

## Quasi-probability representations of quantum theory with applications to quantum information science


To cite this article: Christopher Ferrie 2011 *Rep. Prog. Phys.* **74** 116001

View the [article online](#) for updates and enhancements.

### You may also like


- [Witnessing non-Markovianity by quantum quasi-probability distributions](#)  
Moritz F Richter, Raphael Wiedenmann and Heinz-Peter Breuer
- [Negative quasi-probability as a resource for quantum computation](#)  
Victor Veitch, Christopher Ferrie, David Gross et al.
- [Quasi-probability distributions in loop quantum cosmology](#)  
Jasiel Berra-Montiel and Alberto Molgado

www.hidenanalytical.com  
info@hiden.co.uk




# Instruments for Advanced Science

Mass spectrometers for vacuum, gas, plasma and surface science



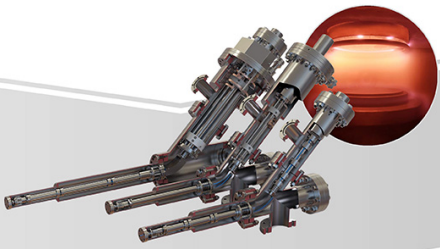
### Residual Gas Analysis

Perform RGA at UHV/XHV. Our RGA configurations include systems for UHV science applications including temperature-programmed desorption and electron/photon stimulated desorption.




### Thin Film Surface Analysis


Conduct both static and dynamic SIMS analysis with a choice of primary ions for full chemical composition and depth profiling. Our SIMS solutions include complete workstations and bolt-on modules.



### Plasma Characterisation

Fully characterise a range of plasmas: RF, DC, ECR and pulsed plasmas, including neutrals and neutral radicals. Extend your analyses to atmospheric pressure processes using the HPR-60, with time-resolved mass/energy analysis.

 www.HidenAnalytical.com

 info@hiden.co.uk

# Quasi-probability representations of quantum theory with applications to quantum information science

Christopher Ferrie

Institute for Quantum Computing and Department of Applied Mathematics, University of Waterloo,  
Waterloo, Ontario, N2L 3G1 Canada

Received 15 October 2010, in final form 20 July 2011

Published 14 October 2011

Online at [stacks.iop.org/RoPP/74/116001](http://stacks.iop.org/RoPP/74/116001)

## Abstract

This paper comprises a review of both the quasi-probability representations of infinite-dimensional quantum theory (including the Wigner function) and the more recently defined quasi-probability representations of finite-dimensional quantum theory. We focus on both the characteristics and applications of these representations with an emphasis toward quantum information theory. We discuss the recently proposed unification of the set of possible quasi-probability representations via frame theory and then discuss the practical relevance of negativity in such representations as a criteria for quantumness.

## Contents

<b>1. Introduction</b>	<b>1</b>	<b>4.6. Discrete Cahill–Glauber formalism</b>	<b>12</b>
<b>2. From probability to quasi-probability</b>	<b>2</b>	<b>4.7. Probability tables</b>	<b>13</b>
2.1. Operational theories, ontological models and non-contextuality	2	<b>4.8. Hardy’s vector representation</b>	<b>14</b>
2.2. Quantum theory, contextuality and quasi-probability	4	<b>4.9. The real density matrix</b>	<b>15</b>
<b>3. Quasi-probability in infinite-dimensional Hilbert space</b>	<b>5</b>	<b>4.10. Symmetric representations</b>	<b>15</b>
3.1. Wigner phase space representation	5	<b>5. Unification of the quasi-probability representations via frames</b>	<b>16</b>
3.2. Other phase space representations	6	5.1. Introduction to frames	16
<b>4. Quasi-probability in finite-dimensional Hilbert space</b>	<b>7</b>	5.2. Unification via the necessity and sufficiency of frames	16
4.1. Spherical phase space	7	5.3. To infinity and beyond	17
4.2. Wootters discrete phase space representation	9	<b>6. Negativity in quasi-probability representation</b>	<b>18</b>
4.3. Extended discrete phase space	10	6.1. Negativity as an indicator of quantumness	18
4.4. Even-dimensional discrete Wigner functions	10	6.2. Traditional contextuality and non-locality	19
4.5. Finite fields discrete phase space representation	11	<b>7. Summary</b>	<b>21</b>
		<b>Acknowledgments</b>	<b>22</b>
		<b>References</b>	<b>22</b>

## 1. Introduction

The quasi-probability representations of quantum theory evolved from the *phase space* representations. Phase space is a natural concept in classical theory since it is equivalent to the state space. The idea of formulating quantum theory in phase space dates back to the early days of quantum theory when the Wigner function was introduced (Wigner 1932). The

term quasi-probability refers to the fact that the function is not a true probability density as it takes on negative values for some quantum states.

The Wigner phase space formulation of quantum theory is equivalent to the usual abstract formalism of quantum theory in the same sense that Heisenberg’s matrix mechanics and Schrödinger’s wave mechanics are equivalent to the abstract formalism. The Wigner function representation

is not the only quasi-probabilistic formulation of quantum theory. However, in most, if not all, of the familiar quasi-probability representations the kinematic or ontic space of the representation is presumed to be the usual canonical phase space of classical physics. In the broader context of attempting to represent quantum mechanics as a classical probability theory, the classical state space need not necessarily correspond to the phase space of some classical canonical variables. But even these more general approaches are not applicable to describing quantum systems with a finite set of distinguishable states.

Recently, analogs of the Wigner function for finite-dimensional quantum systems have been proposed. There are many adequate reviews of infinite-dimensional phase representations<sup>1</sup>, which we complement by reviewing the recently defined quasi-probability representations of finite-dimensional quantum states. Moreover, we review a unifying formalism proposed by Ferrie and Emerson (2008, 2009), and Ferrie *et al* (2010) which encompasses all previously constructed quasi-probability representations and provides a means of defining the class of all possible quasi-probability representations.

While entanglement is the most notable criteria for quantumness, it can only be defined for composite systems. A non-negative quasi-probability function is a true probability distribution, prompting some authors to suggest that the presence of negativity in this function is a defining signature of non-classicality *for a single system*. Thus, a central concept in studies of the quantum–classical contrast in the quasi-probability formalisms of quantum theory is the appearance of negativity. However, it should be understood that any particular representation is non-unique and to some extent arbitrary since a state with negativity in one representation is positive in another. The concept of entanglement arose from the study of *non-locality*. This, and *contextuality*, are perhaps the most famous notions of quantumness. We will discuss the relationship between these and negativity in the final section. Non-locality and contextuality are terms often used as shorthand for the phrases ‘there does not exist a *local* hidden variable model for quantum theory’ and ‘there does not exist a *non-contextual* hidden variable model for quantum theory’. Negativity is a similar concept. Moreover, since the assumptions which go into the meta-concept of the *hidden variable model* are common to each notion, there is a sense in which it is convenient to call them ‘equivalent’ (Spekkens 2008).

The outline of the paper is as follows. In section 2 we see how quasi-probability arises quite naturally when trying to apply standard probabilistic constraints on quantum theory. In section 3 the quasi-probability representations on infinite-dimensional Hilbert space are reviewed. Those defined for finite-dimensional Hilbert spaces are review in section 4. Both are shown to be part of a unified formalism utilizing *frames* in section 5. Negativity as a criterion for quantumness and its relation to contextuality and non-locality are discussed in section 6. The final section summarizes what we have learned.

## 2. From probability to quasi-probability

### 2.1. Operational theories, ontological models and non-contextuality

An *operational theory* (Hardy 2001a, Harrigan and Rudolph 2007, Spekkens 2005) is an attempt to mathematically model a real physical experiment. The concepts in the theory are *preparations*, *systems*, *measurements* and *outcomes*. A preparation  $P$  is a proposition specifying how a real physical system has come to be the object of experimental investigation. We will reason about a set of mutually exclusive preparations  $\mathcal{P}$ . The system is assumed to then be measured according to some measurement procedure which produces an unambiguous answer called the outcome. The measurement procedure is specified by a proposition  $M$  belonging to a set of mutually exclusive possibilities  $\mathcal{M}$ . The outcome is specified by a proposition  $k$  which is assumed to belong to a set of mutually exclusive and exhaustive possibilities  $K$ . This means that, in any given run of the experiment, exactly one of  $K$  is true.

For example,  $P$  could be the setting of the knob on a ‘black box’ which sends objects to another ‘black box’ with knob setting  $M$  which outputs some sensory cue (e.g. an audible ‘click’ or flash of light) labeled  $k$  as each system passes through. Note that  $P$  and  $M$  need not be statements about devices in a laboratory (Schack 2003);  $P$  could be a statement about a photon arriving from the Sun, for example.

For a fixed preparation and measurement the outcome may not be deterministic. The role of the theory is then to describe the probability of each outcome conditional on the various combinations of preparations and measurements. That is, the theory should tabulate the numbers  $\Pr(k|P \wedge M)$  for all  $k, P$  and  $M$ . For fixed  $P$  and  $M$ , the mutually exclusive and exhaustive property of  $K$  implies  $\sum_k \Pr(k|P \wedge M) = 1$ . An operational theory is then a specification  $(\mathcal{P}, \mathcal{M}, K, \{\Pr(k|P \wedge M)\})$ .

Quantum theory is an example of an operational theory where each preparation  $P \in \mathcal{P}$  is associated with a density operator  $\rho_P \in \mathbb{D}(\mathcal{H})$  via the mapping  $\rho : \mathcal{P} \rightarrow \mathbb{D}(\mathcal{H})$  (for a explanation of the notation used in this paper, see the appendix). In general, this mapping is not required to be injective or surjective; different preparations may lead to the same density operator and there may not exist a preparation which leads to every density operator. Similarly, each measurement  $M \in \mathcal{M}$  and outcome  $k \in K$  is associated with an effect  $E_{M,k} \in \mathbb{E}(\mathcal{H})$  via the mapping  $E : \mathcal{M} \times K \rightarrow \mathbb{E}(\mathcal{H})$ . Again this mapping need not be injective or surjective. Quantum theory prescribes the probability distributions  $\Pr(k|P \wedge M) = \text{Tr}(\rho_P E_{M,k})$  which is called the Born rule. Since  $K$  is a set of mutually exclusive and exhaustive possibilities, for fixed  $M$ ,  $\{E_{M,k}\} \in \text{POVM}(\mathcal{H})$ .

Note that an operational theory only specifies the probabilities given the preparations in the set  $\mathcal{P}$ . Suppose, for example, we are told a coin is tossed which determines which of two preparations procedures are implemented in a laboratory experiment. The operational theory does not offer the probability of a given outcome conditional on this information. However, the laws of probability provide us

<sup>1</sup> See, for example, Hillery *et al* (1984), Lee (1995).

with the tools necessary to derive the desired probabilities. In general, the task is, given any disjunction  $\bigvee_i P_i$  of propositions (the set of which is denoted  $D(\mathcal{P})$ ) from the set of preparations  $\mathcal{P}$ , determine the probability distribution  $\Pr(k|\bigvee_i P_i \wedge M)$ . Since the set of preparations is mutually exclusive, the laws of probability dictate

$$\Pr(k|\bigvee_i P_i \wedge M) = \sum_i \Pr(k|P_i \wedge M) \Pr(P_i). \quad (1)$$

Similarly, for any disjunction  $\bigvee_j M_j$  of propositions from the set of measurements  $\mathcal{M}$

$$\Pr(k|P \wedge \bigvee_j M_j) = \sum_j \Pr(k|P \wedge M_j) \Pr(M_j). \quad (2)$$

Putting equations (1) and (2) together yields

$$\Pr(k|\bigvee_i P_i \wedge \bigvee_j M_j) = \sum_{ij} \Pr(k|P_i \wedge M_j) \Pr(P_i) \Pr(M_j). \quad (3)$$

Consider the example of quantum theory again. Let  $\bigvee_i P_i$  be an arbitrary disjunction of preparations from  $\mathcal{P}$ . Then for fixed  $k$  and  $M$

$$\begin{aligned} \Pr(k|\bigvee_i P_i \wedge M) &= \sum_i \Pr(k|P_i \wedge M) \Pr(P_i) \\ &= \sum_j \text{Tr}(\rho_{P_i} E_{M,k}) \Pr(P_i) \\ &= \text{Tr} \left( \sum_i \Pr(P_i) \rho_{P_i} E_{M,k} \right) \end{aligned}$$

suggesting we define the function  $\hat{\rho} : D(\mathcal{P}) \rightarrow \mathbb{D}(\mathcal{H})$  as

$$\hat{\rho} : \bigvee_i P_i \mapsto \sum_i \Pr(P_i) \rho_{P_i}. \quad (4)$$

Similarly, let us define a function  $\hat{E} : D(\mathcal{M}) \times K \rightarrow \mathbb{E}(\mathcal{H})$  as

$$\hat{E} : (\bigvee_j M_j, k) \mapsto \sum_j \Pr(M_j) E_{M_j,k}. \quad (5)$$

Now, using  $\hat{\rho}$  and  $\hat{E}$  and defining  $P_D := \bigvee_i P_i$  and  $M_D = \bigvee_j M_j$  we have

$$\Pr(k|\bigvee_i P_i \wedge \bigvee_j M_j) = \text{Tr}(\hat{\rho}_{P_D} \hat{E}_{M_D,k}). \quad (6)$$

This works for the probability distributions of any operational theory and hence the set  $(D(\mathcal{P}), D(\mathcal{M}), K, \mathcal{H}, \hat{\rho}, \hat{E})$  is an operational theory itself. Unless specified otherwise we will simply assume this analysis has been done when given an operational theory. This is equivalent to assuming the  $\rho$  and  $E$  of an operational theory  $(\mathcal{P}, \mathcal{M}, K, \mathcal{H}, \rho, E)$  have ranges which are convex sets.

An *ontological model* is an attempt at interpreting an operational theory as an effectively epistemic theory of a deeper model describing the *real state of affairs* of the system. An ontological model posits a set  $\Lambda$  of mutually exclusive and exhaustive *ontic states*  $\lambda$ . Each preparation  $P$  is assumed

to output the system in a particular ontic state  $\lambda$ . However, the experimental arrangement may not allow one to have knowledge of which state was prepared. This ignorance is quantified via a conditional probability distribution  $\Pr(\lambda|P)$ . When this probability is viewed as a mapping  $\mathcal{P} \rightarrow \text{Prob}(\Lambda)$ , from preparations to the set of probability distributions on  $\Lambda$ , it is neither injective or surjective. That is, in general, different preparation may give the same probability distribution of  $\Lambda$  and, certainly, not all probability distributions are realized. A measurement  $M$  may not deterministically give an outcome  $k$  which reveals the state  $\lambda$  of the system. Each measurement can then only be associated with an conditional probability distribution  $\Pr(k|M \wedge \lambda) \in \text{Prob}(M)$ . Again, when this probability is viewed as a mapping  $\mathcal{M} \times K \rightarrow \text{Prob}(M)$  it is not assumed to be injective or surjective. An ontological model also requires that  $\Lambda$  be such that knowledge of  $\lambda$  renders knowledge of  $P$  irrelevant. In the language of probability,

CI:  $k$  is conditionally independent of  $P$  given  $\lambda$ .

Summarizing, an ontological model is a set  $(\mathcal{P}, \mathcal{M}, K, \Lambda, \{\Pr(\lambda|P)\}, \{\Pr(k|M \wedge \lambda)\})$  such that CI holds. As a consequence of CI, the law of total probability states

$$\Pr(k|P \wedge M) = \sum_{\lambda} \Pr(k|M \wedge \lambda) \Pr(\lambda|P). \quad (7)$$

Loosely speaking, *non-contextuality* encodes the property that operationally equivalent procedures are represented equivalently in the ontological model (Spekkens 2005). Two preparations are indistinguishable operationally if no measurement exists for which the probability of any outcome is different between the two. An ontological model is non-contextual (with respect to its preparations) if the probability distributions over the ontic space the operationally equivalent preparations produce are equal. Similarly, measurements can be operationally equivalent and the ontological model can be non-contextual with respect to them. A mathematically concise definition of non-contextuality is as follows.

**Definition 1.** Let  $(\mathcal{P}, \mathcal{M}, K, \Lambda, \{\Pr(\lambda|P)\}, \{\Pr(k|M \wedge \lambda)\})$  define an ontological model.

(a) Let  $P, P' \in \mathcal{P}$ . The ontological model is preparation non-contextual if for all  $k \in K$ ,  $M \in \mathcal{M}$  and  $\lambda \in \Lambda$

$$\begin{aligned} \Pr(k|P \wedge M) &= \Pr(k|P' \wedge M) \Rightarrow \Pr(\lambda|P) \\ &= \Pr(\lambda|P'). \end{aligned}$$

(b) Let  $M, M' \in \mathcal{M}$ . The ontological model is measurement non-contextual if for all  $k \in K$ ,  $P \in \mathcal{P}$  and  $\lambda \in \Lambda$

$$\begin{aligned} \Pr(k|P \wedge M) &= \Pr(k|P \wedge M') \Rightarrow \Pr(k|M \wedge \lambda) \\ &= \Pr(k|M' \wedge \lambda). \end{aligned}$$

(c) The model is simply called a non-contextual ontological model if it is both preparation and measurement non-contextual.



## 2.2. Quantum theory, contextuality and quasi-probability

Recall that quantum theory as an operational model is defined via the mappings  $\rho$  and  $E$ . As an example, consider only those preparations associated with pure states. These pure states are determined by a mapping  $\rho =: \psi : \mathcal{P} \rightarrow \mathbb{P}(\mathcal{H})$ . The model of Beltrametti and Bugajski (1995) posits the ontic space  $\Lambda = \mathbb{P}(\mathcal{H})$  and a sharp probability distribution

$$\Pr(\lambda|P) d\lambda = \delta(\lambda - \psi_P) d\lambda \quad (8)$$

for each preparation  $P$ . This model suggests that the quantum state provides a complete specification of reality. Recall each measurement procedure is associated with a POVM via the mapping  $E : \mathcal{M} \times K \rightarrow \mathbb{E}(\mathcal{H})$ . Each measurement procedure  $M$  implies a conditional probability

$$\Pr(k|M \wedge \lambda) = \langle \lambda, E_{M,k} \lambda \rangle. \quad (9)$$

To show this is an ontological model, it remains only to verify that equation (7) is satisfied. It follows that

$$\begin{aligned} \Pr(k|P \wedge M) &= \int_{\Lambda} \Pr(k|M \wedge \lambda) \Pr(\lambda|P) d\lambda \\ &= \int_{\Lambda} \langle \lambda, E_{M,k} \lambda \rangle \delta(\lambda - \psi_P) d\lambda \\ &= \langle \psi_P, E_{M,k} \psi_P \rangle, \end{aligned}$$

which is the Born rule for pure states. The Beltrametti–Bugajski model is an ontological model for pure state quantum theory which is preparation non-contextual. However, the range of the mapping  $\psi$  in this case is not convex. As we will now see, if we looked at all possible logical disjunctions of preparations in this model, so that range of the new mapping  $\rho$  is  $\mathbb{D}(\mathcal{H})$ , quantum theory admits no non-contextual model. The most famous example of an ontological model for pure state quantum theory is the de Broglie–Bohm (or Bohmian) theory, which suffers the same flaw of admitting no non-contextual extension to the full (i.e. including density operators) quantum theory. This model is discussed later in section 6.2.

**Lemma 1.** *Suppose the convex range of  $E$  contains a basis (for  $\mathbb{H}(\mathcal{H})$ ) and the ontological model  $(\mathcal{P}, \mathcal{M}, K, \Lambda, \{\Pr(\lambda|P)\}, \{\Pr(k|M \wedge \lambda)\})$  is preparation non-contextual. Then, there exists an affine mapping (with domain the range of  $\rho$ )  $\mu : \text{Ran}(\rho) \rightarrow \text{Prob}(\Lambda)$  satisfying  $\mu(\lambda|\rho_P) = \Pr(\lambda|P)$ .*

**Proof.** For equivalent disjunctions of preparations

$$\text{Tr}(\rho_P E_{M,k}) = \Pr(k|P \wedge M) = \Pr(k|P' \wedge M) = \text{Tr}(\rho_{P'} E_{M,k}). \quad (10)$$

Since the range of  $E$  is a basis it spans  $\mathbb{H}(\mathcal{H})$  and therefore  $\rho_P = \rho_{P'}$  if and only if  $P$  and  $P'$  are operationally equivalent. Then, from the definition of preparation non-contextuality,  $\Pr(\lambda|P) \neq \Pr(\lambda|P')$  implies  $\rho_P \neq \rho_{P'}$ . Thus the mapping  $\mu(\lambda|\rho_P) = \Pr(\lambda|P)$  is well-defined.

Now we show the mapping  $\mu$  is affine. That is,

$$\mu(\lambda|p\rho_P + (1-p)\rho_{P'}) = p\mu(\lambda|\rho_P) + (1-p)\mu(\lambda|\rho_{P'})$$

for all  $p \in [0, 1]$  and all  $P, P' \in D(\mathcal{P})$ . There is some  $P'' \in D(\mathcal{P})$  such that  $\rho_{P''} = p\rho_P + (1-p)\rho_{P'}$ . That is to say,  $P'' = P \vee P'$  while  $\Pr(P) = p$  and  $\Pr(P') = 1-p$ . From the non-contextuality assumption this implies

$$\begin{aligned} \Pr(\lambda|P'') &= \Pr(\lambda|P \vee P') \\ &= \Pr(P) \Pr(\lambda|P) + \Pr(P') \Pr(\lambda|P'). \end{aligned}$$

Applying the definition of  $\mu$  to this yields

$$\mu(\lambda|\rho_{P''}) = p\mu(\lambda|\rho_P) + (1-p)\mu(\lambda|\rho_{P'}),$$

which proves  $\mu$  is affine.  $\square$

**Lemma 2.** *Suppose the range of  $\rho$  contains a basis (for  $\mathbb{H}(\mathcal{H})$ ) and the ontological model  $(\mathcal{P}, \mathcal{M}, K, \Lambda, \{\Pr(\lambda|P)\}, \{\Pr(k|M \wedge \lambda)\})$  is measurement non-contextual. Then, there exists a unique affine mapping  $\xi : \mathbb{E}(\mathcal{H}) \rightarrow (K)$  satisfying  $\Pr(k|M \wedge \lambda) = \xi(E_{M,k})$ .*

The proof of lemma 2 is similar to that of lemma 1. Together, the mappings  $\mu$  and  $\xi$  are called a *classical representation of quantum theory*.

**Lemma 3.** *A classical representation satisfies, for all  $\lambda \in \Lambda$ , all  $\rho \in \mathbb{D}(\mathcal{H})$  and all  $E \in \mathbb{E}(\mathcal{H})$ ,*

- (a)  $\mu_{\rho}(\lambda) \in [0, 1]$  and  $\sum_{\lambda} \mu_{\rho}(\lambda) = 1$ ,
- (b)  $\xi_E(\lambda) \in [0, 1]$  and  $\xi_{\mathbb{I}}(\lambda) = 1$ ,
- (c)  $\text{Tr}(\rho E) = \sum_{\lambda} \mu_{\rho}(\lambda) \xi_E(\lambda)$ .

In Ferrie and Emerson (2008) the name ‘classical representation’ was defined to be a set of mappings satisfying (a)–(c). Through lemmas 1 and 2 we have shown that the assumption of non-contextuality guarantees that these mappings are well-defined, convex linear and satisfy (a)–(c). However, regardless of how we choose to arrive at a pair of convex-linear mappings satisfying (a)–(c), one (or more) of the assumptions we make will be false as shown by the following theorem.

**Theorem 1.** *A classical representation of quantum theory does not exist.*

This theorem is implied by the equivalence of a classical representation and a non-contextual ontological model of quantum theory (Spekkens 2008) and the impossibility of the latter (Spekkens 2005). This theorem was proven in Ferrie and Emerson (2008) using the notion of a frame and its dual, which we will see later. A direct and more intuitive proof was given in Ferrie and Emerson (2009). The theorem is also implied by an earlier result on a related topic in Busch *et al* (1993).

Now let us relax some assumptions in order to avoid a contradiction. First, we will relax the assumption of non-contextuality.

**Definition 2.** *A contextual representation of quantum theory is a pair of mappings  $\mu : \mathbb{D}(\mathcal{H}) \times C_{\mathcal{P}} \rightarrow P(\Lambda)$  and  $\xi : \mathbb{E}(\mathcal{H}) \times C_{\mathcal{M}} \rightarrow P(\Lambda)$  which satisfy, for all  $\lambda \in \Lambda$ , all  $\rho \in \mathbb{D}(\mathcal{H})$  and all  $E \in \mathbb{E}(\mathcal{H})$ , all  $c_{\mathcal{P}} \in C_{\mathcal{P}}$ , and all  $c_{\mathcal{M}} \in C_{\mathcal{M}}$ ,*

- (a)  $\mu_{\rho, c_{\mathcal{P}}}(\lambda) \in [0, 1]$  and  $\sum_{\lambda} \mu_{\rho, c_{\mathcal{P}}}(\lambda) = 1$ ,
- (b)  $\xi_{E, c_{\mathcal{M}}}(\lambda) \in [0, 1]$  and  $\xi_{\mathbb{I}, c_{\mathcal{M}}}(\lambda) = 1$ ,
- (c)  $\text{Tr}(\rho E) = \sum_{\lambda} \mu_{\rho, c_{\mathcal{P}}}(\lambda) \xi_{E, c_{\mathcal{M}}}(\lambda)$ .

Here  $C_{\mathcal{P}}$  and  $C_{\mathcal{M}}$  are the preparation and measurement contexts.

To the best of our knowledge such a construction does not exist in the literature. In the previous section we looked at the Beltrametti–Bugajski model<sup>2</sup> which was non-contextual for pure states. The model cannot be extended to include all states such that it remains non-contextual. We could artificially create a *preparation contextual* model by assigning a unique context to each convex decomposition of every density operator. It is unclear whether such a construction could be useful. It is more common, however, to relax the assumption of non-negative probability rather than non-contextuality.

**Definition 3.** A quasi-probability representation of quantum theory is a pair of affine mappings  $\mu : \mathbb{D}(\mathcal{H}) \rightarrow L(\Lambda)$  and  $\xi : \mathbb{E}(\mathcal{H}) \rightarrow L(\Lambda)$  which satisfy, for all  $\lambda \in \Lambda$ , all  $\rho \in \mathbb{D}(\mathcal{H})$  and all  $E \in \mathbb{E}(\mathcal{H})$ ,

- (a)  $\mu_{\rho}(\lambda) \in \mathbb{R}$  and  $\sum_{\lambda} \mu_{\rho}(\lambda) = 1$ ,
- (b)  $\xi_E(\lambda) \in \mathbb{R}$  and  $\xi_{\mathbb{I}}(\lambda) = 1$ ,
- (c)  $\text{Tr}(\rho E) = \sum_{\lambda} \mu_{\rho}(\lambda) \xi_E(\lambda)$ .

It is immediately clear that theorem 1 is equivalent to the following ‘negativity theorem’:

**Theorem 2.** A quasi-probability representation of quantum theory must have negativity in either its representation of states or measurements (or both).

There are many instances of mappings  $\mu$  satisfying the first requirement. In Ferrie and Emerson (2008) it was shown by construction how to find a mapping  $\xi$  which, together with  $\mu$ , satisfy all of them. Most of the mappings  $\mu$  are called *phase space functions* as they conform to added mathematical structure not required in definition 3. A phase space representation is then a particular type of quasi-probability representation which we formally define as follows:

**Definition 4.** If there exists a non-trivial symmetry group on  $\Lambda$ ,  $G$ , carrying a unitary representation  $U : G \rightarrow \mathbb{U}(\mathcal{H})$  and a quasi-probability representation satisfying the covariance property  $U_g \rho U_g^\dagger \mapsto \{\mu_{\rho}(g(\lambda))\}_{\lambda \in \Lambda}$  for all  $\rho \in \mathbb{D}(\mathcal{H})$  and  $g \in G$ , then  $\rho \mapsto \mu_{\rho}(\lambda)$  is a phase space representation of quantum states.

It will turn out that all but a few of the representations reviewed below satisfy this stronger criterion.

### 3. Quasi-probability in infinite-dimensional Hilbert space

Here we will review the quasi-probability distributions which have been defined for quantum states living in an infinite-dimensional Hilbert space—the canonical example being a particle moving in one dimension. Since there are a myriad of excellent reviews of the Wigner function and other phase space distributions<sup>3</sup>, our discussion of them will be brief. We will

<sup>2</sup> These comments apply equally to other ontological models for pure states, such as the de Broglie–Bohm model.

<sup>3</sup> See Hillery *et al* (1984) for a classic and Lee (1995) for a more recent review of phase space quasi-probability distributions.

mainly focus on those details which have inspired analogous methods for finite dimensional Hilbert spaces.

First we start with the familiar Wigner function in section 3.1. The other phase space distributions, such as the Husimi function, are bundled up in section 3.2.

#### 3.1. Wigner phase space representation

The position operator  $Q$  and momentum operator  $P$  are the central objects in the abstract formalism of infinite-dimensional quantum theory. The operators satisfy the canonical commutation relations

$$[Q, P] = i.$$

We are looking for a joint probability distribution  $\mu_{\rho}(p, q)$  for the state of the quantum system. From the postulates of quantum mechanics we have a rule for calculating expectation values. In particular, we can compute the characteristic function

$$\phi(\xi, \eta) := \langle e^{i(\xi Q + \eta P)} \rangle = \text{Tr}(e^{i(\xi Q + \eta P)} \rho).$$

Since the characteristic function is just the Fourier transform of the joint probability distribution, we simply invert to obtain

$$\mu_{\rho}(p, q) = \frac{1}{(2\pi)^2} \iint_{\mathbb{R}^2} \text{Tr}(e^{i(\xi Q + \eta P)} \rho) e^{-i(\xi q + \eta p)} d\xi d\eta, \quad (11)$$

which is the celebrated *Wigner function* of  $\rho$  (Wigner 1932). The Wigner function is both positive and negative in general. However, it otherwise behaves as a classical probability density on the classical phase space. For these reasons, the Wigner function and others like it came to be called *quasi-probability functions*.

The Wigner function is the unique representation satisfying the properties (Bertrand and Bertrand 1987)

Wig(1): For all  $\rho$ ,  $\mu_{\rho}(q, p)$  is real.

Wig(2): For all  $\rho_1$  and  $\rho_2$ ,

$$\text{Tr}(\rho_1 \rho_2) = 2\pi \int_{\mathbb{R}^2} dq dp \mu_{\rho_1}(q, p) \mu_{\rho_2}(q, p).$$

Wig(3): For all  $\rho$ , integrating  $\mu_{\rho}$  along the line  $aq + bp = c$  in phase space yields the probability that a measurement of the observable  $aQ + bP$  has the result  $c$ .

We can write the Wigner function as

$$\mu_{\rho}(p, q) = \text{Tr}[F(p, q)\rho],$$

where

$$F(q, p) := \frac{1}{(2\pi)^2} \iint_{\mathbb{R}^2} d\xi d\eta e^{i\xi(Q-q) + i\eta(P-p)}. \quad (12)$$

Thus the properties Wig(1)–(3) can be transformed into properties on a set of operators  $F(q, p)$  which uniquely specify the set in equation (13). These properties are

Wig(4):  $F(q, p)$  is Hermitian.

Wig(5):  $2\pi \text{Tr}(F(q, p)F(q', p')) = \delta(q - q')\delta(p - p')$ .

Wig(6): Let  $P_c$  be the projector onto the eigenstate of  $aQ + bP$  with eigenvalue  $c$ . Then,

$$\int_{\mathbb{R}^2} dq dp F(q, p) \delta(aq + bp - c) = P_c.$$

These six properties are often the basis for generalizing the Wigner function to finite-dimensional Hilbert spaces, as we will soon see.

**3.1.1. Applications: quantum teleportation.** The applications of the Wigner function are far reaching and not limited to physics<sup>4</sup>. A modern application can be found in Caves and Wódkiewicz (2004) where Caves and Wódkiewicz use the Wigner function to obtain a hidden variable model of the continuous-variable teleportation protocol (Braunstein and Kimble 1998, Vaidman 1994). Later, in section 4.6, we will discuss the much simpler discrete-variable teleportation protocol. Here, then, we will avoid the details of the protocol and focus on the result. It suffices to know the following: there are three quantum systems; the goal of the protocol is to transfer a quantum state from system 1 to system 3; the transfer is mediated through the special correlations between system 2 and system 3. The success of the protocol is measured by the average *fidelity*: a measure of the closeness of the initial state  $\rho = |\psi\rangle\langle\psi|$  of system 1 and the average final state  $\rho_{\text{out}}$  of system 3.

Following Caves and Wódkiewicz we define  $v = q + ip$  and index the Wigner function as  $\mu_\rho(v)$ . This is convenient since the protocol is tailored to a quantum optical implementation where the outcomes of measurements are usually expressed as complex numbers. Initially, the state of systems 2 and 3 is described by the joint Wigner function  $\mu_{2,3}(\alpha, \beta)$ .

In terms of the Wigner functions, the average fidelity is

$$\mathcal{F} = \pi \int d^2 v d^2 \beta \mu_\rho(v) \mu_{\rho_{\text{out}}}(\beta).$$

This intuitively measures closeness by quantifying the overlap of the Wigner functions on the classical phase space. The output state, determined by the details of the protocol, is

$$\mu_{\rho_{\text{out}}}(\beta) = \int d^2 v d^2 \alpha \mu_\rho(\beta - v) \mu_{2,3}(\alpha, v - \bar{\alpha}).$$

The initial Wigner function  $\mu_\rho(v)$  and the joint Wigner function  $\mu_{2,3}(\alpha, \beta)$  are determined by the particular implementation of the protocol. The standard quantum optical implementation is done using coherent states of light. It is easy to show that such states have positive Wigner functions<sup>5</sup>. Thus, the Wigner function provides a classical phase space picture of the entire protocol.

A first step toward performing an experiment requiring genuine quantum resources might be to avoid the above classical description by teleporting a non-coherent quantum

state. Caves and Wódkiewicz have devised a classical explanation for this case as well. The new model involves a randomization procedure which transforms the initial non-coherent state into a coherent one thus giving it a positive Wigner function. However, it can be shown that within such a model, the fidelity is bounded:  $\mathcal{F} < 2/3$ . So  $2/3$  emerges as a ‘gold-standard’ since teleporting a non-coherent state with fidelity  $\mathcal{F} \geq 2/3$  avoids this classical phase space description.

### 3.2. Other phase space representations

Another class of solutions to the ordering problem is the association  $e^{i\xi q + i\eta p} \mapsto e^{i\xi Q + i\eta P} f(\xi, \eta)$  for some arbitrary function  $f$ .<sup>6</sup>

Consider again the classical particle phase space  $\mathbb{R}^2$  and the continuous set of operators

$$F(q, p) := \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} d\xi d\eta e^{i\xi(q-Q) + i\eta(p-P)} f(\xi, \eta). \quad (13)$$

The  $f$  dependent distributions

$$\mu_\rho^f(q, p) := \text{Tr}(\rho F(q, p)) \quad (14)$$

define quasi-probability functions on phase space alternative to the Wigner function, which is simply the  $f = 1$  special case of this more general formalism.

Besides the Wigner function, the most popular choices of  $f$  are

$$f(\xi, \eta) = e^{\pm \frac{1}{4}(\xi^2 + \eta^2)},$$

which give, via equation (14), the Glauber–Sudarshan (Glauber 1963, Sudarshan 1963) and Husimi (Husimi 1940) functions, respectively. These two mappings are sometimes referred to as the P- and Q-representations (not to be confused with position and momentum representations). We will follow the usual convention by introducing the annihilation operator

$$a = \frac{1}{\sqrt{2}}(Q + iP) \quad (15)$$

and the *coherent states* defined via  $a|\alpha\rangle = \alpha|\alpha\rangle$ , where we write  $\alpha = q + ip$ . Then the Husimi function can be conveniently written

$$Q(\alpha) := \mu_\rho^f(q, p) = \frac{1}{\pi} \langle \alpha, \rho \alpha \rangle. \quad (16)$$

The Glauber–Sudarshan function  $\rho \rightarrow P(\alpha)$  can be expressed implicitly through the identity

$$\rho = \int d^2 \alpha P(\alpha) |\alpha\rangle\langle\alpha|, \quad (17)$$

where  $d^2 \alpha = (1/2) dq dp$ . Notice that this immediate implies the following *duality* condition between the P- and Q-representation:

$$\text{Tr}(\rho \rho') = \int d^2 \alpha P_\rho(\alpha) Q_{\rho'}(\alpha). \quad (18)$$

These functions are further discussed in section 6.1.

<sup>4</sup> For example, see Dragoman (2005) for a recent review of the applications of the Wigner function in signal processing.

<sup>5</sup> More difficult is to show that coherent states are the *only* states with positive Wigner functions (Hudson 1974).

<sup>6</sup> This is very closely related to the  $s$ -ordered Cahill–Glauber formalism (Cahill and Glauber 1969). See table 1 of Lee (1995) for a concise review of the traditional choices for  $f$ .

**3.2.1. Application: quantumness witness.** Here we will discuss a more recent application of the  $P$  and  $Q$  functions of interest in quantum information and foundations<sup>7</sup>. We will concern ourselves with a particular notion of ‘non-classicality’ defined in Alicki *et al* (2008), Alicki and Ryn (2008), and Korbicz *et al* (2005).

Consider two observables represented by the self-adjoint operators  $R$  and  $S$ . Another observable  $W(R, S)$  written as an ordered power series of  $R$  and  $S$  is a *quantumness witness* if it possesses at least one negative eigenvalue and the function  $w(r, s)$  obtained by replacing  $R$  and  $S$  with its spectral elements is positive:  $w(r, s) \geq 0$  for all  $r \in \text{spec}(R)$  and  $s \in \text{spec}(S)$  ( $\text{spec}(\cdot)$  denotes the spectrum of the observable).

Now consider  $R = Q$  and  $S = P$ , the usual position and momentum operators. Recalling the parametrization in terms of the annihilation operator in equation (15), we define

$$W(a) := \sum_{m,n} c_{mn} (a^\dagger)^m a^n,$$

for any choice of coefficients  $c_{mn}$ , such that

$$w(\alpha) = \sum_{m,n} c_{mn} \bar{\alpha}^m \alpha^n \geq 0$$

and  $W$  possesses at least one negative eigenvalue (an example of such a construction is  $W(a) = (a^\dagger)^2 a^2 - 2ma^\dagger a + m^2$  for  $m \geq 1$ ). Then, in terms of the Q- and P-representations defined above, we have

$$\begin{aligned} \langle W \rangle &= \text{Tr}(\rho W) = \int d^2\alpha \langle \alpha, W\alpha, P \rangle(\alpha) \\ &= \int d^2\alpha w(\alpha) P(\alpha). \end{aligned} \quad (19)$$

In optics especially, coherent states are considered classical. From equation (17) we see that if  $P(\alpha) \geq 0$ , the quantum state is a statistical mixture of coherent states and hence just as classical. So if  $\rho$  is a classical state, equation (19) tells us  $\langle W \rangle \geq 0$ . Therefore, if we measure  $\langle W \rangle \leq 0$ , we can rule out the classical model of statistical mixtures of coherent states; we can say  $W(a)$  detects the quantumness of the states.

#### 4. Quasi-probability in finite-dimensional Hilbert space

Nearly all definitions of quasi-probability distributions for finite-dimensional Hilbert spaces have been motivated by the Wigner function. The earliest such effort was by Stratonovich and is reviewed in section 4.1. The Stratonovich phase space is a sphere and hence continuous. Later, many authors have defined Wigner function analogs on *discrete* phase spaces. A sampling, with a bias towards those which have found application in quantum information theory, is given in sections 4.2–4.6.

There also exist quasi-probability distributions which were introduced to solve various problems far removed from proposing a finite-dimensional analog of the Wigner function.

<sup>7</sup> The  $P$  and  $Q$  functions are powerful visualization tools prominently used in the areas of quantum optics and quantum chaos (Takahashi 1989). See Schleich (2001) for an overview of the applications in quantum optics.

Sections 4.7–4.10 review those quasi-probability distributions which do not have a canonical phase space structure and hence form a somewhat weaker analogy to the Wigner function.

We note that there exists many other quasi-probability distributions defined on discrete phase spaces which are not reviewed here (Hannay and Berry 1980, Cohen and Scully 1986, Feynman 1987, Galetti and de Toledo Piza 1988, Opatrný *et al* 1996, Hakioglu 1998, Luis and Perina 1998, Rivas 1999, Mukunda 2004, Vourdas 2004, Chaturvedi *et al* 2005, 2006, Gross 2006, 2007).

##### 4.1. Spherical phase space

Here we will be concerned with a set of postulates put forth by Stratonovich (1957). Stratonovich’s aim was to find a Wigner function type mapping, analogous to that of a infinite-dimensional system on  $\mathbb{R}^2$ , of a  $d$ -dimensional system on the sphere  $\mathbb{S}^2$ . The first postulate is linearity and is always satisfied if the Wigner functions on the sphere satisfy

$$\mu_\rho(\mathbf{n}) = \text{Tr}(\rho \Delta(\mathbf{n})), \quad (20)$$

where  $\mathbf{n}$  is a point on  $\mathbb{S}^2$ . The remaining postulates on this quasi-probability mapping are

$$\mu_\rho(\mathbf{n})^* = \mu_\rho(\mathbf{n}),$$

$$\frac{d}{4\pi} \int_{\mathbb{S}^2} d\mathbf{n} \mu_\rho(\mathbf{n}) = 1,$$

$$\frac{d}{4\pi} \int_{\mathbb{S}^2} d\mathbf{n} \mu_{\rho_1}(\mathbf{n}) \mu_{\rho_2}(\mathbf{n}) = \text{Tr}(\rho_1 \rho_2),$$

$$\mu_{(g \cdot \rho)}(\mathbf{n}) = \mu_\rho(\mathbf{n})^g, \quad g \in \text{SU}(2),$$

where  $g \cdot \rho$  is the image of  $U_g \rho U_g^\dagger$  and  $U : \text{SU}(2) \rightarrow \mathbb{U}\mathcal{H}$  is an irreducible unitary representation of the group  $\text{SU}(2)$ . These postulates are analogous to Wig(1)–(3) for the Wigner function modulo the second normalization condition (which could have been included in the Wigner function properties).

The continuous set of operators  $\Delta(\mathbf{n})$  is called a *kernel* and we note it plays the role of the more familiar *phase space point operators* in the latter. Requiring that equation (20) holds changes the postulates to new conditions on the kernel

$$\Delta(\mathbf{n})^\dagger = \Delta(\mathbf{n}), \quad (21)$$

$$\frac{d}{4\pi} \int_{\mathbb{S}^2} d\mathbf{n} \Delta(\mathbf{n}) = \mathbb{1}, \quad (22)$$

$$\frac{d}{4\pi} \int_{\mathbb{S}^2} d\mathbf{n} \text{Tr}(\Delta(\mathbf{n}) \Delta(\mathbf{m})) \Delta(\mathbf{n}) = \Delta(\mathbf{m}), \quad (23)$$

$$\Delta(g \cdot \mathbf{n}) = U_g \Delta(\mathbf{n}) U_g^\dagger, \quad g \in \text{SU}(2). \quad (24)$$

These postulates are the spherical analogies of properties Wig(4)–(6) (again, modulo the normalization condition). Defining  $s = \frac{d-1}{2}$  as the *spin*, Heiss and Weigert (2000) provided a concise derivation of  $2^{2s}$  unique kernels satisfying these postulates<sup>8</sup>. They are

$$\Delta(\mathbf{n}) = \sum_{m=-s}^s \sum_{l=0}^{2s} \epsilon_l \frac{2l+1}{2s+1} C_{m0m}^{sls} \phi_m(\mathbf{n}) \phi_m^*(\mathbf{n}), \quad (25)$$

<sup>8</sup> This was also shown earlier (Brif and Mann 1999, Varilly and Graciabondia 1989)—see also Arecchi *et al* (1972), Scully and Wódkiewicz (1994).



where  $C$  denotes the Clebsch–Gordon coefficients. Here,  $\phi_m(\mathbf{n})$  are the eigenvectors of the operator  $\mathbf{S} \cdot \mathbf{n}$ , where  $\mathbf{S} = (X, Y, Z)$ ; and  $\epsilon_l = \pm 1$ , for  $l = 1 \dots 2s$  and  $\epsilon_0 = 1$ .

Heiss and Weigert relax the postulates equations (21)–(24) on the kernel  $\Delta(\mathbf{n})$  to allow for a pair of kernels  $\Delta^n$  and  $\Delta_m$ . The pair individually satisfy equation (21), while one of them satisfies equation (22) and the other equation (24). Together, the pair must satisfy the generalization of equation (23)

$$\frac{d}{4\pi} \int_{\mathbb{S}^2} d\mathbf{n} \text{Tr}(\Delta^n \Delta_m) \Delta^n = \Delta_m. \quad (26)$$

A pair of kernels, together satisfying equation (26), is given by

$$\begin{aligned} \Delta_n &= \sum_{m=-s}^s \sum_{l=0}^{2s} \gamma_l \frac{2l+1}{2s+1} C_{m \ 0 \ m}^{s \ l \ s} \phi_m(\mathbf{n}) \phi_m^*(\mathbf{n}), \\ \Delta^n &= \sum_{m=-s}^s \sum_{l=0}^{2s} \gamma_l^{-1} \frac{2l+1}{2s+1} C_{m \ 0 \ m}^{s \ l \ s} \phi_m(\mathbf{n}) \phi_m^*(\mathbf{n}), \end{aligned}$$

where each  $\gamma_l$  is a finite non-zero real number and  $\gamma_0 = 1$ . The original postulates are satisfied when  $\gamma_l = \gamma_l^{-1} \equiv \epsilon_l$ .

The major contribution of Heiss and Weigert (2000) is the derivation of a *discrete* kernel  $\Delta_v := \Delta_{\mathbf{n}_v}$ , for  $v = 1 \dots d^2$  which satisfies the discretized postulates

$$\Delta_v^\dagger = \Delta_v, \quad (27)$$

$$\frac{1}{d} \sum_{v=1}^{d^2} \Delta_v = \mathbb{1}, \quad (28)$$

$$\frac{1}{d} \sum_{v=1}^{d^2} \text{Tr}(\Delta_v \Delta^\mu) \Delta_v = \Delta^\mu, \quad (29)$$

$$\Delta_{g \cdot v} = U_g \Delta_v U_g^\dagger, \quad g \in \text{SU}(2). \quad (30)$$

The subset of points  $\mathbf{n}_v$  is called a *constellation*. The linearity postulate is not explicitly stated since it is always satisfied under the assumption

$$\rho \rightarrow \mu_\rho(v) = \text{Tr}(\rho \Delta_v). \quad (31)$$

Equation (29) is called a *duality* condition. That is, it is only satisfied if  $\Delta_v$  and  $\Delta^\mu$  are *dual bases* for  $\mathbb{H}(\mathcal{H})$ . In particular,

$$\frac{1}{d} \text{Tr}(\Delta_v \Delta^\mu) = \delta_{v\mu}.$$

Although the explicit construction of a pair of discrete kernels satisfying equations (27)–(30) might be computationally hard, their existence is a trivial exercise in linear algebra. Indeed, so long as  $\Delta_v$  is a basis for  $\mathbb{H}(\mathcal{H})$ , its dual,  $\Delta^\mu$ , is uniquely determined by

$$\Delta^\mu = \sum_{v=1}^{d^2} G_{v\mu}^{-1} \Delta_v,$$

where the Gram matrix  $G$  is given by

$$G_{v\mu} = \text{Tr}(\Delta_v \Delta_\mu).$$

Heiss and Weigert (2000) note that almost any constellation leads to a discrete kernel  $\Delta_v$  forming a basis for  $\mathbb{H}(\mathcal{H})$ . The term *almost any* here means that a randomly selected discrete kernel will form, with probability 1, a basis for  $\mathbb{H}(\mathcal{H})$ .

**4.1.1. Application: NMR quantum computation.** The spherical quasi-probability functions for qubit systems ( $d = 2$ ) were put to use by Schack and Caves for the purpose of obtaining a classical model of NMR experiments designed to perform quantum information tasks (Schack and Caves 1999). For a single qubit we choose the kernels

$$\Delta_n = \frac{1}{2}(\mathbb{1} + \mathbf{n} \cdot \boldsymbol{\sigma}),$$

$$\Delta^n = \frac{1}{4\pi}(\mathbb{1} + 3\mathbf{n} \cdot \boldsymbol{\sigma}),$$

where  $\boldsymbol{\sigma} = (X, Y, Z)$  are the usual Pauli operators. In NMR experiments many qubits are employed to perform quantum information tasks such as error correction and teleportation. Suppose there are  $n$  qubits with total Hilbert space dimension  $2^n$ . We choose an  $n$ -fold tensor product of the qubit kernels. Explicitly, they are

$$\Delta_n = \frac{1}{2^n} \bigotimes_{j=1}^n (\mathbb{1} + \mathbf{n} \cdot \boldsymbol{\sigma}), \quad (32)$$

$$\Delta^n = \frac{1}{(4\pi)^n} \bigotimes_{j=1}^n (\mathbb{1} + 3\mathbf{n} \cdot \boldsymbol{\sigma}). \quad (33)$$

The quasi-probability function is given by

$$\mu_\rho(\mathbf{n}) = \text{Tr}(\rho \Delta^n).$$

As expected, in general, this function is both positive and negative.

The quantum state of an NMR experiment is of the form

$$\rho = (1 - \epsilon) \frac{1}{2^n} \mathbb{1} + \epsilon \rho_1, \quad (34)$$

where  $\rho_1$  is arbitrary but often chosen to be a specific pure state. The parameter scales as

$$\epsilon \propto \frac{n}{2^n}.$$

So we have

$$\mu_\rho(\mathbf{n}) = \frac{1 - \epsilon}{(4\pi)^n} + \epsilon \mu_{\rho_1}(\mathbf{n}).$$

It is easy to determine the lower bound

$$\mu_\rho(\mathbf{n}) \geq \frac{1 - \epsilon}{(4\pi)^n} - \frac{\epsilon 2^{2n-1}}{(4\pi)^n}.$$

Thus, provided

$$\epsilon \leq \frac{1}{1 + 2^{2n-1}},$$

$\mu_\rho(\mathbf{n}) \geq 0$  we have a representation of NMR quantum states in terms of classical probability distributions on a classical phase space. Note however, that this definition of ‘classical’ omits the reasonable requirement that it also be *efficient*. Indeed, the authors note that, for typical experimental values of the scaling parameter in  $\epsilon$ , such a classical representation is valid for  $n < 16$  qubits.

The spherical phase space representations have also been put to good use in visualizing decoherence (Ghose *et al* 2008) and photon squeezing (Shalm *et al* 2009).

#### 4.2. Wootters discrete phase space representation

Wootters (1987) defined a discrete analog of the Wigner function. Associated with each Hilbert space  $\mathcal{H}$  of finite dimension  $d$  is a *discrete phase space*. First assume  $d$  is prime. The *prime phase space*,  $\Phi$ , is a  $d \times d$  array of points  $\alpha = (q, p) \in \mathbb{Z}_d \times \mathbb{Z}_d$ .

A *line*,  $\lambda$ , is the set of  $d$  points satisfying the linear equation  $aq + bp = c$ , where all arithmetic is modulo  $d$ . Two lines are *parallel* if their linear equations differ in the value of  $c$ . The prime phase space  $\Phi$  contains  $d + 1$  sets of  $d$  parallel lines called *striations*.

Assume the Hilbert space  $\mathcal{H}$  has composite dimension  $d = d_1 d_2 \cdots d_k$ . The discrete phase space of the entire  $d$ -dimensional system is the Cartesian product of two-dimensional prime phase spaces of the subsystems. The phase space is thus a  $(d_1 \times d_1) \times (d_2 \times d_2) \times \cdots \times (d_k \times d_k)$  array. Such a construction is formalized as follows: the *discrete phase space* is the multi-dimensional array  $\Phi = \Phi_1 \times \Phi_2 \times \cdots \times \Phi_k$ , where each  $\Phi_i$  is a prime phase space. A *point* is the  $k$ -tuple  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$  of points  $\alpha_i = (q_i, p_i)$  in the prime phase spaces. A *line* is the  $k$ -tuple  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$  of lines in the prime phase spaces. That is, a line is the set of  $d$  points satisfying the equation

$$(a_1 q_1 + b_1 p_1, a_2 q_2 + b_2 p_2, \dots, a_k q_k + b_k p_k) = (c_1, c_2, \dots, c_k),$$

which is symbolically written  $aq + bp = c$ . Two lines are *parallel* if their equations differ in the value  $c$ . As was the case for the prime phase spaces, parallel lines can be partitioned into sets, again called striations; the discrete phase space  $\Phi$  contains  $(d_1 + 1)(d_2 + 1) \cdots (d_k + 1)$  sets of  $d$  parallel lines.

The construction of the discrete phase space is now complete. To introduce Hilbert space into the discrete phase space formalism, Wootters chooses the following special basis for the space of Hermitian operators. The set of operators  $\{A_\alpha : \alpha \in \Phi\}$  acting on an  $d$ -dimensional Hilbert space are called *phase point operators* if the operators satisfy

Woo(4): For each point  $\alpha$ ,  $A_\alpha$  is Hermitian.

Woo(5): For any two points  $\alpha$  and  $\beta$ ,  $\text{Tr}(A_\alpha A_\beta) = d \delta_{\alpha\beta}$ .

Woo(6): For each line  $\lambda$  in a given striation, the operators

$$P_\lambda = \frac{1}{d} \sum_{\alpha \in \lambda} A_\alpha$$

form a projective valued measurement (PVM): a set of  $d$  orthogonal projectors which sum to identity.

Note that these properties of the phase point operators Woo(4)–(6) are discrete analogs of the properties Wig(4)–(6) of the function  $F$  defining the original Wigner function. This definition suggests that the lines in the discrete phase space should be labeled with states of the Hilbert space. Since each striation is associated with a PVM, each of the  $d$  lines in a striation is labeled with an orthogonal state. For each  $\Phi$ , there is a unique set of phase point operators up to unitary equivalence.

Although the sets of phase point operators are unitarily equivalent, the induced labeling of the lines associated with the chosen set of phase point operators are not equivalent. This

is clear from the fact that unitarily equivalent PVMs do not project onto the same basis.

The choice of phase point operators in Wootters (1987) will be adopted. For  $d$  prime, the phase point operators are

$$A_\alpha = \frac{1}{d} \sum_{j,m=0}^{d-1} \omega^{pj-qm+\frac{jm}{2}} X^j Z^m, \quad (35)$$

where  $\omega$  is a  $d$ th root of unity and  $X$  and  $Z$  are the generalized Pauli operators (see the appendix). For composite  $d$ , the phase point operator in  $\Phi$  associated with the point  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$  is given by

$$A_\alpha = A_{\alpha_1} \otimes A_{\alpha_2} \otimes \cdots \otimes A_{\alpha_k}, \quad (36)$$

where each  $A_{\alpha_i}$  is the phase point operator of the point  $\alpha_i$  in  $\Phi_i$ .

The  $d^2$  phase point operators are linearly independent and form a basis for the space of Hermitian operators acting on an  $d$ -dimensional Hilbert space. Thus, any density operator  $\rho$  can be decomposed as

$$\rho = \sum_{q,p} \mu_\rho(q, p) A(q, p),$$

where the real coefficients are explicitly given by

$$\mu_\rho(q, p) = \frac{1}{d} \text{Tr}(\rho A(q, p)). \quad (37)$$

This discrete phase space function is the Wootters *discrete Wigner function*. This discrete quasi-probability function satisfies the following properties which are the discrete analogs of the properties Wig(1)–(3) that the original continuous Wigner function satisfies.

Woo(1): For all  $\rho$ ,  $\mu_\rho(q, p)$  is real.

Woo(2): For all  $\rho_1$  and  $\rho_2$ ,

$$\text{Tr}(\rho_1 \rho_2) = d \sum_{q,p} \mu_{\rho_1}(q, p) \mu_{\rho_2}(q, p).$$

Woo(3): For all  $\rho$ , summing  $\mu_\rho$  along the line  $\lambda$  in phase space yields the probability that a measurement of the PVM associated with the striation which contains  $\lambda$  has the outcome associated with  $\lambda$ .

**4.2.1. Application: entanglement characterization.** Franco and Penna (2006) relate the negativity of Wootters's discrete Wigner function to entanglement. Recall that a bipartite density matrix  $\rho$  is *separable* if it can be written as a convex combination of the form

$$\rho = \sum_k p_k \rho_k^{(1)} \otimes \rho_k^{(2)},$$

for all  $k$ , where  $\rho_k^{(1)}$  and  $\rho_k^{(2)}$  are states on the individual subsystems. Let  $\Phi_1$  and  $\Phi_2$  be the DPS associated with  $\rho_k^{(1)}$  and  $\rho_k^{(2)}$ , respectively.

The Wootters representation of a density matrix of the form  $\rho = \rho^{(1)} \otimes \rho^{(2)}$  is given by  $\mu_\rho(\alpha) = \mu_\rho^{(1)}(\alpha_1) \mu_\rho^{(2)}(\alpha_2)$ ,

where  $\alpha = (\alpha_1, \alpha_2) \in \Phi_1 \times \Phi_2$ . This can be shown as follows:

$$\begin{aligned}\mu_\rho(\alpha_1, \alpha_2) &= \frac{1}{d} \text{Tr}(\rho^{(1)} A_{\alpha_1} \otimes \rho^{(2)} A_{\alpha_2}) \\ &= \frac{1}{d} \sum_{\beta_1 \in \Phi_1, \beta_2 \in \Phi_2} \mu_\rho^{(1)}(\beta_1) \mu_\rho^{(2)}(\beta_2) \\ &\quad \times \text{Tr}(A_{\beta_1} A_{\alpha_1} \otimes A_{\beta_2} A_{\alpha_2}) \\ &= \frac{1}{d} \sum_{\beta_1 \in \Phi_1, \beta_2 \in \Phi_2} \mu_\rho^{(1)}(\beta_1) \mu_\rho^{(2)}(\beta_2) d \delta_{\beta_1 \alpha_1} \delta_{\beta_2 \alpha_2} \\ &= \mu_\rho^{(1)}(\alpha_1) \mu_\rho^{(2)}(\alpha_2).\end{aligned}$$

Thus, separability can be recast entirely in terms of the discrete phase space. That is, a discrete Wigner function is *separable* if it can be written

$$\mu_\rho(\alpha) = \sum_k p_k \mu_\rho^{(1)}(\alpha_1)_k \mu_\rho^{(2)}(\alpha_2)_k, \quad (38)$$

or else it is *entangled*.

The two qubit product state  $\mu_\rho(\alpha) = \mu_\rho^{(1)}(\alpha_1) \mu_\rho^{(2)}(\alpha_2)$  with  $\mu_\rho^{(1)}(\alpha_1) = \frac{1}{2}$  for some  $\alpha_1$  and  $\mu_\rho^{(2)}(\alpha_2) = \frac{1-\sqrt{3}}{4}$  for some  $\alpha_2$  will have the most negative value for a separable state, namely  $\frac{1-\sqrt{3}}{8}$ . Thus, if a two qubit Wigner function has a value strictly less than  $\frac{1-\sqrt{3}}{8}$ , it is entangled. Since entanglement is considered non-classical, negativity of the Wigner function (below some threshold) is associated with non-classicality. However, even if a Wigner function is positive on all of phase space, it can still be entangled. Therefore, Franco and Penna have found a new sufficient condition for entanglement in two qubits.

For a necessary condition, Franco and Penna (2006) turn to the positive partial transpose condition (Horodecki 1997, Peres 1996). The result is that a two qubit state  $\rho$  is separable if and only if both the discrete Wigner function of  $\rho$  and the discrete Wigner function of  $\rho^{T_2}$  (the partial transpose) are non-negative everywhere on the discrete phase space.

Wootters' discrete Wigner function has also found application in quantum teleportation (Koniarczyk *et al* 2001). The authors have found the discrete phase space representation of the teleportation protocol much clearer especially when considering quantum systems with much larger than two-dimensional Hilbert spaces.

#### 4.3. Extended discrete phase space

Cohendet *et al* (1988) define a discrete analog of the Wigner function which is valid for integer spin<sup>9</sup>. That is,  $\dim(\mathcal{H}) = d$  is assumed to be odd. Whereas Wootters builds up a discrete phase space before defining a Wigner function, Cohendet *et al* (1988) implicitly define a discrete phase space through the definition of their Wigner function.

Consider the operators

$$W_{mn} \phi_k = \omega^{2n(k-m)} \phi_{k-2m},$$

<sup>9</sup> This difficulty was overcome in a later paper (Cohendet *et al* 1990).

with  $m, n \in \mathbb{Z}_d$  and  $\phi_k$  are the eigenvectors of  $Z$ . Then, the *discrete Wigner function* of a density operator  $\rho$  is

$$\mu_\rho^{\text{odd}}(q, p) = \frac{1}{d} \text{Tr}(\rho W_{qp} P), \quad (39)$$

where  $P$  is the parity operator.

The authors call the operators  $\Delta_{qp} = W_{qp} P$  *Fano operators* and note that they satisfy

$$\Delta_{qp}^\dagger = \Delta_{qp},$$

$$\text{Tr}(\Delta_{qp} \Delta_{q'p'}) = d \delta_{qq'} \delta_{pp'},$$

$$W_{xk}^\dagger \Delta_{qp} W_{xk} = \Delta_{q-2x, p-2k}.$$

The Fano operators play a role similar to Wootters' phase point operators; they form a complete basis of the space of Hermitian operators. The phase space implicitly defined through the definition of the discrete Wigner function (39) is  $\mathbb{Z}_d \times \mathbb{Z}_d$ . When  $d$  is an odd prime, this phase space is equivalent to Wootters' discrete phase space. In this case the Fano operators are  $\Delta_{qp} = A_{(-q,p)}$ . This can be seen by writing the Wootters phase point operators as

$$A_{(q,p)} = \frac{1}{d} X^{2q} Z^{2p} P \omega^{2qp}.$$

Let  $\sigma \in \{\pm 1\}$ . The *extended* phase space is  $\mathbb{Z}_d \times \mathbb{Z}_d \times \{\pm 1\}$ . Define the new Wigner function

$$\mu_\rho(q, p, \sigma) = \frac{1}{4d} \left( \frac{2}{d} + \sigma \mu_\rho^{\text{odd}}(q, p) \right).$$

This function satisfies the positivity and normalization requirements of a true probability distribution.

**4.3.1. Application: Master equation for an integer spin.** In the same paper, Cohendet *et al* show the quantum dynamical equation of motion can be represented in the extended phase space as a classical stochastic process. This is achieved by showing the time evolution of the discrete Wigner function is

$$\begin{aligned}\frac{\partial}{\partial t} \mu_\rho(q, p, \sigma; t) \\ = \sum_{q', p', \sigma'} A(q, p, \sigma | q', p', \sigma') \mu_\rho(q', p', \sigma'; t),\end{aligned}$$

for a suitable choice of jump moments  $A$ . This is in the form of the master equation of a Markov process. The authors interpret this result as follows: 'Quantum mechanics of an integer spin appears as the mixture of two classical schemes of a spin. However, at random times the schemes are exchanged'.

#### 4.4. Even-dimensional discrete Wigner functions

Leonhardt (1995) defines discrete analogs of the Wigner function for both odd- and even-dimensional Hilbert spaces. In a later paper Leonhardt (1996) discusses the need for separate definitions for the odd and even dimension cases. Naively applying his definition, or that of Cohendet *et al*, of the discrete Wigner function for odd dimensions to even

dimensions yields unsatisfactory results. The reason for this is the discrete Wigner function carries redundant information for even dimensions which is insufficient to specify the state uniquely. The solution is to enlarge the phase space until the information in the phase space function becomes sufficient to specify the state uniquely.

Suppose  $\dim(\mathcal{H}) = d$  is odd. Leonhardt defines the discrete Wigner function as

$$\mu_{\rho}^{\text{Leo}}(q, p) = \frac{1}{d} \text{Tr}(\rho X^{2q} Z^{2p} P \omega^{2qp}).$$

Leonhardt's definition of an odd-dimensional discrete Wigner function is unitarily equivalent to the Cohendet *et al* definition. That is,  $\mu_{\rho}^{\text{Leo}}(q, p) = \mu_{\rho}^{\text{odd}}(-q, p)$ . To define a discrete Wigner function for even dimensions, Leonhardt takes half-integer values of  $q$  and  $p$ . This amounts to enlarging the phase space to  $\mathbb{Z}_{2d} \times \mathbb{Z}_{2d}$ . Thus the *even-dimensional* discrete Wigner function is

$$\mu_{\rho}^{\text{even}}(q, p) = \frac{1}{2d} \text{Tr}(\rho X^q Z^p P \omega^{\frac{qp}{2}}),$$

where the operators

$$\Delta_{qp}^{\text{even}} = \frac{1}{2d} X^q Z^p P \omega^{\frac{qp}{2}}$$

could be called the even-dimensional Fano or phase point operators. Of course, these operators do not satisfy all the criteria which the Fano operators (in the case of Cohendet *et al*) or the phase point operators (in the case of Wootters) satisfy; they are not orthogonal, for example. Moreover, they are not even linearly independent which can easily be inferred since there are  $4d^2$  of them and a set of linearly independent operators contains a maximum of  $d^2$  operators.

**4.4.1. Application: quantum computation.** Leonhardt's discrete Wigner function has been used to visualize and gain insights for algorithms expected to be performed on a quantum computer (Bianucci *et al* 2002, Miquel *et al* 2002a, 2002b). For each step in a quantum algorithm the state  $\rho(t)$  of the quantum computer is updated via some unitary transformation

$$\rho(t+1) = U\rho(t)U^{\dagger}.$$

This can be represented in the discrete phase space as

$$\mu_{\rho}(q, p; t) = \sum_{q'p'} Z(p, q|p', q') \mu_{\rho}(q', p'; t),$$

where  $Z$  can be easily obtained from  $U$ . This resembles the update map for the probabilities of classical stochastic variables. However, the properties of  $Z$  imply that not all admissible maps are classical; they do not connect single points in phase space and hence are 'non-local'. In Miquel *et al* (2002a) the authors identify a family of classical maps which can be efficiently implemented on a quantum computer. The authors admit that the ultimate usefulness of this approach is uncertain but speculate that the phase space representation may inspire improvement and innovation in quantum algorithms. It certainly makes for some inspiring pictures!

The Leonhardt phase space formalism has also been applied to study decoherence in quantum walks (López and Paz 2003). For large system, numerics are often employed to study the main features. The phase space method offers an intuitive and visual alternative. It allows one to visually see the quantum interference and its disappearance under decoherence. Related to these is a hybrid approach between the Wootters and Leonhardt discrete phase spaces used to analyze various aspects of quantum teleportation (Paz 2002).

#### 4.5. Finite fields discrete phase space representation

Recall that when  $\dim(\mathcal{H}) = d$  is prime, Wootters defines the discrete phase space as a  $d \times d$  lattice indexed by the group  $\mathbb{Z}_d$ . Wootters (2004) generalizes his original construction of a discrete phase space to allow the  $d \times d$  lattice to be indexed by a finite field  $\mathbb{F}_d$  which exists if and only if  $d = p^n$  is an integer power of a prime number. This approach is discussed at length in a paper (Gibbons *et al* 2004) authored by Gibbons, Hoffman and Wootters (GHW).

Similar to his earlier approach, Wootters defines the *phase space*,  $\Phi_d$ , as a  $d \times d$  array of points  $\alpha = (q, p) \in \mathbb{F}_d \times \mathbb{F}_d$ . A *line*,  $\lambda$ , is the set of  $d$  points satisfying the linear equation  $aq + bp = c$ , where all arithmetic is done in  $\mathbb{F}_d$ . Two lines are *parallel* if their linear equations differ in the value of  $c$ .

The mathematical structure of  $\mathbb{F}_d$  is appealing because lines defined as above have the following useful properties: (i) given any two points, exactly one line contains both points, (ii) given a point  $\alpha$  and a line  $\lambda$  not containing  $\alpha$ , there is exactly one line parallel to  $\lambda$  that contains  $\alpha$  and (iii) two non-parallel lines intersect at exactly one point. Note that these are usual properties of lines in Euclidean space. As before, the  $d^2$  points of the phase space  $\Phi_d$  can be partitioned into  $d+1$  sets of  $d$  parallel lines called *striations*. The line containing the point  $(q, p)$  and the origin  $(0, 0)$  is called a *ray* and consists of the points  $(sq, sp)$ , where  $s$  is a parameter taking values in  $\mathbb{F}_d$ . We choose each ray, specified by the equation  $aq + bp = 0$ , to be the representative of the striation it belongs to.

A translation in phase space,  $\mathcal{T}_{\alpha_0}$ , adds a constant vector,  $\alpha_0 = (q_0, p_0)$ , to every phase space point:  $\mathcal{T}_{\alpha_0}\alpha = \alpha + \alpha_0$ . Each line,  $\lambda$ , in a striation is invariant under a translation by any point contained in its ray, parametrized by the points  $(sq, sp)$ . That is,

$$\tau_{(sq, sp)}\lambda = \lambda. \quad (40)$$

The discrete Wigner function is

$$\mu_{\rho}^{\text{field}}(q, p) = \frac{1}{d} \text{Tr}(\rho A_{(q,p)}),$$

where now the Hermitian *phase point operators* satisfy the following properties for a projector valued function  $Q$ , called a *quantum net*, to be defined later.

GHW(4): For each point  $\alpha$ ,  $A$  is Hermitian.

GHW(5): For any two points  $\alpha$  and  $\beta$ ,  $\text{Tr}(A_{\alpha}A_{\beta}) = d\delta_{\alpha\beta}$ .

GHW(6): For any line  $\lambda$ ,  $\sum_{\alpha \in \lambda} A_{\alpha} = dQ(\lambda)$ .

The projector valued function  $Q$  assigns quantum states to lines in phase space. This mapping is required to satisfy the special property of *translational covariance*, which is defined after



a short, but necessary, mathematical digression. Notice first that properties GHW(4) and GHW(5) are identical to Woo(4) and Woo(5). Also note that if GHW(6) is to be analogous to Woo(6), the property of translation covariance must be such that the set  $\{Q(\lambda)\}$  when  $\lambda$  ranges over a striation forms a PVM.

The set of elements  $E = \{e_0, \dots, e_{n-1}\} \subset \mathbb{F}_d$  is called a *field basis* for  $\mathbb{F}_d$  if any element,  $x$ , in  $\mathbb{F}_d$  can be written

$$x = \sum_{i=0}^{n-1} x_i e_i, \quad (41)$$

where each  $x_i$  is an element of the prime field  $\mathbb{Z}_p$ . The *field trace*<sup>10</sup> of any field element is given by

$$\text{tr}(x) = \sum_{i=0}^{n-1} x^{p^i}. \quad (42)$$

There exists a unique field basis,  $\tilde{E} = \{\tilde{e}_0, \dots, \tilde{e}_{n-1}\}$ , such that  $\text{tr}(\tilde{e}_i e_j) = \delta_{ij}$ . We call  $\tilde{E}$  the *dual* of  $E$ .

The construction presented in Gibbons *et al* (2004) is physically significant for a system of  $n$  objects (called *particles*) having a  $p$ -dimensional Hilbert space. A translation operator,  $T_\alpha$  associated with a point in phase space  $\alpha = (q, p)$  must act independently on each particle in order to preserve the tensor product structure of the composite system's Hilbert space. We expand each component of the point  $\alpha$  into its field basis decomposition as in equation (41)

$$q = \sum_{i=0}^{n-1} q_i e_i \quad (43)$$

and

$$p = \sum_{i=0}^{n-1} p_i f \tilde{e}_i, \quad (44)$$

with  $f$  any element of  $\mathbb{F}_d$ . Note that the basis we choose for  $p$  is a multiple of the dual of that chosen for  $q$ . Now, the translation operator associated with the point  $(q, p)$  is

$$T_{(q,p)} = \bigotimes_{i=0}^{n-1} X^{q_i} Z^{p_i}. \quad (45)$$

Since  $X$  and  $Z$  are unitary,  $T_\alpha$  is unitary.

We assign with each line in phase space a pure quantum state. The quantum net  $Q$  is defined such that for each line,  $\lambda$ ,  $Q(\lambda)$  is the operator which projects onto the pure state associated with  $\lambda$ . As a consequence of the choice of basis for  $p$  in equation (44), the state assigned to the line  $\tau_\alpha \lambda$  is obtained through

$$Q(\tau_\alpha \lambda) = T_\alpha Q(\lambda) T_\alpha^\dagger. \quad (46)$$

This is the condition of translational covariance and it implies that each striation is associated with an orthonormal basis of the Hilbert space. To see this, recall the property in equation (40). From equation (46), this implies that, for each  $s \in \mathbb{F}_d$ ,  $T_{(sq,sp)}$

must commute with  $Q(\lambda)$ , where the line  $\lambda$  is any line in the striation defined by the ray consisting of the points  $(sq, sp)$ . That is, the states associated with the lines of the striation must be common eigenstates of the unitary translation operator  $T_{(sq,sp)}$ , for each  $s \in \mathbb{F}_d$ . Thus, the states are orthogonal and form a basis for the Hilbert space. That is, their projectors form a PVM which makes GHW(6) identical to Woo(6) when  $d$  is prime.

In Gibbons *et al* (2004), the authors note that, although the association between states and vertical and horizontal lines is fixed, the quantum net is not unique. In fact, there are  $d^{d+1}$  quantum nets which satisfy equation (40). When  $d$  is prime, one of these quantum nets corresponds exactly to the original discrete Wigner function defined by Wootters in section 4.2.

**4.5.1. Application: quantum computation.** As conjectured by Galvão (2005), the authors of Cormick *et al* (2006) have shown the only quantum states having a non-negative discrete Wigner function<sup>11</sup> are convex combinations of stabilizer states, which are simultaneous eigenstates of the generalized Pauli operators (Gottesman 1997). Working only with stabilizer states is ‘classical’ in the sense that that one can represent them with only a polynomial number of classical bits whereas an arbitrary quantum state requires an exponential number of bits (Nielsen and Chuang 2000).

Strengthening the connection between negativity and non-classicality, it was also shown that the unitary operators preserving the non-negativity of the discrete Wigner function are a subset of the Clifford group, which are those unitaries which preserve Pauli operators under a conjugate mapping. According to the Gottesman–Knill theorem, a quantum computation using only operators from the Clifford group and stabilizer states can be efficiently simulated on a classical computer (Gottesman 1997). Thus, as noted in Galvão (2005) for a particular computational model, negativity of the Wigner function is necessary for quantum computational speedup. A direct generalization of this result to continuous variables appears in Bartlett and Sanders (2002) where the authors generalize the stabilizer states to the familiar Gaussian states used in quantum optics settings.

This discrete Wigner function was also used to analyze quantum error correcting codes in Paz *et al* (2005). The aim was to gain insights and intuition for various quantum maps by studying their pictorial representation in the discrete phase space.

#### 4.6. Discrete Cahill–Glauber formalism

Ruzzi *et al* (2005) have discretized the Cahill–Glauber phase space formalism. The set of operators  $\{S(\eta, \xi)\}$ , where  $\eta, \xi \in [-l, l]$  and  $l = \frac{d-1}{2}$  ( $d$  odd), is called the *Schwinger basis* and is explicitly given by

$$S(\eta, \xi) = \frac{1}{\sqrt{d}} X^\eta Z^\xi \omega^{\frac{\eta\xi}{2}}.$$

<sup>10</sup> Note that we will distinguish the field trace,  $\text{tr}(\cdot)$ , from the usual trace of a Hilbert space operator,  $\text{Tr}(\cdot)$ , by the case of the first letter.

<sup>11</sup> Note that it is assumed the discrete Wigner function is non-negative for all definitions—that is, for all quantum nets.

These  $d^2$  operators form an orthonormal basis for the space of linear operators. In analogy with the Cahill–Glauber formalism, the basis is generalized to

$$S^{(s)}(\eta, \xi) = S(\eta, \xi) \mathcal{K}(\eta, \xi)^{(-s)},$$

where  $|s| \leq 1$  is any complex number and  $\mathcal{K}(\eta, \xi)$  is a (relatively) complicated expression of Jacobi  $\vartheta$ -functions (see the appendix of Ruzzi *et al* (2005)). Next we take the Fourier transform

$$T^{(s)}(q, p) = \frac{1}{\sqrt{d}} \sum_{\eta, \xi=-l}^l S^{(s)}(\eta, \xi) \omega^{-(\eta q + \xi p)}.$$

The set operators  $\{T^{(s)}(q, p)\}$  is the discrete analog of the  $s$ -ordered mapping kernel of the Cahill–Glauber formalism. Moreover, the authors of Ruzzi *et al* (2005) have shown that the continuous limit of this set is indeed the Cahill–Glauber mapping kernel.

Suppose  $s$  is real. Then, the operators  $\{T^{(s)}(q, p)\}$  enjoy the following familiar properties:

$$T^{(s)}(q, p)^\dagger = T^{(s)}(q, p),$$

$$\text{Tr}(T^{(s)}(q, p) T^{(-s)}(q', p')) = d \delta_{qq'} \delta_{pp'}.$$

Thus, similarly to the discrete kernel of Heiss and Weigert,  $\{T^{(s)}(q, p)\}$  and  $\{T^{(-s)}(q, p)\}$  are dual bases for the space of Hermitian operators.

In the now familiar way, we can define a quasi-probability function on the  $(q, p)$  phase space as

$$\mu_\rho^{(s)}(q, p) = \text{Tr}(T^{(s)}(q, p) \rho). \quad (47)$$

Cahill and Glauber showed, for their  $s$ -ordered formalism, that  $s = 0$  corresponds to the Wigner function;  $s = 1$  corresponds to the Husimi function; and  $s = -1$  corresponds to the Glauber–Sudarshan function. Using equation (47), we call, for example, the function obtained when  $s = 0$  the discrete Wigner function.

**4.6.1. Application: quantum teleportation.** Marchioli *et al* (2005) have applied this formalism to quantum tomography and teleportation. The teleportation protocol was analyzed for arbitrary  $s$  but, for brevity, we will consider the  $s = 0$  case (which is now assumed so the superscript can be ignored). The teleportation protocol utilizes entanglement to transfer a quantum state between two parties through the exchange of only a small amount of classical information (Nielsen and Chuang 2000). Consider the tripartite system

$$\rho = \rho^{(1)} \otimes \rho^{(2,3)},$$

where one party possesses subsystems 1 and 2 and the other possesses subsystem 3. The goal is for  $\rho^{(1)}$  to be transferred from subsystem 1 to subsystem 3 without simply swapping them. It is essential that the shared state  $\rho^{(2,3)}$  be entangled. In particular, assume it is a maximally entangled pure Bell-state (Nielsen and Chuang 2000). We choose, following Wootters (1987), to construct the global phase space to be a Cartesian

product of the phase spaces of the individual subsystems. The discrete Wigner function of the whole system is then

$$\begin{aligned} \mu_\rho(q_1, q_2, q_3; p_1, p_2, p_3) \\ = \mu_{\rho^{(1)}}(q_1, p_1) \mu_{\rho^{(2,3)}}(q_2, q_3; p_2, p_3). \end{aligned}$$

A Bell-measurement is performed on the first two subsystems, which in phase space is interpreted as a measurement of the total momentum and relative coordinate of the subsystem composed of subsystems 1 and 2. Marginalizing over subsystems 1 and 2 gives

$$\mu_{\rho^{(3)}}(q_3, p_3) = \mu_{\rho^{(1)}}(q_3 - \alpha, p_3 + \beta),$$

where  $\alpha$  and  $\beta$  parametrize the result of the Bell-measurement (note  $\rho^{(3)}$  can be identified as the reduced state of subsystem 3). Thus, the final state of subsystem 3 is simply a displacement in the phase space and communicating only the measurement result  $(\alpha, \beta)$  leads to the recovery of the initial state.

The discrete Husimi function ( $s = 1$ ) was used to define a discrete analog of squeezed states (Marchioli *et al* 2007) and to analyze spin tunneling effects in a particular toy model of interacting fermions (Marchioli *et al* 2009).

#### 4.7. Probability tables

In 1986, before introducing the discrete Wigner function, Wootters represented the quantum state as a ‘probability table’ which was simply a list of outcome probabilities for a complete set of measurements (Wootters 1986). The complete set of interest was that of *mutually unbiased bases* (MUBs). We call  $n$  bases  $\{\psi_k^n\}$  mutually unbiased if they satisfy

$$|\langle \psi_{k'}^{n'}, \psi_k^n \rangle|^2 = \delta_{kk'} \delta_{nn'} + \frac{1}{d} (1 - \delta_{nn'}). \quad (48)$$

Wootters noted for  $d$  prime, a set of  $n = d + 1$  MUBs could be explicitly constructed via a prescription in Ivonovic (1981). Wootters also posed many questions of MUBs, some of which have now been answered. It is now known that for any dimension  $3 \leq n \leq d + 1$ , where the upper bound can be achieved, by construction, for any dimension which is a power of a prime<sup>12</sup>.

Here we will consider the case when  $d$  is prime and all probability tables for non-prime dimensions can be built up from those for their prime factors, in much the same way as was done in section 4.2 for the discrete phase spaces.

Consider the generalized Pauli operator  $Z$  and its eigenbasis  $\{\phi_k\}$  and the projectors onto these vectors  $P_k := \phi_k \phi_k^*$ . Define the finite Fourier transform

$$F = \frac{1}{\sqrt{d}} \sum_{k, k'=0}^{d-1} \omega^{kk'} \phi_k \phi_{k'}^* \quad (49)$$

and the operator

$$V = \sum_{k=0}^{d-1} \omega^{\frac{k^2}{2}} F P_k F^\dagger. \quad (50)$$

<sup>12</sup> For a recent review of the MUB problem see Combes *et al* (2006).

Here, as before, division by two represents the multiplicative inverse of the element 2. For Hilbert space dimension  $d = 2$ , this operator requires the special definition

$$V = \frac{1}{2} \begin{pmatrix} 1+i & 1-i \\ 1+i & 1-i \end{pmatrix}. \quad (51)$$

Now we can construct  $d + 1$  MUBs via

$$\begin{aligned} \psi_k^0 &= \phi_k, \\ \psi_k^n &= V^n \phi_k, \quad n = 1, \dots, d. \end{aligned}$$

We will denote the projectors onto these basis vectors  $P(n, k) := \psi_k^n \psi_k^{n*}$ . Then the probability of obtaining the  $k$ th outcome when measuring in the  $n$ th basis is

$$\mu_\rho(n, k) = \text{Tr}(\rho P(n, k)). \quad (52)$$

This can be viewed as a matrix in which the columns index the measurements while the rows index the outcomes. This can also be viewed as a mapping whose inverse is given by

$$\rho = \sum_{n=0}^d \sum_{k=0}^{d-1} \mu_\rho(n, k) P(n, k) - \mathbb{1}. \quad (53)$$

#### 4.7.1. Application: quantum mechanics without amplitudes.

The purpose of Wootters (1986) was not to introduce a new representation of the quantum state *per se*, but to show that the whole of the operational formalism of quantum mechanics can be done rather simply without complex numbers.

Wootters notes first:

‘It is obviously possible to devise a formulation of quantum mechanics without probability amplitudes. One is never forced to use any quantities in one’s theory other than the raw results of measurements. However, there is no reason to expect such a formulation to be anything other than extremely ugly.’

To our surprise, the rule for transitioning between the probability tables turn out to be remarkably simple. In quantum theory the transition probability from state  $\rho$  to  $\rho'$  is the probability of preparing the state  $\rho$ , performing the measurement  $\{\rho', \mathbb{1} - \rho'\}$  and obtaining  $\rho'$ . This transition probability is  $\text{Pr}(\rho \rightarrow \rho') = \text{Tr}(\rho \rho')$ . If we work with the probability tables and call the tables  $\mu$  and  $\mu'$ , Wootters obtains

$$\text{Pr}(\mu \rightarrow \mu') = \sum_{n=0}^d \sum_{k=0}^{d-1} \mu(n, k) \mu'(n, k) - 1. \quad (54)$$

Unfortunately, as Wootters notes, it is not easy to ignore the density matrix altogether. We have yet to specify which probability tables are valid and which do not correspond to quantum states. The simplest characterization of valid probability tables is to say those for which equation (53) is a unit trace positive semi-definite matrix. This is unsatisfying as we would like a characterization independent of the density matrix.

#### 4.8. Hardy’s vector representation

Hardy (2001a) showed that five axioms are sufficient to imply a special vector representation which is equivalent to an operational form of quantum theory. We first describe the vector representation.

Consider a basis for a  $d$ -dimensional Hilbert space  $\{\phi_k\}$  (the eigenbasis of  $Z$ , say) and the following set of  $d^2$  projectors:

$$P_{kj} := \begin{cases} \phi_k \phi_k^* & \text{if } k = j, \\ (\phi_k + \phi_j)(\phi_k + \phi_j)^* & \text{if } k < j, \\ (\phi_k + i\phi_j)(\phi_k + i\phi_j)^* & \text{if } k > j. \end{cases} \quad (55)$$

These projectors span the space of linear operators on the Hilbert space spanned by  $\{\phi_k\}$ . Now we vectorize by choosing an arbitrary but fixed ordering convention. For definiteness, we choose to stack the rows on top of one another. To this end, define  $\alpha := dk + j$  and  $P(\alpha) := P_{kj}$ . Then, the vector representation of the state  $\rho$  is given by

$$\mu_\rho(\alpha) = \text{Tr}(\rho P(\alpha)). \quad (56)$$

Now the outcome of any quantum measurement can be assigned a positive operator  $E$ . Call this ‘outcome  $E$ ’. Define the vector  $\xi_E(\alpha)$  implicitly through

$$E = \sum_{\alpha} \xi_E(\alpha) P(\alpha).$$

Then, the probability of ‘outcome  $E$ ’ is given by

$$\text{Pr}(\text{‘outcome } E\text{’}) = \sum_{\alpha} \xi_E(\alpha) \mu_\rho(\alpha),$$

which, in vector notation, we can write as the dot product  $\vec{\xi} \cdot \vec{\mu}$ .

We define the sets  $M$  and  $\Xi$  as the set of vectors obtainable through the mappings  $\rho \mapsto \mu$  and  $E \mapsto \xi$  defined above. More precise statements, in the form of inequalities, which make no recourse to the usual quantum mechanical objects, can be made to define these sets. Assuming this has been done, we can rephrase the axioms of quantum mechanics, without mention of Hermitian operators and the like, in this vector representation succinctly as follows: states are represented by vectors  $\vec{\mu} \in M$ ; measurement outcomes are represented by vectors  $\vec{\xi} \in \Xi$ ; the probability of ‘outcome  $\vec{\xi}$ ’ in state  $\vec{\mu}$  is given by  $\text{Pr}(\text{‘outcome } \vec{\xi}\text{’}) = \vec{\xi} \cdot \vec{\mu}$ .

**4.8.1. Application: quantum axiomatics.** As was the case in the previous section, this vector representation was not introduced as such. Hardy (2001a, 2001b) has shown that five axioms are sufficient to imply the real vector formalism of quantum mechanics. The frequency interpretation of probability was given its own axiom. However, if we take our everyday intuitive notion of probability (Porta Mana 2007), we no longer require this first axiom, which is independent of the rest (Schack 2003).

We will make use of the following definitions:

- The number of *degrees of freedom*,  $K$ , is defined as the minimum number of yes-no measurements whose outcome probabilities are needed to determine the state (of belief in the mind of a reasonable agent), or, more roughly, as the number of real parameters required to specify the state.

- The *dimension*,  $d$ , is defined as the maximum number of states that can be reliably distinguished from one another in a single shot measurement.

Axiom 1 defines probability as limiting frequencies and is not required (Schack 2003). The remainder of the axioms are as follows:

- Axiom 2: *Simplicity*.  $K$  is determined by a function of  $N$  (i.e.  $K = K(d)$ ) where  $d = 1, 2, \dots$  and where, for each given  $d$ ,  $K$  takes the minimum value consistent with the axioms.
- Axiom 3: *Subspaces*. A system whose state is constrained to belong to an  $n$ -dimensional subspace (i.e. have support on only  $n$  of a set of  $d$  possible distinguishable states) behaves like a system of dimension  $n$ .
- Axiom 4: *Composite systems*. A composite system consisting of subsystems  $A$  and  $B$  satisfies  $d = d_A d_B$  and  $K = K_A K_B$ .
- Axiom 5: *Continuity*. There exists a continuous reversible transformation on a system between any two pure states of that system.

These four axioms are sufficient for a derivation of the vector representation of quantum theory defined above. This axiomatization is also important for contrasting quantum theory with classical probability theory. As Hardy has shown, discrete classical probability theory (of dice, coins and so on) can be derived from only axioms 2, 3 and 4. That is, the only difference between quantum and classical theory is the existence of a *continuous* reversible transformation between pure states.

#### 4.9. The real density matrix

Havel (2002) defined the ‘real density matrix’ which, not surprisingly, is a particular real-valued matrix representation of the quantum state.

For  $d = 2$ , define the  $2 \times 2$  matrix of Pauli operators as

$$P = \begin{pmatrix} \mathbb{1} & X \\ Y & Z \end{pmatrix}. \quad (57)$$

For  $d = 2^n$ , denote the bits in the binary expansion of  $k$  as

$$k = \sum_{a=1}^n k_a \times 2^{n-a},$$

and similarly for  $j$ . Then, the  $d \times d$  matrix of Pauli operators is given by

$$P_{kj} = P_{k_1 j_1} \otimes \dots \otimes P_{k_n j_n}. \quad (58)$$

These  $d^2$  operators are orthogonal, and hence form a basis for the space of linear operators on the  $d$ -dimensional Hilbert space. Therefore, each density matrix can be expressed as

$$\rho = \frac{1}{d} \sum_{k,j=0}^{d-1} \sigma_{kj} P_{kj},$$

where the coefficients  $\sigma_{kj}$ , are explicitly given by

$$\sigma_{kj} = \text{Tr}(\rho P_{kj}), \quad (59)$$

form the *real density matrix*.

**4.9.1. Application: NMR pedagogy** Since the observables measured in NMR experiments are elements of the matrix of Pauli operators (58), the elements of the real density matrix are the experimentally measurable values. There is no need to reconstruct the density matrix. This is also a convenient fix to the problem of reporting or visualizing a quantum state. Since the density matrix contains  $d^2$  complex values, it is often graphically displayed as two  $d \times d$  matrices of the real and imaginary parts. Not only is this redundant, it is conceptually awkward. On the other hand, the real density matrix can be displayed as a single  $d \times d$  matrix of real values. Havel offers the real density matrix as a useful teaching device in such situations.

#### 4.10. Symmetric representations

Consider the unitary group

$$U_{(p,q)} = \omega^{\frac{pq}{2}} X^p Z^q, \quad (60)$$

where  $(p, q) \in \mathbb{Z}_d \times \mathbb{Z}_d$ . In Renes *et al* (2004), the authors conjecture<sup>13</sup> that the set  $\{U_{(p,q)}\phi\}$  for some fiducial  $\phi \in \mathcal{H}$  forms a *symmetric informationally complete positive operator valued measure* (SIC-POVM). The defining condition of a SIC-POVM is a set of  $d^2$  vectors  $\{\phi_k\}$  such that

$$|\langle \phi_k, \phi_j \rangle|^2 = \frac{\delta_{kj} d + 1}{d + 1}. \quad (61)$$

The set is called symmetric since the vectors have equal overlap. The POVM is formed by taking the projectors onto the one-dimensional subspaces spanned by the vectors. It is informationally complete since these  $d^2$  projectors span the space of linear operators acting on  $\mathcal{H}$ .

As of writing, it is an open question whether SIC-POVMs exist in every dimension; although numerical evidence suggests this to be the case (Scott and Grassl 2010).

For the remainder of this section we assume, for any dimension  $d$ , a SIC-POVM exists. Define the operators  $P_k := \frac{1}{d} \phi_k \phi_k^*$ . Then, define the *symmetric-representation* of a quantum state  $\rho$  as

$$\mu_\rho(k) = \text{Tr}(\rho P_k). \quad (62)$$

This is a probability distribution and in particular it is *the* probability distribution for the POVM measurement formed by the effects  $\{P_k\}$ . As we have noted, this is an informationally complete measurement. Therefore, the density matrix can be reconstructed from the probabilities via

$$\rho = d(d+1) \sum_{k=0}^{d^2-1} \mu_\rho(k) P_k - \mathbb{1}. \quad (63)$$

When viewed as a mapping, this representation is a bijection from the convex set of density matrices to a convex subset of the  $d^2$ -dimensional probability simplex.

<sup>13</sup> Apparently this was conjectured earlier by Zauner in a *PhD Thesis* not available in English. See Zauner (1999) and [www.imaph.tu-bs.de/qi/problems/23.html](http://www.imaph.tu-bs.de/qi/problems/23.html).



**4.10.1. Application: Quantum Bayesianism.** Quantum Bayesianism (Caves *et al* 2002a, Fuchs 2002, Fuchs 2010, Fuchs and Schack 2009, Schack *et al* 2001) is an interpretation of quantum theory which sheds new light on not only the tradition ‘foundational’ problems (the ‘measurement problem’, for example) but also many concepts in quantum theory, such as the ‘unknown quantum state’ (Caves *et al* 2002b). A key realization is the mathematical and conceptual sufficiency of viewing quantum states as the probability distribution via the Born rule for a fixed POVM  $\{E_k\}$ . The only remaining freedom is which one.

One ideal is to have the POVM elements orthogonal:  $\text{Tr}(P_k P_j) = \delta_{kj}$ . The statement that is not possible is equivalent to theorem 1. Next, then, we desire them to be as close to orthogonal as possible. Formally, we want to minimize the quantity

$$F = \sum_{kj} (\text{Tr}(P_k P_j) - \delta_{kj})^2.$$

This expression is minimized if and only if the  $\{P_k\}$  form a SIC-POVM. Using the reconstruction formula in equation (63) it can be shown that, in terms of this SIC-representation, the Born rule for a measurement  $\{E_j\}$  given state  $\mu(k)$  is

$$\text{Pr}(\text{outcome } j) = \sum_k \left( \mu(k) - \frac{1}{d} \right) \xi(j|k), \quad (64)$$

where  $\xi(j|k) = \text{Tr}(E_j P_k)$ . In the same sense as the SIC-POVM being as close to orthogonal as possible, equation (64) is as close as possible to the classical law of total probability. Effort is being made to use equation (64) as a starting point for a natural set of axioms which would single out quantum theory.

## 5. Unification of the quasi-probability representations via frames

### 5.1. Introduction to frames

A *frame* can be thought of as a generalization of an orthonormal basis (Christensen 2003). However, the particular Hilbert space under consideration here is not  $\mathcal{H}$ . Considered here is a generalization of a basis for  $\mathbb{H}(\mathcal{H})$ , which is the set of Hermitian operators on a complex Hilbert space of dimension  $d$ . With the trace inner product  $\langle A, B \rangle := \text{Tr}(AB)$ ,  $\mathbb{H}(\mathcal{H})$  forms a *real* Hilbert space itself of dimension  $d^2$ . Let  $\Lambda$  be some set of cardinality  $d^2 \leq |\Lambda| < \infty$ .

A frame<sup>14</sup> for  $\mathbb{H}(\mathcal{H})$  is a set of operators  $\mathcal{F} := \{F(\lambda)\} \subset \mathbb{H}(\mathcal{H})$  which satisfies

$$a\|A\|^2 \leq \sum_{\lambda \in \Lambda} \text{Tr}[F(\lambda)A]^2 \leq b\|A\|^2, \quad (65)$$

for all  $A \in \mathbb{H}(\mathcal{H})$  and some constants  $a, b > 0$ . This definition generalizes a defining condition for an orthogonal basis  $\{B_k\}_{k=1}^{d^2}$

$$\sum_{k=1}^{d^2} \text{Tr}[B_k A]^2 = \|A\|^2, \quad (66)$$

for all  $A \in \mathbb{H}(\mathcal{H})$ . The mapping  $A \mapsto \text{Tr}[F(\lambda)A]$  is called a *frame representation* of  $\mathbb{H}(\mathcal{H})$ .

A frame  $\mathcal{D} := \{D(\lambda)\}$  which satisfies

$$A = \sum_{\lambda \in \Lambda} \text{Tr}[F(\lambda)A] D(\lambda), \quad (67)$$

for all  $A \in \mathbb{H}(\mathcal{H})$ , is a *dual frame* (to  $\mathcal{F}$ ). The *frame operator* associated with the frame  $\mathcal{F}$  is defined as

$$S(A) := \sum_{\lambda \in \Lambda} \text{Tr}[F(\lambda)A] F(\lambda).$$

If the frame operator is proportional to the identity *superoperator*,  $S = a\mathbb{I}$ , the frame is called *tight*. The frame operator is invertible and thus every operator has a representation

$$A = S^{-1}SA = \sum_{\lambda \in \Lambda} \text{Tr}[F(\lambda)A] S^{-1}F(\lambda). \quad (68)$$

The frame  $S^{-1}\mathcal{F}$  is called the *canonical dual frame*. When  $|\Lambda| = d^2$ , the canonical dual frame is the unique dual, otherwise there are infinitely many choices for a dual frame. A tight frame is ideal from the perspective that its canonical dual is proportional to the frame itself. Hence, the reconstruction is given by the convenient formula

$$A = S^{-1}SA = \frac{1}{a} \sum_{\lambda \in \Lambda} \text{Tr}[F(\lambda)A] F(\lambda)$$

which is to be compared with

$$A = \sum_{k=1}^{d^2} \text{Tr}[B_k A] B_k$$

which defines  $\{B_k\}_{k=1}^{d^2}$  as an orthonormal basis.

### 5.2. Unification via the necessity and sufficiency of frames

Recalling the formal definition (3) of a quasi-probability representation, we have the following theorem.

**Theorem 3.** *Two functions  $\mu$  and  $\xi$  constitute a quasi-probability representation if and only if*

$$\begin{aligned} \mu_\rho(\lambda) &= \text{Tr}[\rho F(\lambda)], \\ \xi_E(\lambda) &= \text{Tr}[E D(\lambda)], \end{aligned}$$

where  $\{F(\lambda)\}$  is a frame and  $\{D(\lambda)\}$  is one of its duals.

This was proven in Ferrie and Emerson (2009)<sup>15</sup>. This theorem allows us to make the following statement which is equivalent to the no-classical-representation theorem (1) and negativity theorem (2): there does not exist two frames of positive operators which are dual to each other.

With this result we can create quasi-probability representations of the whole operational formalism of quantum theory, not just states. First, we choose one of the discrete quasi-probability functions described in section 4. Second, we identify the frame which gives rise to it. Lastly, we compute its dual frame to obtain the part of the quasi-probability

<sup>14</sup> Frames have been considered in the context of quantum theory for other purposes in Bisio *et al* (2009) and Scott (2006).

<sup>15</sup> Compare this with a similar result in a more operational setting in Porta Mana (2004a, 2004b).

representation mapping,  $\xi$ , which takes measurements to functions on the space  $\Lambda$ .

Suppose instead the functions  $\mu$  and  $\xi$  are defined via

$$\mu_\rho(\lambda) = \text{Tr}[\rho F(\lambda)],$$

$$\xi_E(\lambda) = \text{Tr}[E F(\lambda)],$$

where  $\{F(\lambda)\}$  is a frame. Then,

- (a)  $\mu_\rho(\lambda) \in [0, 1]$  and  $\sum_\lambda \mu_\rho(\lambda) = 1$ ,
- (b)  $\xi_E(\lambda) \in [0, 1]$  and  $\xi_{\mathbb{1}}(\lambda) = 1$ ,
- (c)  $\text{Tr}(\rho E) = \sum_{\lambda, \lambda'} \mu_\rho(\lambda) \xi_E(\lambda') \text{Tr}[D(\lambda) D(\lambda')]$ .

In Ferrie and Emerson (2008) this representation was called a *deformed* probability representation since states and measurements are represented as true probabilities but the law of total probability is deformed.

The frame formalism also provides a convenient transformation matrix to map between representations. We have

$$\begin{aligned} \mu_\rho(\lambda) &= \text{Tr}[\rho F(\lambda)] \\ &= \sum_{\lambda'} \text{Tr}[\rho F'(\lambda')] \text{Tr}[D'(\lambda') F(\lambda)] \\ &= \sum_{\lambda'} T_{\lambda\lambda'} \mu'(\lambda'), \end{aligned}$$

where the matrix  $T_{\lambda\lambda'}$  is the symmetric matrix which takes the  $\mu$  representation to the  $\mu'$  representation.

### 5.3. To infinity and beyond

Given a quasi-probability representation note that the frame satisfies

$$\sum_{\lambda \in \Lambda} F(\lambda) = \mathbb{1}.$$

Thus, if the quasi-probability representation satisfies  $0 \leq \mu(\lambda) \leq 1$ , the frame is an informationally complete positive operator valued measure (IC-POVM)<sup>16</sup>. Similarly, the dual frame satisfies

$$\text{Tr}[D(\lambda)] = 1,$$

for all  $\lambda \in \Lambda$ . Thus, if the quasi-probability representation satisfies  $0 \leq \xi(\lambda) \leq 1$ , the dual frame is a set of density operators. The definitions and results we have so far considered are tailored to the case  $d < \infty$  – that is, finite-dimensional quantum theory. Now we will extend them to infinite dimensions as was done in Ferrie *et al* (2010).

We now allow the dimension of the Hilbert space  $\mathcal{H}$  to be arbitrary and let  $(\Omega, \Sigma)$  be a measurable space, where  $\Sigma$  is a  $\sigma$ -algebra. Over this space,  $\mathcal{M}_{\mathbb{R}}(\Omega, \Sigma)$  denotes the bounded signed measures while  $\mathcal{F}_{\mathbb{R}}(\Omega, \Sigma)$  denotes the bounded measurable functions. A signed measure generalizes the usual notion of measure to allow negative values. The classical states are the probability measures  $\mathcal{S}(\Omega, \Sigma) \subset \mathcal{M}_{\mathbb{R}}(\Omega, \Sigma)$ . We define the generalization of a frame to the space of trace-class operators  $\mathcal{T}_s(\mathcal{H})$ .

<sup>16</sup> Frames have also been used in the definition of informational completeness in the context of tomography in Bisio *et al* (2009).

We generalize the definition of informationally complete observable for infinite dimensions and define an *operator valued measure* as a map  $F : \Sigma \rightarrow \mathcal{B}_s(\mathcal{H})$  satisfying  $F(\emptyset) = 0$ ,  $F(\Omega) = 1$  and

$$F\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} B_i,$$

where the sets  $B_i \in \Sigma$  are mutually disjoint and the sum converges in the weak sense. A *frame* for  $\mathcal{T}_s(\mathcal{H})$  is an operator valued measure  $F$  for which the map  $T : \mathcal{T}_s(\mathcal{H}) \rightarrow \mathcal{M}_{\mathbb{R}}(\Omega, \Sigma)$ ,

$$T(W)(B) := \text{Tr}(WF(B)),$$

is injective. The map  $T$  is called a *frame representation* of  $\mathcal{T}_s(\mathcal{H})$ .

Similarly, generalizing the reconstruction formula defining the dual frame for finite dimensions yields the generalized notion of a dual. That is, given a frame  $F$ , a *dual frame* to  $F$  is a map  $D : \Omega \rightarrow \mathcal{B}_s(\mathcal{H})^*$ , the dual space, for which the function

$$(SA)(\omega) := (D(\omega))(A)$$

is measurable and satisfies

$$A = \int_{\Omega} SA \, dF, \quad (69)$$

for all  $A \in \mathcal{B}_s(\mathcal{H})$ .

On the other hand, we define a quasi-probability representation of quantum mechanics as a pair of mappings  $T : \mathcal{S}(\mathcal{H}) \rightarrow \mathcal{M}_{\mathbb{R}}(\Omega, \Sigma)$  and  $S : \mathcal{E}(\mathcal{H}) \rightarrow \mathcal{F}_{\mathbb{R}}(\Omega, \Sigma)$  such that

- (1)  $T$  and  $S$  are affine,  $T$  is bounded.
- (2)  $T\rho(\Omega) = 1$ .
- (3)  $S(0) = 0$ .
- (4) For all  $\rho \in \mathcal{S}(\mathcal{H})$  and  $E \in \mathcal{E}(\mathcal{H})$ ,

$$\text{Tr}(\rho E) = \int_{\Omega} (SE) \, d(T\rho). \quad (70)$$

In Ferrie *et al* (2010) it was shown that for these generalized definitions frame representations and quasi-probability representations remain equivalent and the following ‘negativity theorem’ was proven.

**Theorem 4.** *A quasi-probability representation of quantum mechanics must have, for some  $\rho \in \mathcal{S}(\mathcal{H})$ ,  $E \in \mathcal{E}(\mathcal{H})$  either  $(T\rho)(B) < 0$  for some  $B \in \Sigma$  or  $(SE)(\omega) \notin [0, 1]$  for some  $\omega \in \Omega$ .*

Being more general, this theorem implies the previous three equivalent ‘negativity theorems’ and says essentially the same; a classical representation does not exist and within a quasi-probability representation negativity must appear in the representation of the states or measurements or both.

## 6. Negativity in quasi-probability representation

There have been a variety of approaches to the problem of characterizing what is non-classical about quantum theory. In the previous section one such notion of non-classicality was considered: the requirement of ‘negative probability’, or simply *negativity*. However, the negativity theorem leaves open the question of the interpretation of negativity in any *particular* representation. In section 6.1 below we discuss the ways in which negativity can be applied as a criterion for quantumness with respect to particular choices of representation. In section 6.2 we discuss the traditional ideas of *contextuality* and *non-locality* in relation to negativity<sup>17</sup>.

### 6.1. Negativity as an indicator of quantumness

There is a strong tradition in physics of considering negativity of the Wigner function as an indicator of non-classical features of quantum states. The non-classical features attributed to negativity of the *original* Wigner function include quantum non-locality (Bell 2004, Kalev *et al* 2009), quantum chaos (Paz *et al* 1993) and quantum coherence (Paz *et al* 1993). In quantum optics, however, tradition has been to use the  $Q$  and  $P$  functions (section 3.2) to define quantumness (Mandel 1986). Recall equation (17), where the  $P$ -function of  $\rho$  was defined implicitly through

$$\rho = \int d^2\alpha P(\alpha) |\alpha\rangle\langle\alpha|.$$

If  $P$  has the properties of a probability distribution, then the state is a mixture of coherent states. Coherent states are minimum uncertainty states and this fact is often cited when it is stated that such a state is ‘the most classical’ of the quantum states of light. More specifically, if  $P$  is a probability distribution then the quantum field cannot display genuine quantum optical effects and can be simulated by a stochastic classical electromagnetic field (Walls and Milburn 1995). Technically, however,  $P$  is not a function but a distribution which can be highly singular. Thus  $P$  functions which are not classical distributions are difficult to experimentally prepare and verify; although, recent progress has been made (Kiesel *et al* 2008).

Effort has been extended beyond qualitatively defining negativity as quantumness to *quantifying* quantumness via negativity. In terms of the Wigner function, the *volume* of the negative parts of the represented quantum state has been suggested as the appropriate measure of quantumness (Kenfack and Życzkowski 2004). The *distance* (in some preferred norm on  $\mathbb{H}(\mathcal{H})$ ) to the convex subset of positive Wigner functions was suggested to quantify quantumness in Mari *et al* (2010). This was also done in Giraud *et al* (2008, 2010) for finite analogs of the  $P$ - and  $Q$ -functions rather than the Wigner function.

The main difficulty with interpreting negativity in a particular quasi-probability representation as a criterion for

or definition of quantumness is the non-uniqueness of that particular quasi-probability representation. We can always find a new representation in which any given state admits a non-negative quasi-probability representation. Recall, in fact, that in some representations all states are non-negative. Thus, negativity of some state  $\rho$  in one particular arbitrary representation is a meaningless notion of quantumness *per se*.

An alternative approach to establishing a connection between quantumness and negativity is to start by assuming some criterion for quantumness and then finding a choice of representation in which this criterion is expressed via negativity. This approach has been applied in the context of multi-partite systems for which entanglement is presumed to provide a criterion for quantumness. Entanglement is a kind of correlation between two quantum systems which cannot be achieved for classical variables and is one of the central ingredients in quantum information theory<sup>18</sup>. Recall that a density matrix  $\rho$  is *entangled* if it *cannot* be written as a convex combination of the form  $\rho = \sum_k p_k \rho_k^{(1)} \otimes \rho_k^{(2)}$ , for all  $k$ , where  $\rho_k^{(1)}$  and  $\rho_k^{(2)}$  are states on the individual subsystems. Consider a *product-state frame* constructed out of frames for two subsystems. That is, consider the frame  $\{F^{(1)}(\lambda) \otimes F^{(2)}(\lambda')\}$ , where each  $\{F^{(j)}(\lambda)\}$  is a set of density matrices composing a frame. Then, if we represent a quantum state using the dual frame, we have a quasi-probability representation in which states with negativity are entangled. Explicit constructions of such quasi-probability representations were developed by Schack and Caves (2000). An optimization procedure to find the representation with the *minimum* amount of negativity was given in Sperling and Vogel (2009).

An obvious limitation with the above approach is that entanglement cannot capture any notion of quantumness for single quantum systems. A second, more subtle, issue is that identification of entanglement as ‘the’ crucial non-classical resources is problematic in certain branches of quantum information science. The most striking example questioning the role of entanglement in quantum information theory is DQC1 (deterministic quantum computing with one clean qubit). DQC1 (Knill and Laflamme 1998) is a model of computation which refers to any algorithm which satisfies the following (or a modification not requiring exponentially more resources)<sup>19</sup>:

- (1) its input consists of a single pure state in the first *control* register and the remaining  $n$  registers are in the maximally mixed state  $\rho = 2^{-n}\mathbb{1}$ ;
- (2) the input state is subjected to a unitary  $U_n$  controlled by the state of the first register;
- (3) the output is a statistical estimate of  $2^{-n}\text{Tr}(U_n)$  (achieved by measuring the average of control bit in the  $Z$  basis).

DQC1 appears to be a non-trivial computational model which has been shown to have exponential advantages over (known) classical algorithms in the following areas: simulation of

<sup>18</sup> For a recent review of entanglement, see Horodecki *et al* (2009).

<sup>19</sup> This model of computation has served as the basis for various definitions of complexity classes, also called DQC1. There are many open questions in this line of research and the interested reader should consult Ambainis *et al* (2006), Shepherd (2006) or Shor and Jordan (2008).

<sup>17</sup> These are diverse and rich fields of study in their own right. A starting point for the interested reader on contextuality is Held (2009) and on quantum non-locality is Redhead (1989).

quantum systems (Knill and Laflamme 1998), quadratically signed weight enumerators (Knill and Laflamme 2001), evaluating the local density of states (Emerson *et al* 2004), estimating the average fidelity decay under quantum maps (Poulin *et al* 2004) and estimating the value of Jones polynomials (Shor and Jordan 2008).

In the DQC1 model, the bipartite split between the control qubit and the rest contains no entanglement and in Datta *et al* (2005) it was shown that there is a vanishingly small amount of entanglement across any other bipartite splitting. This suggests it is unlikely that entanglement is responsible for the speedup provided by DQC1 (Datta 2008). Conceptually, computation is a local task with complex dynamics and may not require the non-local, Bell-inequality-violating correlations of entanglement (Laflamme *et al* 2002). A sentiment issued in Laflamme *et al* (2002) and reiterated recently by Vedral (2010) is that no one single criteria can capture quantumness and perhaps even the resources necessary for the quantum advantage must be studied on a case-by-case basis.

An important consideration for all of the above approaches is that the notion of a quantum state, considered in isolation, is operationally meaningless. Comparison with experiment always requires specifying both a state (a preparation procedure) and a measurement. Consider two experiments, one of which prepares a product state and measures the state by projecting onto an entangled basis, and a second which prepares an entangled state and measures that state in a product basis. Both experiments produce the same statistical predictions, but only the second is considered non-classical when considering the state in isolation. As emphasized in Ferrie and Emerson (2008) and Spekkens (2008) we can overcome this obvious deficiency if we consider the whole operational setup—states *and* measurements. In this way, the existence of a positive quasi-probability representation implies the existence of a non-contextual ontological model and vice versa.

The formalism of Ferrie and Emerson (2008, 2009) and Ferrie *et al* (2010) shows the necessity of negativity when considering a representation of the full quantum formalism. That is, the negativity theorem (theorem 2) applies to quasi-probability representations of quantum theory as a whole. However, the negativity theorem may not apply if we consider a specific experiment, device, or protocol which may not faithfully reproduce the full power of quantum theory. This work suggests the following promising approach: define a classical representation of an experiment as the existence of a frame and dual for which the convex hull of the experimentally accessible states and measurements have positive representation. Then, we can conclude that negativity, taken to mean the absence of any representation satisfying the above conditions, corresponds to quantumness.

The above criterion for classicality was considered in Schack and Caves (1999) to question the quantum nature of proposed NMR quantum computers. However, as noted there, the immediate objection is the following: the states and measurements can be represented by classical probabilities while the transformation between them may not be represented by classical stochastic maps. That is, a truly classical model

must represent each applied transformation in a experiment as a classical stochastic mapping. In Schack and Caves (1999), such stochastic maps were identified for the set of NMR experiments reported at the time<sup>20</sup>.

The scope of quantum theory that has been considered thus far can be thought of as *kinematical*; only the description of experimental configurations is of concern. The traditional approach to quantum theory (quantum *mechanics*) focuses on how and why quantum systems change in time. Using the Wigner function formalism to describe the dynamical transformations predicted by quantum mechanics yields the dynamical law

$$\frac{\partial \mu_\rho}{\partial t} = \{H, \mu_\rho\} + \sum_{n=1}^{\infty} \frac{1}{2^{2n}(2n+1)!} \frac{\partial^{2n+1} H}{\partial q^{2n+1}} \frac{\partial^{2n+1} \mu_\rho}{\partial p^{2n+1}}, \quad (71)$$

where  $\mu_\rho$  is the Wigner function and  $H$  is the classical Hamiltonian and  $\{H, \mu_\rho\}$  is the classical Poisson bracket. Note then that equation (71) is of the form ‘classical evolution’ + ‘quantum correction terms’. Using this formalism, one can then do more than discuss which experimental procedures are classical. Now one can discuss the *transitions* between quantum and classical descriptions, a process known as *decoherence* (Habib *et al* 1998, Joos 2003, Paz *et al* 1993). The representation of the dynamics was also studied for the spherical phase space in Kalmykov *et al* (2008) and Zueco and Calvo (2007). The goal is to compare the representation of the quantum dynamics (be it the Schrödinger equation or the more general *master equation*) to the natural classical dynamics of the representation’s phase space. The challenge for finite-dimensional systems is that no natural notion of discrete phase space exists for classical system. This problem has been recently studied by Livine (2010) by introducing a *discrete differential calculus* for the discrete phase spaces of Wootters. However, beyond these few examples, transformations and dynamics have not been studied anywhere near to the extent that states have for quasi-probability representations and presents itself as an open problem.

## 6.2. Traditional contextuality and non-locality

The traditional definition of contextuality evolved from a theorem which appears in a paper by Kochen and Specker (1967). The Kochen–Specker theorem concerns the standard quantum formalism: physical systems are assigned states in a complex Hilbert space  $\mathcal{H}$  and measurements are made of observables represented by Hermitian operators. The theorem establishes a contradiction between a set of plausible assumptions which together imply that quantum systems possess a consistent set of pre-measurement values for observable quantities. Let  $\mathcal{H}$  be the Hilbert space associated with a quantum system and  $A \in \mathbb{H}(\mathcal{H})$  be the operator associated with some observable. The function  $f_\psi(A)$  represents the value of the observable when the system is in state  $\psi$ . One assumption used to derive the contradiction is that for any function  $F$ ,  $f_\psi(F(A)) = F(f_\psi(A))$ . This is plausible

<sup>20</sup> Note, however, that the reasonable requirement of an *efficient* classical model was not met.



because, for example, we would expect that the value of  $A^2$  could be obtained in this way from the value of  $A$ .

The conclusion of the Kochen–Specker theorem implies the following counterintuitive example (Isham 1995). Suppose three operators  $A$ ,  $B$  and  $C$  satisfy  $[A, B] = 0 = [A, C]$ , but  $[B, C] \neq 0$ . Then, the value of the observable  $A$  will depend on whether observable  $B$  or  $C$  is chosen to be measured as well. That is, assuming that physical systems do possess values which can be revealed via measurements, the value of  $A$  depends on the *context* of the measurement.

What the Kochen–Specker theorem establishes then is that the mathematical framework of quantum theory does not allow for a *non-contextual* model for pre-measurement values. This fact is often expressed via the phrase ‘quantum theory is contextual’.

The original notion of contextuality is lacking in the sense that it only applies to the standard formalism of quantum theory and does not apply to general operational models. This problem was addressed by Spekkens as discussed in section 2. The notion of contextuality defined by Spekkens is more general; one can recover the original assumptions of Kochen–Specker by assuming that the projector valued measures in the spectral resolutions of observables are represented by dispersion free (0–1 valued) conditional probabilities (these are also called *sharp indicator functions*)<sup>21</sup>. Since the set of fewer assumptions already contains a contradiction when taken in conjunction, the addition of the assumption of Kochen–Specker is unnecessary. Thus, we need only consider the more general notion of contextuality we have already defined. This more general notion of contextuality has also recently been subject to experimental tests (Spekkens *et al* 2009).

A hidden variable theory originally formulated by de Broglie and later by Bohm (1989) is perhaps the most famous example of an ontological model of quantum theory. The model assumes that for a given experimental configuration, there exists a particle with well-defined trajectory and a quantum state  $\psi$ . The hidden variable is the position of the particle in real space. That is, the classical state space is  $\Lambda = \mathbb{R}^3 \times \mathcal{H}$ . The Hilbert space is included in the state space as it serves as a wave which guides the particle. If at any time the particle is distributed according to quantum probability distribution  $|\psi|^2$ , it remains so. Thus, so long as it is assumed that the particle is prepared according to this distribution, the model provides the same predictions as the standard formulation of quantum theory.

Note that this model does not fit into the framework of quasi-probability representations. Exactly as it was for the Beltrametti–Bugajski model, the de Broglie–Bohm model does not consider the entire range of possible quantum states. Where a classical representation contains a convex-linear mapping  $\rho \mapsto \mu(\lambda)$  for all  $\rho \in \mathbb{D}(\mathcal{H})$ , the de Broglie–Bohm model considers only a mapping with domain  $\mathcal{H}$ . Bell (2004) notes that ‘in the de Broglie–Bohm theory a fundamental

significance is given to the wavefunction, and it cannot be transferred to the density matrix’.

Bell *does not* claim that the situation is such that the de Broglie–Bohm model *cannot* be extended to include density operators. The key words in his comment are ‘fundament significance’. Indeed, the de Broglie–Bohm model *can* be extended to include density operators provided this extension is either contextual or contains negativity. In either case, the pure states (wavefunctions) retain their significance while the density operators possess non-classical features. As an example, the de Broglie–Bohm model could be such that  $(\rho, c_P) \mapsto \mu_{c_P}(\lambda)$  where each preparation consists of a density operator  $\rho$  supplied with a context  $c_P$  which specifies a particular convex decomposition of  $\rho$  into pure states. Such a model would be preparation contextual.

The non-locality debate was initiated by a paper by Einstein, Podolsky and Rosen (EPR) (Einstein *et al* 1935) where it was argued that quantum mechanics is *incomplete* (each element of physical reality does not have a counterpart in quantum theory) if special relativity remains valid. The latter means physical causation must be local or events cannot have causes outside of their past light cones. Using a particular spatially separated quantum system, and some standard quantum theory, EPR concluded that quantum mechanics is either incomplete or non-local (or both!). Locality was such a desired property of any theory that quantum mechanics was concluded to be incomplete. That is, there must be elements of physical reality (hidden variables) which quantum mechanics does not account for.

The argument of EPR was reformulated by Bohm (1989) for two qubits. The argument is built around the following hypothetical experiment. Two parties, Alice and Bob, are at distant locations with a source midway between them creating quantum systems described by the quantum state

$$\psi = \frac{1}{\sqrt{2}}(\phi_1 \otimes \phi_2 - \phi_2 \otimes \phi_1), \quad (72)$$

where  $\{\phi_1, \phi_2\}$  is an orthonormal basis for a qubit. One particle is sent to Alice and the other to Bob. Alice performs the projective two-outcome measurement  $\{P_1, P_2\}$  on the particle which was sent to her. The state in equation (72) is such that Alice, once she performs her measurement, can predict with certainty the outcome Bob receives when he performs the same measurement at his side of the experiment regardless of whether or not the measurement events are spacelike separated (i.e. non-local). For example, Alice could perform the measurement  $\{\phi_1\phi_1^*, \phi_2\phi_2^*\}$ . According to the collapse postulate, if Alice registers the first outcome, Bob’s particle will immediately collapse to  $\phi_2$  and he is certain to obtain the second outcome if he were to make the same measurement. Therefore, unless there exist hidden variables which pre-determine the possible outcomes when the particles are created, quantum theory is *non-local*.

Bell later investigated the possibility of finding the hidden variables Einstein thought to exist (Bell 2004). He noted immediately that the de Broglie–Bohm theory was such a theory yet it contained an astonishingly non-local character. He was able to prove that any hidden variable theory of

<sup>21</sup> Cabello has also generalized the notion of contextuality to POVMs (Cabello 2003). Again, however, the additional assumption of dispersion free condition probabilities is used. See Grudka and Kurzyński (2008) for an elaboration on this point. For a more broad discussion on contextuality see Morris (2009) and Harrigan and Rudolph (2007).

quantum phenomena must possess non-local features. This is now called Bell's theorem.

The proof is by contraction and follows the general line of reasoning which leads to the negativity theorem: build a mathematical model with assumptions that can be identified with (or motivated by) some notion of classicality then prove that quantum theory does not satisfy these assumptions. Consider the EPR experimental setup. Alice and Bob can each perform a two-outcome measurement with outcomes labeled  $A$  and  $B$ , respectively. Without loss of generality, the outcomes can be assigned numerical values  $A, B = \pm 1$ .

Suppose there exist a classical state space  $\Lambda$  (i.e. a set of hidden variables or, as we have called it above, an ontology) which serves to determine the outcomes  $A$  and  $B$ . Probabilistic knowledge of the state is represented by a density  $\mu(\lambda) \geq 0$  which is normalized

$$\int_{\Lambda} d\lambda \mu(\lambda) = 1.$$

The different measurements Alice and Bob can perform are parametrized by detector settings  $a$  and  $b$ , respectively. Locality is enforced by assuming that the outcomes  $A$  and  $B$  depend only the local detector settings and the global state. That is  $A = A(a, \lambda)$  is allowed but  $A = A(a, b, \lambda)$  is not. Define the correlation function

$$C(a, b) = \int_{\Lambda} d\lambda A(a, \lambda) B(b, \lambda) \mu(\lambda). \quad (73)$$

Bell's theorem states that the correlations obtained in the EPR experiment (i.e. a particular quantum experiment) cannot satisfy this equation. The proof follows by deriving an inequality from equation (73) such as

$$|C(a, b) - C(a, c)| \leq 1 + C(b, c). \quad (74)$$

This inequality holds for any hidden variable model which satisfies the locality assumption. For the quantum state in equation (72), the inequality is violated. This is the contradiction between the quantum theory and a local hidden variable model which proves Bell's theorem.

It was noted that the assumptions which go into the hidden variable models first considered by Bell imply those models are *deterministic*. That is, the theorem did not exclude models which suggested quantum theory only provides *stochastic* (or probabilistic) information of the possible outcomes of measurements. Bell later extended the theorem to include such models. For the EPR experimental setup, let the conditional probability of outcome  $A = 1$ , for Alice, given the state (hidden variable) is  $\lambda \in \Lambda$  be denoted  $M_A(\lambda)$  and similarly define  $M_B(\lambda)$  for Bob. Now denote the conditional joint probability of the simultaneous outcomes  $A, B = 1$  by  $M_{AB}(\lambda)$ . Fine (1982) defines a *stochastic hidden variable model* as one which satisfies

$$\Pr(A = 1) = \int_{\Lambda} d\lambda M_A(\lambda) \mu(\lambda) \quad (75)$$

and

$$\Pr(A = 1, B = 1) = \int_{\Lambda} d\lambda M_{AB}(\lambda) \mu(\lambda). \quad (76)$$

If  $M_{AB}(\lambda) = M_A(\lambda)M_B(\lambda)$ , then the model is *factorizable*. Bell claimed this also encoded the assumption of locality. Again, it can be shown that quantum theory is in contradiction with an inequality derived from these assumptions. Fine showed that a factorizable stochastic hidden variable model exists for the EPR-type correlation experiment if and only if a deterministic hidden variable model exists for the experiment. Since the latter is ruled out, the former is also ruled out.

It is often stated that *the* consequence of Bell's theorem is that 'quantum theory is non-local'. However the theorem only states that quantum theory does not satisfy the assumptions which go into a classical model *Bell defines as local*. It is not necessary that the locality assumption is violated. Nor is it unanimously agreed that the mathematical condition Bell refers to as locality in the stochastic hidden variable models reflects any physical significance (Fine 1982). Next, we will see how such a claim can be supported by appealing to the notions of negativity.

Note that a (non-factorizable) stochastic hidden variable model is exactly a classical representation. Then factorizability (Bell locality) is an additional requirement. However, the negativity theorem rules out a classical representation independently of any assumption of locality for the most general quantum experiment. The negativity theorem implies that one (or more) of the constraints which go into the definition of a classical representation are false. Indeed, relaxing the assumptions of positive probability yields a quasi-probability representation which is not in conflict with quantum theory. The existence of *positive* probability distributions in the hidden variables Bell attempts to rule out is often considered an unquestionable assumption. However, if it is the case that negative probability encodes something about Nature that is independent of locality, then it is not necessarily the case that Bell's theorem implies non-locality (Han 1996, Cereceda 2000, Rothman and Sudarshan 2001).

## 7. Summary

In this paper, we have seen how quasi-probability representations arise naturally out of the search for an ontological model of quantum theory. A non-contextual ontological model implies the existence of a classical representation, which was shown to be in contradiction with the quantum framework. However, by relaxing the requirement of positive probabilities we have seen that many *quasi*-probability representations are consistent with quantum theory, and indeed useful for quantum information theory.

We have shown how the theory of frames unifies the quasi-probability representations into a single formalism. This frame formalism allows two canonical methods for incorporating measurements into the framework. In the first, we use a *dual* frame to represent measurements. This gives rise to framework respecting the usual dot product between the represented functions (which is the classical rule of total probability) and is equivalent to the formal definition of a quasi-probability representation. In the second method, the *same* frame is used to represent states and measurements. This gives rise to the *deformed* probability representation. Here, negativity need not

appear in the representations of states and measurements but the Born rule is replaced by a deformation of the classical rule of total probability.

Finally, we have noted that there is significant interest surrounding the presence and interpretation of negativity in the quasi-probability representations. In particular, the use of negative probability allows us to avoid the contradictions which had lead to conclusions that quantum theory is contextual (via the Kochen–Specker theorem) and non-local (via Bell’s theorem).

We have discussed how negativity of *states alone* in some particular representation is a meaningless criterion for quantumness *per se*. Alternatively, we can accept some well-defined notion of a truly quantum state (such as an entangled state) and show that negativity is necessary for such a state in some representation. If we are ultimately to make use of our notion of quantumness, however, it needs to be defined for an operational task and hence incorporate the full operational formalism of quantum theory. That is, it must incorporate measurements (which we have done here) and *dynamical transformation* (which we have identified as a largely unexplored area of research).

## Acknowledgments

I thank Joseph Emerson, Chris Fuchs, Ryan Morris, Cozmin Ududec, Rob Spekkens, Victor Veitch and Joel Wallman for many helpful discussions. In particular, I thank Joseph Emerson for a careful reading of an early draft and suggestions for its improvement. This work was financially supported by the Canadian government through NSERC.

## Appendix. Notation and conventions

Quantum theory makes use of complex Hilbert spaces. If these spaces are finite-dimensional, then they are equivalent to inner product spaces. Unless otherwise noted, a Hilbert space  $\mathcal{H}$  will be assumed to have dimension  $d < \infty$  and (for  $\psi, \phi \in \mathcal{H}$ ) its inner product will be denoted  $\langle \psi, \phi \rangle$ . The following list defines some special sets of linear operators acting on  $\mathcal{H}$ .

- (1) An operator  $U$  satisfying

$$U^\dagger U = U U^\dagger = \mathbb{1}$$

is called *unitary*. The set of all unitary operators is denoted  $\mathbb{U}$ .

- (2) An operator  $A$  satisfying

$$\langle A\psi, \phi \rangle = \langle \psi, A\phi \rangle,$$

for all  $\psi, \phi \in \mathcal{H}$ , is called *Hermitian*. The set of all Hermitian operators is denoted  $\mathbb{H}(\mathcal{H})$ .

- (3) An operator  $E$  satisfying

$$0 \leq \langle \psi, E\psi \rangle \leq 1,$$

for all  $\psi \in \mathcal{H}$ , is called an *effect*. The set of all effects is denoted  $\mathbb{E}(\mathcal{H})$ .

- (4) An effect  $\rho$  satisfying

$$\text{Tr}(\rho) = 1$$

is called a *density operator*. The set of all density operators is denoted  $\mathbb{D}(\mathcal{H})$ .

- (5) A set of effects  $\{E_k\}$  which satisfy

$$\sum_k E_k = \mathbb{1}$$

is called a POVM (positive operator valued measure). The set of all POVMs is denoted  $\text{POVM}(\mathcal{H})$ .

- (6) An effect  $P$  satisfying

$$P^2 = P$$

is called a *projector*. The set of all projectors is denoted  $\mathbb{P}(\mathcal{H})$ .

- (7) A set of projectors  $\{P_k\}$ , each of rank 1, which satisfy

$$\sum_k P_k = \mathbb{1}$$

is called a PVM (projector valued measure). The set of all PVMs is denoted  $\text{PVM}(\mathcal{H})$ .

Note that from these definitions we have  $\mathbb{P}(\mathcal{H}) \subset \mathbb{D}(\mathcal{H}) \subset \mathbb{E}(\mathcal{H}) \subset \mathbb{H}(\mathcal{H})$  and  $\text{PVM}(\mathcal{H}) \subset \text{POVM}(\mathcal{H})$ . The set  $\mathbb{H}(\mathcal{H})$  defines its own Hilbert space with inner product (for  $A, B \in \mathbb{H}(\mathcal{H})$ )  $\langle A, B \rangle := \text{Tr}(AB)$ . The dimension of  $\mathbb{H}(\mathcal{H})$  is  $d^2$ .

The following are some examples of important operators used in this paper. Consider the operator  $Z$  whose spectrum is  $\text{spec}(Z) = \{\omega^k : k \in \mathbb{Z}_d\}$ , where  $\omega^k = e^{-\frac{2\pi i}{d}k}$ . The eigenvectors form a basis for  $\mathcal{H}$  and are denoted  $\{\phi_k\}$ . Consider also the operator defined by  $X\phi_k = \phi_{k+1}$ , where all arithmetic is modulo  $d$ . Define  $Y$  implicitly through  $[X, Z] = 2iY$ . The operators  $Z, X$  and  $Y$  are often called *generalized Pauli operators* since they are indeed the usual Pauli operators when  $d = 2$ . The *parity operator* is defined by  $P\phi_k = \phi_{-k}$ .

## References

- Alicki R, Piani M and Ryn N V 2008 *J. Phys. A: Math. Theor.* **41** 495303  
Alicki R and Ryn N V 2008 *J. Phys. A: Math. Theor.* **41** 062001  
Ambainis A, Schulman L J and Vazirani U 2006 *J. ACM* **53** 507  
Arecchi F T, Courtens E, Gilmore R and Thomas H 1972 *Phys. Rev. A* **6** 2211  
Bell J S 2004 *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy* (Cambridge: Cambridge University Press)  
Beltrametti E G and Bugajski S 1995 *J. Phys. A: Math. Gen.* **28** 3329  
Bertrand J and Bertrand P 1987 *Found. Phys.* **17** 397  
Bianucci P, Miquela C, Paz J P and Saraceno M 2002 *Phys. Lett. A* **297** 353  
Bisio A, Chiribella G, D’Ariano G M, Facchini S and Perinotti P 2009 *Phys. Rev. Lett.* **102** 010404  
Bohm D 1989 *Quantum Theory* (New York: Dover)  
Braunstein S L and Kimble H J 1998 *Phys. Rev. Lett.* **80** 869  
Brif C and Mann A 1999 *Phys. Rev. A* **59** 971  
Busch P, Hellwig K E and Stulpe W 1993 *Int. J. Theor. Phys.* **32** 399  
Cabello A 2003 *Phys. Rev. Lett.* **90** 190401  
Cahill K E and Glauber R J 1969 *Phys. Rev.* **177** 1882



- Caves C M, Fuchs C A and Schack R 2002a *Phys. Rev. A* **65** 022305
- Caves C M, Fuchs C A and Schack R 2002b *J. Math. Phys.* **43** 4537
- Caves C M and Wódkiewicz K 2004 *Phys. Rev. Lett.* **93** 040506
- Cereceda J L 2000 arXiv:[quant-ph/0010091](#)
- Chaturvedi S, Ercolessi E, Marmo G, Morandi G, Mukunda N and Simon R 2005 *Pramana* **65** 981
- Chaturvedi S, Ercolessi E, Marmo G, Morandi G, Mukunda N and Simon R 2006 *J. Phys. A: Math. Gen.* **39** 1405
- Christensen O 2003 *An Introduction to Frames and Riesz Bases* (Boston, MA: Birkhäuser)
- Cohen L and Scully M 1986 *Found. Phys.* **16** 295
- Cohendet O, Combe P and Collin S M 1990 *J. Phys. A: Math. Gen.* **23** 2001
- Cohendet O, Combe P, Sirugue M and Collin S M 1988 *J. Phys. A: Math. Gen.* **21** 2875
- Combescur M 2006 arXiv:[quant-ph/0605090](#)
- Cormick C, Galvão E F, Gottesman D, Paz J P and Pittenger A O 2006 *Phys. Rev. A* **73** 012301
- Datta A 2008 Studies on the role of entanglement in mixed-state quantum computation *PhD Thesis* University of New Mexico (arXiv:[0807.4490](#))
- Datta A, Flammia S T and Caves C M 2005 *Phys. Rev. A* **72** 042316
- Dragoman D 2005 *EURASIP J. Appl. Signal Process.* **2005** 1520
- Einstein A, Podolsky B and Rosen N 1935 *Phys. Rev.* **47** 777
- Emerson J, Lloyd S, Poulin D and Cory D 2004 *Phys. Rev. A* **69** 050305
- Ferrie C and Emerson J 2008 *J. Phys. A: Math. Theor.* **41** 352001
- Ferrie C and Emerson J 2009 *New J. Phys.* **11** 063040
- Ferrie C, Morris R and Emerson J 2010 *Phys. Rev. A* **82** 044103
- Feynman R P 1987 Negative probability *Quantum Implications: Essays in Honour of David Bohm* ed F D Peat and B Hiley (London: Routledge and Kegan Paul) pp 235–48
- Fine A 1982 *Synthese* **50** 279
- Franco R and Penna V 2006 *J. Phys. A: Math. Gen.* **39** 5907
- Fuchs C A 2002 arXiv:[quant-ph/0205039](#)
- Fuchs C A 2010 arXiv:[1003.5209](#)
- Fuchs C A and Schack R 2009 arXiv:[0906.2187](#)
- Galetti D and de Toledo Piza A F R 1988 *Physica A* **149** 267
- Galvão E F 2005 *Phys. Rev. A* **71** 042302
- Ghose S, Stock R, Jessen P, Lal R and Silberfarb A 2008 *Phys. Rev. A* **78** 042318
- Gibbons K S, Hoffman M J and Wootters W K 2004 *Phys. Rev. A* **70** 062101
- Giraud O, Braun P and Braun D 2008 *Phys. Rev. A* **78** 042112
- Giraud O, Braun P A and Braun D 2010 arXiv:[1002.2158](#)
- Glauber R J 1963 *Phys. Rev.* **131** 2766
- Gottesman D 1997 Stabilizer codes and quantum error correction *PhD Thesis* California Institute of Technology (arXiv:[quant-ph/9705052v1](#))
- Gross D 2006 *J. Math. Phys.* **47** 122107
- Gross D 2007 *Appl. Phys. B* **86** 367
- Grudka A and Kurzyński P 2008 *Phys. Rev. Lett.* **100** 160401
- Habib S, Shizume K and Zurek W H 1998 *Phys. Rev. Lett.* **80** 4361
- Hakioglu T 1998 *J. Phys. A: Math. Gen.* **31** 6975
- Han Y 1996 *Phys. Lett. A* **221** 283
- Hannay J and Berry M V 1980 *Physica D* **1** 267
- Hardy L 2001a arXiv:[quant-ph/0101012](#)
- Hardy L 2001b arXiv:[quant-ph/0111068](#)
- Harrigan N and Rudolph T 2007 arXiv:[0709.4266](#)
- Havel T F 2002 *Quantum Inform. Process.* **1** 511
- Heiss S and Weigert S 2000 *Phys. Rev. A* **63** 012105
- Held C 2009 Kochen–Specker theorem *Compendium of Quantum Physics* ed D Greenberger, K Hentschel and F Weinert (Berlin: Springer) pp 331–5
- Hillery M, O’Connell R F, Scully M O and Wigner E P 1984 *Phys. Rep.* **106** 121
- Horodecki P 1997 *Phys. Lett. A* **232** 333
- Horodecki R, Horodecki P, Horodecki M and Horodecki K 2009 *Rev. Mod. Phys.* **81** 865
- Hudson R 1974 *Rep. Math. Phys.* **6** 249
- Husimi K 1940 *Proc. Physico-Math. Soc. Japan* **22** 264
- Isham C 1995 *Lectures on Quantum Theory* (London: Imperial College Press)
- Ivonovic I D 1981 *J. Phys. A: Math. Gen.* **14** 3241
- Joos E 2003 Decoherence through interaction with the environment *Decoherence and the Appearance of a Classical World in Quantum Theory* 2nd edn, ed E Joos *et al* (Berlin: Springer) pp 41–180, chapter 3
- Kalev A, Mann A, Mello P A and Revzen M 2009 *Phys. Rev. A* **79** 014104
- Kalmykov Y P, Coffey W T and Titov S V 2008 *J. Phys. A: Math. Theor.* **41** 105302
- Kenfack A and Życzkowski K 2004 *J. Opt. B: Quantum Semiclassical Opt.* **6** 396
- Kiesel T, Vogel W, Parigi V, Zavatta A and Bellini M 2008 *Phys. Rev. A* **78** 021804
- Knill E and Laflamme R 1998 *Phys. Rev. Lett.* **81** 5672
- Knill E and Laflamme R 2001 *Inform. Process. Lett.* **79** 173
- Kochen S and Specker E 1967 *J. Math. Mech.* **17** 59
- Koniorczyk M, Bužek V and Janszky J 2001 *Phys. Rev. A* **64** 034301
- Laflamme R, Cory D G, Negrevergne C and Viola L 2002 *Quantum Inform. Comput.* **2** 166 (arXiv:[quant-ph/0110029](#))
- Lee H 1995 *Phys. Rep.* **259** 147
- Leonhardt U 1995 *Phys. Rev. Lett.* **74** 4101
- Leonhardt U 1996 *Phys. Rev. A* **53** 2998
- Livine E R 2010 *J. Phys. A: Math. Theor.* **43** 075303
- López C C and Paz J P 2003 *Phys. Rev. A* **68** 052305
- Luis A and Perina J 1998 *J. Phys. A: Math. Gen.* **31** 1423
- Mandel L 1986 *Phys. Scr.* **T12** 34
- Marchioli M A, Ruzzi M and Galetti D 2005 *Phys. Rev. A* **72** 042308
- Marchioli M A, Ruzzi M and Galetti D 2007 *Phys. Rev. A* **76** 032102
- Marchioli M A, Silva E C and Galetti D 2009 *Phys. Rev. A* **79** 022114
- Mari A, Kieliing K, Nielsen B M, Polzik E S and Eisert J 2010 arXiv:[1005.1665](#)
- Miquel C, Paz J P and Saraceno M 2002a *Phys. Rev. A* **65** 062309
- Miquel C, Paz J P, Saraceno M, Knill E, Laflamme R and Negrevergne C 2002b *Nature* **418** 59
- Morris R 2009 Topics in quantum foundations: ontological models, and distinguishability as a resource *Master’s Thesis* University of Waterloo
- Mukunda N 2004 *Phys. Lett. A* **321** 160
- Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* 1st edn (Cambridge: Cambridge University Press)
- Opatrný T, Welsch D G and Bužek V 1996 *Phys. Rev. A* **53** 3822
- Paz J P 2002 *Phys. Rev. A* **65** 062311
- Paz J P, Habib S and Zurek W H 1993 *Phys. Rev. D* **47** 488
- Paz J P, Roncaglia A J and Saraceno M 2005 *Phys. Rev. A* **72** 012309
- Peres A 1996 *Phys. Rev. Lett.* **77** 1413
- Porta Mana P G L 2004a arXiv:[quant-ph/0403084](#)
- Porta Mana P G L 2004b arXiv:[quant-ph/0305117](#)
- Porta Mana P G L 2007 Studies in plausibility theory, with applications to physics *PhD Thesis* KTH Royal Institute of Technology, Stockholm, Sweden
- Poulin D, Kohout R B, Laflamme R and Ollivier H 2004 *Phys. Rev. Lett.* **92** 177906
- Redhead M 1989 *Incompleteness, Nonlocality, and Realism* (Oxford: Oxford University Press)
- Renes J M, Blume-Kohout R, Scott A J and Caves C M 2004 *J. Math. Phys.* **45** 2171
- Rivas A 1999 *Ann. Phys.* **276** 223
- Rothman T and Sudarshan E C G 2001 *Int. J. Theor. Phys.* **40** 1525



- Ruzzi M, Marchiolli M A and Galetti D 2005 *J. Phys. A: Math. Gen.* **38** 6239
- Schack R 2003 *Found. Phys.* **33** 1461
- Schack R, Brun T A and Caves C M 2001 *Phys. Rev. A* **64** 014305
- Schack R and Caves C M 1999 *Phys. Rev. A* **60** 4354
- Schack R and Caves C M 2000 *J. Mod. Opt.* **47** 387
- Schleich W P 2001 *Quantum Optics in Phase Space* 1st edn (New York: Wiley-VCH)
- Scott A J 2006 *J. Phys. A: Math. Gen.* **39** 13507
- Scott A J and Grassl M 2010 *J. Math. Phys.* **51** 042203
- Scully M and Wódkiewicz K 1994 *Found. Phys.* **24** 85
- Shalm L K, Adamson R B A and Steinberg A M 2009 *Nature* **457** 67
- Shepherd D 2006 arXiv:[quant-ph/0608132](https://arxiv.org/abs/quant-ph/0608132)
- Shor P W and Jordan S P 2008 *Quantum Inform. Comput.* **8** 681 (arXiv:[0707.2831v3](https://arxiv.org/abs/0707.2831v3))
- Spekkens R W 2005 *Phys. Rev. A* **71** 052108
- Spekkens R W 2008 *Phys. Rev. Lett.* **101** 020401
- Sperling J and Vogel W 2009 *Phys. Rev. A* **79** 042337
- Stratonovich R L 1957 *J. Exp. Theor. Phys.* **4** 891
- Sudarshan E C G 1963 *Phys. Rev. Lett.* **10** 277
- Takahashi K 1989 *Prog. Theor. Phys. Suppl.* **98** 109
- Vaidman L 1994 *Phys. Rev. A* **49** 1473
- Varilly J and Graciabondia J 1989 *Ann. Phys.* **190** 107
- Vedral V 2010 *Found. Phys.* **40** 1141
- Vourdas A 2004 *Rep. Prog. Phys.* **67** 267
- Walls D F and Milburn G J 1995 *Quantum Opt.* (Berlin: Springer)
- Wigner E 1932 *Phys. Rev.* **40** 749
- Wootters W 1986 *Found. Phys.* **16** 391
- Wootters W 1987 *Ann. Phys.* **176** 1
- Wootters W K 2004 *IBM J. Res. Dev.* **48** 99
- Zauner G 1999 Quantum designs: foundations of a non-commutative theory of designs *PhD Thesis* University of Vienna
- Zueco D and Calvo I 2007 *J. Phys. A: Math. Theor.* **40** 4635
- Bartlett S D and Sanders B C 2002 *Phys. Rev. Lett.* **89** 207903
- Korbicz J K, Cirac J I, Wehr J and Lewenstein M 2005 *Phys. Rev. Lett.* **94** 153601
- Spekkens R W, Buzacott D H, Keehn A J, Toner B and Pryde G J 2009 *Phys. Rev. Lett.* **102** 010401