\{U|0\rangle\langle 0|U^\dagger : U \in \mathcal{C}_d\right\},$$

and we take the full set of stabilizer states to be the convex hull of this set:

$$\mathrm{STAB}(\mathcal{H}_d) = \left\{\sigma \in L(\mathcal{H}_d) : \ \sigma = \sum_i p_i S_i \text{ for some probability distribution } p_i\right\}.$$

The pure stabilizer measurements are the projective valued measurements composed of projectors onto stabilizer states. The full set of stabilizer measurements are convex combinations of the pure stabilizer measurements.

Finally, we define the stabilizer subtheory of quantum theory to be the collection of all stabilizer operations.

1 For qubits an alternative definition where $T^{\text{usual}}_{(1,1)} = Y \equiv -\imath ZX$ is often used. From the perspective of the present work this definition is confusing bordering on wrong, as explained in detail below.

2.1.2 *Phase Space Techniques*

The Clifford group on a Hilbert space of dimension p^n has a close relationship with the group of symplectic matrices of size $2n$ over the finite field $\mathbb{Z}_p$. This connection is the source of much of the interesting structure of the stabilizer subtheory, including the fact that it can be efficiently classically simulated. Understanding this connection is critical for much of the content to follow. Here we restrict ourselves to odd p because the treatment is slightly simpler and this is the case that we are primarily interested in for this thesis. The qubit case is nearly identical except that some equations must carry a physically irrelevant global phase.

For the Clifford group $\mathcal{C}_p$ on a Hilbert space with prime dimension p the basic object of study is the phase space,

$$V := \mathbb{Z}_p \times \mathbb{Z}_p,$$

which should be thought of as a finite size analogue of the usual $\mathbb{R} \times \mathbb{R}$ phase space of classical mechanics. We imbue this set with a symplectic inner product:

$$\left\langle \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \right\rangle = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

where -1 is $p-1 \mod p$. The connection with the Heisenberg-Weyl operators on $\mathcal{H}_p$ comes from their commutation relations[2]:

$$T_{\boldsymbol{a}} T_{\boldsymbol{b}} = \omega^{\langle \boldsymbol{a}, \boldsymbol{b} \rangle} T_{\boldsymbol{b}} T_{\boldsymbol{a}}.$$

A 2×2 matrix F with entries from $\mathbb{Z}_p$ is *symplectic* if

$$F^T \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} F = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

ie. if it preserves the symplectic inner product. Notice that the mapping

$$T_{\boldsymbol{a}} \to T_{F\boldsymbol{a}}$$

defines an automorphism of the Heisenberg-Weyl operators since the commutation relations are preserved. In fact, it is always possible to find a unitary operator U_F on $\mathcal{H}_p$ such that,[3]

$$U_F T_{\boldsymbol{a}} U_F^\dagger = T_{F\boldsymbol{a}}.$$

This is a unitary representation of the group of all 2×2 symplectic matrices with entries in $\mathbb{Z}_p$, $\mathrm{Sp}(2, p)$. Obviously these symplectic unitary operators are in $\mathcal{C}_p$. It is also clear that the Heisenberg-Weyl operators are also in $\mathcal{C}_p$. What is not so obvious is that this is the whole story: up to a phase every operator $U \in \mathcal{C}_p$ can be written as:

$$U = U_F T_{\boldsymbol{a}},$$

2 Actually, the fact that this has such a nice form is because of the clever choice of phases in definition (2.1.1). The connection would still be there with any other choice of phases, but it would not be so clearly manifest.

3 for $p = 2$ this is only true up to phase, see below.

where $F \in \mathrm{Sp}(2,p)$, $a \in \mathbb{Z}_p \times \mathbb{Z}_p$.

For power of prime dimension p^n the story turns out to be much the same. Notice however that we have some freedom for how we define the symplectic group. We could choose either,

$$\begin{aligned} V &= \left(\mathbb{Z}_p \times \mathbb{Z}_p\right)^n \text{ or} \\ \tilde{V} &= \mathbb{F}_{p^n} \times \mathbb{F}_{p^n}, \end{aligned}$$

where $\mathbb{F}_{p^n}$ is the finite field with p^n elements. There is a bijection between these objects considered as sets. Our definition of the Heisenberg-Weyl operators for power of prime dimension meshes most easily with the first choice, but in fact we could have just as well indexed the operators with the elements of $\tilde{V}$. However, the 'natural' symplectic groups are not the same. We can work with either,

$$\begin{aligned} \mathrm{Sp}(2n,p) &= \left\{2n \times 2n \text{ symplectic matrices with entries in } \mathbb{Z}_p\right\} \text{ or} \\ \mathrm{Sp}(2,p^n) &= \left\{2 \times 2 \text{ symplectic matrices with entries in } \mathbb{F}_{p^n}\right\}. \end{aligned}$$

Where a $2n \times 2n$ matrix F is symplectic if $F^T \begin{pmatrix} \mathbf{0}_n & -\mathbb{I}_n \\ \mathbb{I}_n & \mathbf{0}_n \end{pmatrix} F = \begin{pmatrix} \mathbf{0}_n & -\mathbb{I}_n \\ \mathbb{I}_n & \mathbf{0}_n \end{pmatrix}$. It turns out that,

$$\mathrm{Sp}(2,p^n) \lneq \mathrm{Sp}(2n,p).$$

In fact $\mathrm{Sp}(2n,p)$ is exponentially larger. Again, for every element $F \in \mathrm{Sp}(2n,p)$ there is a unitary operator U_F such that $U_F T_{\boldsymbol{a}} U_F^\dagger = T_{F\boldsymbol{a}}$. Moreover, it is again true that every Clifford operator in $\mathcal{C}_{p^n}$ may be written as,

$$U = U_F T_{\boldsymbol{a}},$$

where $F \in \mathrm{Sp}(2n,p)$, $a \in \mathbb{Z}_{p^n} \times \mathbb{Z}_{p^n}$.

It is an interesting fact that there is a whole hierarchy of symplectic groups,

$$\mathrm{Sp}(2,p^n) \lneq \mathrm{Sp}\left(4,p^{n-1}\right) \lneq \cdots \lneq \mathrm{Sp}\left(2(n-1),p^2\right) \lneq \mathrm{Sp}(2n,p).$$

Each of these admits a faithful representation on H_{p^n} and preserve much of the interesting structure of the full Clifford group. In particular, the group of operators $\{U_F T_{\boldsymbol{a}}\}$ is a unitary 2-design for F taken from any of these groups. This may have useful applications for quantum process tomography using randomized benchmarking[51, 49].

For our purposes the major significance of these phase space techniques is that every Clifford operation admits a description that is linear in n. This is at the heart of simulation results about the Clifford group.

For further reading, particularly on different choices of symplectic group for power of prime, see section IV of [36].

2.1.3 *Magic State Distillation*

It is possible to implement stabilizer operations fault tolerantly, but these operations are not sufficient for universal quantum computation. To promote stabilizer computation to universal quantum computation some additional resource is required. One possible form for this

resource is a pure non-stabilizer state $|\psi\rangle$. Using only stabilizer operations this resource state can be consumed in order to implement some non-Clifford gate $U_{|\psi\rangle}$ and using this technique it can be shown that the stabilizer operations supplemented with the ability to prepare an arbitrary non-stabilizer pure state $|\psi\rangle$ are universal for quantum computation[11]. However, since our motivating scenario is one in which only stabilizer operations have a fault tolerant implementation we generally expect that any physically accessible non-stabilizer state preparation procedure will be highly noisy.

In this context there are two critical questions that must be answered: which non-stabilizer resource states can be used to promote stabilizer computation to universal quantum computation and how can this be done? The first of these questions is one of the main subjects of this thesis. The second question finds a particularly elegant solution in the form of magic state distillation [6] (see also [2, 11] for the extension from qubits to qudits).

Magic state distillation protocols aim to consume a large number of copies of a noisy non-stabilizer qudit input state ρ_{in} to produce a single non-stabilizer qudit output state σ_{out} with higher fidelity to some non-stabilizer pure state. This output state is then consumed to implement some non-Clifford unitary gate. These protocols typically have the following structure:

- Prepare a number of copies of the input state $\rho_{\text{in}}^{\otimes n}$.
- Perform some Clifford gate on $\rho_{\text{in}}^{\otimes n}$.
- Make a computational basis measurement on the last $n-1$ registers and post select on the outcome.

When these protocols succeed the first register will be the output state σ_{out}. Typically these protocols work iteratively, repeatedly consuming $\rho_{\text{in}}^{\otimes n}$ until n copies of $\sigma_{\text{out},1}$ have been produced and then using $\sigma_{\text{out},1}^{\otimes n}$ as the input to the protocol to produce $\sigma_{\text{out},2}$ with even higher fidelity to the target pure state. This is repeated until a sufficiently high fidelity is reached, ultimately producing a single high fidelity output state $\sigma_{\text{out},k}$ by consuming n^k resource states ρ_{in}, where k is the number of iterations.

In this thesis we are interested in a broader set of distillation protocols than what is encompassed above. We study the conversion of input states $\rho_{\text{in}}^{\otimes n}$ to output states $\sigma_{\text{out}}^{\otimes m}$ using arbitrary stabilizer operations combined in any fashion. Moreover, we are interested in the conversion between arbitrary non-stabilizer states. In this context it is natural to define a magic state to be any non-stabilizer state in the same way that an entangled state is any non-separable state.

Definition 1. A state is magic if it is not a stabilizer state.

The most general kind of stabilizer operation possible is of the following type:

Definition 2. A stabilizer protocol is any map from $\mathcal{S}(\mathcal{H}_{d^n})$ to $\mathcal{S}(\mathcal{H}_{d^m})$ composed from the following operations:

1. Clifford unitaries, $\rho \to U\rho U^\dagger$
2. Composition with stabilizer states, $\rho \to \rho \otimes S$ where S is a stabilizer state

3. Post selected computational basis measurement on the final qudit,

$$\rho \to (\mathbb{I} \otimes |i\rangle\langle i|)\, \rho\, (\mathbb{I} \otimes |i\rangle\langle i|) \,/\, \mathrm{Tr}\,(\rho \mathbb{I} \otimes |i\rangle\langle i|)$$

with probability $\mathrm{Tr}\,(\rho \mathbb{I} \otimes |i\rangle\langle i|)$

4. Partial trace of the final qudit, $\rho \to \mathrm{Tr}_n\,(\rho)$

and classical randomness.

These operations may take place in any order and the $(n+1)$th operation may depend on the outcomes of any measurements in the first n operations in any fashion.

The goal of magic state distillation is to consume resource magic states ρ_{res} to produce some (very nearly) pure state $|\psi\rangle\langle\psi|_{\mathrm{target}}$. We will say a state is magic state distillable if it is possible to convert any number of copies of it to at least one copy of any non-stabilizer pure state in the asymptotic limit, ie.

Definition 3. A state $\rho \in \mathcal{H}_d$, $d = p^k$ is magic state distillable if there is some family of stabilizer protocols $\Lambda_n\,(\rho^{\otimes n}) = \tilde{\sigma}_n$ and some pure non-stabilizer state $|\psi\rangle\langle\psi|$ such that $\tilde{\sigma}_n$ becomes arbitrarily close to $|\psi\rangle\langle\psi|$ in the 1-norm, i.e., $\lim_{n\to\infty} \|\tilde{\sigma}_n - |\psi\rangle\langle\psi|\|_1 \to 0$.

We will often use the terms "magic state distillable" and simply "distillable" interchangeably. One of the main results of this thesis is the demonstration that not all magic states are distillable.

2.1.4 *Why the Heisenberg-Weyl operators?*

One important reason for studying the stabilizer formalism is its importance for fault tolerant quantum computation. However, the formalism has a remarkable amount of depth that is quite surprising if we adopt this perspective. For instance, we know that the stabilizer formalism supplemented with any non-stabilizer pure state is universal for quantum computation[11] and that the stabilizer formalism supplemented with any (mixed) non-stabilizer state suffices to demonstrate quantum contextuality[43]. It seems strange that there should be any connection at all between error correcting codes with high threshold and the "maximal classical subtheory" afforded by these observations. This leads us to look for some alternative motivation for the stabilizer subtheory. In effect, this amounts to finding some deeper motivation for the Heisenberg-Weyl operators. Such a motivation does exist and in fact predates quantum computation entirely.

We begin by recalling some facts about quantum mechanics on the real line $\mathbb{R}$. In this case we may begin with position $\hat{q}$ and momentum $\hat{p}$ operators satisfying the canonical commutation relations,

$$[\hat{p}, \hat{q}] = \imath\hbar.$$

It is easy to see that this expression has no finite dimensional analogue (just take the trace of both sides). However, we can equivalently study the group obtained by exponentiating the

Lie algebra defined by the commutation relation. This is the Heisenberg group $H_1(\mathbb{R}) \subset \mathrm{GL}(\mathbb{R})$,

$$H_1(\mathbb{R}) = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}, \; a, b, c \in \mathbb{R} \right\}.$$

At this juncture we are thinking of both this group and its associated lie algebra as abstract entities; to recover the familiar quantum formalism we must move to some concrete representation of the Heisenberg group. The familiar way to do this is via:

$$\begin{aligned} \left(\exp\left(\imath \hat{p} a\right) \psi\right)(x) &= \psi(x + a) \\ \left(\exp\left(\imath \hat{q} b\right) \psi\right)(x) &= e^{\imath b x} \psi(x), \end{aligned}$$

where $\psi \in L^2(\mathbb{R})$[4]. One might expect that there would be many physically inequivalent unitary representations of $H_3(\mathbb{R})$, in which case the usual (Schrödinger) picture of quantum theory would be only part of the whole story. This is not the case: the usual representation is essentially unique, which is the primary content of the Stone-von Neumann theorem.

Instead of dealing with positions on $\mathbb{R}$ we could study positions on a finite (toroidal)[5] set of positions indexed by $\mathbb{Z}_p$, for p a prime. Although the canonical commutation relations do not have an analogue in this scenario we can easily think of one for the Heisenberg group:

$$H_1(\mathbb{Z}_p) = \left\{ \begin{pmatrix} 1 & x & c \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix}, \; x, z, c \in \mathbb{Z}_p \right\}.$$

This can be interpreted as the group of position and momentum translations in a finite phase space $\mathbb{Z}_p \times \mathbb{Z}_p$. It is important to be clear that at this point the finite Heisenberg group (also called extra special groups[78]) has been defined by fiat; our strategy of simply replacing $\mathbb{R}$ by $\mathbb{Z}_p$ carries an undisguisable arbitrariness. Happily, we can offer some post-hoc justification in the form of the discrete Stone-von Neumann theorem (see eg. [34, 78]):

Theorem 4. *The unitarily inequivalent irreducible representations of $H_1(\mathbb{Z}_p)$ are:*

1. *p^2 irreducible representations of dimension 1. These are the representations of the abelian group obtained by modding out the center of the group ($H_1(\mathbb{Z}_p) / Z(H_1(\mathbb{Z}_p)) \simeq \mathbb{Z}_p \times \mathbb{Z}_p$).*

4 It might help intuition for how to pass from infinite dimension to finite dimension to point out that $\mathcal{H}_{p^n} \simeq L^2\left(\mathbb{Z}_p^n\right)$ by $|\psi\rangle \leftrightarrow \psi(q) = \langle q|\psi\rangle$.

5 This is just a fancy way of saying that counting 'loops back around', eg. $(p-1)+3 = 2 \mod p$.

2. *$p-1$ irreducible representations of dimension p. The $k = 1 \ldots p-1$ representations for the generators of the group may be given as:*

$$\begin{aligned} U^{(k)}\left(\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) &= X^k \\ U^{(k)}\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}\right) &= Z \\ U^{(k)}\left(\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) &= \omega^k \mathbb{I}_p. \end{aligned}$$

That is, this group affords only 1 and p dimensional irreducible representations and the distinction between the unitarily inequivalent p dimensional representations is not physically meaningful, since it amounts only to a reordering of the computational basis and a change in the physically irrelevant global phase.

If we now allow ourselves n systems with p possible positions each the appropriate Heisenberg group is:

$$H_n\left(\mathbb{Z}_p\right) = \left\{ \begin{pmatrix} 1 & \boldsymbol{x}^T & c \\ \mathbf{0} & \mathbb{I}_n & \boldsymbol{z} \\ 0 & \mathbf{0} & 1 \end{pmatrix}, \ \boldsymbol{x}, \boldsymbol{z} \in \mathbb{Z}_p^n, \ c \in \mathbb{Z}_p \right\}$$

and again it is true that the Heisenberg-Weyl operators of dimension p^n form (with phases) an irreducible representation of the group and that this representation is essentially unique.

We have been led to the Heisenberg-Weyl operators in prime dimension by looking for an analogue of the canonical infinite dimensional position and momentum operators. We can now understand the stabilizer subtheory as a natural quantization of a finite size classical phase space. From this perspective the close connection between stabilizers and quantum phenomena (such as computational speedup) seems very natural. This also makes the relationship between the formalism of linear optics and the stabilizer formalism manifest.

2.1.5 *What's up with qubits?*

It is very common in the study of finite fields $\mathbb{F}_{p^n} \simeq \mathbb{Z}_p^n$ to find that $p = 2$ is an exceptional case[6]. In the context of the stabilizer formalism this fact manifests itself by the fact that qubits require a slightly different treatment than qu*p*its where $p \neq 2$. An enormous amount of confusion has arisen from this point, including the totally inappropriate name "Clifford

6 As the only even prime, two is the oddest prime of all.

group" for $\mathcal{C}_d$. Here we try to resolve some of this confusion by giving a clear articulation of the difference between $p = 2$ and $p \neq 2$.

If p is an odd prime then the Heisenberg group $H_3(\mathbb{Z}_p)$ has exponent p[78], meaning that for all $g \in H_3(\mathbb{Z}_p)$ it holds that $g^p = \mathbb{I}$, where $\mathbb{I}$ is the identity element of the group. This is reflected by the fact that every generalized Heisenberg-Weyl operator acting on $\mathcal{H}_p$ satisfies $T^p_{(a_1,a_2)} = \left(\omega^{-2^{-1}a_1a_2} Z^{a_1} X^{a_2}\right)^p = \mathbb{I}$. For $p = 2$ this is not true: $(XZ)^2 = -\mathbb{I}$, i.e., the Heisenberg-Weyl operator has order 4 instead of order 2. In fact, this is as it should be. For $p = 2$ the Heisenberg group is isomorphic to the dihedral group of order 8, $H_3(\mathbb{Z}_2) \simeq D_8$, and this group has two elements of order 4. Although it may seem strange that $p = 2$ is an exceptional case in this sense *there is no problem, either conceptual or mathematical, with our definition of the Heisenberg-Weyl operators for qubits*. The operators are elements of a faithful irreducible representation of the Heisenberg group.

Nevertheless, it is very common to see an alternative definition used for qubits where $T^{\text{usual}}_{(1,1)} = \imath ZX$[7]. The root of the confusion is due to the fact that the Heisenberg-Weyl operators have (at least) two distinct roles:

1. The qubit Heisenberg-Weyl operators defined by equation (2.1.1) are (elements of) the 2 dimensional irreducible representation of $H_3(\mathbb{Z}_2)$ given by the representation of the group generators,

$$\begin{aligned} U\left(\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) &= X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ U\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}\right) &= Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ U\left(\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) &= -\mathbb{I}_2. \end{aligned}$$

   As we have seen this is the view that is most appropriate for the study of the stabilizer formalism.

2. The qubit (Pauli) operators X, Z and $Y = \imath XZ$ are a representation of the Clifford algebra of $\mathbb{R}^3$ (the quaternions). This is equivalent to their role as a representation of generators of rotation in $\mathbb{R}^3$.

For a physicist already familiar with the Pauli matrices in their second role it is very easy to go astray. This is particularly true since $(\imath XZ)^2 = \mathbb{I}$, i.e., imposing the extra factor $\imath$ has the effect of forcing the operators to have exponent $p = 2$ in accordance with our wrong

7 In fact, the earliest publications on the stabilizer formalism actually used the definition $Y = XZ$ we use here, eg. [28]. The convention $Y = \imath XZ$ was only introduced later, apparently an error induced by buggy error correction of the earlier work.

intuition that this should hold for $p = 2$ as well as $p \neq 2$! Moreover, since the matrices differ only by phases that are physically irrelevant in quantum mechanics the representations of the two things are virtually identical even though the abstract objects and their conceptual bases are totally different. This confusion is almost certainly the reason that $\mathcal{C}_{p^n}$ was named the Clifford group.

The distinction for the full Clifford group is somewhat less benign. Recall that we defined the Clifford group to be the unitary operators that map Heisenberg-Weyl operators to Heisenberg-Weyl operators *up to a phase*. For odd prime dimension we can say something more precise. We define $\mathcal{W}_{p^n}$ to be the cannonical p^n dimensional representation of the finite Heisenberg group (see Theorem 4 for the prime case), i.e., $\mathcal{W}_{p^n}$ is the Heisenberg-Weyl operators supplemented with the complex phases $\omega = e^{\frac{2\pi}{p}\imath}$,

$$\begin{aligned}\mathcal{W}_{p^n} \quad = \quad & \langle X \otimes \mathbb{I} \otimes \cdots \otimes \mathbb{I},\ \mathbb{I} \otimes X \otimes \cdots \otimes \mathbb{I}, \ldots, \mathbb{I} \otimes \mathbb{I} \otimes \cdots \otimes X, \\ & Z \otimes \mathbb{I} \otimes \cdots \otimes \mathbb{I},\ \mathbb{I} \otimes Z \otimes \cdots \otimes \mathbb{I}, \ldots, \mathbb{I} \otimes \mathbb{I} \otimes \cdots \otimes Z, \\ & \omega\mathbb{I}\rangle.\end{aligned}$$

For odd prime p we can give a sharper definition of the Clifford group[8]:

$$\mathcal{C}_{p^n} = \left\{ U \in U\left(p^n\right) \mid UT_{\boldsymbol{u}}U^\dagger \in \mathcal{W}_{p^n}\ \forall T_{\boldsymbol{u}} \in \mathcal{W}_{p^n} \right\},\ p \text{ odd}.$$

That is, the complex phases are just the roots of unity ω required for $\mathcal{W}_{p^n}$ to be a group. This is not the case for qubits. Consider the phase gate $P = \begin{pmatrix} 1 & 0 \\ 0 & \imath \end{pmatrix} \in \mathcal{C}_2$, then:

$$PXP^\dagger = -\imath ZX,$$

but X, Z and $-\mathbb{I}$ are all real matrices so $\imath XZ$ can not possibly belong to $\mathcal{W}_2 = \langle X, Z, -\mathbb{I}\rangle$. The phase gate takes $\mathcal{W}_2$ to itself only up to a factor of $\imath$. In fact, this is generic for any gate in $\mathcal{C}_{2^n}$: if we define $\overline{\mathcal{W}}_{2^n} = \{1, \imath\} \times \mathcal{W}_{2^n}$ then

$$\mathcal{C}_{2^n} = \left\{ U \in U\left(2^n\right) \mid UT_{\boldsymbol{u}}U^\dagger \in \overline{\mathcal{W}}_{2^n}\ \forall T_{\boldsymbol{u}} \in \mathcal{W}_{2^n} \right\}.$$

Actually, it's quite common (possibly ubiquitous) in the quantum information literature to see the Pauli group to be defined as the set $\overline{\mathcal{W}}_{2^n}$ equipped with matrix multiplication (eg. [27, 29, 56]). $\overline{\mathcal{W}}_2$, for instance, has 16 elements and is thus *not* isomorphic to the (2 dimensional irrep of the) Heisenberg group, $\overline{\mathcal{W}}_2 \not\simeq \mathcal{W}_2 \simeq H_1\left(\mathbb{Z}_2\right)$ (nor is $\overline{\mathcal{W}}_2$ a representation of the Clifford algebra of $\mathbb{R}^3$ so this group is neither of the roles of the Pauli operators described above).

The symplectic structure

I've separated out the discussion of the symplectic structure for qubits vs qudits because I'm signficantly less confident about how it works; the following should be taken with a grain of salt.

8 The equivalence with our earlier definition is implicit in the proof of theorem 3 of [34].

For every $U \in \mathcal{C}_{p^n}$ there is some symplectic matrix $F \in \mathrm{Sp}(2n, p)$ such that:

$$UT_{\boldsymbol{v}}U^{\dagger} = e^{\imath\phi_U(\boldsymbol{v})}T_{F\boldsymbol{v}} \ \forall \boldsymbol{v} \in \mathbb{Z}_{p^n},$$

where the phase $\phi_U(\boldsymbol{v})$ is a function of $\boldsymbol{v}$ and the particular function might depend on U. Thus, up to the phases, a Clifford operator is specified by a symplectic matrix F. Moreover, for every $F \in \mathrm{Sp}(2n, p)$ there is an element of the Clifford group $U_F \in \mathrm{Sp}(2n, p)$ such that[34],

$$U_F T_{\boldsymbol{v}} U_F^{\dagger} = e^{\imath\phi_{U_F}(\boldsymbol{v})}T_{F\boldsymbol{v}} \ \forall \boldsymbol{v} \in \mathbb{Z}_{p^n}.$$

The Heisenberg-Weyl operators are obviously in $\mathcal{C}_{p^n}$. From the Heisenberg-Weyl commutation relations we have that $T_{\boldsymbol{a}}T_{\boldsymbol{v}}T_{\boldsymbol{a}}^{\dagger} = \omega^{\langle \boldsymbol{a},\boldsymbol{v}\rangle}T_{\boldsymbol{v}}$. Thus for any operator $U_F \in \mathcal{C}_{p^n}$ all operators $U_F T_{\boldsymbol{a}}$, $T_{\boldsymbol{a}} \in \mathcal{W}_{p^n}$ implement the same symplectic transformation on $\mathcal{W}_{p^n}$, differing only in the phase they assign the operators. Notice also that any operator of the form $U_F T_{\boldsymbol{a}} e^{\imath\xi}$ will obviously have the same action as $U_F T_{\boldsymbol{a}}$ on $\mathcal{W}_{p^n}$. In fact, these cover all the possibilities:

The Clifford group modulo the Heisenberg Weyl group is isomorphic to the symplectic group, i.e., $\left(\mathcal{C}_{p^n}/\mathcal{W}_{p^n}\right)/U(1) \simeq \mathbf{Sp}(2n, p)$.[9] This holds for any prime p.

The distinction between the $p = 2$ and $p \neq 2$ case come from the way in which the matrix multiplication in $\mathcal{C}_{p^n}$ reflects group multiplication in $\mathrm{Sp}(2n, p)$. For any $F, G \in \mathrm{Sp}(2n, p)$ we take the corresponding $U_F, V_G \in \left(\mathcal{C}_{p^n}/\mathcal{W}_{p^n}\right)$

$$\begin{aligned} U_F V_G T_{\boldsymbol{v}} V_G^{\dagger} U_F^{\dagger} &= e^{\imath\phi_{V_G}(\boldsymbol{v})} U_F T_{G\boldsymbol{v}} U_{F^{\dagger}} \\ &= e^{\imath\phi_{U_F}(G\boldsymbol{v})} e^{\imath\phi_{U_G}(\boldsymbol{v})} T_{FG\boldsymbol{v}}, \end{aligned}$$

and for $U_{FG} \in \left(\mathcal{C}_{p^n}/\mathcal{W}_{p^n}\right)$

$$W_{FG} T_{\boldsymbol{v}} W_{FG}^{\dagger} = e^{\imath\phi_{W_{FG}}(\boldsymbol{v})} T_{FG\boldsymbol{v}},$$

so we conclude that, for some phase ϕ:

$$W_{FG} = U_F V_G e^{\imath\phi}.$$

That is, the group multiplication of $\mathrm{Sp}(2n, p)$ is reproduced by matrix multiplication of the operators $\{U_F\}$ up to a phase. The discussion of this paragraph may be summarized as:

The operators $\mathcal{C}_{p^n}/\mathcal{W}_{p^n}$ are a *projective* representation of $\mathbf{Sp}(2n, p)$. This holds for any prime p.

For odd primes $p \neq 2$ we can say more. For any $U \in \mathcal{C}_{p^n}$ there is some symplectic matrix $F \in \mathrm{Sp}(2n, p)$ such that

$$UT_{\boldsymbol{v}}U^{\dagger} = e^{\imath\phi_U(\boldsymbol{v})}T_{F\boldsymbol{v}} \ \forall T_{\boldsymbol{v}} \in \mathcal{W}_{p^n},$$

9 The $U(1)$ is just the freedom $U_F T_{\boldsymbol{a}} \to U_F T_{\boldsymbol{a}} e^{i\xi}$.

where the phase $e^{\imath\phi_U(v)} = \omega^{\langle a_U, v\rangle}$ for some $a_U \in \mathbb{Z}_p^n \times \mathbb{Z}_p^n$[34]. The distinction is that we now have an explicit description for the function $\phi_U(v)$. What's more, since $T_a T_v T_a^\dagger = \omega^{\langle a,v\rangle}$ this means that any Clifford operator may be written as:

$$U_{F,a} = U_F T_a.$$

Thus for any $F \in \mathrm{Sp}(2n,p)$ there is a unique operator $U_F \in \mathcal{C}_{p^n}/\mathcal{W}_{p^n}$ such that $U_F T_v U_F^\dagger = T_{Fv}\ \forall T_v \in \mathcal{W}_{p^n}$. Since now there are no mystery phases to muck things up this establishes:

For odd prime p the operators $\mathcal{C}_{p^n}/\mathcal{W}_{p^n}$ are a *linear* representation of Sp $(2n,p)$.

For $p = 2$ we know already that the situation must be different. For odd primes it holds that for any $U \in \mathcal{C}_p$ with action $UT_u U^\dagger = e^{\imath\phi} T_{Fu}$ it is possible to find $T_a \in \mathcal{W}_p$ such that $U_F \equiv UT_a^\dagger$ has the action $U_F T_u U_F^\dagger = T_{Fu}$; it is possible to kill the phase by the action of a Heisenberg-Weyl operator. Again consider the phase gate $P = \begin{pmatrix} 1 & 0 \\ 0 & \imath \end{pmatrix} \in \mathcal{C}_2$. The action of this gate on the $X = T_{(0,1)}$ operator is:

$$PT_{(0,1)}P^\dagger \quad = \quad -\imath T_{(1,1)},$$

Taking $\tilde{P} = PT_a$ we see that

$$\tilde{P}T_{(0,1)}\tilde{P}^\dagger = (-1)^{\langle a,(0,1)\rangle} \cdot \left(-\imath T_{(1,1)}\right).$$

The point is that the complex phase persists for any choice of a so the argument that was used in for $p \neq 2$ can't possibly apply here. Unfortunately, I don't have a clean statement of what the situation actually is. Using that $\mathcal{C}_2$ is generated by H and P one can check by brute force computation that $|\mathcal{C}_2| = 192$ so that $\left|\mathcal{C}_2/\overline{\mathcal{W}}_2\right| = 192/16 = 2 \cdot |\mathrm{Sp}(2,2)|$. Thus even modding out the "extra" phases $\pm\imath$ there are still too many elements for the quotient to even be isomorphic to the symplectic group, much less a faithful representation. Determining exactly how to describe the Cifford group for $p = 2$ would be a major step towards generalizing the discrete Wigner function to work for qubits. One hint might come from the pure mathematics literature on the automorphism groups of the extra special groups.

2.1.6 *Disambiguation of Heisenberg and Clifford Groups*

One of the very confusing aspects of these mathematics is that there exist in the literature several different objects called the Heisenberg-Weyl operators and several different objects called the Clifford group. These include:

A. The Heisenberg-Weyl operators that we use. These are irreducible representations of the finite Heisenberg groups $H_n(\mathbb{Z}_p)$. From these we get our definition of the Clifford group as the unitaries that map Heisenberg operators to Heisenberg operators up to a phase. This is what is normally meant by Clifford group in quantum information.

B. Let $V = \mathbb{Z}_p^n \times \mathbb{Z}_p^n$ be a vector space with symplectic inner product $\langle \cdot, \cdot \rangle$. Sometimes the Heisenberg group is defined as the set $V \times \mathbb{Z}_p$ equipped with the group product $(\boldsymbol{v}, a)(\boldsymbol{u}, b) = \left(\boldsymbol{v} + \boldsymbol{u}, a + b + 2^{-1} \langle \boldsymbol{v}, \boldsymbol{u} \rangle\right)$. This agrees with our definition *except when* $p = 2$, because 2^{-1} is not defined in $\mathbb{Z}_2$. Consequently our definition of the Heisenberg-Weyl operators and the Clifford group will agree for $p \neq 2$. This is the definition used by David Gross in his discrete Wigner function paper.[34]

C. For Hilbert space $\mathcal{H}_d$ define the X and Z Heisenberg-Weyl operators to be:

$$\begin{aligned} X|j\rangle &= |j+1 \mod d\rangle \\ Z|j\rangle &= \exp\left(\frac{2\pi\imath}{d}\right)|j\rangle \end{aligned}$$

and take the Heisenberg group to be the group generated by X and Z. The Heisenberg-Weyl operators are the elements of the group. This agrees with our definition in prime dimension, but is otherwise not the same. The Clifford group is then defined to be the unitaries that map these Heisenberg-Weyl operators to themselves up to a phase. Again, this definition agrees with ours in prime Hilbert space dimension but not otherwise. These are the definitions used by the SIC-POVM crowd, eg.[3].

D. A group related to Clifford algebras. This is what is meant by the Clifford group outside of quantum information.

2.2 HIDDEN VARIABLE THEORIES AND QUASI-PROBABILITY REPRESENTATION

2.2.1 *A Word on Classical Probabilistic Theories*

A great deal of the content of this thesis deals with the question of what can be inferred from the failure of quantum theory to behave as a classical probabilistic theory. As such, the first step in the exposition must be to make precise what we mean by a classical probabilistic theory.

We consider such a theory to have two main components: descriptions of our system given by probability distributions over some sample space and descriptions of our measurement apparati given by probability distributions of measurement outcomes such that the true definite state of the system is known.

The basic building block of the theory is a sample space Λ composed of the possible beliefs we might hold about the system. For example, in classical stochastic mechanics Λ is phase space ($\simeq \left(\mathbb{R}^2\right)^{\times n}$ for n particles) or for n flips of a coin we might take $\Lambda = \{\text{heads, tails}\}^{\times n}$ or we might take

$$\begin{aligned} \Lambda = \{&\text{mass of the coin, the details of how it was flipped,} \\ &\text{the rate of expansion of the universe, etc.}\} \end{aligned}$$

Once we have fixed the things we are describing our description of the state of the system is given by some probability distribution over the variables in Λ, ie.

$$\text{system description} \leftrightarrow p(\lambda). \tag{2.2.1}$$

We emphasize that the system is described by the distribution $p(\cdot)$ and *not* the value of the distribution at λ. Since it is always clear from context whether we are referring to a distribution or the value of a distribution at a point we will allow ourselves this abuse of notation in order to emphasize the role of the space Λ.

The description of measurement is slightly more complicated so we will build it up somewhat carefully to make it obvious. If we have access to a measurement of the form

$$M_\lambda = \begin{cases} 1 & \text{if the definite state is } \lambda \\ 0 & \text{otherwise } \left(\bar{\lambda}\right) \end{cases},$$

then the probability of getting outcome 1 is given by:

$$P\left(\text{outcome of } M_\lambda = 1\right) = 1 \cdot p(\lambda) + 0 \cdot p\left(\bar{\lambda}\right).$$

In general we must consider less well behaved measurements. Suppose that our apparatus M_λ^{noisy} is imperfect so that it will only click with some probability ($q(1|\lambda)$) even when the true state is λ, ie.

$$P\left(\text{outcome of } M_\lambda^{\text{noisy}} = 1 | \text{true state is } \lambda\right) = q(1|\lambda),$$

and that it will sometimes click even when the true state is $\tilde{\lambda} \neq \lambda$, ie.

$$P\left(\text{outcome of } M_\lambda^{\text{noisy}} = 1 | \text{true state is } \tilde{\lambda}\right) = q\left(1|\tilde{\lambda}\right).$$

If our description of the system is given by $p(\lambda)$ then,

$$\begin{aligned} P\left(\text{outcome of } M_\lambda^{\text{noisy}} = 1\right) &= \sum_\lambda P\left(\text{outcome of } M_\lambda^{\text{noisy}} = 1 | \text{true state } \lambda\right) \cdot P\left(\text{true state } \lambda\right) \\ &= \sum_\lambda p(\lambda)\, q(1|\lambda). \end{aligned}$$

More generally we may wish to consider continuous sample spaces (eg. the phase space of classical mechanics) in which case we just replace the sum by an integral,

$$P\left(\text{outcome of } M_\lambda^{\text{noisy}} = 1\right) = \int_\lambda p(\lambda)\, q(1|\lambda)\, d\lambda$$

The final, obvious, generalization we must consider is to allow measurements to have more than two outcomes. In this case our description of a measurement apparatus M is simply the probability distribution of outcomes k such that the true state is λ, ie.

$$\text{measurement description } \leftrightarrow q_M(k|\lambda). \tag{2.2.2}$$

Again, the measurement description is given by the distribution and *not* by the value of the distribution at k, λ, despite our abuse of notation.

The last (implicit) ingredient of the probabilistic classical theory is the use of the law of total probability to assign probabilities to measurement outcomes.[10]

$$\text{law of total probability} \leftrightarrow \Pr(k) = \int_{\lambda} p(\lambda)\, q(k|\lambda)\, d\lambda. \tag{2.2.3}$$

At this point we have given a sufficient description of the operational use of probability theory to explain the technical results contained within this thesis. However, there are a great deal of extremely important questions that we have swept under the rug. For example: what sort of objects can comprise Λ? What do the probability assignments mean? Why are we using probabilities at all? A careful treatment of these points is far beyond the scope of this thesis, but the reader is strongly urged not to dismiss them. The failure to understand the role of probability theory in physical reasoning may well be the most common source of errors in modern science. The clearest exposition of the role of probabilistic reasoning in science that I am aware of is ET Jaynes' excellent book *Probability Theory: The Logic of Science*[44].[11] The recommendation of this book is almost certainly the most valuable contribution of this thesis.

A Digression of the Mind Projection Fallacy

It is important to be aware of the distinction between our description of a system and the definite reality of that system. In the context of probability theory there is an increasing awareness that probability distributions constitute a description of our state of knowledge and not of any definite, intrinsic feature of reality. This often misleadingly called a "subjective" description, which implies a certain degree of arbitrariness in probability assignments. Although it is true that probability assignments can differ between different rational agents there is in fact essentially no freedom in how any agent manipulates his probability assignments. To understand this consider the following situation:

1. Some physical principal tells us $A \implies B$, both Alice and Bob are aware of this physical principal.
2. Alice does an experiment in which she finds B.
3. Bob does what he believes is the same experiment and finds not B ($\bar{B}$).

At this point Alice and Bob hold different beliefs about A. Alice remains agnostic but Bob is certain that A is false. The point is that neither of them exercised any subjective judgment in deciding what they believed about A, their conclusions were totally determined by the laws

10 Notice that at some point we have assumed that all of the relevant measures are absolutely continuous with respect to some privileged measure $d\lambda$. It is interesting to ask if anything could be gained by dropping this assumption. As far as I know the only 'no go' type theorem that shows any awareness of this possibility is [20]).

11 Those who have not given it much thought often dismiss the question of interpretation of probability as a matter of personal preference and the proponents of each school of thought as religious fanatics. If interpretation of probability is to be thought of as a religious affiliation then I suppose I am a adherent of Jayneism.

of Aristotelian logic; their differing beliefs are because they are reasoning from different premises.[12] *This is exactly the same "subjectivity" as the "subjectivity" in probability theory.*

The practical importance of understanding that probability assignments correspond to the beliefs of some particular agent is that this dictates how we incorporate physical reasoning into probability assignments. For example, consider the following game played by Alice and Bob:

1. Two coins are prepared by the same coin manufacturing apparatus, which is described by the parameters $\boldsymbol{\lambda}$.

2. One coin is given to Alice and the other coin is given to Bob. They both independently choose how they will flip the coin. A coin may be flipped in two distinct ways. Flip style F_1 corresponds to tossing a coin into the air with a flick of the thumb in the most common way a coin is tossed (eg. the way a coin is tossed to determine who goes first in a sporting event). Flip style F_2 corresponds to setting the coin on its edge on a table and spinning it like a top.

3. Alice must guess whether Bob got the same coin flip result as she did.

If we are promised that Alice and Bob are a large distance from each other when the coins are flipped we might expect the physical principal of local causality to take the following form for Alice's probability assignment:

$$P_A\left(O_B|\ \boldsymbol{\lambda}, F_A, F_B, O_A\right) = P_A^{\text{naive}}\left(O_B|\ \boldsymbol{\lambda}, F_B\right), \tag{2.2.4}$$

where $O_A,\ O_B \in \{H, T\}$ denote Alice and Bob's flip outcome and $F_A,\ F_B \in \{F_1, F_2\}$ denote their coin flipping choices. The interpretation of this first guess assignment is that Bob's coin flip outcome can't possibly depend on how Alice flipped her coin or what outcome she found. The problem is that we have justified the assumption of *physical* independence but equation 2.2.4 actually assumes *logical* independence. Generally, when Alice flips the coin she will learn something about the relationship between the way the coin was produced $\boldsymbol{\lambda}$, the way the coin is flipped F and the outcome O. For instance, suppose Alice flips the coin by spinning it ($F_A = F_2$) and gets a head, as a result she changes her beliefs about the flip outcome O:

$$P_A\left(O = H|\ \boldsymbol{\lambda}, F_2,\ \text{new information}\right) > P_A\left(O = H|\ \boldsymbol{\lambda}, F_2\right). \tag{2.2.5}$$

We now see that the correct way for Alice to make her probability assignment is:

$$P_A\left(O_B|\ \boldsymbol{\lambda}, F_A, F_B, O_A\right) = P_A^{\text{locality}}\left(O_B|\ \boldsymbol{\lambda}, F_B, \{F_A = F_B,\ A\}\right).$$

Where the notation $\{F_A = F_B,\ A\}$ is meant to suggest that the dependence on F_A and A is only due to Alice's changed beliefs (equation 2.2.5). Thus the role of local causality in this

12 A professor giving a lecture during an air raid continued her derivation even as the building shook until at last she was stopped by a colleague who, fearful the building would collapse, said, "You must stop. Our premises can not support your derivation!"
This joke is not my own creation, but I don't know the original source for proper attribution.

example is not to eliminate dependence on spatially separated events but rather to restrict the form that dependence[13].

The similarity of the above example to Bell-type non-locality arguments is not accidental. The point that we are trying to drive home is that to draw conclusions about how "quantum" behavior differs from "classical" behavior we must be hyper aware of the physical assumptions motivating our mathematical assumptions and that to understand the relationship between the two requires a careful consideration of the distinction between logical reasoning and physical ontology. We should emphasize that this is not to be taken as a repudiation, or even a criticism, of Bell-type results. These results give very interesting insight into the physical world and have interesting applications (eg. quantum cryptography). We are merely urging due diligence in their interpretation.

Actually, it is possible to formulate a Bell-type argument entirely without the explicit use of probability theory[33]. In this case the assumption of local causality is imposed directly on the sample space Λ. In order to do this we must understand the relationship between the set of admissible logical propositions Λ and the space of definite, ontological features of the physical world. It is very common in the quantum foundations literature to see Λ referred to as the ontic space, i.e., to make the assumption that there is an exact correspondence between admissible logical propositions about a system and the ontology of the system. This is orders of magnitude less dangerous than identifying probability distributions as ontological features, but it is still important to recognize that an assumption has been made.

2.2.2 *Quasi-Probability Representations for Quantum Theory*

Suppose that there were a deeper hidden variable theory underlying quantum theory. In this case it must be possible to recover the usual quantum formalism from the deeper theory,

$$\begin{aligned} \text{hidden variable theory} &\twoheadrightarrow \text{quantum theory, by:} \\ \text{system description} \leftrightarrow \text{probability distributions} &\twoheadrightarrow \text{quantum states} \\ \text{measurement description} \leftrightarrow \text{conditional probability distributions} &\twoheadrightarrow \text{POVMs} \\ \text{law of total probability} &\leftrightarrow \text{Born rule.} \end{aligned}$$

Generally, the hidden variable theory could depend on physical considerations not available to quantum theory. For example, the hidden variable system description might depend on gravitational effects, some particular details of the preparation procedure or what you had for breakfast on the day of any experiment you perform. Moreover, it is possible quantum theory might not correspond exactly to the post-quantum theory in any parameter regime but instead only be recovered as some kind of "classical" limit, in the same way that non-relativistic mechanics is recovered as the approximation $v/c \to 0$.

13 Actually, there is another more subtle way that locality is coming into play. When Alice makes her own coin flip she assumes that her outcome is logically independent of what Bob is doing, i.e., $P_A\left(O_A|\ \boldsymbol{\lambda}, F_A, F_B, O_B\right) = P_A\left(O_A|\ \boldsymbol{\lambda}, F_A\right)$. This assumption is critical for her to make inferences about the behavior of Bob's coin based on her own outcome.

An extraordinarily naive possible form for a hidden variable theory is to assume that all of the relevant information is already included in the quantum description, ie.

$$\begin{aligned} \text{probability distributions} &\leftrightarrow \text{quantum states} \\ \text{conditional probability distributions} &\leftrightarrow \text{POVMs} \\ \text{law of total probability} &\leftrightarrow \text{Born rule.} \end{aligned}$$

Models of this kind are nearly non-contextual, in that they don't have any explicit dependence on anything outside of the quantum formalism. However, we are still implicitly allowing some super-quantum dependence in the form of our measurement mapping. We have not yet disallowed the possibility that the conditional distribution corresponding to a POVM element depends on the entire POVM, ie.

$$E_k \in \{E_k\}^{(M)} \rightarrow q_{E_k,\{E_k\}^{(M)}}(k|\lambda)\,.$$

Since the predictions of quantum theory (via $\Pr(k) = \mathrm{Tr}(\rho E_k)$) don't depend on what POVM M the element E_k belongs to any mapping from POVMs to conditional probability distributions that does have such a dependence must be making implicit use of some non-quantum context. If we disavow models of this type then we are left with:

$$\begin{aligned} \text{probability distributions} &\leftrightarrow \text{quantum states} \\ \text{conditional probability distributions} &\leftrightarrow \text{POVM } \textbf{elements} \\ \text{law of total probability} &\leftrightarrow \text{Born rule.} \end{aligned} \tag{2.2.6}$$

Hidden variable theories of this type are *non-contextual* in the sense of Spekkens[66]. In fact, no Spekkens non-contextual hidden variable theory can reproduce quantum theory. Roughly speaking, the reason for this is that the convexity of quantum theory forces the bijections of (2.2.6) to be *linear* mappings and it can be shown that any such mapping will assign negative values to some of the 'probability' distributions corresponding to the quantum states or measurements[8, 20, 18]. Since (2.2.6) specifies an exact mapping between quantum theory and the 'negative' probability model the latter gives a representation for quantum theory. Representations of this type are *quasi-probability representations* and (2.2.6) can be taken to be the defining properties of such representations.

To see how this works we will follow [66] and begin by considering a quantum state of the form $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$. Suppose each state is represented in the Spekkens non-contextual model as $|\psi_j\rangle\langle\psi_j| \leftrightarrow \mu_{|\psi_j\rangle\langle\psi_j|}(\lambda)$[14]. This state can be prepared by sampling the integer j from the distribution p_j and preparing the state $|\psi_j\rangle\langle\psi_j|$, which in the Spekkens non-contextual model corresponds to preparing the distribution $\mu_{|\psi_j\rangle\langle\psi_j|}$ with probability p_j. That is, in the Spekkens non-contextual model we must have:

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j| \implies \mu_\rho(\lambda) = \sum_j p_j \mu_{|\psi_j\rangle\langle\psi_j|}(\lambda)\,. \tag{2.2.7}$$

14 The argument λ is included to emphasize the space the measure is over. The formula should be read as a mapping between quantum states and distributions, *not* as a mapping between quantum states and values of the distribution at a point.

By identical reasoning we can also infer that for a positive operator E it holds:

$$E = \sum_j p_j E_j \implies \xi_E(\lambda) = \sum_j p_j \xi_{E_j}(\lambda) \tag{2.2.8}$$

where $\xi_{E_j}(\lambda) = q_M(j|\lambda) \leftrightarrow E_j$ is written in this form to emphasize both the dependence on λ and the independence from the associated POVM. Equation (2.2.7) means that the mapping from quantum states ρ to the value of the measure $\mu_\rho(\lambda)$ at any λ is a convex linear function from the space of density matrices to $\mathbb{R}$. Similarly, the mapping from positive operators to the value of $\xi_E(\lambda)$ at any λ (equation (2.2.8)) is a convex linear function on the space of positive operators less than $\mathbb{I}$. A convex linear function $f : L(\mathcal{H}) \to \mathbb{R}$ on a convex subset of $L(\mathcal{H})$ that includes a basis for $L(\mathcal{H})$ can be uniquely extended to a linear function on this space (with the requirement that $f(\mathbf{0}) = 0$). Since we are now dealing with linear functionals on a Hilbert space we may invoke the Riesz representation theorem to find:

$$\mu_\rho(\lambda) = \mathrm{Tr}(\rho F(\lambda)) \text{ and } \xi_E(\lambda) = \mathrm{Tr}(\rho G(\lambda)) \tag{2.2.9}$$

where $F(\lambda)$, $G(\lambda)$ are Hermitian matrices associated to each point $\lambda \in \Lambda$.

We can now define quasi-probability representations for quantum theory. We first give an informal definition to communicate the idea clearly:

Definition 5. (Informal) Let $\mathcal{H}$ be a Hilbert space. Let Λ be a sample space and let $\{F(\lambda)\}_{\lambda\in\Lambda}$ and $\{G(\lambda)\}_{\lambda\in\Lambda}$ be spanning sets for $\mathcal{H}$. A quasi-probability representation is a pair of mappings T, S where T maps quantum states to measures over Λ via:

$$\begin{aligned} T:\ \rho &\to \mu_\rho \text{ by} \\ \mu_\rho(\lambda) &= \mathrm{Tr}(\rho F(\lambda)) \end{aligned}$$

and S maps POVMs to conditional probability distributions via:

$$\begin{aligned} S:\ M = \{E_k\} &\to \{q_M(k|\cdot)\} \text{ by} \\ q_M(k|\lambda) &= \mathrm{Tr}(\rho G(\lambda)), \end{aligned}$$

such that the Born rule is reproduced by the law of total probability:

$$\Pr(\text{outcome k}|\ \rho, M) = \mathrm{Tr}(\rho E_k) = \int_\Lambda q_M(k|\lambda)\, \mu_\rho(\lambda)\, d\lambda.$$

See Figure 2.3.1 on page 28 for a concrete example.

The main takeway is that if we want a Spekkens' non-contextual representation (equation (2.2.6)) then the representation is specified by the pair of frames $\{F(\lambda)\}, \{G(\lambda)\}$, an approach suggested in [8, 20, 18]. The point being that physical or operational significance of the representation is given by the requirement of Spekkens' non-contextuality while its mathematical manifestation is captured by equation (2.2.9).

It is possible to show that the dual of any frame of positive operators can not also be a frame of positive operators, which implies that every quasi-probability representation must admit negativity on at least some measurements or states.

We conclude with a formal and very general definition of quasi-probability representations following [20].

Definition 6. Let $\mathcal{S}(\mathcal{H})$ be the quantum states on Hilbert space $\mathcal{H}$ and $\mathcal{E}(\mathcal{H})$ be the set of positive operators bounded above by $\mathbb{I}$. Let Λ be a set and Σ a σ-algebra over Λ, so (Λ, Σ) is a measurable space. Let $\mathcal{S}_{\mathbb{R}}^{\pm}(\Lambda, \Sigma)$ be the bounded, signed measures on (Λ, Σ) and let $\mathcal{E}^{\pm}(\Lambda, \Sigma)$ be the bounded measurable functions. A quasi-probability representation for quantum theory is a pair of maps $T : \mathcal{S}(\mathcal{H}) \to \mathcal{S}_{\mathbb{R}}^{\pm}(\Lambda, \Sigma)$ and $S : \mathcal{E}(\mathcal{H}) \to \mathcal{E}_{\mathbb{R}}^{\pm}(\Lambda, \Sigma)$ such that

$$\begin{aligned} T : \rho &\to \mu_\rho \text{ by} \\ \mu_\rho(\lambda) &= \operatorname{Tr}(\rho F(\lambda)), \end{aligned}$$

where $\{F(\lambda)\}$ is a frame (spanning set) for $L(\mathcal{H})$ and

$$\begin{aligned} S : E &\to \xi_E \text{ by} \\ \xi_E(\lambda) &= \operatorname{Tr}(\rho G(\lambda)), \end{aligned}$$

where $\{G(\lambda)\}$ is a frame (spanning set) for $L(\mathcal{H})$ such that:

$$\operatorname{Tr}(\rho E) = \int_\Lambda \xi_E(\lambda)\, d\mu_\rho(\lambda).$$

The main change in the definition is that instead of mapping POVMs to conditional probability distributions we map POVM elements to measurable functions. Although this notion is easier to formalize it is less easily digested.

Aside on the size of the sample space

From the perspective of studying classical subtheories of quantum theory we would like the positively represented subtheory of a quasi-probability representation to be as large as possible. Notice that since d^2 real numbers are required to specify a Hermitian operator $H \in L(\mathcal{H}_d)$ the minimum number of elements of μ_H required to reproduce quantum theory is d^2. Thus we must have $d^2 \leq |\Lambda|$. States are assigned to quasi-probability distributions over Λ via

$$\mu_\rho(\lambda) = \operatorname{Tr}(\rho F(\lambda))$$

where $F(\lambda)$ is a Hermitian operator. A state ρ has positive representation if $\mu_\rho(\lambda) > 0\ \forall \lambda \in \Lambda$. Thus each additional element of Λ corresponds to an additional half-space constraint that ρ must satisfy to be in the positively represented subtheory. Ergo for any quasi-probability representation with $|\Lambda| > d^2$ there is another quasi-probability representation with $|\Lambda| = d^2$ such that any state that is positive with respect to the first representation is also positive with respect to the second representation, but not vice versa. Nevertheless it may sometimes be desirable to study representations with $|\Lambda| > d^2$, for example if we are interested in $\Lambda = G$ for some group with $|G| > d^2$.

2.2.3 *Quasi-Probability Representations for Subtheories of Quantum Theory*

Negative quasi-probability has a long history as a largely nonsensical indicator of quantum behavior. The most notable examples come from quantum optics where the Wigner

function[77] and the Q and P functions [52] play prominent roles. The typical approach is to declare negative probabilities in some particular quasi-probability representation to be an indicator of quantum behaviour. Such approaches suffer in significance due to the problem of non-uniqueness of the choice of representation. While a quantum state may correspond to a negative-valued quasi-probability function in one choice of quasi-probability representation, in another choice that same state can be positive, and hence a valid classical probability density. The negativity of the representation of a state or measurement in any particular choice of quasi-probability distribution is essentially meaningless. Nevertheless, it is possible to give meaning to negative quasi-probability by considering the representation of *subtheories* of quantum theory, rather than just individual states and measurements.

We define a subtheory of quantum theory as a collection of states and measurements:

$$S_{\mathrm{QM}} = \left\{ \{\rho_i\}, \{M_j\} \right\}.$$

For example, we might consider the subtheory of all stabilizer projectors and measurements or the subtheory of all separable states and measurements. We are primarily interested in studying subtheories that are distinguished by some natural physical or operational restriction, such as fault tolerance or space-like separation in the examples just given. The idea is that if we can find a classical hidden variable theory that reproduces S_{QM} then we expect that it is not possible to build devices that outperform classical systems using only elements from S_{QM}. For instance, if S_{QM} admits a classical hidden variable explanation then we would guess it is not possible to use elements from S_{QM} to execute Shor's algorithm or win a non-local game. We have just seen that it is possible to find a non-contextual hidden variable theory for S_{QM} if and only if there is some quasi-probability representation such that the elements of S_{QM} all have positive representation. Thus,

If a subtheory S_{QM} of quantum theory admits a quasi-probability representation where all elements of S_{QM} are positively represented then S_{QM} should offer no super-classical advantage.

Suppose that the subtheory S_{QM} does admit some quasi-probability representation $\left(\mu_\rho(\cdot), q_M(k|\cdot)\right)$ such that all elements of S_{QM} have non-negative representation. We can use this to define $\widehat{S_{\mathrm{QM}}} = \left\{ \left\{ \rho : \ \mu_\rho(\lambda) \geq 0 \ \forall \lambda \right\}, \left\{ M : \ q_M(k|\lambda) \geq 0 \ \forall k, \lambda \right\} \right\}$, the subtheory of all states and measurements that are positively represented for this special choice of quasi-probability representation. Notice that $S_{\mathrm{QM}} \subset \widehat{S_{\mathrm{QM}}}$ and this inclusion may be strict. That is, the 'classical' subtheory of quantum theory accounted for by the non-contextual hidden variable model afforded by the quasi-probability representation may include elements outside the scope of the original subtheory of interest. This is exactly how bound magic states arise, and it's an intriguing open problem to determine if bound entangled states can be understood in exactly the same fashion.

With the preceding discussion in hand we are now able to give a cogent interpretation of negative quasi-probability. The states and measurements with negative representation are precisely those that can not be accounted for by the non-contextual hidden variable model that explains the subtheory $\widehat{S_{\mathrm{QM}}}$. We expect that there is some negatively represented state ρ such that by supplementing $\widehat{S_{\mathrm{QM}}}$ with ρ we will be able to exhibit a performance improvement over what is possible classically. Moreover, it is very natural to guess that the

degree of negativity in the representation of the extra resource ρ should act as a measure of how useful it is for this purpose. That is,

We expect that the negativity in a quasi-probability representation of a state ρ (or measurement M) serves as a measure of the degree to which the non-contextual hidden variable theory associated to that quasi-probability representation fails to account for the quantum behavior possible when the subtheory is supplemented with ρ (or measurement M).

The major contribution of this thesis is to transform the qualitative observations of this section into precise statements in the context of the stabilizer subtheory of quantum theory.

Why non-contextuality?

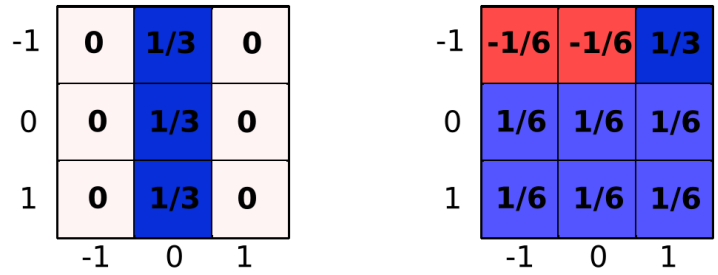
We have just argued that a positive quasi-probability representation for a subtheory of quantum theory is interesting precisely because it corresponds to a non-contextual hidden variable explanation for this subtheory. Of course, it is not obvious why we should restrict ourselves to *non-contextual* hidden variable models. One possible answer is that a contextual hidden variable theory might hide some super-classical advantage in the extra context. For example, it might be possible to find a subtheory S_{QM} containing the elements of quantum theory required to violate a Bell-type inequality that admits a hidden variable explanation by making use of some non-local context; in this case the existence of a hidden variable model does not capture the relevant notion of classicality. Similarly, we might be able to explain a subtheory that is universal for quantum computation with a hidden variable model where the hidden variables include a classical computer with a size that grows exponentially with the number qudits of the subtheory. Insisting on non-contextuality amounts to a rather draconian way of eliminating this sort of possibility. Generally we expect that the existence of a non-contextual hidden variable theory for S_{QM} implies that S_{QM} is not useful for super-classical tasks, but that the converse need not hold. A subtheory of quantum theory is defined to be *(Spekkens') contextual* if it does not admit a (Spekkens') non-contextual hidden variable explanation.

Qualitatively, (Spekkens') contextuality is a necessary but not sufficient condition for a subtheory of quantum theory to be useful for super-classical tasks.
Equivalently, a quasi-probability representation where all elements of a subtheory of quantum theory are positively represented is a sufficient but not necessary condition for that subtheory to be useless for super-classical tasks.

One final example will help make this point clear. To study quantum communication in the same way we study quantum computation in this thesis we would like to find a hidden variable theory for the states and measurements corresponding to the restriction to LOCC:

$$S_{\mathrm{QM}}^{\mathrm{sep}} = \{\text{separable states},\ \text{separable measurements}\}\,.$$

However, a single qubit suffices to demonstrate contextuality in Spekkens sense. Thus there is no possible *non-contextual* hidden variable explanation for $S_{\mathrm{QM}}^{\mathrm{sep}}$. However, in this case we



(a) Wigner representation of the qutrit $|1\rangle$ state. This is a stabilizer state.

(b) Wigner representation of state $\frac{1}{\sqrt{6}}(-|0\rangle + 2|1\rangle - |2\rangle)$. This state has negative representation.

Figure 2.3.1: Wigner Representation of Quantum States. The value of the Wigner function for qutrit state ρ at a point $\boldsymbol{u} \in \mathbb{Z}_3 \times \mathbb{Z}_3$ is given by $W_\rho(\boldsymbol{u}) = \frac{1}{3}\mathrm{Tr}(\rho A_{\boldsymbol{u}})$.

see no reason to refuse to admit a *local* context. For instance, suppose we had a hidden variable theory explanation for a single qubit owned by Alice:

$$\begin{aligned} \text{Sample space } \Lambda_A & & \\ \text{probability distributions } \mu_A(\lambda_A) & \twoheadrightarrow & \text{Alice's quantum states } \rho_{\mu_A} \\ \text{conditional probability distributions } \{q_A(k|\lambda_A)\} & \twoheadrightarrow & \text{Alice's POVMs } \left\{E_k^{(q_A)}\right\} \\ \text{law of total probability} & \leftrightarrow & \text{Born rule.} \end{aligned}$$

and something analogous for a single qubit owned by Bob. In this case one reasonable hidden variable model for the joint Alice, Bob system might be constrained to look like[15]:

$$\begin{aligned} \Lambda_{AB} &= \Lambda_A \times \Lambda_B \\ \mu_A(\lambda_A) \otimes \nu_B(\lambda_B) &\to \rho_{\mu_A} \otimes \rho_{\nu_B} \\ \{q_A(k|\lambda_A)\} \otimes \{m_B(k'|\lambda_B)\} &\to \left\{E_k^{(q_A)}\right\} \otimes \left\{E_{k'}^{(m_B)}\right\}. \end{aligned}$$

A model of this kind is powerful enough to explain $S_{\mathrm{QM}}^{\mathrm{sep}}$ but not powerful enough to explain the 2-qubit Alice-Bob system. Even though any such model must be contextual we feel comfortable saying that anything within the purview of the model is not useful for promoting $S_{\mathrm{QM}}^{\mathrm{sep}}$ to something useful for quantum communication tasks.

2.3 THE DISCRETE WIGNER REPRESENTATION

As the attentive reader has likely inferred by now:

- We wish to study the stabilizer formalism and,

15 Or it might look nothing like this. Contemplation of 2.2.1 shows we have made some probably unnecessarily strong restrictions.

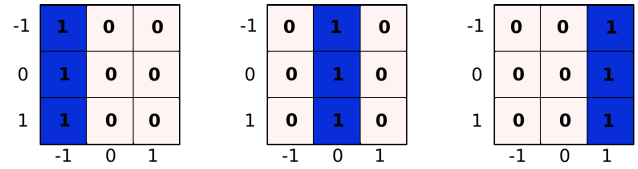


Figure 2.3.2: Wigner Representation of Quantum Measurement. Pictured is Wigner Representation of qutrit Stabilizer PVM $\{|-1\rangle\langle -1|, |0\rangle\langle 0|, |1\rangle\langle 1|\}$. The Wigner representation of POVM elements $\{E_k\}$ give a weighted partitioning of the discrete phase space. The Wigner representation of POVM elements is $W_{E_k}(\boldsymbol{u}) = \mathrm{Tr}(E_k A_{\boldsymbol{u}})$. If all POVM elements are positively represented we interpret $W_{E_k}(\boldsymbol{u}) = \Pr(\text{outcome } k| \text{ true state } \boldsymbol{u})$. In the case of projective measurement the outcome is determined by $\boldsymbol{u}$. For example, $\Pr(\text{outcome } -1| \text{ true state } (0,-1)) = 1$.

- We are going to use quasi-probability representation to do it.

We want a quasi-probability representation where the stabilizer formalism is positively represented. For odd dimensional quantum systems such a representation already exists: this is the discrete Wigner function. This is an analogue of the usual infinite dimensional Wigner function[77]. The main contribution of this thesis is to use this tool to formalize the intuitions of the previous section and establish a precise connection between quantum computational speedup and negativity of the discrete Wigner function representation.

The discrete Wigner function we use was first written down by Wooters[79] but was brought to maturity when Gross[34, 35] discovered its close relation to the stabilizer formalism 20 years later.

2.3.1 *Definition and Properties*

The discrete Wigner representation of a state $\rho \in L(\mathcal{H}_{p^n})$ is a quasi-probability distribution over $(\mathbb{Z}_p \times \mathbb{Z}_p)^n$, which can be thought of as p^n by p^n grid (see Figure 2.3.1 on page 28). The mapping assigning quantum states ρ to Wigner functions $\{W_\rho(\boldsymbol{u})\}$ is given by

$$W_\rho(\boldsymbol{u}) = \frac{1}{p^n}\mathrm{Tr}(A_{\boldsymbol{u}}\rho),$$

where $\{A_{\boldsymbol{u}}\}$ are the *phase space point operators*. These are defined in terms of the Heisenberg-Weyl operators as,

$$A_{\mathbf{0}} = \frac{1}{p^n}\sum_{\boldsymbol{u}} T_{\boldsymbol{u}},\ A_{\boldsymbol{u}} = T_{\boldsymbol{u}} A_{\mathbf{0}} T_{\boldsymbol{u}}^\dagger.$$

These operators are Hermitian so the discrete Wigner representation is real-valued. There are $(p^n)^2$ such operators for p^n-dimensional Hilbert space, corresponding to the $(p^n)^2$ points of discrete phase space.

A quantum measurement with POVM $\{E_k\}$ is represented by assigning conditional quasi-probability functions over the phase space to each measurement outcome,

$$W_{E_k}(\boldsymbol{u}) = \mathrm{Tr}(A_{\boldsymbol{u}} E_k).$$

In the case where $W_{E_k}(\boldsymbol{u}) \geq 0 \ \forall \boldsymbol{u}$, this can be interpreted classically as the probability of getting outcome k given that the system is actually at point $\boldsymbol{u}$, $W_{E_k}(\boldsymbol{u}) = \Pr(\text{outcome } k|\text{location } \boldsymbol{u})$. If both $W_\rho(\boldsymbol{u})$ and $W_{E_k}(\boldsymbol{u})$ are positive then the law of total probability gives the probability of getting outcome k from a measurement of state ρ,

$$\Pr(k) = \sum_{\boldsymbol{u}} W_\rho(\boldsymbol{u}) W_{E_k}(\boldsymbol{u}).$$

In fact, this holds even when $W_\rho(\boldsymbol{u})$ or $W_{E_k}(\boldsymbol{u})$ take on negative values.

We say a state ρ has positive representation if $W_\rho(\boldsymbol{u}) \geq 0 \ \forall \boldsymbol{u} \in \mathbb{Z}_d^n \times \mathbb{Z}_d^n$ and negative representation otherwise. We will say a measurement with POVM $M = \{E_k\}$ has positive representation if $W_{E_k}(\boldsymbol{u}) \geq 0 \ \forall \boldsymbol{u} \in \mathbb{Z}_p^n \times \mathbb{Z}_p^n, \ \forall E_k \in M$ and negative representation otherwise.

The connection between the stabilizer formalism and the discrete Wigner function comes from the following set of properties[34]:

1. (Discrete Hudson's theorem) if $|S\rangle$ is a stabilizer state then $\mathrm{Tr}(A_{\boldsymbol{u}}|S\rangle\langle S|) \geq 0 \ \forall \boldsymbol{u}$, and the stabilizer states are the only pure states satisfying this property. That is, a pure state has positive representation if and only if it is a stabilizer state.

2. A Clifford operator $U_{F,\boldsymbol{a}}, \ F \in \mathrm{Sp}(2n, p), \boldsymbol{a} \in \mathbb{Z}_d^n \times \mathbb{Z}_d^n$ has the action $U_{F,\boldsymbol{a}} A_{\boldsymbol{u}} U_{F,\boldsymbol{a}}^\dagger = A_{F\boldsymbol{u}+\boldsymbol{a}}$. This means that

   $$W_{U_{F,\boldsymbol{a}}\rho U_{F,\boldsymbol{a}}^\dagger}(v) = W_\rho(F^{-1}(\boldsymbol{v} - \boldsymbol{a}))$$

   so that Clifford transformations act as symplectic permutations of the underlying phase space.

3. The phase space point operators in dimension p^n are tensor products of n copies of the p dimension phase space point operators. That is, the sample space for the joint system AB is $\Lambda_{AB} = \Lambda_A \times \Lambda_B$.

4. The phase space point operators are orthogonal, $\mathrm{Tr}(A_{\boldsymbol{u}} A_{\boldsymbol{v}}) = d\delta_{\boldsymbol{u},\boldsymbol{v}}$. As a consequence, $\rho = \sum_{\boldsymbol{u}} W_\rho(\boldsymbol{u}) A_{\boldsymbol{u}}$ for any $\rho \in L(\mathcal{H}_d)$.

A great deal of additional information about the discrete Wigner function is available in [34], but we will not need any of it. The Wigner representation of a number of quantum operations is summarized in Table 2.3.1 on page 31.

Aside on Lüders' Rule

Since so much of the usual quantum formalism is accounted for in a natural way in the Wigner representation it is perhaps surprising that the projection postulate is not. To see the problem consider the qutrit ($d = 3$) projective valued measure (PVM):

$$M = \{|0\rangle\langle 0|, |1\rangle\langle 1| + |2\rangle\langle 2|\}.$$

It is easy to see that this measurement has positive discrete Wigner representation since the PVM elements are positive combinations of stabilizer projectors. However, if we perform

Vanilla Quantum Formalism	Discrete Wigner Function
Hermitian Operators $\rho \in L(\mathcal{H}_d)$	Quasi-Probability Distribution $W_\rho(\boldsymbol{u}) = \frac{1}{d}\mathrm{Tr}(\rho A_{\boldsymbol{u}})$
POVM Elements $E_k \in \{E_k\}_k$	Conditional Quasi-Probability Distribution $W_{E_k}(\boldsymbol{u}) = \mathrm{Tr}(E_k A_{\boldsymbol{u}})$
Inner Product $\mathrm{Tr}(\rho\sigma)\ \ \rho,\sigma \in L(\mathcal{H}_d)$	Law of Total Probability $d\sum_{\boldsymbol{u}} W_\rho(\boldsymbol{u}) W_\sigma(\boldsymbol{u})$
Tensor Product $\rho \otimes \sigma \in L(\mathcal{H}^A \otimes \mathcal{H}^B)$	Tensor Product $W_{\rho\otimes\sigma}(\boldsymbol{u}_A, \boldsymbol{u}_B) = W_\rho(\boldsymbol{u}_A) \otimes W_\sigma(\boldsymbol{u}_B)$
Partial Trace $\mathrm{Tr}_B(\rho),\ \rho \in L(\mathcal{H}^A \otimes \mathcal{H}^B)$	Marginalization $W_{\mathrm{Tr}_B(\rho)}(\boldsymbol{u}_A) = \sum_{\boldsymbol{u}_B} W_\rho(\boldsymbol{u}_A, \boldsymbol{u}_B)$
Matrix Multiplication $\rho\sigma\ \rho,\sigma \in L(\mathcal{H}_d)$	Moyal Product $W_{\rho\sigma}(\boldsymbol{u}) = \frac{1}{d}\sum_{\boldsymbol{v},\boldsymbol{w}} W_\rho(\boldsymbol{u}+\boldsymbol{v}) W_\sigma(\boldsymbol{u}+\boldsymbol{w}) e^{-\imath\frac{2\pi}{d}[\boldsymbol{v},\boldsymbol{w}]}$

Table 2.3.1: Wigner representation of some quantum operations, see [34, 25] for anything that doesn't look trivial.

the measurement on the stabilizer state $|\psi\rangle = |0\rangle + |1\rangle + |2\rangle$ and get the second outcome Luders' rule implies:

$$\begin{aligned} |\psi'\rangle\langle\psi'| &\propto (|1\rangle\langle 1| + |2\rangle\langle 2|)\,|\psi\rangle\langle\psi|\,(|1\rangle\langle 1| + |2\rangle\langle 2|) \\ &= |1\rangle\langle 1| + |2\rangle\langle 2| + |1\rangle\langle 2| + |2\rangle\langle 1| \end{aligned}$$

and this state has *negative* discrete Wigner representation. As far as I know this is the only quantum rule involving elements with positive Wigner representation that does not have a clear "hidden variable" explanation in discrete phase space.

Notice that this is not a problem for stabilizer PVMs.

2.3.2 *The discrete Wigner function and the usual Wigner function*

There are essentially two ways to think about the discrete Wigner function:

1. As the finite dimensional analogue of the usual Wigner function.
2. As the particular quasi-probability representation that is useful for studying the stabilizer formalism.

In this thesis we take the second view, but essentially all the previous work on the subject [79, 34, 35, 25] took the former view and the candidate discrete Wigner functions they constructed were designed to behave as much like the infinite dimensional Wigner function as possible. This close analogy affords us a sort of rosetta stone for translating between odd

Finite Odd Prime Dimension	Infinite Dimension
Hilbert Space $L^2\left(\mathbb{Z}_p^n\right)$	Hilbert Space $L^2\left(\mathbb{R}^n\right)$
Discrete Phase Space $\mathbb{Z}_p^n \times \mathbb{Z}_p^n$	Phase Space $\mathbb{R}^n \times \mathbb{R}^n$
Finite Heisenberg Group $H_n\left(\mathbb{Z}_p\right)$	Heisenberg Group $H_n\left(\mathbb{R}\right)$
Stabilizer States $\langle q\|\psi\rangle \propto \exp\left(\imath\frac{2\pi}{p^n}\left(q\theta q\right)+xq\right)$ where $x, q \in \mathbb{Z}_p^n$ and θ a	Gaussian States $\langle q\|\psi\rangle \propto \exp\left(\imath\frac{2\pi}{p^n}\left(q\theta q\right)+xq\right)$ where $x, q \in \mathbb{R}^n$ and θ a
symmetric matrix with entries in $\mathbb{Z}_p$	symmetric matrix with entries in $\mathbb{R}$
Clifford Unitaries permutation of phase space	Quadratic Hamilitonian Evolution permutation of phase space
Discrete Wigner Function $W_\rho\left(q,p\right)=\mathcal{F}\left(\frac{1}{p^n}\mathrm{Tr}\left(T_{(\xi,\eta)}^\dagger\rho\right)\right)$ where $T_{(\xi,\eta)}$ are Heisenberg-Weyl operators and $\mathcal{F}$ is the discrete Fourier transform	Wigner Function $W_\rho\left(q,p\right)=\mathcal{F}\left(\mathrm{Tr}\left(w_{\xi,\eta}^\dagger\rho\right)\right)$ where $w_{(\xi,\eta)}$ are Weyl operators[21] and $\mathcal{F}$ is the (usual) Fourier transform

Table 2.3.2: Correspondence between stabilizer operations in finite odd dimension and linear optics in infinite dimension, see [34] for anything that doesn't look trivial.

power of prime dimensional Hilbert space and infinite dimensional Hilbert space, summarized in Table 2.3.2 on page 32. Actually, the analogy works equally well for translating between arbitrary odd dimensional Hilbert space and infinite dimensional Hilbert space, but this adds a good deal of mathematical complication without bringing any additional conceptual clarity so as usual we will neglect it.

In this thesis we describe a finite dimension simulation protocol for stabilizer operations supplemented by states with positive discrete Wigner representation (from [71]). Using the analogy described in this section this simulation protocol has already been extended to a simulation protocol for linear optics supplemented by states with positive (usual) Wigner function[73, 53].

2.3.3 *What's up with Qubits?*

The major drawback of the discrete Wigner function we use here is that it is not defined for qubits. This oddity is inherited from the strange behaviour of the stabilizer formalism for qubits, which is in turn a consequence of the fact that $\mathbb{Z}_2^n$ behaves differently than $\mathbb{Z}_p^n,\ p \neq 2$. More concretely, for any $\mathcal{H}_p,\ p \neq 2$ we can identify an operator $A_\mathbf{0} \in L\left(\mathcal{H}_p\right),\ A_\mathbf{0} \neq \mathbb{I}$ that is invariant under the symplectic component of the Clifford group, ie.

$$U_F A_\mathbf{0} U_F^\dagger = A_\mathbf{0}\ \forall U_F \in \mathcal{C}_{p^n}/\mathcal{W}_{p^n},$$

where $\mathcal{W}_{p^n}$ denotes the group of Heisenberg-Weyl operators supplemented by the phases $\omega\mathbb{I}$. The problem is that for $\mathcal{H}_2$ there is no operator with this property.

This may give the impression that the reason the Wigner function does not exist for qubits is a minor mathematical bug to be expunged with a little bit of additional cleverness. This view is particularly tempting in light of the fact that for David Gross the issue was that his abstract definition of the Heisenberg group does not extend to $p = 2$[37], a problem we have already shown how to resolve. It may yet turn out to be true that there is a simple mathematical resolution, but the problem is more subtle than it first appears. Using 3 or more qubits it is possible to violate a contextuality inequality using only stabilizer operations[33]. Thus there can be no non-contextual hidden variable model for the qubit stabilizer formalism. As we explained at length in Section 2.2 this means that any quasi-probability representation for $\mathcal{H}_{2^n},\ n \geq 3$ will assign negative probability to at least some of the stabilizer states or measurements. However, as we saw, the physical significance of the necessity of negative quasi-probability representation for a subtheory is merely that any hidden variable theory explaining that subtheory must be contextual. This leaves open the possibility that there exists a perfectly acceptable contextual classical hidden variable model for the qubit stabilizer formalism. This would allow the results of this thesis to be extended to the qubit case.

Part I

POSITIVE WIGNER FUNCTIONS AND STABILIZER COMPUTATION

This part of the thesis covers the results of [71]. The first result characterizes the geometry of the region of quantum states with positive Wigner representation. This establishes the existence of a large class of mixed magic (non-stabilizer) states with positive Wigner representation. The next result is an efficient classical simulation protocol for quantum circuits using only stabilizer operations and states with positive Wigner representation. This establishes that there are mixed magic states that are useless for promoting stabilizer computation to universal quantum computation. This part of the thesis ends with a direct proof of the fact that states with positive Wigner representation can not be distilled to pure magic states using stabilizer operations, thus establishing the existence of bound magic states.

The Gottesman-Knill theorem[27, 1] provides an efficient classical simulation protocol for circuits of Clifford unitaries acting on stabilizer states. This result deals with pure qubit stabilizer state inputs and simulates the evolution of the full quantum state. The simulation scheme given in this thesis deals with odd dimensional systems, makes no distinction between mixed state and pure state input, and allows the simulation of a large class of non-stabilizer states. However, the present scheme only simulates the distribution of measurement outcomes rather than the evolution of the full quantum state.

A number of papers have addressed the question of which ancilla states enable universal quantum computation for the magic state model in qubit systems [12, 10, 9, 60, 61, 62, 55, 69]. The most directly comparable result is the demonstration by Campbell and Browne [12] that for any magic state distillation protocol consuming a fixed number n of resource states ρ there exists a ρ outside the convex hull of stabilizer states that maps to a convex combination of stabilizers. As n grows these states are known to exist only within some arbitrarily small distance ϵ of the convex hull of stabilizer states. By contrast, the present result implies the existence of states a fixed distance from the hull which are not distillable by any protocol.

The results of this part of the thesis are complementary to previous work connecting negativity in discrete Wigner function type representations to quantum computational speedup [14, 24, 69]. In particular, van Dam and Howard [68] have used techniques of this type to derive a bound on the amount of depolarizing noise a state can withstand before entering the stabilizer polytope. Their work deals only with prime dimensional systems, and in this case it turns out that the noise threshold they derive is the same as the amount of noise required for their "maximally robust" state to enter the region of states with positive discrete Wigner representation.

3

THE GEOMETRY OF STATES WITH POSITIVE WIGNER REPRESENTATION

Since the only pure states with positive discrete Wigner representation are stabilizer states it is natural to wonder if every positively represented state is a mixture of stabilizer states. Remarkably this is *false*: there are a large class of states with positive Wigner representation that can not be written as a convex combination of stabilizer states. To establish this we will clarify the geometry of the region of state space which has positive representation and show that it strictly contains the set of mixtures of stabilizer states.

The set of convex combinations of stabilizer states is a convex polytope with the stabilizer states as vertices. Any polytope can be defined either in terms of its vertices or as a list of half space inequalities called facets. Intuitively, these correspond to the faces of 3 dimensional polyhedrons. We show that in power of prime dimension each of the d^2 phase space point operators define a facet of the stabilizer polytope. These are only a proper subset of the faces of the stabilizer polytope, implying the existence of states with positive representation which are not convex combinations of stabilizer states. See Figure 3.0.3 for a cartoon capturing the intuition for this result.

The stabilizer polytope may be thought of as a bounded convex polytope living in $\mathbb{R}^{d^2-1}$, the space of d dimensional mixed quantum states. A minimal half space description for a polytope in $\mathbb{R}^D$ is a finite set of bounding equalities called facets $\{F_i, f_i\}$ with $F_i \in \mathbb{R}^D$ and $f_i \in \mathbb{R}$. $X \in \mathbb{R}^D$ is in the polytope if and only if $X \cdot F_i \leq f_i \ \forall i$. In the usual quantum state space the vectors X of interest are density matrices, the inner product is the trace inner product and facets may be defined as $\{\hat{A}_i, a_i\}$ where $\hat{A}_i$ are Hermitian matrices and

$$\rho \in \text{polytope} \iff \text{Tr}(\rho \hat{A}_i) \leq a_i \ \forall i.$$

Our objective is to show that $\{-A_u, 0\}$ are facets of the polytope defined by stabilizer state vertices.

It is possible to explicitly compute a facet description for a polytope given the vertex description, but the complexity of this computation scales polynomially in the number of vertices. Since the number of stabilizer states grows super-exponentially with the number of qudits [34] the conversion is generally impractical. The analytic proof given here circumvents this issue. We also remark that the work of Cormick *et al* [14] implies that the phase space point operators considered here are facets for the case of *prime* dimension.

Theorem 7. *The d^2 phase space point operators $\{A_u\}$ with the inequalities $Tr(\rho A_u) > 0$ define facets of the stabilizer polytope.*

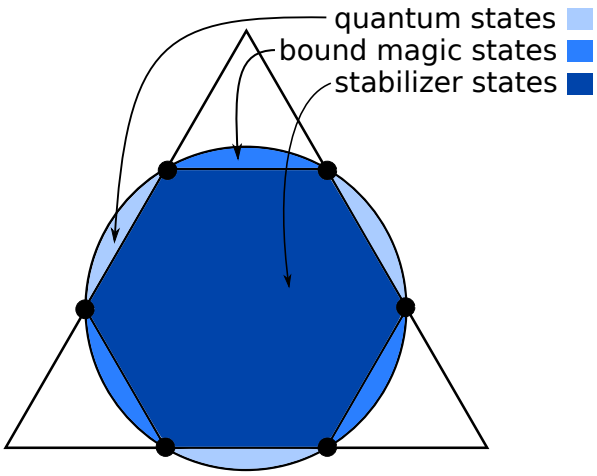


Figure 3.0.3: A cartoon of the intersection of the discrete Wigner probability simplex (the triangular region) with the quantum state space (the circle). The simplex intersects the boundary at stabilizer states (bold dots). The region of convex combinations of stabilizer states is strictly contained within the set of quantum states that also lie inside the simplex. The quantum states outside the simplex are the bound states. Finally, the quantum states with negative discrete Wigner representation are those lying outside the positive discrete Wigner simplex. We show that the half space inequalities defining the facets of the discrete Wigner simplex also define the facets of the stabilizer polytope; a fact reflected in this cartoon.

Proof. To establish that a halfspace inequality for a polytope in $\mathbb{R}^D$ is a facet there are two requirements: every vertex must satisfy the inequality and there must be a set of vertices saturating the inequality which span a space of dimension D [81].

The requirement that all vertices satisfy the half space inequality is $\mathrm{Tr}(A_u S) \geq 0$ for every stabilizer state S, and this the discrete Hudson's theorem.

We consider the stabilizer polytope as an object in $\mathbb{R}^{d^2-1}$ and look for a set of d^2-1 linearly independent vertices which satisfy $\mathrm{Tr}(A_u S) = 0$. Since we are restricting to power of prime dimension we may choose a complete set of mutually unbiased bases of $d(d+1)$ states from the full set of stabilizer states. Suppose more than $d+1$ states V_i from this set satisfy $\mathrm{Tr}(V_i A_u) > 0$. Then a counting arguments shows that there must be two distinct states V_0, V_1 belonging to an orthonormal basis which satisfy this criterion. But then

$$\begin{aligned}\mathrm{Tr}(V_0 V_1) &= \frac{1}{d}\sum_v \mathrm{Tr}(V_0 A_v)\mathrm{Tr}(V_1 A_v) \\ &\geq \frac{1}{d}\mathrm{Tr}(V_0 A_u)\mathrm{Tr}(V_1 A_u) \neq 0,\end{aligned}$$

which contradicts the orthonormality. Thus at least $d(d+1)-(d+1)$ states in the mutually unbiased bases satisfy $\mathrm{Tr}(A_u V_i) = 0$. These are the required a set of d^2-1 linearly independent vertices. □

The phase space point operators considered here give only a proper subset of the defining halfspace inequalities for the stabilizer polytope. This means that there are states that

may not be written as a convex combination of stabilizer states which nevertheless satisfy $\mathrm{Tr}(A_u\rho) \geq 0$ for all phase space point operators. That is, there are positive states which are not in the convex hull of stabilizer states. An explicit example of such a state for the qutrit is given in [34]. These regions can be visualized by taking two and three dimensional slices of the qutrit state space. Such slices are depicted in Figures 3.0.4 and 3.0.5.

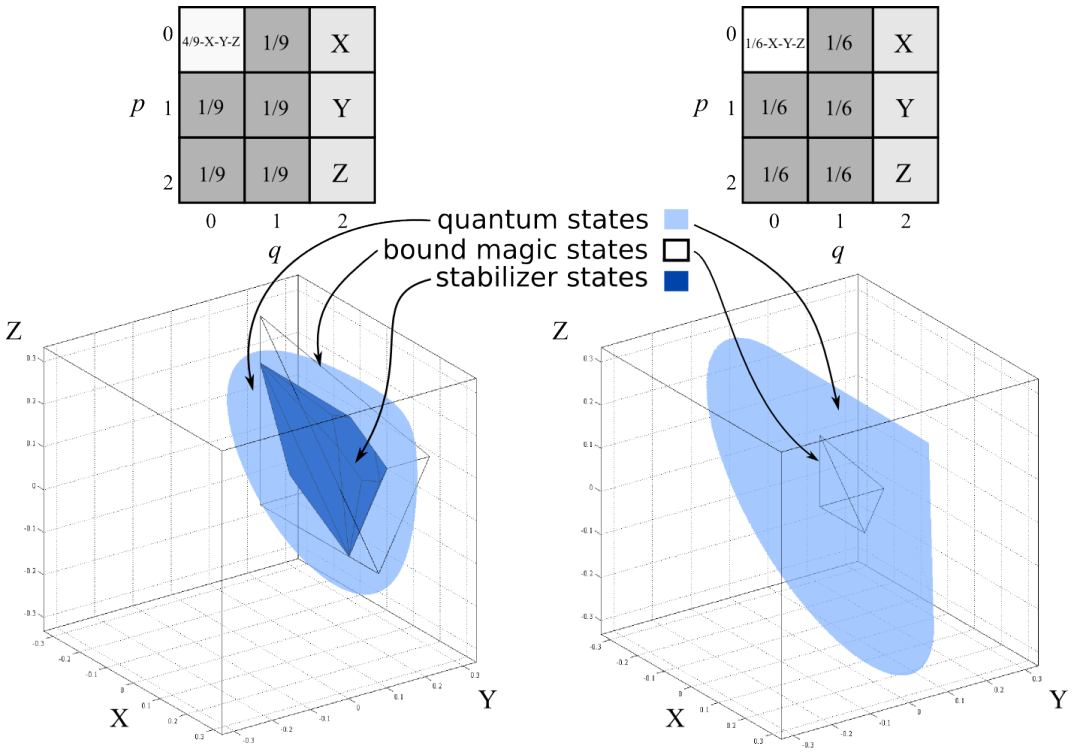


Figure 3.0.4: Orthogonal 3-dimensional slices of qutrit state space. Above each slice is the five values of the Wigner function which are fixed at a value of $1/9$ (left) and $1/6$ (right). Three of the remaining four values are allowed to vary and carve out regions depicted in the graphs. The final value is fixed by $\mathrm{Tr}(\rho) = 1$. Note that the slice on the right does not cut through the stabilizer polytope but does contain a region of bound states. See also Figure 3.0.5 for 2-dimensional slice of the figure on the left.

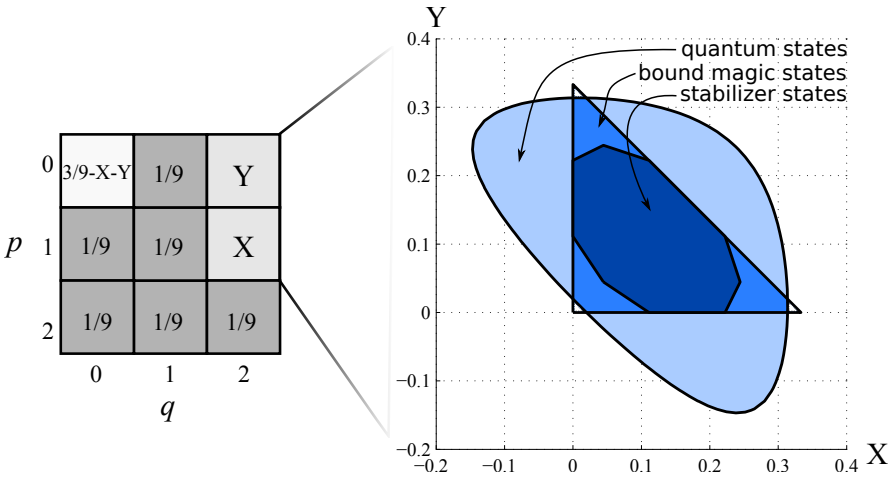


Figure 3.0.5: Orthogonal 2-dimensional slice of qutrit state space. On the left are the six values of the Wigner function which are fixed at a value of $1/9$. Two of the remaining three are allowed to vary with the third fixed by $\mathrm{Tr}(\rho) = 1$. The maximally mixed state is the point $(X, Y) = (1, 1)/9$. The various regions carved out by varying these values are shown on the right. Note the similarity to the caricature in 3.0.3, remarkable since this cartoon was merely the intersection of the simplest simplex (a triangle) with the simplest continuous state space (a circle).

4

EFFICIENT CLASSICAL SIMULATION USING POSITIVE DISCRETE WIGNER FUNCTION

We now establish that any quantum computation consisting of stabilizer operations acting on product input states with positive representation can not produce an exponential computational speed-up. This is accomplished by giving an explicit efficient classical simulation protocol for such circuits. Like the Gottesman-Knill protocol this scheme allows for the simulation of pure state stabilizer inputs to circuits composed of Clifford transformations and stabilizer measurements. However, our simulation scheme extends the Gottesman-Knill result in several ways. First, it applies to systems of qudits rather than qubits. Second, it applies to mixed state inputs. Thirdly, and most remarkably, it applies to some non-stabilizer resources - namely those with positive discrete Wigner representation.

Any particular run of a quantum algorithm on n registers will produce a string $\boldsymbol{k}$ of n measurement outcomes. These outcomes occur at random and we assign the random variable K_{quant} to be the algorithm output. The algorithm can then be considered as a way of sampling outcomes according to the distribution $\Pr\left(K_{\text{quant}} = k\right)$. To simulate a quantum algorithm it suffices to give a simulating algorithm which samples from the distribution $\Pr\left(K_{\text{quant}} = \boldsymbol{k}\right)$, which is what we do here. Notice that this form of simulation does not allow us to actually infer the distribution of outcomes, but it does suffice for many important tasks (for example, estimating the expected outcome).

The type of algorithms we treat here take the following form (see Figure 4.0.6 for an example):

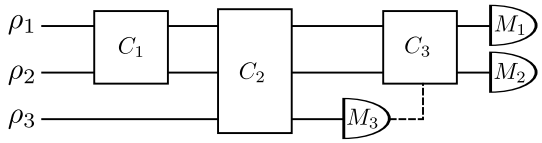


Figure 4.0.6: Example of a stabilizer circuit: ρ_i have positive representation, C_i are Clifford gates and M_i have positive representation. The choice of gate C_3 can be conditioned on the outcome of measurement M_3.

Algorithm 4.1 Family of Simulable Quantum Algorithms
Algorithms in this class sample strings of measurement outcomes $\boldsymbol{k}$ according to the distribution $\Pr\left(K_{\text{quant}} = \boldsymbol{k}\right)$ determined by the Born rule.

---

1. Prepare an initial n qudit input state $\rho_1 \otimes \cdots \otimes \rho_n \in \rho \in L(\mathcal{H}_{p^n})$ where $\rho_1, \ldots, \rho_n$ have positive discrete Wigner representation.

2. Until all registers have been measured:

   a) Apply a Clifford unitary gate U_F, labeled by the symplectic transformation $F \in \mathrm{Sp}(2n, p)$.

   b) Measure the final qudit register using a measurement with positive discrete Wigner representation. Record the outcome k_j of measurement the jth register. Further steps in the computation may be conditioned on the outcome of this measurement.

Notice that there is no loss of generality in considering only symplectic Clifford transformations as the Heisenberg-Weyl component can be rolled into the measurement.

The essential idea for the simulation is to take seriously the hidden variable model the restrictions allow us. In the discrete Wigner picture the system begins at point $\boldsymbol{u}$ in the discrete phase space, which is unknown but definite and fixed. The effect of $U_{\boldsymbol{F}}$ is to move the system from the point $\boldsymbol{u}$ to the point $F\boldsymbol{u}$, and measurement amounts to checking some region of the phase space to see if it contains the system. Since the vector $\boldsymbol{u}$ and matrix F are size $2n$ with entries from $\mathbb{Z}_d$ it is computationally efficient to classically store and update the system's location. Of course, a (positively represented) quantum state corresponds to a probability density over the space so we must treat this a little more carefully. The simulation protocol is:

Algorithm 4.2 Classical Simulation Algorithm
Algorithms in this class sample strings of measurement outcomes $\boldsymbol{k}$ according to the distribution $\Pr(K_{\text{class}} = \boldsymbol{k})$

1. Sample $\boldsymbol{u} \in \mathbb{F}_d^{2n}$ according to the distribution $W_{\rho_1 \otimes \cdots \otimes \rho_n}(\boldsymbol{u}) = W_{\rho_1}(\boldsymbol{u}_1) W_{\rho_2}(\boldsymbol{u}_2) \ldots W_{\rho_n}(\boldsymbol{u}_n)$.
2. Repeat until all registers have been measured:
   a) If the unitary U_F is applied then update $\boldsymbol{u} \to F\boldsymbol{u}$.
   b) If the measurement M with corresponding POVM $\{E_k\}$ is made on the last register of the quantum circuit then report outcome k with probability $W_{E_k}(\boldsymbol{u}_m)$ where $\boldsymbol{u}_m$ is the ontic position of the last qudit system, defined by $\boldsymbol{u} = \boldsymbol{u}_1 \oplus \boldsymbol{u}_2 \cdots \oplus \boldsymbol{u}_m$. If the quantum algorithm conditions further steps on the outcome of measurement on this register then condition further steps of the simulation on measurement outcome k.

Our claim is that the classical algorithm in Algorithm Class 4.2 efficiently simulates the corresponding quantum algorithm in Algorithm Class 4.1. More precisely,

Theorem 8. *An n qudit quantum algorithm belonging to Algorithm Class 4.1 is simulable by the corresponding $2n$ dit classical algorithm in Algorithm Class 4.2 in the sense that the distribution of outcomes $\boldsymbol{k}$ is the same for both algorithms, $Pr(K_{class} = \boldsymbol{k}) = Pr(K_{quant} = \boldsymbol{k})$.*

Proof. The input to the classical circuit is a $2n$ dit string and the transformations are all matrices of size $2n$ with entries in $\mathbb{Z}_d$ so the $2n$ dit portion of the claim is obvious.

To show that this protocol genuinely simulates the circuit it suffices to show any string of measurement outcomes $\boldsymbol{k} = (k_1 k_2 \ldots k_n)$ occurs with the same probability for both the original circuit and the simulation. Lets first consider probability distribution $\Pr(k_n)$ of the outcomes of the first measurement. In the quantum circuit the preparation $\rho_1 \otimes \cdots \otimes \rho_n$ is passed to the (possibly identity) gate U_F and measurement M_n with corresponding POVM $\{E_{k_n}\}$ is applied to the nth register. The probability of getting outcome k_n is then:

$$\begin{aligned} \Pr_{\text{quant}}(k_n) &= \operatorname{Tr}(U_F \rho_1 \otimes \cdots \otimes \rho_n U_F^\dagger \mathbb{I} \otimes \cdots \otimes \mathbb{I} \otimes E_{k_n}) \\ &= \sum_{\boldsymbol{v} \in \mathbb{Z}_d^{2n}} W_{U_F \rho_1 \otimes \cdots \otimes \rho_n U_F^\dagger}(\boldsymbol{v}) W_{\mathbb{I} \otimes \cdots \otimes \mathbb{I} \otimes E_{k_n}}(\boldsymbol{v}) \\ &= \sum_{\boldsymbol{v} \in \mathbb{Z}_d^{2n}} W_{\rho_1 \otimes \cdots \otimes \rho_n}(F^{-1}\boldsymbol{v}) W_{\mathbb{I} \otimes \cdots \otimes \mathbb{I} \otimes E_{k_n}}(\boldsymbol{v}). \end{aligned} \tag{4.0.1}$$

Where we have recast the inner product into the discrete Wigner form for convenience of comparison. We must now establish that the classical circuit has the same distribution.

Classically, if the system is initially at point $\boldsymbol{v}$ on the discrete phase space then probability of getting outcome k_n from the simulation circuit is given by:

$$\begin{aligned} \Pr_{\text{class}}(k_n | \boldsymbol{v} \text{ sampled initially}) &= \Pr_{\text{class}}(k_n | F\boldsymbol{v} \text{ final location}) \\ &= W_{\mathbb{I} \otimes \cdots \otimes \mathbb{I} \otimes E_{k_n}}(F\boldsymbol{v}). \end{aligned}$$

Which just says that the system is moved from point $\boldsymbol{v}$ to point $F\boldsymbol{v}$ and the probability of outcome k_n is the probability we see the system when we look at the region of phase space measured by E_{k_n}, which is $W_{E_{k_n}}(F\boldsymbol{v})$ by definition. The total probability of outcome k_n is then:

$$\begin{aligned}\Pr_{\text{class}}(k_n) &= \sum_{\boldsymbol{v}\in\mathbb{Z}_d^{2n}} \Pr_{\text{class}}(\boldsymbol{k}|\boldsymbol{v} \text{ sampled init.})\Pr(\boldsymbol{v} \text{ sampled initially}) \\ &= \sum_{\boldsymbol{v}\in\mathbb{Z}_d^{2n}} W_{E_{k_1}\otimes\cdots\otimes E_{k_n}}(F\boldsymbol{v})W_{\rho_1\otimes\cdots\otimes\rho_n}(\boldsymbol{v}) \\ &= \sum_{\boldsymbol{v}\in\mathbb{Z}_d^{2n}} W_{E_{k_1}\otimes\cdots\otimes E_{k_n}}(\boldsymbol{v})W_{\rho_1\otimes\cdots\otimes\rho_n}(F^{-1}\boldsymbol{v}). \qquad (4.0.2)\end{aligned}$$

Comparing Algorithm Class (4.0.1), the distribution of measurement outcomes on the last register for the quantum circuit, and Algorithm Class (4.0.2), the simulated distribution of measurement outcomes on the last register, we see they are the same.

If the quantum algorithm is independent of the measurement outcomes then simply applying the above argument to each register would suffice to complete the proof. However, in general adaptive schemes are possible, such as the algorithm illustrated in Figure 4.0.6 on page 41 where the final gate applied depends on the outcome of the measurement on the third qudit. Using the assumption that the registers are measured from last to first we can factor the distribution of outcome strings as

$$\Pr(\boldsymbol{k}) = \Pr(k_1|k_2\ldots k_n)\Pr(k_2|k_3\ldots k_n)\ldots\Pr(k_{n-1}|k_n)\Pr(k_n).$$

Since the simulation conditions on measurement outcome in exactly the same way as the original quantum algorithm a simple inductive argument shows that the distribution of outcomes must be the same for the quantum algorithm and its classical simulator. □

Corollary 9. *Quantum algorithms belonging to Algorithm Class 4.1 offer no super linear advantage over classical computation.*

Proof. We have seen that if it is computationally efficient (linear in the number of qudits) to sample from the classical distributions corresponding to the input state and the measurements then such quantum circuits are efficiently simulable. Since we have assumed separability of the input and measurements and the discrete Wigner function factors this efficient sampling is guaranteed. □

A couple of remarks are in order. We have restricted ourselves to separable inputs and measurements, but this is not strictly necessary for efficient simulation. Any positively represented preparation or measurement can be accommodated provided it is possible to classically efficiently sample from the corresponding distribution. Since it is exponentially difficult to even write down general quantum states this is a rather strong restriction.

Finally, we note that in some situations it may be natural to increase the size of the input register conditional on measurement outcomes. This can be accounted for in the simulation protocol above by simply increasing the size of the phase space accordingly and sampling from the new additional positive Wigner functions.

5

BOUND MAGIC STATES

A corollary of the simulation result just given is that states with positive Wigner representation are not distillable by any stabilizer protocol within the broad class encompassed in algorithm 4.1. Since there are mixed magic states with positive Wigner representation this implies the existence of bound magic states. However, the simulation aspect of this argument actually obfuscates the conceptual role of negativity of the Wigner function as a resource for magic state computation. This chapter gives a direct proof of this fact that clarifies things considerably.

First, recall that the most general way to map magic states to each other via stabilizer operations is:

Definition. A stabilizer protocol is any map from $\rho \in \mathcal{S}(\mathcal{H}_{d^n})$ to $\sigma \in \mathcal{S}(\mathcal{H}_{d^m})$ composed from the following operations:

1. Clifford unitaries, $\rho \to U\rho U^\dagger$
2. Composition with stabilizer states, $\rho \to \rho \otimes S$ where S is a stabilizer state
3. Post selected computational basis measurement on the final qudit,
$$\rho \to (\mathbb{I} \otimes |i\rangle\langle i|)\, \rho\, (\mathbb{I} \otimes |i\rangle\langle i|) \,/ \mathrm{Tr}\,(\rho \mathbb{I} \otimes |i\rangle\langle i|) \text{ with probability } \mathrm{Tr}\,(\rho \mathbb{I} \otimes |i\rangle\langle i|)$$
4. Partial trace of the final qudit, $\rho \to \mathrm{Tr}_n(\rho)$

and classical randomness.

To show that states with positive Wigner functions are not distillable we can just show that none of the above operations can create negative Wigner representation, that is:

Theorem 10. *States with positive Wigner representation are not distillable.*

Proof. The only pure states with positive representation are stabilizer states so it suffices to show that if Λ is an arbitrary stabilizer protocol and ρ is a state with postive Wigner representation then $\Lambda(\rho)$ also has positive Wigner representation. We establish this by showing that each step of any stabilizer protocol preserves positivity. Suppose $\rho \in L\left(\mathcal{H}_{p^n}\right)$ has positive Wigner representation, then:

1. $U\rho U^\dagger$ has positive Wigner representation for any Clifford operator U. This follows since U acts as a permutation of the discrete phase space.

2. $\rho \otimes S$ has positive Wigner representation for any stabilizer state S. This follows since $W_{\rho\otimes S}(\boldsymbol{u},\boldsymbol{v}) = W_\rho(\boldsymbol{u}) W_S(\boldsymbol{v})$ and S has positive Wigner representation by discrete Hudson's theorem.

3. $(\mathbb{I}\otimes|i\rangle\langle i|)\rho(\mathbb{I}\otimes|i\rangle\langle i|)/\mathrm{Tr}(\rho\mathbb{I}\otimes|i\rangle\langle i|)$ has positive Wigner representation. First, notice that $\mathrm{Tr}(\rho\mathbb{I}\otimes|i\rangle\langle i|) > 0$[1] so we can just consider $(\mathbb{I}\otimes|i\rangle\langle i|)\rho(\mathbb{I}\otimes|i\rangle\langle i|)$. For any computational basis state $|i\rangle \in \mathcal{H}_p$ it holds that $|i\rangle\langle i| = \sum_{\boldsymbol{u}} c_{\boldsymbol{u}} A_{\boldsymbol{u}}$ where $c_{\boldsymbol{u}}$ is either 0 or $\frac{1}{d}$. Armed with this fact:

$$\begin{aligned} p^n W_{(\mathbb{I}\otimes|i\rangle\langle i|)\rho(\mathbb{I}\otimes|i\rangle\langle i|)}(\boldsymbol{u},\boldsymbol{v}) &= \mathrm{Tr}\left(A_{\boldsymbol{u}}\otimes A_{\boldsymbol{v}}(\mathbb{I}\otimes|i\rangle\langle i|)\rho(\mathbb{I}\otimes|i\rangle\langle i|)\right) \\ &= \langle i|A_{\boldsymbol{u}}|i\rangle \mathrm{Tr}(A_{\boldsymbol{v}}\otimes|i\rangle\langle i|\rho) \\ &= \langle i|A_{\boldsymbol{u}}|i\rangle \sum_{\boldsymbol{w}} c_{\boldsymbol{w}} \mathrm{Tr}(A_{\boldsymbol{v}}\otimes A_{\boldsymbol{w}}\rho) \\ &\geq 0 \end{aligned}$$

   where the last line follows since $\langle i|A_{\boldsymbol{u}}|i\rangle \geq 0$, $c_{\boldsymbol{w}} \geq 0$ and $\mathrm{Tr}(A_{\boldsymbol{v}}\otimes A_{\boldsymbol{w}}\rho) \propto W_\rho(\boldsymbol{v},\boldsymbol{w}) \geq 0$.

4. $\mathrm{Tr}_n(\rho)$ has positive Wigner representation. This follows since $W_{\mathrm{Tr}_n(\rho)}(\boldsymbol{u}) = \sum_{\boldsymbol{v}} W_\rho(\boldsymbol{u},\boldsymbol{v})$.

Thus any stabilizer protocol preserves positive Wigner representation.

□

The significance of this proof is that it gives a very strong indicator of how we can move beyond the binary question of whether a magic state is at all useful for quantum computation to *quantifying* how useful a state is.

1 In the case $\mathrm{Tr}(\rho\mathbb{I}\otimes|i\rangle\langle i|) = 0$ the measurement outcome i could not have occured.

Part II

QUANTIFYING MAGIC

This part of the thesis covers the results of [72]. The major practical obstacle to a physical implementation of magic state computation is not the inability to produce distillable states but rather the enormous number of resource states required. To do magic state distillation efficiently we must determine how to best use stabilizer operations to transform the resource magic states we are able to produce into the target magic states that will be consumed to implement non-Clifford gates. The transformation of resource states using a restricted set of operations is the province of resource theories[40]. In this part of the thesis we develop the resource theory of magic: a resource theory for quantum computation using stabilizer operations supplemented with magic states.

To develop a resource theory for magic state computation we will divide the set of quantum states into those that can be prepared using the stabilizer formalism, the *stabilizer* states, and those that can not, the *magic* states. This is analogous to the division that is imposed in the resource theory of quantum communication between the states that can be prepared using local operations and classical communication, the separable states, and those that can not, the entangled states. The question we wish to answer is how to best use stabilizer operations to transform resource magic states ρ_{res} into the target magic states σ_{target} required for the implementation of non-stabilizer gates. This is best considered as two distinct problems:

1. Is it possible to produce even a single copy of σ_{target} from any number of copies of ρ_{res}?

2. Assuming distillation is possible, how efficiently can it be done? That is, how many copies of ρ_{res} are required to produce m copies σ_{target}?

The known protocols are able to distill some, but not all, resource magic states ρ_{res} to target states useful for quantum computation. Until very recently it wasn't even known whether some distillation protocol could be found to take any magic state to a nearly pure magic state. As we have seen in the previous part of the thesis the answer to this question (at least in odd dimensions) is no: there is a large class of *bound magic states* that are not distillable to pure magic states using any protocol. There has also been some intriguing work on this problem in the qubit case [12, 64, 63]. The second question is the primary focus of this part of the thesis. The key insight is the introduction of a quantitative measure of how magic a quantum state is, which in particular allows us to upper bound the distillation efficiency. Roughly speaking, suppose the target state is five times as magical as the resource state according to some measure, then we can immediately infer that at least five resource states will be required for each copy of the target state.[2]

To quantify the amount of magic resource in a quantum state we introduce the notion of a *magic monotone*. This is any function mapping quantum states to real numbers that is non-increasing under stabilizer operations. This is just the common sense requirement that the amount of non-stabilizer resource available can not be increased using only stabilizer operations. Magic monotones are valid measures of the magic of a quantum state in exactly the same way entanglement monotones are valid measures of the entanglement of a quantum state. The main contribution of this part of the thesis is the introduction and study of two magic monotones: the *relative entropy of magic* and the *mana*.

The requirement that a magic monotone be non-increasing under stabilizer operations can be formalized as:

Definition 11. Let $\mathcal{M}_d : \mathcal{S}(\mathcal{H}_d) \to \mathbb{R}$ be a mapping from the space of density operators on $\mathcal{H}_d \cong \mathbb{C}^d$ to the real numbers. Define $\mathcal{M}(\rho) \equiv \mathcal{M}_d(\rho) \ \forall \rho \in \mathcal{S}(\mathcal{H}_d)$ so that $\mathcal{M}(\cdot)$ is defined for all finite dimensional Hilbert spaces. If, on average, $\mathcal{M}(\Lambda(\rho)) \leq \mathcal{M}(\rho)$ for any stabilizer protocol Λ then we say $\mathcal{M}(\cdot)$ is a magic monotone.

2 Since magic monotones need not be additive and are only non-increasing *on average* this is not literally true, but the intuition is sound.

There are two important points to notice here. The first is that we require operations to be magic non-increasing only on average; if $\Lambda(\rho) = \sigma_i$ with probability p_i then we only require $\mathcal{M}(\rho) \geq \sum_i p_i \mathcal{M}(\sigma_i)$. In particular this means that post selected measurement can increase magic in the sense that we allow $\mathcal{M}\left(\left(\mathbb{I} \otimes |i\rangle\langle i|\right) \rho \left(\mathbb{I} \otimes |i\rangle\langle i|\right) / \mathrm{Tr}\left(\rho \mathbb{I} \otimes |i\rangle\langle i|\right)\right) \geq \mathcal{M}(\rho)$ as long as measurement outcome i is obtained with sufficiently small probability. This allows us to analyze non-deterministic protocols. The second point is that we do not require convexity, i.e. it can happen that $\mathcal{M}(p\rho + (1-p)\sigma) \geq p\mathcal{M}(\rho) + (1-p)\mathcal{M}(\sigma)$. Although convexity is a desirable feature it is possible to have interesting monotones that are not convex (for example, the logarithmic entanglement negativity[59]).

Notice also that because Clifford gates and composition with stabilizer states are reversible within the stabilizer formalism (by the inverse gate and the partial trace respectively) any monotone must actually be invariant under these operations, as opposed to merely non-increasing.

With the formal definition in hand we can now understand how magic monotones allow us to put bounds on the efficiency of magic state distillation. Suppose we wish to distill m copies of σ from n copies of ρ. If $\mathcal{M}(\cdot)$ is a magic monotone then for any stabilizer protocol Λ such that $\Lambda(\rho^{\otimes n}) = \sigma_i^{\otimes m_i}$ with probability p_i we will have $\mathcal{M}(\rho^{\otimes n}) \geq \sum p_i \mathcal{M}(\sigma_i)$. In particular this means that n must be large enough so that $\mathcal{M}(\rho^{\otimes n}) \geq p_\sigma \mathcal{M}(\sigma)$ where p_σ is the probability that the protocol produces σ.

6

THE RELATIVE ENTROPY OF MAGIC

Generic resource theories can, and usually do, admit more than one valid choice of monotone. Requiring a function to be non-increasing under stabilizer operations is the minimal imposition for it to be a sensible measure of magic. However, there is no guarantee that all monotones will be equally interesting or useful. This leads us to wonder if some further natural conditions could be imposed to eliminate some of these measures and pick out especially interesting monotones. Resource theories are concerned with the problem of using restricted operations to convert between different types of resource states, for example distilling pure magic states from mixed ones or changing one type of pure magic state to another type of pure magic state. Most often this conversion is studied in the asymptotic regime (eg. [31, 40, 41, 39, 32]) where an infinite number of resource states are assumed to be available to conversion protocols and the task is to determine the rate at which one type of resource can be converted into another. In this regime it turns out that for many resource theories the monotone zoo can be cut down in a rather spectacular fashion: there is a monotone that *uniquely* specifies the rate at which the asymptotic interconversion of resource states can take place. Because of the importance of asymptotic interconversion of resource states this measure is often called the unique measure of the resource[40]. For magic theory this vaunted measure is the regularized relative entropy of magic; the purpose of this chapter is to introduce this quantity and explain in what sense it is unique.

The relative entropy distance between quantum states is $S(\rho\|\sigma) \equiv \mathrm{Tr}(\rho\log\rho) - \mathrm{Tr}(\rho\log\sigma)$. This is a measure of how distinguishable ρ is from σ. Qualitatively, we might expect that a measure of how distinguishable ρ is from all stabilizer states to be a good measure of magic, which inspires the definition:

Definition 12. Let $\rho \in \mathcal{S}(\mathcal{H}_d)$, then the relative entropy of magic is $\mathrm{r}_{\mathcal{M}}(\rho) \equiv \min_{\sigma\in\mathrm{STAB}(\mathcal{H}_d)} S(\rho\|\sigma)$.

The intuition that this should be a magic measure is immediately validated:

Theorem 13. *The relative entropy of magic is a magic monotone.*

Proof. This is essentially a consequence of the nice properties of the relative entropy and holds for the same reasons that the relative entropy is a monotone for other resources theories. See Section 6.5 for details. □

The main importance of the relative entropy of magic is in the asymptotic regime where a slightly modified definition is required. We will find that the relative entropy of

magic is subadditive in the sense $\mathrm{r}_{\mathcal{M}}(\rho^{\otimes n}) < n\mathrm{r}_{\mathcal{M}}(\rho)$. This is because in general there can be some entangled stabilizer state $\sigma_{AB} \in \mathcal{S}(\mathcal{H}_d \otimes \mathcal{H}_d)$ such that $S(\rho \otimes \rho \| \sigma_{AB}) < \min_{\sigma_A, \sigma_B \in \mathrm{STAB}(\mathcal{H}_d)} S(\rho \otimes \rho \| \sigma_A \otimes \sigma_B)$. In particular this means that the amount of magic added by producing a single copy of ρ depends on how many copies of ρ we already have access to. In the asymptotic limit an appropriate measure should give the amount of magic in ρ when an infinite number of copies of ρ are available. This prompts us to introduce the asymptotic variant of the relative entropy measure:

Definition 14. Let $\rho \in \mathcal{S}(\mathcal{H}_d)$, then the regularized relative entropy of magic is $\mathrm{r}^{\infty}_{\mathcal{M}}(\rho) \equiv \lim_{n\to\infty} \frac{1}{n} \mathrm{r}_{\mathcal{M}}(\rho^{\otimes n})$.

We do not have an analytic expression for the relative entropy of magic and thus we do not have an analytic expression for the asymptotic version. Moreover, because of the infinite limit in the definition we do not even know how to numerically approximate $\mathrm{r}^{\infty}_{\mathcal{M}}(\rho)$ in general. This is the same as the situation in entanglement theory where it remains a famous open problem to find a "single letter" expression for the regularized relative entropy of entanglement. Nevertheless, the (regularized) relative entropy of magic is useful for the holistic study of magic theory. For instance, we will use it as a tool to show that a non-zero amount of pure magic states are always required to produce mixed magic states via stabilizer protocols, even when the mixed state is a bound magic state.

6.1 RELATIVE ENTROPY OF MAGIC

One of the major difficulties with the study of resource monotones is that the actual computation of the value of the monotone for a particular state is often an intractable problem. Although we do not know a simple analytic expression for the relative entropy of magic it can be computed numerically. For systems with low Hilbert space dimension this is reasonably straightforward. The relative entropy is a convex function and we want to perform minimization over the convex set of stabilizer states. This means that a simple numerical gradient search will succeed in finding $\min_{\sigma \in \mathrm{STAB}(\mathcal{H}_d)} S(\rho \| \sigma)$. Each qudit stabilizer state can be written as a convex combination of the N pure qudit stabilizer states. A simple strategy for finding the relative entropy of magic is to do a numerical search over the $N-1$ values that define the probability distribution over the stabilizer states. Unfortunately, for a system of n qudits the number of pure stabilizer states is[34]

$$N = d^n \prod_{i=1}^{n} \left(d^i + 1\right),$$

and this grows too quickly for a numerical search to be viable in general. For example, the original H-type magic state distillation protocol[6] consumes 15 resource states ρ_{input} to produce an output magic state σ_{output} with higher purity. In principle we can bound the quantity of the resource required via

$$\mathrm{r}_{\mathcal{M}}\left(\rho_{\mathrm{input}}^{\otimes 15}\right) \geq p\mathrm{r}_{\mathcal{M}}\left(\rho_{\mathrm{output}}\right),$$

where p is the success probability of the protocol, but this would require a numerical optimization over more than 2^{136} parameters using the approach just outlined, which is not viable.

For arbitrary resource states it is not clear how to avoid the numerical optimization. However, the states typically used in magic state distillation protocols have a great deal of additional structure that can be exploited. In particular, many protocols have a "twirling" step where a random Clifford unitary is applied to the resource state to ensure it has the form,

$$\rho_\epsilon = (1-\epsilon)\left|M\right\rangle\!\left\langle M\right| + \epsilon\frac{\mathbb{I}}{d}.$$

If the twirling map is $\mathcal{T} : \rho_{\text{resource}} \to \sum_i p_i U_i \rho_{\text{resource}} U_i^\dagger$ for some subset $\{U_i\}$of the Clifford operators then,

$$\begin{aligned}
\min_{\sigma\in\text{STAB}} S\left((1-\epsilon)\left|M\right\rangle\!\left\langle M\right| + \epsilon\frac{\mathbb{I}}{d}\|\sigma\right) &\geq \min_{\sigma\in\text{STAB}} S\left(\mathcal{T}\left((1-\epsilon)\left|M\right\rangle\!\left\langle M\right| + \epsilon\frac{\mathbb{I}}{d}\right)\|\mathcal{T}(\sigma)\right) \\
&= \min_{p\leq p_T} S\left((1-\epsilon)\left|M\right\rangle\!\left\langle M\right| + \epsilon\frac{\mathbb{I}}{d}\|p\left|M\right\rangle\!\left\langle M\right| + (1-p)\frac{\mathbb{I}}{d}\right) \\
&= S\left((1-\epsilon)\left|M\right\rangle\!\left\langle M\right| + \epsilon\frac{\mathbb{I}}{d}\|p_T\left|M\right\rangle\!\left\langle M\right| + (1-p_T)\frac{\mathbb{I}}{d}\right),
\end{aligned}$$

where p_T is the largest value such that $p_T\left|M\right\rangle\!\left\langle M\right| + (1-p_T)\frac{\mathbb{I}}{d}$ is a stabilizer state. This means that the relative entropy of magic can be computed exactly for states of this form by finding p_T. Unfortunately the twirling is only applied to individual qudits so this does not by itself resolve the numerical problems. Nevertheless, it is possible to give naive bounds according the following observation:

$$\begin{aligned}
\mathrm{r}_{\mathcal{M}}\left(\rho_{\text{output}}\right) &\leq \mathrm{r}_{\mathcal{M}}\left(\rho_{\text{input}}^{\otimes n}\right) \\
&\leq n\mathrm{r}_{\mathcal{M}}\left(\rho_{\text{input}}\right)
\end{aligned}$$

where we have used the obvious fact that the relative entropy of magic is subadditive.

This bound might not seem naive at all. One might suspect that the relative entropy of magic is genuinely additive so $\mathrm{r}_{\mathcal{M}}\left(\rho_{\text{input}}^{\otimes n}\right) \stackrel{?}{=} n\mathrm{r}_{\mathcal{M}}\left(\rho_{\text{input}}\right)$. This seems like a very desirable feature for a monotone to have: n copies of a resource state should contain n times as much resource as a single copy. The relative entropy of magic does not have this feature, it can be the case that $\mathrm{r}_{\mathcal{M}}\left(\rho^{\otimes 2}\right) \lneq 2\mathrm{r}_{\mathcal{M}}\left(\rho\right)$. To establish this we consider the qutrit *Strange* state $\left|\mathrm{S}\right\rangle\!\left\langle \mathrm{S}\right|$ defined as the pure qutrit state invariant under the symplectic component of the Clifford group (see Section 7.4). Twirling by the symplectic subgroup $\mathrm{Sp}\left(2,3\right)$[1] of the Clifford group has the effect

$$\sum_F \frac{1}{\left|\mathrm{Sp}\left(2,3\right)\right|} U_F \rho U_F^\dagger = \left(1-\epsilon_\rho\right)\left|\mathrm{S}\right\rangle\!\left\langle \mathrm{S}\right| + \epsilon_\rho\frac{\mathbb{I}_3}{3},$$

so we can use our twirling argument to find $\mathrm{r}_{\mathcal{M}}\left(\left|\mathrm{S}\right\rangle\!\left\langle \mathrm{S}\right|\right)$ exactly. A numerical search over the two qutrit stabilizer states turns up a state $\sigma \in \mathrm{STAB}\left(\mathcal{H}_9\right)$ such that $S\left(\left|\mathrm{S}\right\rangle\!\left\langle \mathrm{S}\right|^{\otimes 2}\|\sigma\right) < 2\mathrm{r}_{\mathcal{M}}\left(\left|\mathrm{S}\right\rangle\!\left\langle \mathrm{S}\right|\right)$.

1 this is the Clifford group modulo the Heisenberg-Weyl (Pauli) group.

The relative entropy of entanglement is also subadditive. However, there is a very important difference between the entanglement and magic measures: for *pure* states the relative entropy of entanglement is an additive measure. This fact is at the heart of the importance of the relative entropy distance for the theory of entanglement. As we have just seen this is not true for the relative entropy of magic.

6.2 THE (REGULARIZED) RELATIVE ENTROPY OF MAGIC IS FAITHFUL

The relative entropy $S(\rho\|\sigma)$ is 0 if and only if $\rho = \sigma$. It is easy to see that this implies that $\mathrm{r}_{\mathcal{M}}(\rho)$ is *faithful* in the sense that $\mathrm{r}_{\mathcal{M}}(\rho) > 0$ if and only if ρ is magic. Since $\mathrm{r}_{\mathcal{M}}(\cdot)$ is a magic monotone, if it is possible to create a magic state σ from a pure resource state $|\psi\rangle\langle\psi|$ using a stabilizer protocol it must be the case that $\mathrm{r}_{\mathcal{M}}(|\psi\rangle\langle\psi|) \geq \mathrm{r}_{\mathcal{M}}(\sigma)$. Together these facts imply that to create any magic state by consuming pure magic states a non-zero number of pure magic state states are required. However, we have already established that there are bound magic states that can not be distilled to pure magic states. This means that the amount of magic that can be distilled from a resource state is not equal to the amount of magic required to create it; this is the analogue in the present case of the famous result in entanglement theory that the entanglement of creation is not equal to the entanglement of distillation.

Because the relative entropy of magic is subadditive it could be that $\mathrm{r}^{\infty}_{\mathcal{M}}(\rho) = \lim_{n\to\infty} \frac{1}{n}\mathrm{r}_{\mathcal{M}}(\rho^{\otimes n}) = 0$ for some magic state ρ; it is not automatic that the regularized relative entropy of magic is faithful. For example, in the resource theory of asymmetry[31] the regularized relative entropy measure is 0 for all states. Happily, for magic theory the relative entropy is well behaved in the asymptotic regime:

Lemma 15. *The regularized relative entropy of magic is faithful in the sense that $r^{\infty}_{\mathcal{M}}(\rho) = 0$ if and only if ρ is a stabilizer state.*

Proof. The proof of this fact is a straightforward application of a theorem of Piani[58] showing that the regularized relative entropy measure is faithful for all resource theories where the set of restricted operations includes tomographically complete measurements and the partial trace. The idea is to define a variant of the relative entropy distance that quantifies the distinguishability of states using only stabilizer measurements. This quantity lower bounds the usual relative entropy of magic so by showing that its regularization is faithful we get the claimed result. See Subsection 6.5 for details. □

We will need this result for the proof of corollary 17 showing that the regularized relative entropy gives the optimal rate of asymptotic interconversion. It also represents a strengthening of our earlier claim that a non-zero amount of pure state magic is required to create any magic state. For finite size protocols this followed from the faithfulness of the relative entropy of magic, as just explained. The faithfulness of the regularized relative entropy implies that the creation of magic states by an asymptotic stabilizer protocol requires resource magic states to be consumed at a non-zero rate. The analogous problem in entanglement theory was the main motivation for proving that the regularized relative entropy of entanglement is faithful[58][5].

We shall give an additional requirement for a magic monotone to be a magic measure that is well motivated in the scenario of asymptotic state conversion. To understand this it is simplest to first consider the case of finite state interconversion. Suppose there is some stabilizer protocol Λ that maps n copies of resource state ρ to m copies of a target magic state σ. Then it must be the case that $\mathcal{M}\left(\rho^{\otimes n}\right) \geq \mathcal{M}\left(\sigma^{\otimes m}\right)$ for any magic monotone $\mathcal{M}\left(\cdot\right)$. If there is also some other stabilizer protocol that maps σ to ρ then it must be the case that $\mathcal{M}\left(\rho^{\otimes n}\right) = \mathcal{M}\left(\sigma^{\otimes m}\right)$, which conceptually just means that if $\rho^{\otimes n}$ and $\sigma^{\otimes m}$ are equivalent resources then they have the same magic according to any magic measure. It is rarely possible to exactly interconvert between resource states with only a finite number of copies available. The condition that picks out the regularized relative entropy of magic as the unique measure of magic is, roughly speaking, the requirement that asymptotically reversibly interconvertible states have the same resource value.

Typically if we try to convert $\rho^{\otimes n}$ into m copies of σ, the stabilizer protocol $(\Lambda_n : \mathcal{H}_{d^n} \to \mathcal{H}_{d^m})$ we use will depend on the number of input states n. When converting ρ to σ it is thus necessary to consider a family of stabilizer protocols Λ_n taking $\rho^{\otimes n}$ as input and producing $m(n)$ approximate copies of σ with an error $\left\|\Lambda_n\left(\rho^{\otimes n}\right) - \sigma^{\otimes m(n)}\right\|_1 = \epsilon_n$. In the case that the approximation becomes arbitrarily good in the asymptotic limit (ie. $\lim_{n\to\infty} \epsilon_n \to 0$) we say ρ is *asymptotically convertible* to σ at a rate $R\left(\rho \to \sigma\right) = \lim_{n\to\infty} \frac{m(n)}{n}$. The additional condition we impose is that if $\rho^{\otimes n}$ is asymptotically convertible to $\sigma^{\otimes m(n)}$ then,

$$\lim_{n\to\infty} \frac{1}{n}\left[\mathcal{M}\left(\rho^{\otimes n}\right) - \mathcal{M}\left(\sigma^{\otimes m(n)}\right)\right] \geq 0. \tag{6.3.1}$$

That is, if asymptotic conversion is possible then on average we must put in at least as much magic as we get out, up to some $o(n)$ discrepancy.

If it is possible to interconvert between σ and ρ at rates $R\left(\sigma \to \rho\right) = R\left(\rho \to \sigma\right)^{-1}$ then we say the two resources are *asymptotically reversibly interconvertible*. Any magic monotone satisfying this additional condition gives the rate of asymptotic interconversion according to the following theorem:

Theorem 16. *Let $\mathcal{M}\left(\cdot\right)$ be a magic monotone satisfying the condition given by (6.3.1) and define its asymptotic variant $\mathcal{M}^\infty\left(\rho\right) = \lim_{n\to\infty} \mathcal{M}\left(\rho^{\otimes n}\right)/n$. Then if it is possible to asymptotically reversibly interconvert between magic states ρ and σ and $\mathcal{M}^\infty\left(\rho\right)$ is non-zero then the rate of conversion is given by $R\left(\rho \to \sigma\right) = \mathcal{M}^\infty\left(\rho\right)/\mathcal{M}^\infty\left(\sigma\right)$.*

Proof. This is a special case of broader theorem that says this result holds in any resource theory. The result was first proved in [41]. That paper dealt primarily with entanglement and missed the requirement that the regularization of the monotone needs to be non-zero. This was pointed out in [31], and the theorem we state here is essentially the application of their Theorem 4 to magic theory. The only subtlety is that instead of condition (6.3.1) they require the monotone to be *asymptotically continuous*, which means $\lim_{n\to\infty} \left\|\Lambda_n\left(\rho^{\otimes n}\right) - \sigma^{\otimes m(n)}\right\|_1 \to 0$ implies

$$\lim_{n\to\infty} \frac{\mathcal{M}\left(\Lambda_n\left(\rho^{\otimes n}\right)\right) - \mathcal{M}\left(\sigma^{\otimes m(n)}\right)}{1+n} \to 0.$$

The first step of their proof is to show that this cryptic condition implies (6.3.1) so we start with the weaker, interpretable requirement directly. □

Corollary 17. *The regularized relative entropy of magic is a unique measure of magic in the sense that if it is possible to asymptotically reversibly interconvert between magic states ρ and σ the rate at which this can be done is $R(\rho \to \sigma) = r^{\infty}_{\mathcal{M}}(\sigma) / r^{\infty}_{\mathcal{M}}(\rho)$.*

Proof. In [67] it is shown that the relative entropy distance to any convex set of quantum states is asymptotically continuous. Since asymptotic continuity implies (6.3.1) and the stabilizer states are a convex set the relative entropy of magic is a magic monotone satisfying condition (6.3.1). Moreover, we showed in Theorem 15 that the regularized relative entropy is non-zero for all magic states. □

Notice that the relative entropy is only one example of a monotone satisfying the conditions of Theorem 16. There could be other monotones for which this result holds. In fact it is conceivable that this result holds for every magic monotone. Really what is meant by the use of the word 'unique' here is that the conversion rate predicted by the measure is unique. This implies that for any magic monotone with this property if it is possible to asymptotically interconvert between ρ and σ it must be the case that $\mathcal{M}^{\infty}(\rho) = C\mathrm{r}^{\infty}_{\mathcal{M}}(\rho) \implies \mathcal{M}^{\infty}(\sigma) = C\mathrm{r}^{\infty}_{\mathcal{M}}(\sigma)$ so $\mathrm{r}^{\infty}_{\mathcal{M}}(\sigma) / \mathrm{r}^{\infty}_{\mathcal{M}}(\rho) = \mathcal{M}^{\infty}(\sigma) / \mathcal{M}^{\infty}(\rho)$. Ie. the regularization of such magic measures can differ only up to a multiplicative factor that can vary between sets of quantum states where asymptotic interconversion is possible.

If we have a resource measure $\mathcal{M}(\cdot)$ that is additive then it will be equal to its own regularization, $\mathcal{M}(\cdot) = \mathcal{M}^{\infty}(\cdot)$. If this measure also satisfies 6.3.1 then it will tell us how to compute the asymptotic interconversion rate whenever asymptotic inversion is possible. In the particular case that we have a resource theory where asymptotic interconversion is possible between any two resource states then it is easy to see that up to a constant factor there really is a single unique measure of magic. Much of the work on resource theories has been either specifically focused on or inspired by the theory of bipartite entanglement. In the case of bipartite pure entangled states there is an additive resource measure which satisfies our condition, namely the entanglement entropy[4]. Moreover, every bipartite pure entangled state is asymptotically interconvertible using LOCC. Thus the entanglement entropy is the genuinely unique measure of pure state bipartite entanglement. One of the special features of the relative entropy of entanglement is that it reduces to the entanglement entropy on pure states. It is this feature which is ultimately responsible for the privileged status of the relative entropy of entanglement. In the case of magic theory the relative entropy of magic does not reduce to an additive measure on pure states so *there is no apparent reason to prefer the relative entropy of magic over any other monotone satisfying the conditions of Theorem 16*. This stands in contrast to the claim that the relative entropy distance to the set of free states is the unique measure of the resource (eg. [40]).

6.4 DISCUSSION

The privileged status of the relative entropy of magic comes from its role in the asymptotic regime. Since the assumption of infinite state preparations is unreasonable for an actual

physical system one might expect that the measure would be of limited practical value. This suspicion is lent additional weight by the fact that, like the relative entropy distance in other resource theories, actually computing $\mathrm{r}_{\mathcal{M}}(\rho)$ appears to be intractably difficult. The regularized relative entropy distance is essentially useless for analyzing the performance of particular distillation protocols. Nevertheless, the monotone is a useful tool for the holistic study of the resource theory of magic. This is the role of the regularized relative entropy distance in the theory of entanglement, where it is a well studied object. We had a taste of this already in our demonstration that the amount of pure state magic required to create a magic state does not equal the amount of pure state magic that can be distilled from that state. It is an exciting direction for future work to see what other insights can be gleaned from the relative entropy of magic and its asymptotic variant.

It is also important to understand exactly what corollary 17 says. The statement is that *if* asymptotic interconversion is possible then the rate is given by $\mathrm{r}^{\infty}_{\mathcal{M}}(\rho)/\mathrm{r}^{\infty}_{\mathcal{M}}(\sigma)$. This if clause is a deceptively strong requirement: it is not guaranteed that asymptotic interconversion will always be possible, or even that it will ever be possible. In particular, every currently known magic state distillation protocol has rate 0 and it is an important open problem to determine if a positive rate distillation protocol exists.

6.5 PROOFS

This section presents the details of the proofs that were omitted from the main text of the chapter in order to improve readability.

We begin by showing that the relative entropy is a valid measure of magic.

Relative entropy of magic is a monotone

Theorem. *The relative entropy of magic is a magic monotone.*

Proof. We need to verify that this function is non-increasing under stabilizer operations.

1. Invariance under Clifford unitaries. For any unitary, $S\left(U\rho U^{\dagger}\|U\sigma U^{\dagger}\right) = S(\rho\|\sigma)$. If U is a Clifford and σ is a stabilizer state then $U\sigma U^{\dagger}$ will also be a stabilizer state, ergo $\mathrm{r}_{\mathcal{M}}\left(U\rho U^{\dagger}\right) = \min_{\sigma} S\left(U\rho U^{\dagger}\|\sigma\right) = \min_{\sigma} S\left(U\rho U^{\dagger}\|U\sigma U^{\dagger}\right) = \min_{\sigma} S(\rho\|\sigma) = \mathrm{r}_{\mathcal{M}}(\rho)$.

2. Non-increasing on average under stabilizer measurement. We consider computational basis measurement on the final qudit. Let $\{V_i\} = \{\mathbb{I}\otimes|i\rangle\langle i|\}$ be the measurement POVM and label outcome probabilities $p_i = \mathrm{Tr}\left(V_i\rho\right)$, $q_i = \mathrm{Tr}\left(V_i\sigma\right)$ as well as post-measurement states $\rho_i = V_i\rho V_i^{\dagger}$ and $\sigma_i = V_i\sigma V_i^{\dagger}$. In reference [70] it is shown that

$$\sum_i p_i S\left(\frac{\rho_i}{p_i}\|\frac{\sigma_i}{q_i}\right) \leq S(\rho\|\sigma).$$

   Since σ_i/q_i is a stabilizer state whenever σ is a stabilizer state this implies measurement does not increase the relative entropy of magic on average.

3. Invariance under composition with stabilizer states. $S(\rho \otimes A \| \sigma \otimes A) = S(\rho \| \sigma)$ for any quantum state A, from which it follows $\mathrm{r}_{\mathcal{M}}(\rho \otimes A) \leq \mathrm{r}_{\mathcal{M}}(\rho)$. Equality follows because (4) the relative entropy of magic is non-increasing under the partial trace, ie. $\mathrm{r}_{\mathcal{M}}(\rho) \leq \mathrm{r}_{\mathcal{M}}(\rho \otimes A)$.

4. Non-increasing under partial trace. From the strong subadditivity property of the von Neumann Entropy[50] we have $S(\mathrm{Tr}_B(\rho) \| \mathrm{Tr}_B(\sigma)) \leq S(\rho \| \sigma)$ from which the result follows immediately.

□

We now turn to the asymptotic variant of the relative entropy of magic, $\mathrm{r}^{\infty}_{\mathcal{M}}(\rho) = \lim_{n\to\infty} \mathrm{r}_{\mathcal{M}}(\rho^{\otimes n})/n$. We show that this quantity is non-zero if and only if ρ is a magic state. We will need this result for the proof of Theorem 16.

Regularized relative entropy of magic is faithful.

Theorem. *The regularized relative entropy of magic is faithful in the sense that $r^{\infty}_{\mathcal{M}}(\rho) = 0$ if and only if ρ may be written as a convex combination of stabilizer states.*

Proof. We recover this result as a special case of the main theorem of reference [58]. This paper introduces a variant of the relative entropy measure that quantifies the distinguishability of a quantum state from the set of free states using a restricted set of measurements. Let $\{M_i\}$ be a measurement POVM and define the map,

$$\mathcal{M}(\rho) = \sum_i p_i(\rho) |i\rangle\langle i|, \; p_i(\rho) = \mathrm{Tr}(\rho M_i),$$

where $\{|i\rangle\}$ is any orthonormal set and $\mathcal{M}$ is a map associated to measurement $\{M_i\}$. Letting $\mathbb{M}$ be the set of restricted measurements we can define,

$$\mathbb{M}S(\rho \| \sigma) \equiv \max_{\mathcal{M}\in\mathbb{M}} S(\mathcal{M}(\rho) \| \mathcal{M}(\sigma)).$$

The significance of this quantity is from theorem 1 of [58]:

Theorem. *Consider a restricted set of operations inducing a resource theory. Let $\mathbb{M}$ be the restricted set of measurements (here the stabilizer measurements) and P the set of free states (here the stabilizer states). If the set of free states is closed under restricted measurement and the partial trace then it holds that the regularization of the relative entropy distance to the set of free states $r^{\infty}_{P}(\rho)$ satisfies*

$$r^{\infty}_{P}(\rho) \geq \min_{\sigma\in P} \mathbb{M}S(\rho \| \sigma).$$

The stabilizer formalism satisfy the conditions of the theorem. Moreover, since the stabilizer measurements contain an informationally complete measurement it holds that $\mathbb{M}S(\rho \| \sigma) > 0$ whenever ρ is a magic state. This implies $\mathrm{r}^{\infty}_{\mathcal{M}}(\rho) > 0$ whenever ρ is a magic state. $\mathrm{r}^{\infty}_{\mathcal{M}}(\rho) = 0$ for all stabilizer states ρ is clear, so the claimed result follows.

□

7

NEGATIVITY OF THE WIGNER FUNCTION AS A COMPUTABLE MEASURE OF MAGIC

The results of the previous chapter deal primarily with reversible conversion of magic states in the limit where infinite copies are available, but for magic state distillation we are interested in the one way distillation of resources magic states to pure target magic states in the regime where only a finite number of resource states are available. Because of this there is no reason to prefer the (regularized) relative entropy of magic over any other monotone. Nevertheless, the relative entropy, like any monotone, gives non-trivial bounds on distillation efficiency. There is a more fundamental problem: it is generally computationally intractable to compute the relative entropy and we have no idea how to compute the regularized relative entropy. In particular this means we are unable to find explicit upper bounds for the efficiency of distillation. In this chapter we resolve this issue by introducing a computable measure of magic.

7.1 SUM NEGATIVITY AND MANA

The work in the previous chapter establishing the existence of bound magic states provides a starting place in the search for a computable monotone. The fundamental tool in that construction is the discrete Wigner function. There it was found that a necessary condition for a magic state to be distillable is that it have negative Wigner representation. However, that work is purely binary in the sense that it does not distinguish degrees of negative representation. It is natural to suspect that a state that is "nearly" positively represented is less magic than a state with a large amount of negativity in its representation. Here we

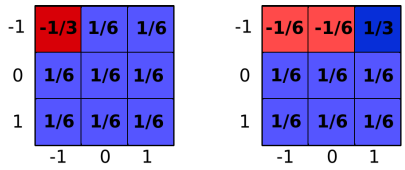


Figure 7.1.1: The Wigner representations of two qutrit states, $|\mathbb{S}\rangle = \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle)$ (left) and $|\mathbb{N}\rangle = \frac{1}{\sqrt{6}}(-|0\rangle + 2|1\rangle - |2\rangle)$ (right). $|\mathbb{S}\rangle$ has sum negativity $\left|-\frac{1}{3}\right|$ and the $|\mathbb{N}\rangle$ has sum negativity $\left|-\frac{1}{6} - \frac{1}{6}\right| = \frac{1}{3}$.

formalize this intuition by showing that the absolute value of the sum of the negative entries of the discrete Wigner representation of a quantum state is a magic monotone.

Definition 18. The sum negativity of a state ρ is the sum of the negative elements of the Wigner function, $\mathrm{sn}(\rho) \equiv \sum_{u:W_\rho(u)<0} |W_\rho(u)| \equiv \frac{1}{2}\left(\sum_u |W_\rho(u)| - 1\right)$.

The right hand side of this expression follows because the normalization of quantum states ($\mathrm{Tr}\rho = 1$) implies $\sum_u W_\rho(u) = 1$. The advantage of writing the expression in this form is that $\|\rho\|_W \equiv \sum_u |W_\rho(u)|$ is a multiplicative norm and is thus very nice to work with. By this we mean that the composition law is given as:

$$\begin{aligned} \|\rho \otimes \sigma\|_W &= \sum_{u,v} |W_{\rho\otimes\sigma}(u,v)| \\ &= \sum_{u,v} |W_\rho(u) W_\sigma(v)| \\ &= \left(\sum_u |W_\rho(u)|\right)\left(\sum_v |W_\rho(v)|\right). \end{aligned} \tag{7.1.1}$$

Since the sum negativity is a linear function of this quantity we can establish that the former is a magic monotone by showing this for the latter:

Theorem 19. *The sum negativity is a magic monotone.*

Proof. It suffices to show $\sum_u |W_\rho(u)|$ is a magic monotone by verifying the required properties. The main components are the use of $\rho = \sum_u W_\rho(u) A_u$ and the composition identity (7.1.1), which is the main motivation for working with this quantity rather than with the sum negativity directly. See subsection 7.7.1 for details. □

The sum negativity is an intuitively appealing way of using the Wigner function to define a magic monotone, but it has some irritating features. The worst of these is the composition law,

$$\mathrm{sn}\left(\rho^{\otimes n}\right) = \frac{1}{2}\left[(2\mathrm{sn}(\rho)+1)^n - 1\right],$$

which has the troubling feature that a linear increase in the number of resource states implies an exponential increase the amount of resource according to the measure. Happily there is a simple resolution to this problem suggested by the composition law (7.1.1), we define a new monotone by a particular function of the sum negativity:

Definition 20. The mana of a quantum state ρ is $\mathscr{M}(\rho) \equiv \log\left(\sum_u |W_\rho(u)|\right) = \log(2\mathrm{sn}(\rho)+1)$.

Theorem 21. *The mana is a magic monotone.*

Proof. Most of the monotone requirements follow because log is a monotonic function, but there is a small subtlety here. Consider a stabilizer protocol that sends $\rho \to \sigma_i$ with probability p_i (eg. post selected computational basis measurement), then we require $\log(\|\rho\|_W) \geq \sum_i p_i \log(\|\sigma_i\|_W)$. This need not be true for arbitrary monotonic functions of $\|\rho\|_W$ but it is easy to see that it follows from the concavity of log and $\|\rho\|_W \geq \sum_i p_i \|\sigma_i\|_W$. □

From equation (7.1.1) this monotone is additive in the sense,

$$\mathscr{M}(\rho \otimes \sigma) = \mathscr{M}(\rho) + \mathscr{M}(\sigma).$$

Beyond its intuitive appeal, additivity is a nice feature for a monotone to have because it makes the associated bound on distillation efficiency take an especially nice form. How many copies n of a resource magic state ρ are required to distill m copies of a magic state σ? Suppose we have a stabilizer protocol $\Lambda(\rho^{\otimes n}) \to \sigma_i$ with probability p_i, then the monotone condition combined with additivity shows:

$$\sum_i p_i \mathscr{M}(\sigma_i) \leq n \mathscr{M}(\rho).$$

Taking $\sigma_0 = \sigma$ and $p_0 = p$, the above discussion lets us see:

Theorem 22. *Suppose Λ is a stabilizer protocol that consumes resource states ρ to produce m copies of target state σ, succeeding probabilistically. Any such protocol requires at least $\mathbb{E}[n] \geq m\frac{\mathscr{M}(\sigma)}{\mathscr{M}(\rho)}$ copies of ρ on average.*

Proof. Suppose $\Lambda\left(\rho^{\otimes k}\right) = \sigma^{\otimes m}$ with probability p. The fact that the mana is an additive magic monotone implies:

$$k\mathscr{M}(\rho) \geq pm\mathscr{M}(\sigma) \implies \frac{k}{p} \geq m\frac{\mathscr{M}(\rho)}{\mathscr{M}(\sigma)}$$

Letting l be the number of times we must run the protocol to get a success we have $n = kl$ and,

$$\mathbb{E}[l] = \frac{1}{p},$$

from which it follows that $\mathbb{E}[n] = \frac{k}{p} \geq m\frac{\mathscr{M}(\rho)}{\mathscr{M}(\sigma)}$. □

We can only bound the average number of copies required because the monotone is only non-increasing on average under stabilizer operations.

Most currently known magic state distillation protocols are built around 'primitive' distillation protocols that consume k input states to produce a single output: $\Lambda_{\text{primitive}}\left(\rho_0^{\otimes k}\right) \to \rho_1$. If ρ_1 is not adequately pure then the protocol $\Lambda_{\text{primitive}}$ is repeated k times to produce $\rho_1^{\otimes k}$ and this is consumed by the primitive protocol to produce an output state with higher purity, $\Lambda_{\text{primitive}}\left(\rho_1^{\otimes k}\right) \to \rho_2$. The primitive distillation protocol is concatenated in this fashion until a final output with the required purity is reached. The bound we have derived here can be applied to the analysis of either the primitive protocol or the effective protocol that maps $\rho_1^{\otimes k^l} \to \rho_l$, where l is the number of concatenations of the primitive protocol.

Indeed, this bound covers a broader set of protocols than it might first appear. One might have expected to do better by "recycling" the output states of the failed protocols. For instance, if

$$\Lambda\left(\rho^{\otimes k}\right) = \begin{cases} \sigma^{\otimes m} & \text{with probability } p \\ \tau & \text{with probability } 1-p \end{cases},$$

then one expects to reduce the overhead of the total number of copies ρ required by introducing a second stabilizer protocol to be invoked whenever the first protocol produces τ:

$$\mathcal{E}\left(\tau \otimes \rho^{\otimes k'}\right) = \sigma^{\otimes m} \text{ with probability } q.$$

However, by just combining the two steps we have a new protocol $\tilde{\Lambda}\left(\rho^{\otimes(k+k')}\right) = \sigma^{\otimes m}$ succeding with probability $\tilde{p} = p + (1-p)\,q$ and our theorem applies.

Computing the mana of a quantum state is straight forward: we find the Wigner function by taking the trace of ρ with the d^2 phase space point operators and compute $\log\left(\|\rho\|_W\right)$. This means that the mana provides a simple way to numerically upper bound the efficiency of distillation protocols, fulfilling the major promise of this chapter.

7.2 UNIQUENESS OF SUM NEGATIVITY

Quantifying the magic of a state by the negativity in its Wigner representation is an intuitively appealing idea, but it is not clear that the sum of the negative elements is the best way to do this. For example, we might have instead looked at the maximally negative element of the Wigner function, $\text{maxneg}(\rho) = -\min_{\boldsymbol{u}} W_\rho(\boldsymbol{u})$. It is not immediately obvious that the sum negativity is a better way to quantify the magic of a quantum state than the maximal negativity just defined. It turns out that the maximal negativity is not a magic monotone so it is inappropriate a measure of usefulness for stabilizer computation. In fact, we will now show that *any* magic monotone that is determined solely by the values of the negative entries of the Wigner function (and in particular not by the positions in phase space of the negative entries) can be written as a function of only the sum negativity.

The reason that the maximally negative entry is not a magic monotone is that it is not invariant under composition with stabilizer states. Suppose we have some resource state ρ and we compose it with the maximally mixed state on a qudit $\mathbb{I}_d/d$, then $\text{maxneg}(\rho \otimes \mathbb{I}_d/d) = -\min_{\boldsymbol{u},\boldsymbol{v}} W_\rho(\boldsymbol{u}) \cdot W_{\mathbb{I}/d}(\boldsymbol{v}) = -\min_{\boldsymbol{u},\boldsymbol{v}} W_\rho(\boldsymbol{u}) \cdot \frac{1}{d^2} = \text{maxneg}(\rho)/d^2$ so this function can decrease under composition with stabilizer states and thus can increase under partial trace: it is a poor measure of the amount of resource in ρ. The requirement that magic monotones must be invariant under composition with arbitrary stabilizer states is an extremely strong one; it forms the backbone of our proof of the uniqueness of the sum negativity.

Theorem 23. *Assume $\mathcal{M}(\rho)$ is a function on quantum states that satisfies the following conditions: 1. $\mathcal{M}(\rho)$ is a magic monotone, 2. $\mathcal{M}(\rho)$ is determined only by the negative values of the Wigner function and 3. $\mathcal{M}(\rho)$ is invariant under arbitrary permutations of discrete phase space (that is, even under permutations that do not correspond to quantum transformations). Then $\mathcal{M}(\rho)$ may be written as a function of only $\text{sn}(\rho)$.*

Proof. Consider two quantum states ρ and ρ' that have Wigner representations with different negative entries but $\text{sn}(\rho) = \text{sn}(\rho')$. The idea is to construct stabilizer ancilla states A and A' such that $\rho \otimes A$ and $\rho' \otimes A'$ have the same negative Wigner function entries. In this case conditions 2 and 3 imply $\mathcal{M}(\rho \otimes A) = \mathcal{M}(\rho' \otimes A')$ and since magic monotones are invariant under composition with stabilizer states this means $\mathcal{M}(\rho) = \mathcal{M}(\rho')$, i.e., $\mathcal{M}(\rho)$ is entirely determined by the sum negativity. For details see 7.7.2. □

For our proof of Theorem 23 to succeed it is critical that the value of the monotone does not depend on the locations of the negative entries. All magic monotones must be invariant under Clifford unitaries, $\mathcal{M}\left(U\rho U^\dagger\right) = \mathcal{M}\left(\rho\right)\ \forall U \in \mathcal{C}_{p^n}$, and these operations correspond to permutations of the phase space. Thus the monotone condition already implies invariance under a subset of possible permutations (namely those that preserve the symplectic inner product). However, we require invariance under arbitrary permutations and there is no compelling reason to expect magic monotones to have this feature in general. It is not clear whether this additional assumption was really necessary; it was introduced because actually working with only the symplectic transformations is extremely challenging. It remains an interesting open problem to either prove uniqueness without this assumption or give a counterexample in the form of a magic monotone that is determined by the negative entries of the Wigner representation and does depend on their position. Even if such a monotone is found Theorem 23 is useful because it at least shows that sum negativity is the unique "simple" monotone, in the sense that its value does not depend on the detailed symplectic structure of phase space. Simplicity of computation is one of our primary motivations for the study of Wigner function monotones so this is a significant advantage.

In Chapter 6 we showed that (the regularization of) any monotone satisfying a certain natural asymptotic condition uniquely specifies the rate at which asymptotic interconversion of resource states is possible. Since the mana is additive it is clearly equal to its own regularization. Thus if it satisfied the condition given by (6.3.1) we would be able to compute the conversion rates explicitly. Typically it is usually a stronger property that is demanded: asymptotic continuity of the monotone. In 7.7.3 we show that the mana is not asymptotically continuous. However, our counter example leaves open the possibility that the weaker condition actually required by the theorem holds. It would be very exciting to either prove or disprove this.

7.3 NUMERICAL ANALYSIS OF MAGIC STATE DISTILLATION PROTOCOLS

To illustrate the use of mana in the evaluation of magic state distillation protocols we have computed the input and output mana of single steps of several (qudit) magic distillation protocols from the literature over a large parameter range. Figures 7.3.1 and 7.3.2 present qutrit codes from [2] and [11] respectively. Figure 7.3.3 presents a ququint ($d = 5$) code from [11]. Notice that none of the protocols come close to meeting the mana bound, which is illustrated as a red line in all three figures.

7.4 THE QUTRIT CASE

To build some intuition we compute the qutrit states with maximal sum negativity. Since,

$$\begin{aligned} \mathrm{sn}\left(\rho\right) &= -\sum_{u:\mathrm{Tr}\left(\rho A_u\right)<0} \mathrm{Tr}\left(\rho A_u\right) \\ &= -\mathrm{Tr}\left(\rho \sum_{u:\mathrm{Tr}\left(\rho A_u\right)<0} A_u\right), \end{aligned}$$

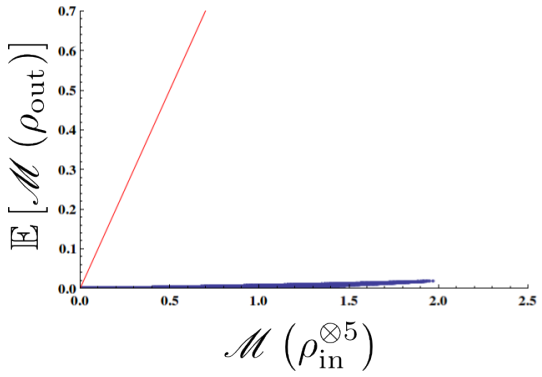


Figure 7.3.1: Efficiency of the $[5,1,3]$ qutrit code of [2]. We generate 50000 inputs of the form $\rho_{\text{in}} = (1-p_1-p_2)\,|H_+\rangle\langle H_+| + p_1|H_-\rangle\langle H_-| + p_2|H_i\rangle\langle H_i|$, where this is the form ρ_{in} takes after the twirling step of the protocol. The mana of the 5 input states is computed and plotted against the effective mana output following one round of the protocol, $\mathbb{E}\left[\mathscr{M}\left(\rho_{\text{out}}\right)\right] = \Pr\left(\text{protocol succeeds}\right)\cdot\mathscr{M}\left(\rho_{\text{out}}\right)$. We used $p_1 \in_R [0, 0.4]$ and $p_2 \in_R [0, 0.3]$ and the twirling basis states are the eigenstates of the qutrit Hadamard operator[2], with eigenvalues $\{1, -1, \imath\}$..

it is easy to see that the states with maximal sum negativity must be eigenstates of operators $\sum_{u\in S} A_u$ where S is some subset of the discrete phase space. An exhaustive search over such subsets reveals two classes of maximally sum negative states.

1. The *Strange* states with 1 negative Wigner function entry equal to $-1/3$. There are $\binom{9}{1} = 9$ such states, eg.
$$|\mathbb{S}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\-1\end{pmatrix}$$
2. The *Norrell* states with 2 negative Wigner function entries equal to $-1/6$. There are $\binom{9}{2} = 36$ such states, eg.
$$|\mathbb{N}\rangle = \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\2\\-1\end{pmatrix}.$$

The maximum value is $\text{sn}\left(|\mathbb{S}\rangle\langle\mathbb{S}|\right) = \text{sn}\left(|\mathbb{N}\rangle\langle\mathbb{N}|\right) = -1/3$. An example of both types of states are plotted in Figure 7.1.1 on page 57.

Geometrically each Strange state lies outside a single face of the Wigner simplex and each Norrell state lies outside the intersection of two faces, analogous to the qubit T-type (outside a face) and H-type (outside an edge) states. This intuition is further strengthened since the Norrell states are also the generalized H-type states of [42] and [2].

Note that the states with maximal resource value do not need to agree between monotones. In particular,
$$\frac{\text{r}_{\mathcal{M}}\left(|\mathbb{S}\rangle\langle\mathbb{S}|\right)}{\text{r}_{\mathcal{M}}\left(|\mathbb{N}\rangle\langle\mathbb{N}|\right)} \approx 1.71.$$

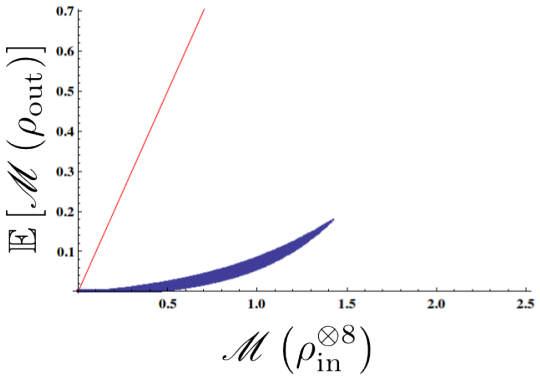


Figure 7.3.2: Efficiency of the $[8,1,3]$ qutrit code of [11].We generate 50000 inputs of the form $\rho_{\text{in}} = (1-p_1-p_2)\,|M_0\rangle\langle M_0| + p_1|M_1\rangle\langle M_1| + p_2|M_2\rangle\langle M_2|$, where this is the form ρ_{in} takes after the twirling step of the protocol. The mana of the 8 input states is computed and plotted against the effective mana output following one round of the protocol, $\mathbb{E}\left[\mathscr{M}\left(\rho_{\text{out}}\right)\right] = \Pr\left(\text{protocol succeeds}\right)\cdot\mathscr{M}\left(\rho_{\text{out}}\right)$. We used $p_1 \in_R [0,0.3]$, $p_2 \in_R [0,0.3]$ and the twirling basis states are $|M_0\rangle = \frac{1}{\sqrt{3}}\left(e^{\frac{4}{9}\pi i}|0\rangle + e^{\frac{2}{9}\pi i}|1\rangle + |2\rangle\right)$, $|M_1\rangle = \frac{1}{\sqrt{3}}\left(e^{\frac{16}{9}\pi i}|0\rangle + e^{\frac{8}{9}\pi i}|1\rangle + |2\rangle\right)$, $|M_2\rangle = \frac{1}{\sqrt{3}}\left(e^{\frac{10}{9}\pi i}|0\rangle + e^{\frac{14}{9}\pi i}|1\rangle + |2\rangle\right)$.

Of course this still leaves open the possibility that $\mathrm{r}^{\infty}_{\mathcal{M}}\left(|\mathbb{S}\rangle\langle\mathbb{S}|\right) = \mathrm{r}^{\infty}_{\mathcal{M}}\left(|\mathbb{N}\rangle\langle\mathbb{N}|\right)$.

7.5 How well motivated is the mana?

Our main motivation for studying the mana is that it can be computed explicitly to give concrete bounds on the rate at which magic states can be converted. However, one might suspect that this bound, although non-trivial, is rather arbitrary. For example, it is not clear a priori if the bound given by Theorem 22 can ever be saturated, or under what circumstances this might occur. The mana arose very naturally from the discrete Wigner function, but it is not immediately clear that the discrete Wigner representation is an appropriate tool for the study of magic theory. In fact, a number of recent results show that the use of discrete Wigner representation is extremely well motivated in this context.

It is natural to wonder if we could have started with some other notion of the discrete Wigner function and defined a monotone from that. Recent work[43] has shown this is not the case. Suppose we have a subtheory of quantum theory consisting of the stabilizer measurements and some subset of the quantum states, then if there exists a non-contextual hidden variable theory capable of reproducing the measurement statistics of this subtheory then every state in the subtheory has positive Wigner representation. That is, the subtheory with positive Wigner representation is the largest possible subtheory of quantum theory that includes the stabilizer measurements and admits a non-contextual hidden variable model. In particular this means that any other choice of discrete Wigner function would have a positively represented region that is strictly contained within the discrete Wigner function we use here.

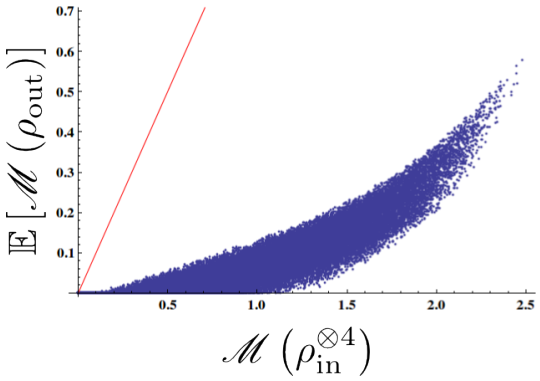


Figure 7.3.3: Efficiency of the $[4,1,2]$ ququint code of [11].We generate 50000 inputs of the form $\rho_{\text{in}} = (1-p_1-p_2-p_3-p_4)\,|M_0\rangle\langle M_0| + \sum_{i=1}^{4} p_i |M_i\rangle\langle M_i|$, where this is the form ρ_{in} takes after the twirling step of the protocol. The mana of the 4 input states is computed and plotted against the effective mana output following one round of the protocol, $\mathbb{E}\left[\mathscr{M}\left(\rho_{\text{out}}\right)\right] = \Pr\left(\text{protocol succeeds}\right)\cdot\mathscr{M}\left(\rho_{\text{out}}\right)$. We used $p_i \in_R [0, 0.2]$ and the twirling basis states are the eigenstates of the CM ququint operator defined at [11].

The subtheory of quantum theory consisting of elements with positive discrete Wigner representation is the maximal classical subtheory in the sense of non-classicality given by contextuality. For the purposes of magic state distillation we are more interested in the notion of non-classicality given by universal quantum computation. The results of the first part of this thesis (also [71, 53]) show that there is an intimate connection: the hidden variable model afforded by the discrete Wigner function leads naturally to an efficient classical simulation scheme for quantum circuits with positive Wigner representation. It is not known if access to any negatively represented state suffices to promote stabilizer computation to universal quantum computation, but it is at least apparent that the known classical simulation protocols can not be extended to deal with this case. In the context of magic state computation it is desirable for the magic measures to give an indication of how useful a state is for quantum computation. In this sense the fact that the mana is not faithful can be considered a feature rather than a bug.

Although the mana is essentially the unique monotone arising from the *negativity* of the Wigner function it is not the only choice of monotone that reflects the *geometry* of the Wigner function. In particular, one very natural choice is the relative entropy distance to the set of states with positive Wigner representation, $\mathrm{r}_{\mathcal{W}}(\rho) = \min_{\sigma: W_\sigma(\boldsymbol{u})\geq 0 \forall \boldsymbol{u}} S(\rho\|\sigma)$. It is easy to check that all of the results of Chapter 6 go through for this new monotone, subject to obvious modifications in the statement of the theorems.

7.6 DISCUSSION

The major inspiration for the monotones of this chapter was earlier work showing that states with positive Wigner representation can not be distilled by stabilizer protocols. In the theory of entanglement it is known that states with positive partial transpose (ppt) can not be distilled by LOCC protocols[38], and this inspired the introduction of the entanglement

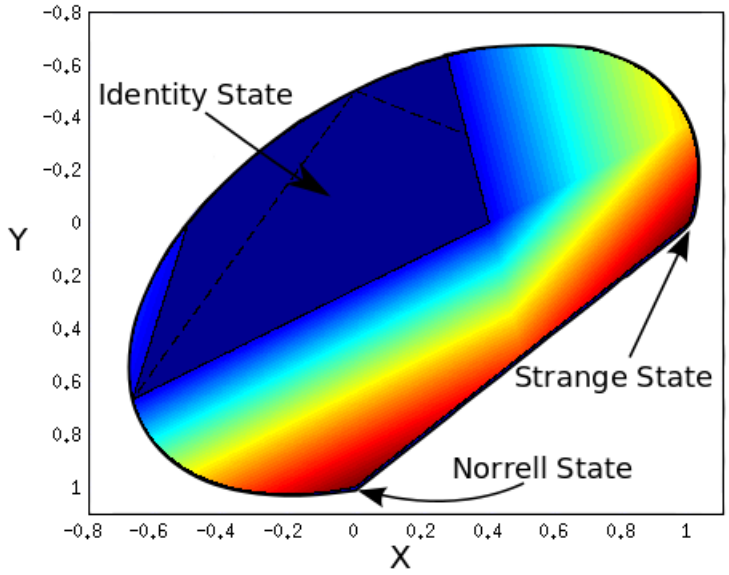


Figure 7.4.1: The plane $(1-x-y)\frac{\mathbb{I}}{3}+x|\mathrm{S}\rangle\langle\mathrm{S}|+y|\mathbb{N}\rangle\langle\mathbb{N}|$. The heat map shows the value of the mana. The dark blue (0 mana) region is the simplex of states with positive Winger representation. The stabilizer polytope is delineated by a dashed line.

negativity $\mathcal{N}(\rho)$, an measure of the violation of the ppt, as a measure of entanglement[74]. As with the sum negativity the major advantage of this measure is that it is computable, allowing for explicit upper bounds on the efficiency of entanglement distillation. The entanglement negativity grows exponentially in the number of resource states, prompting the definition of an additive variant $\mathcal{LN}(\rho) \equiv \log(2\mathcal{N}(\rho)+1)$ - exactly as in the present case. Like the mana this measure has the strange feature that is neither convex nor asymptotically continuous[1]. The close analogy we have uncovered suggests that it may be possible to adapt much of the work on entanglement negativity to the magic case: this is an interesting direction for future work.

There is at least one way in which the sum negativity is better behaved than the entanglement negativity. All separable states are local, but it is not known whether all entangled states are non-local in the sense that they enable violation of a Bell type inequality. In [57] Peres conjectured that any ppt state should admit a local hidden variable model; proving or disproving this conjecture is one of the major outstanding problems in the study of entanglement. In our case the equivalent conjecture would be that any state with positive Wigner representation admits a non-contextual hidden variable model. This is obvious: the Wigner itself is this non-contextual hidden variable theory! Moreover, recent work[43] has shown that (at least for small prime dimension) magic states admit such a model *only* if they have positive Wigner representation. The easy resolution of this potentially difficult problem is a consequence of our use of the Wigner function quasi-probability technology. The quasi-probability techniques used in this chapter have no known analogue in other

1 In fact it is now known that these two features are closely related[59].

resource theories. Exporting this technology to the study of other resource theories, in particular entanglement theory, is a fascinating and important problem.

A closely related problem is to determine a qubit analogue for the mana. Because it is possible to violate a contextuality inequality (eg. a GHZ inequality[33]) using qubit stabilizers there can be no qubit analogue for the discrete Wigner function (see also [76]). This is because the discrete Wigner function is a non-contextual hidden variable theory. Nevertheless, it may be possible to find a computable monotone of a similar flavour.

7.7 PROOFS

This section presents the details of the proofs that were omitted from the main text of the chapter in order to improve readability.

7.7.1 *Wigner function 1 norm is a magic monotone.*

The main ingredient in establishing both $\mathrm{sn}(\rho)$ and $\mathscr{M}(\rho)$ as magic monotones is to show that $\|\rho\|_W = \sum_{\boldsymbol{u}} \left|W_\rho(\boldsymbol{u})\right|$ is a magic monotone.

Theorem. $\|\rho\|_W = \sum_{\boldsymbol{u}} \left|W_\rho(\boldsymbol{u})\right|$ *is a magic monotone.*

Proof. We need to verify that this function is non-increasing under stabilizer operations:

1. Invariance under Clifford unitaries: the action of Clifford unitaries on the phase space of the Wigner function is a permutation, $\boldsymbol{u} \to F\boldsymbol{u}$. Thus, $\|U\rho U^\dagger\|_W = \sum_{\boldsymbol{u}} \left|W_{U\rho U^\dagger}(\boldsymbol{u})\right| = \sum_{\boldsymbol{u}} \left|W_\rho(F\boldsymbol{u})\right| = \sum_{\boldsymbol{u}} \left|W_\rho(\boldsymbol{u})\right| = \|\rho\|_W$.

2. Non-increasing on average under stabilizer measurement: we consider computational basis measurement on the final qudit. The expected value of $\|\tilde{\rho}\|_W$ for the post measurement state $\tilde{\rho}$ is:

$$\begin{aligned}\mathbb{E}\left[\|\tilde{\rho}\|_W\right] &= \sum_i \mathrm{Tr}\left(\rho\mathbb{I}\otimes|i\rangle\langle i|\right)\|\mathbb{I}\otimes|i\rangle\langle i|\rho\mathbb{I}\otimes|i\rangle\langle i|/\mathrm{Tr}\left(\rho\mathbb{I}\otimes|i\rangle\langle i|\right)\|_W \\ &= \sum_i \|\mathbb{I}\otimes|i\rangle\langle i|\rho\mathbb{I}\otimes|i\rangle\langle i|\|_W,\end{aligned}$$

   and by writing $\mathbb{I}\otimes|i\rangle\langle i|\rho\mathbb{I}\otimes|i\rangle\langle i|$ as:

$$\begin{aligned}\mathbb{I}\otimes|i\rangle\langle i|\rho\mathbb{I}\otimes|i\rangle\langle i| &= \sum_{\boldsymbol{u},\boldsymbol{v}} W_\rho(\boldsymbol{u},\boldsymbol{v})\langle i|A_{\boldsymbol{v}}|i\rangle\cdot A_{\boldsymbol{u}}\otimes|i\rangle\langle i| \\ &= \sum_{\boldsymbol{u}}\left(\sum_{\boldsymbol{v}} W_\rho(\boldsymbol{u},\boldsymbol{v})\langle i|A_{\boldsymbol{v}}|i\rangle\right)A_{\boldsymbol{u}}\otimes\sum_{\boldsymbol{w}}\left(\frac{1}{d}\langle i|A_{\boldsymbol{w}}|i\rangle\right)A_{\boldsymbol{w}}\end{aligned}$$

we find,

$$\begin{aligned}
\mathbb{E}\left[\|\tilde{\rho}\|_W\right] &= \sum_i \sum_{\boldsymbol{u},\boldsymbol{w}} \left| \left(\sum_{\boldsymbol{v}} W_\rho\left(\boldsymbol{u},\boldsymbol{v}\right) \langle i| A_{\boldsymbol{v}} |i\rangle \right) \left(\frac{1}{d} \langle i| A_{\boldsymbol{w}} |i\rangle \right) \right| \\
&= \sum_i \sum_{\boldsymbol{u}} \left(\sum_{\boldsymbol{w}} \frac{1}{d} \langle i| A_{\boldsymbol{w}} |i\rangle \right) \left| \left(\sum_{\boldsymbol{v}} W_\rho\left(\boldsymbol{u},\boldsymbol{v}\right) \langle i| A_{\boldsymbol{v}} |i\rangle \right) \right| \quad (\because \langle i| A_{\boldsymbol{w}} |i\rangle \geq 0) \\
&\leq \sum_i \sum_{\boldsymbol{u}} \sum_{\boldsymbol{v}} \left| W_\rho\left(\boldsymbol{u},\boldsymbol{v}\right) \langle i| A_{\boldsymbol{v}} |i\rangle \right| \quad (\because \text{triangle inequality and } \sum_{\boldsymbol{w}} \frac{1}{d} \langle i| A_{\boldsymbol{w}} |i\rangle = 1) \\
&= \sum_{\boldsymbol{u},\boldsymbol{v}} \left(\sum_i \langle i| A_{\boldsymbol{v}} |i\rangle \right) \left| W_\rho\left(\boldsymbol{u},\boldsymbol{v}\right) \right| \quad (\because \langle i| A_{\boldsymbol{w}} |i\rangle \geq 0) \\
&= \|\rho\|_W \quad (\because \sum_i \langle i| A_{\boldsymbol{v}} |i\rangle = 1).
\end{aligned}$$

3. Invariance under composition with stabilizer states: let σ be any state with positive Wigner representation. Then,

$$\begin{aligned}
\|\rho \otimes \sigma\|_W &= \|\rho\|_W \|\sigma\|_W \\
&= \|\rho\|_W
\end{aligned}$$

   Since $\|\sigma\|_W = \sum_{\boldsymbol{u}} |W_\sigma(\boldsymbol{u})| = \sum_{\boldsymbol{u}} W_\sigma(\boldsymbol{u}) = 1$ for positively represented states. All stabilizer states are positively represented so they are included as a special case.

4. Non-increasing under partial trace: we consider tracing out the final qudit of the system. If $\rho = \sum_{\boldsymbol{u},\boldsymbol{v}} W_\rho\left(\boldsymbol{u} \oplus \boldsymbol{v}\right) A_{\boldsymbol{u}} \otimes A_{\boldsymbol{v}}$ then $\mathrm{Tr}_B\left(\rho\right) = \sum_{\boldsymbol{u}} \left(\sum_{\boldsymbol{v}} W_\rho\left(\boldsymbol{u} \oplus \boldsymbol{v}\right) \right) A_{\boldsymbol{u}}$ so,

$$\begin{aligned}
\|\mathrm{Tr}_B\left(\rho\right)\|_W &= \sum_{\boldsymbol{u}} \left| \sum_{\boldsymbol{v}} W_\rho\left(\boldsymbol{u},\boldsymbol{v}\right) \right| \\
&\leq \|\rho\|_W,
\end{aligned}$$

   by the triangle inequality.

5. Convexity: this is not strictly required, but it's a nice feature so we note it here.

$$\begin{aligned}
\|p\rho + (1-p)\,\sigma\|_W &= \sum_{\boldsymbol{u}} \left| pW_\rho\left(\boldsymbol{u}\right) + (1-p)\, W_\sigma\left(\boldsymbol{u}\right) \right| \\
&\leq p\|\rho\|_W + (1-p)\,\|\sigma\|_W,
\end{aligned}$$

   by the triangle inequality.

□

We next establish that this was essentially the only choice we could have made to (simply) quantify the magic of a quantum state via its Wigner representation,

7.7.2 *Sum negativity is the unique phase space measure of magic.*

Theorem. *Assume $\mathcal{M}(\rho)$ is a function on quantum states that satisfies the following conditions: 1. $\mathcal{M}(\rho)$ is a magic monotone, 2. $\mathcal{M}(\rho)$ is determined only by the negative values of the Wigner function and 3. $\mathcal{M}(\rho)$ is invariant under arbitrary permutations of discrete phase space (that is, even under permutations that do not correspond to quantum transformations). Then $\mathcal{M}(\rho)$ may be written as a function of only sn (ρ).*

Proof. Let ρ have negative entries $-N_1, -N_2, \ldots, -N_k$ and ρ' have negative entries $-N_1', -N_2', \ldots, -N_{k'}'$, with

$$N \equiv \text{sn}(\rho) = \sum N_i = \sum N_i' = \text{sn}(\rho').$$

A and A' will be ancilla states acting on m qudits, with $m = \max\{\lceil \log_d k \rceil, \lceil \log_d k' \rceil\}$; d is the size of each qudit.

$$\begin{aligned} A &= \sum_{i=1}^{k'} (N_i'/N)|i\rangle\langle i| \\ A' &= \sum_{i=1}^{k} (N_i/N)|i\rangle\langle i|. \end{aligned}$$

These are valid states since the sum of the N_i and N_i' is the same. The Wigner function of A consists of columns labeled by i with entries N_i'/rN, with $r = d^m$; each column contains r such elements. It also has $d^m - k'$ columns filled with zeros. Similarly for A', except it has $d^m - k$ zero columns and the non-zero columns have r copies of N_i/rN instead.

The negative Wigner function entries for the state $\rho \otimes A$ are of the form $-N_i N_j'/rN$, for all i and j. Each of these appears r times. The negative Wigner function entries for $\rho' \otimes A'$ are of the form $-N_j' N_i/rN$, for all i and j. Again, each appears r times. These entries could be in different locations, but since the function we are calculating does not depend on location of negative entries, only their values, it follows that

$$\mathcal{M}(\rho) = \mathcal{M}(\rho \otimes A) = \mathcal{M}(\rho' \otimes A') = \mathcal{M}(\rho').$$

Therefore, $\mathcal{M}(\rho)$ is a function only of sn (ρ). □

7.7.3 *Continuity and Asymptotic Continuity*

In practice a perfect conversion is generally not possible, $\|\Lambda(\rho^{\otimes m}) - \sigma^{\otimes n}\|_1 > 0$ for even the best choice of stabilizer protocol Λ. A state $\tilde{\sigma}_n$ that is close enough to $\sigma^{\otimes n}$ can be used in place of $\sigma^{\otimes n}$ in information theoretic tasks so a better notion of conversion would be: how many copies of ρ are required to produce a state $\Lambda(\rho^{\otimes m}) = \tilde{\sigma}_n$ that is "close enough" to $\sigma^{\otimes n}$. A natural notion of closeness is $\|\tilde{\sigma}_n - \sigma^{\otimes n}\|_1 < \epsilon$ for some operationally relevant ϵ. It is conceivable that there is some choice of $\tilde{\sigma}_n$ in the epsilon ball around $\sigma^{\otimes n}$ such that $\mathscr{M}(\tilde{\sigma}_n) \ll \mathscr{M}(\sigma^{\otimes n})$, in which case $\mathscr{M}(\sigma)$ would have little operational significance. Happily, it is not

difficult to show that $\mathscr{M}(\rho)$ is continuous with respect to the 1-norm in the sense that for a sequence of states $\rho_k, \sigma_k \in \mathcal{S}(\mathcal{H}_d)$ $\{\|\rho_k - \sigma_k\|\}_k \to 0 \implies \{|\mathscr{M}(\rho_k) - \mathscr{M}(\sigma_k)|\}_k \to 0$ so for a target state *on fixed dimension* there is some well defined sense in which closeness in the 1-norm implies that the mana of two states is close.

In the case of asymptotic conversion of states this notion needs some massaging. Formally, let $\Lambda_n : \mathcal{S}(\mathcal{H}_{d^{m(n)}}) \to \mathcal{S}(\mathcal{H}_{d^n})$ be stabilizer protocols satisfying

$$\lim_{n\to\infty} \|\Lambda_n\left(\rho^{\otimes m(n)}\right) - \sigma^{\otimes n}\| \quad \to \quad 0$$

In particular we would like to avoid a situation where $\lim_{n\to\infty} \mathscr{M}\left(\Lambda_n\left(\rho^{\otimes m(n)}\right)\right) \ll \mathscr{M}(\sigma^{\otimes n})$, which would mean that states that are nearly equivalent (in the one norm) could have radically different mana values. One way that this requirement can be formalized is the property of asymptotic continuity. A function is said to be asymptotically continuous if for sequences ρ_n, σ_n on $\mathcal{H}_n$, $\lim_{n\to\infty} \|\rho_n - \sigma_n\| \to 0$ implies:

$$\lim_{n\to\infty} \frac{f(\rho_n) - f(\sigma_n)}{1 + \log(\dim \mathcal{H}_n)} \to 0.$$

This notion is the commonly accepted generalization of continuity to the asymptotic regime and is of particular importance because if the mana could be shown to be asymptotically continuous it would be the unique measure of magic in the sense of Theorem 16. Unhappily, it is very difficult to show this. This is mostly because it is false,

Theorem. *$\mathscr{M}(\sigma)$ is not asymptotically continuous.*

Proof. Define $\tilde{\sigma}_n = (1-\delta_n)\,\sigma^{\otimes n} + \delta_n \eta_n$. Asymptotic continuity would imply

$$\lim_{n\to\infty} \delta_n \to 0 \implies \lim_{n\to\infty} \frac{\mathscr{M}(\tilde{\sigma}_n) - \mathscr{M}(\sigma^{\otimes n})}{n} \to 0,$$

we will show this need not be the case. Suppose σ is negative on points $\mathcal{N} = \{\boldsymbol{u} : \ W_\sigma(\boldsymbol{u}) < 0\}$. Let η be the state with maximal sum negativity satisfying $W_\eta(\boldsymbol{u}) < 0 \iff \boldsymbol{u} \in \mathcal{N}$ (i.e. η is negative on the same points as σ). Then,

$$\begin{aligned}
\|\tilde{\sigma}\|_W &= \sum_{\boldsymbol{u}} |\,(1-\delta_n)\,W_{\sigma^{\otimes n}}(\boldsymbol{u}) + \delta_n W_{\eta^{\otimes n}}(\boldsymbol{u})| \\
&= \sum_{\boldsymbol{u}} \left((1-\delta_n)\,|W_{\sigma^{\otimes n}}(\boldsymbol{u})| + \delta_n |W_{\eta^{\otimes n}}(\boldsymbol{u})|\right) \\
&= (1-\delta_n)\,\|\sigma^{\otimes n}\|_W + \delta_n \|\eta^{\otimes n}\|_W \\
&= (1-\delta_n)\,\|\sigma\|_W^n + \delta_n \|\eta\|_W^n.
\end{aligned}$$

Here we have exploited that the sign of $W_{\eta^{\otimes n}}(\boldsymbol{u})$ and the sign of $W_{\sigma^{\otimes n}}(\boldsymbol{u})$ are always the same. Subbing this in,

$$\frac{\mathscr{M}(\tilde{\sigma}_n) - \mathscr{M}(\sigma^{\otimes n})}{n} = \frac{1}{n}\log\left((1-\delta_n) + \delta_n\left(\frac{\|\eta\|_W}{\|\sigma\|_W}\right)^n\right),$$

but by assumption $\|\eta\|_W > \|\sigma\|_W$ unless $\|\sigma\|_W$ is maximal for all states that are negative on $\mathcal{N}$, so the limit need not go to 0. Thus asymptotic continuity can not hold generally. □

This result is not actually terribly surprising. Suppose we have a preparation apparatus that always prepares $\sigma^{\otimes n}$. Now further suppose that we rebuild our apparatus so that with probability δ_n it will instead produce $\eta^{\otimes n}$ *with a far greater amount of negativity*. Then it is intuitively obvious that we should be able to extract more negativity from the new apparatus just by sacrificing a few copies of the output state to determine whether we have produced σ or η. Of course as n goes to infinity this will only work if δ_n goes to zero slowly enough, but this argument does clarify the physical irrelevance of asymptotic continuity.

Essentially, asymptotic continuity fails because it is possible that access to a very large amount of resource, even with small probability, can dramatically improve our preparation procedure. Notice that the opposite is not (obviously) true: if our machine fails with a very small probability this does not make it useless. Indeed, if we had a promise of the form $\tilde{\sigma}_n = (1-\delta)\,\sigma^{\otimes n} + \delta\eta^{\otimes n}$ then we could just sacrifice some small number of registers to check that that the output state was in fact $\sigma^{\otimes n}$.

Part III

CONCLUDING THOUGHTS

In this thesis we have introduced the resource theory of magic, showing how the tools of resource theories can be applied to study the extra resources required to promote stabilizer computation to universal quantum computation. Further, we have established negativity of the discrete Wigner function as a resource for quantum computation in this paradigm by showing both that negativity is necessary for quantum computational speedup and that it gives a quantitative measure of how useful a state is for promoting stabilizer computation to universal quantum computation. We have thus demonstrated a precise relationship between the traditional, qualitative notion of "quantumness" given by negative quasi-probability representation and the modern, concrete notion given by computational speedup.

We have given an explicit simulation protocol for quantum circuits using stabilizer operations acting on states with positive discrete Wigner representation; this shows that for systems of power of odd prime dimension a necessary condition for computational speedup is negativity of the discrete Wigner representation of the inputs. This result is immediately relevant in the context of magic state distillation, where it shows that a necessary condition for distillability is negative discrete Wigner representation of the ancilla preparation. We have also shown that the phase space point operators defining the discrete Wigner function correspond to a privileged set of facets of the stabilizer polytope. Together the two results imply the existence of non-stabilizer resources that do not promote Clifford computation to universal quantum computation; and in particular this establishes the existence of bound states for magic state distillation, or bound magic states.

By casting magic state computation as a resource theory we have been able to quantify how useful a quantum state is for promoting stabilizer computation to universal quantum computation. The key step here is the introduction of magic monotones: functions assigning quantum states to real numbers that are non-increasing under stabilizer operations. We have discovered two very interesting magic measures of this type.

The relative entropy of magic, and it's asymptotic variant, are useful tools for the holistic study of magic theory. In particular, we saw that (even asymptotically) a non-zero amount of magic is required to create any magic state establishing, in conjunction with the existence of bound magic states, that generally the amount of pure magic states that can be extracted from a mixed magic state is not equal to the amount pure magic states required to create it: the magic of creation does not equal the magic of distillation. The main motivation for studying the relative entropy of magic was that its asymptotic regularization is in some sense a unique measure of magic. However, as we have seen, this is not a special feature of the relative entropy of magic but a (potentially) common feature among magic monotones. Indeed, the relative entropy of magic has some serious drawbacks. Foremost among these are the lack of a closed form expression and that it is subadditive, even for pure magic states. The combination of these two irritants implies that computing the relative entropy of magic generally requires a numerical search that is computationally infeasible.

To address this short coming we introduced the mana, a computable monotone. We have shown this monotone has the appealing feature that it is additive, $\mathscr{M}(\rho \otimes \sigma) = \mathscr{M}(\rho) + \mathscr{M}(\sigma)$. As a consequence we may give explicit lower bounds on the number of resource states ρ required to produce m copies of a resource state σ. This is an explicit, absolute upper bound on the efficiency of magic state distillation protocols. This monotone is in some sense

the unique measure of magic arising from the negativity of the discrete Wigner function. Since the discrete Wigner function itself is essentially the unique classical representation for the stabilizer formalism[43] there is some reason to believe that the mana has some privileged status among all possible monotones. Determining if and how this intuition can be formalized is a very important open problem.

The resource theory of quantum (magic state) computation that we have developed here is closely analogout to the resource theory of quantum communication, i.e., entanglement theory. It is known that there are slightly entangled mixed states that can not be consumed by distillation routines to produce highly entangled states[38]. The non-stabilizer but positively represented quantum states are exactly analogous to these bound entangled states. Similarly, it is known that for pure states large amounts of entanglement are required for quantum computational speedup[75], but for mixed states this is still an open question. However, for negative discrete Wigner representation there is no relevant distinction between mixed states and pure states.

There are a number of directions for future work, many of which have already been discussed in the main body of the text. Other resource theories admit a wealth of monotones. This is especially true in the theory of entanglement where a large number of entanglement measures have been developed to solve specialized problems. One obvious direction for future work is the creation of additional magic monotones to address particular problems in magic resource theory. It is also important to develop the parts of the resource theory that are not encapsulated by magic monotones. For example, analogues of entanglement catalysis and activation are discussed in [10]. The most urgent outstanding problem of this type is to find a criterion for determining if it is possible to (asymptotically) reversibly convert between particular resource states using stabilizer operations. Concretely, it is always possible to use LOCC to reversibly convert pure bipartite entangled states but this is not true for tripartite entanglement; we would like to know which situation holds for magic theory. Even a partial result of this type would be very powerful, offering deep insight into the structure of stabilizer protocols.

Much of this work has been showing that much of the technology from other resource theories can be imported to the resource theory of magic. It is very interesting to ask if we can go in the other direction and export the insights of magic theory to the study of generic resource theories and quantum theory broadly. One obvious extension of this type is to the setting of linear optics, which is the infinite dimensional analogue of the stabilizer formalism. Some progress on this front has already been made in [46]. This paper examined the volume of the negative region of the infinite dimensional Wigner function as a measure of non-classicality but missed the resource theoretic connection. Furthermore, the simulation result of Chapter 4 has already been extended to the linear optical regime[73, 53].

The most interesting outstanding question raised by this work is whether the ability to prepare any state with negative discrete Wigner representation is sufficient to promote Clifford computation to universal quantum computation. In prime dimension the discrete Wigner construction is the *unique* choice of quasi-probability representation covariant under the action of Clifford operations [34]. On this basis we conjecture that the condition here is sufficient. From the work of Campbell, Anwar and Browne [11] it is already known that

access to any non-stabilizer *pure* state (or equivalently any negatively represented pure state) suffices. If this conjecture is true then this implies an equivlance of two previously unrelated concepts of non-classicality, namely, quantum computational speedup and negative quasi-probability representation.

BIBLIOGRAPHY

[1] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328+, November 2004. (Cited on pages 2 and 35.)

[2] Hussain Anwar, Earl T Campbell, and Dan E Browne. Qutrit magic state distillation. *New Journal of Physics*, 14(6):063006, 2012. (Cited on pages ix, 10, 61, and 62.)

[3] DM Appleby. SIC-POVMs and the Extended Clifford Group, 2004. (Cited on page 18.)

[4] Charles H. Bennett, Herbert J. Bernstein, Sandu Popescu, and Benjamin Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53:2046–2052, Apr 1996. (Cited on page 54.)

[5] Fernando G.S.L. Brandao and Martin B. Plenio. A generalization of quantum stein's lemma. *Communications in Mathematical Physics*, 295:791–828, 2010. (Cited on page 52.)

[6] Sergei Bravyi and Alexei Kitaev. Universal Quantum Computation with ideal Clifford gates and noisy ancillas. *Physical Review A*, 71(2):022316+, December 2004. (Cited on pages 2, 10, and 50.)

[7] Sergey Bravyi and Jeongwan Haah. Magic-state distillation with low overhead. *Phys. Rev. A*, 86:052329, Nov 2012. (Cited on page 3.)

[8] J. Emerson C. Ferrie. Framed hilbert space: hanging the quasi-probability pictures of quantum theory. *New J. Phys.*, 11:063040, 2009. (Cited on pages 23 and 24.)

[9] Earl Campbell and Dan Browne. On the structure of protocols for magic state distillation. In Andrew Childs and Michele Mosca, editors, *Theory of Quantum Computation, Communication, and Cryptography*, volume 5906 of *Lecture Notes in Computer Science*, pages 20–32. Springer Berlin / Heidelberg, 2009. (Cited on page 35.)

[10] Earl T. Campbell. Catalysis and activation of magic states in fault-tolerant architectures. *Phys. Rev. A*, 83:032317, Mar 2011. (Cited on pages 35 and 73.)

[11] Earl T. Campbell, Hussain Anwar, and Dan E. Browne. Magic-state distillation in all prime dimensions using quantum reed-muller codes. *Phys. Rev. X*, 2:041021, Dec 2012. (Cited on pages ix, 3, 10, 11, 61, 63, 64, and 73.)

[12] Earl T. Campbell and Dan E. Browne. Bound states for magic state distillation in fault-tolerant quantum computation. *Phys. Rev. Lett.*, 104:030503, Jan 2010. (Cited on pages 35 and 47.)

[13] Xie Chen, Hyeyoun Chung, Andrew W. Cross, Bei Zeng, and Isaac L. Chuang. Subsystem stabilizer codes cannot have a universal set of transversal gates for even one encoded qudit. *Physical Review A*, 78:012353+, July 2008. (Cited on page 2.)

[14] Cecilia Cormick, Ernesto F. Galvão, Daniel Gottesman, Juan P. Paz, and Arthur O. Pittenger. Classicality in discrete Wigner functions. *Physical Review A (Atomic, Molecular, and Optical Physics)*, 73(1):012301+, 2006. (Cited on pages 35 and 36.)

[15] Animesh Datta, Steven T. Flammia, and Carlton M. Caves. Entanglement and the power of one qubit. *Physical Review A*, 72(4):042316+, October 2005. (Cited on page 1.)

[16] Bryan Eastin. Distilling one-qubit magic states into Toffoli states. 2013. (Cited on page 3.)

[17] Bryan Eastin and Emanuel Knill. Restrictions on Transversal Encoded Quantum Gate Sets. *Physical Review Letters*, 102:110502+, March 2009. (Cited on page 2.)

[18] C. Ferrie and J. Emerson. Frame representations of quantum mechanics and the necessity of negativity in quasi-probability representations. *Journal of Physics A: Mathematical and Theoretical*, 41:352001, 2008. (Cited on pages 4, 23, and 24.)

[19] Christopher Ferrie and Joseph Emerson. Framed Hilbert space: hanging the quasi-probability pictures of quantum theory. *New Journal of Physics*, 11(6):063040+, 2009. (Cited on page 4.)

[20] Christopher Ferrie, Ryan Morris, and Joseph Emerson. Necessity of negativity in quantum theory. *Phys. Rev. A*, 82:044103, 2010. (Cited on pages 4, 20, 23, and 24.)

[21] Gerald Folland. *Harmonic Analysis in Phase Space*. Princeton University Press, 1989. (Cited on page 32.)

[22] Austin Fowler, Matteo Mariantoni, John Martinis, and Andrew Cleland. A primer on surface codes: Developing a machine language for a quantum computer, 2012. (Cited on page 3.)

[23] M Freedman, A. Kitaev, M. Larsen, and Z. Wang. Topological quantum computation. *Bull. Amer. Math. Soc.*, 40, 2003. (Cited on page 2.)

[24] Ernesto F. Galvão. Discrete Wigner functions and quantum computational speedup. *Physical Review A*, 71(4):042302+, April 2005. (Cited on page 35.)

[25] Kathleen S. Gibbons, Matthew J. Hoffman, and William K. Wootters. Discrete phase space based on finite fields. *Physical Review A*, 70(6):062101+, December 2004. (Cited on pages xi and 31.)

[26] Daniel Gottesman. *Stabilizer Codes and Quantum Error Correction*. PhD thesis, California Institute of Technology, 1997. (Cited on pages 1, 2, and 6.)

[27] Daniel Gottesman. The heisenberg representation of quantum computers. *arXiv:quant-ph/987006*, 1998. (Cited on pages 15 and 35.)

[28] Daniel Gottesman. Theory of fault-tolerant quantum computation. *Phys. Rev. A*, 57:127–137, Jan 1998. (Cited on page 14.)

[29] Daniel Gottesman. *Quantum Error Correction and Fault-Tolerance*, pages 196–201. Elsevier, July 2006. (Cited on pages 2 and 15.)

[30] Daniel Gottesman. An introduction to quantum error correction and fault-tolerant quantum computation. In *Quantum Information Science and Its Contributions to Mathematics*. 2009. (Cited on page 6.)

[31] Gilad Gour, Iman Marvian, and Robert W. Spekkens. Measuring the quality of a quantum reference frame: The relative entropy of frameness. *Phys. Rev. A*, 80:012307, Jul 2009. (Cited on pages 49, 52, and 53.)

[32] Gilad Gour and Robert W Spekkens. The resource theory of quantum reference frames: manipulations and monotones. *New Journal of Physics*, 10(3):033023, 2008. (Cited on page 49.)

[33] Daniel M. Greenberger, Michael A. Horne, Abner Shimony, and Anton Zeilinger. Bell's theorem without inequalities. *American Journal of Physics*, 58(12):1131–1143, 1990. (Cited on pages 22, 33, and 66.)

[34] D. Gross. Hudson's theorem for finite-dimensional quantum systems. *Journal of Mathematical Physics*, 47(12):122107, 2006. (Cited on pages xi, 4, 6, 12, 15, 16, 17, 18, 29, 30, 31, 32, 36, 38, 50, and 73.)

[35] D. Gross. Non-negative Wigner functions in prime dimensions. *Applied Physics B*, 86(3):367–370, February 2007. (Cited on pages 4, 29, and 31.)

[36] D. Gross, K. Audenaert, and J. Eisert. Evenly distributed unitaries: On the structure of unitary designs. *Journal of Mathematical Physics*, 48(5):052104+, 2007. (Cited on pages 6 and 9.)

[37] David Gross. *Computational power of quantum many-body states and some results on discrete phase spaces*. PhD thesis, Imperial College London, 2008. (Cited on page 33.)

[38] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Mixed-state entanglement and distillation: Is there a "bound" entanglement in nature? *Phys. Rev. Lett.*, 80:5239–5242, Jun 1998. (Cited on pages 64 and 73.)

[39] Michał Horodecki, Paweł Horodecki, and Jonathan Oppenheim. Reversible transformations from pure to mixed states and the unique measure of information. *Phys. Rev. A*, 67:062104, Jun 2003. (Cited on page 49.)

[40] Michal Horodecki and Jonathan Oppenheim. (Quantumness in the context of) Resource Theories. 2012. (Cited on pages 5, 46, 49, and 54.)

[41] Michał Horodecki, Jonathan Oppenheim, and Ryszard Horodecki. Are the laws of entanglement theory thermodynamical? *Phys. Rev. Lett.*, 89:240403, Nov 2002. (Cited on pages 49 and 53.)

[42] Mark Howard and Jiri Vala. Qudit versions of the qubit $\pi/8$ gate. *Phys. Rev. A*, 86:022316, Aug 2012. (Cited on page 62.)

[43] Mark Howard, Victor Veitch, and Joseph Emerson. Negativity,contextuality and universal quantum computation. In Preparation. (Cited on pages 11, 63, 65, and 73.)

[44] Edwin Thompson Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003. (Cited on page 20.)

[45] Cody Jones. Distillation protocols for Fourier states in quantum computing. 2013. (Cited on page 3.)

[46] Anatole Kenfack and Karol Życzkowski. Negativity of the wigner function as an indicator of non-classicality. *Journal of Optics B: Quantum and Semiclassical Optics*, 6(10):396, 2004. (Cited on page 73.)

[47] A.Yu. Kitaev. Fault-tolerant quantum computation by anyons. *Annals of Physics*, 303(1):2 – 30, 2003. (Cited on page 2.)

[48] E. Knill and R. Laflamme. Power of One Bit of Quantum Information. *Phys. Rev. Lett.*, 81(25):5672–5675, 1998. (Cited on page 1.)

[49] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Phys. Rev. A*, 77:012307, Jan 2008. (Cited on page 9.)

[50] Elliott H. Lieb and Mary Beth Ruskai. A fundamental property of quantum-mechanical entropy. *Phys. Rev. Lett.*, 30:434–436, Mar 1973. (Cited on page 56.)

[51] Easwar Magesan, J. M. Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.*, 106:180504, May 2011. (Cited on page 9.)

[52] L. Mandel. Non-Classical States of the Electromagnetic Field. *Physica Scripta*, 1986(T12):34, January 1986. (Cited on page 26.)

[53] A. Mari and J. Eisert. Positive wigner functions render classical simulation of quantum computation efficient, 2012. (Cited on pages 32, 64, and 73.)

[54] Adam M. Meier, Bryan Eastin, and Emanuel Knill. Distillation protocols for Fourier states in quantum computing. 2013. (Cited on page 3.)

[55] S. Virmani N. Ratanje. Generalised state spaces and non-locality in fault tolerant quantum computing schemes. January 2011. (Cited on page 35.)

[56] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 1 edition, October 2004. (Cited on page 15.)

[57] Asher Peres. All the bell inequalities. *Foundations of Physics*, 29:589–614, 1999. (Cited on page 65.)

[58] M. Piani. Relative entropy of entanglement and restricted measurements. *Phys. Rev. Lett.*, 103:160504, Oct 2009. (Cited on pages 52 and 56.)

[59] M. B. Plenio. Logarithmic negativity: A full entanglement monotone that is not convex. *Phys. Rev. Lett.*, 95:090503, Aug 2005. (Cited on pages 48 and 65.)

[60] Ben W. Reichardt. Quantum Universality from Magic States Distillation Applied to CSS Codes. *Quantum Information Processing*, 4(3):251–264, August 2005. (Cited on page 35.)

[61] Ben W. Reichardt. Error-Detection-Based Quantum Fault-Tolerance Threshold. *Algorithmica*, 55(3):517–556, November 2009. (Cited on page 35.)

[62] Ben W. Reichardt. Quantum universality by state distillation. *Quantum Information & Computation*, 9:1030–1052, July 2009. (Cited on page 35.)

[63] B.W. Reichardt. *Quantum Inf. Proc.*, 4:251, 2005. (Cited on page 47.)

[64] B.W. Reichardt. *Quantum Inf. Comput.*, 9:1030, 2009. (Cited on page 47.)

[65] Robert W. Spekkens. Negativity and Contextuality are Equivalent Notions of Nonclassicality. *Physical Review Letters*, 101(2):020401+, 2008. (Cited on page 4.)

[66] Robert W. Spekkens. Negativity and contextuality are equivalent notions of nonclassicality. *Phys. Rev. Lett.*, 101:020401, Jul 2008. (Cited on page 23.)

[67] Barbara Synak-Radtke and Michal Horodecki. On asymptotic continuity of functions of quantum states. *Journal of Physics A: Mathematical and General*, 39(26):L423, 2006. (Cited on page 54.)

[68] Wim van Dam and Mark Howard. Tight noise thresholds for quantum computation with perfect stabilizer operations. *Phys. Rev. Lett.*, 103:170504, Oct 2009. (Cited on page 35.)

[69] Wim van Dam and Mark Howard. Noise thresholds for higher-dimensional systems using the discrete wigner function. *Phys. Rev. A*, 83:032310, Mar 2011. (Cited on page 35.)

[70] V. Vedral and M. B. Plenio. Entanglement measures and purification procedures. *Phys. Rev. A*, 57:1619–1633, Mar 1998. (Cited on page 55.)

[71] Victor Veitch, Christopher Ferrie, David Gross, and Joseph Emerson. Negative quasi-probability as a resource for quantum computation. *New Journal of Physics*, 14(11):113011, 2012. (Cited on pages 5, 6, 32, 34, and 64.)

[72] Victor Veitch, Ali Hamed, Daniel Gottesman, and Joseph Emerson. The resource theory of stabilizer computation. In Preparation. (Cited on pages 5, 6, and 46.)

[73] Victor Veitch, Nathan Wiebe, Christopher Ferrie, and Joseph Emerson. Efficient simulation scheme for a class of quantum optics experiments with non-negative wigner representation. *New Journal of Physics*, 15(1):013037, 2013. (Cited on pages 6, 32, and 73.)

[74] G. Vidal and R. F. Werner. Computable measure of entanglement. *Phys. Rev. A*, 65:032314, Feb 2002. (Cited on page 65.)

[75] Guifré Vidal. Efficient Classical Simulation of Slightly Entangled Quantum Computations. *Physical Review Letters*, 91:147902+, October 2003. (Cited on pages 1 and 73.)

[76] Joel J. Wallman and Stephen D. Bartlett. Non-negative subtheories and quasiprobability representations of qubits. *Phys. Rev. A*, 85:062121, Jun 2012. (Cited on page 66.)

[77] E. Wigner. On the quantum correction for thermodynamic equilibrium. *Phys. Rev.*, 40:749–759, 1932. (Cited on pages 26 and 29.)

[78] Wikipedia. Extra special group, 2013. [Online; accessed 22-June-2013]. (Cited on pages 12 and 14.)

[79] William K Wootters. A wigner-function formulation of finite-state quantum mechanics. *Annals of Physics*, 176(1):1 – 21, 1987. (Cited on pages 4, 29, and 31.)

[80] Bei Zeng, Andrew Cross, and Isaac L. Chuang. Transversality versus Universality for Additive Quantum Codes, September 2007. (Cited on page 2.)

[81] Günter M. Ziegler. *Lectures on Polytopes*. Springer, 1995. (Cited on page 37.)