## OPEN ACCESS

## A derivation of quantum theory from physical requirements

To cite this article: Lluís Masanes and Markus P Müller 2011 New J. Phys. 13063001

You may also like
Democratization of quantum technologies Zeki C Seskir, Steven Umbrello, Christopher Coenen et al.

Improvement of wear and hot melt loss resistance of metal carbide layers on H 13 steel
Zhi-Zhong Zeng, Dan Lin and Jian Shang
Quantum teleportation in terms of creation operators of photons M V Fedorov and A A Sysoeva

View the article online for updates and enhancements.

# A derivation of quantum theory from physical requirements 

Lluís Masanes ${ }^{1,5}$ and Markus $\mathbf{P}$ Müller ${ }^{2,3,4}$<br>${ }^{1}$ ICFO-Institut de Ciencies Fotoniques, Mediterranean Technology Park, 08860 Castelldefels, Barcelona, Spain<br>${ }^{2}$ Institute of Mathematics, Technical University of Berlin, 10623 Berlin, Germany<br>${ }^{3}$ Institute of Physics and Astronomy, University of Potsdam, 14476 Potsdam, Germany<br>${ }^{4}$ Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, ON, N2L 2Y5, Canada<br>E-mail: lluis.masanes@icfo.es

New Journal of Physics 13 (2011) 063001 (29pp)
Received 10 February 2011
Published 1 June 2011
Online at http://www.njp.org/
doi:10.1088/1367-2630/13/6/063001


#### Abstract

Quantum theory (QT) is usually formulated in terms of abstract mathematical postulates involving Hilbert spaces, state vectors and unitary operators. In this paper, we show that the full formalism of QT can instead be derived from five simple physical requirements, based on elementary assumptions regarding preparations, transformations and measurements. This is very similar to the usual formulation of special relativity, where two simple physical requirements-the principles of relativity and light speed invarianceare used to derive the mathematical structure of Minkowski space-time. Our derivation provides insights into the physical origin of the structure of quantum state spaces (including a group-theoretic explanation of the Bloch ball and its three dimensionality) and suggests several natural possibilities to construct consistent modifications of QT.


[^0]
## Contents

1. Introduction ..... 2
2. Generalized probabilistic theories ..... 4
2.1. States ..... 4
2.2. Measurements ..... 5
2.3. Transformations ..... 6
2.4. Composite systems ..... 6
2.5. Equivalent state spaces ..... 8
2.6. Instances of generalized probabilistic theories ..... 8
3. The requirements ..... 8
4. Characterization of all theories satisfying the requirements ..... 10
4.1. The maximally mixed state ..... 10
4.2. The generalized bit ..... 11
4.3. Capacity and dimension ..... 12
4.4. Recovering classical probability theory ..... 14
4.5. Reversible transformations for the generalized bit ..... 14
4.6. Two generalized bits ..... 15
4.7. Consistency in the subspaces of two generalized bits ..... 16
4.8. The Hermitian representation ..... 19
4.9. Reconstructing quantum theory ..... 20
5. Conclusion ..... 22
Acknowledgments ..... 22
Appendix. Lemmas ..... 22
References ..... 28

## 1. Introduction

Quantum theory (QT) is usually formulated by postulating the mathematical structure and representation of states, transformations and measurements. The general physical consequences that follow (such as the violation of Bell-type inequalities [1], the possibility of performing state tomography with local measurements or the factorization of integers in polynomial time [2]) come as theorems that use the postulates as premises. In this work, this procedure is reversed: we impose five simple physical requirements, and this suffices to single out QT and derive its mathematical formalism uniquely. This is very similar to the usual formulation of special relativity, where two simple physical requirements-the principles of relativity and light speed invariance-are used to derive the mathematical structure of Minkowski space-time and its transformations.

The requirements can be schematically stated thus:

1. In systems that carry one bit of information, each state is characterized by a finite set of outcome probabilities.
2. The state of a composite system is characterized by the statistics of measurements on the individual components.
3. All systems that effectively carry the same amount of information have equivalent state spaces.
4. Any pure state of a system can be reversibly transformed into any other.
5. In systems that carry one bit of information, all mathematically well-defined measurements are allowed by the theory.

These requirements are imposed on the framework of generalized probabilistic theories [3-9], which already assumes that some operational notions (preparation, mixture, measurement and counting relative frequencies of measurement outcomes) make sense. Owing to its conceptual simplicity, this framework leaves room for an infinitude of possible theories, allowing for weaker- or stronger-than-quantum non-locality [6], [10-14]. In this paper, we show that QT and classical probability theory (CPT) are very special among those theories: they are the only general probabilistic theories that satisfy the five requirements stated above.

The non-uniqueness of the solution is not a problem, since CPT is embedded in QT; thus QT is the most general theory satisfying the requirements. One can also proceed as Hardy in [4]: if requirement 4 is strengthened by imposing continuity of the reversible transformations, then CPT is ruled out and QT is the only theory satisfying the requirements. This strengthening can be justified by the continuity of time evolution of physical systems.

It is conceivable that in the future, another theory may replace or generalize QT. Such a theory must violate at least one of our assumptions. The clear meaning of our requirements allows one to straightforwardly explore potential features of such a theory. The relaxation of each of our requirements constitutes a different way to go beyond QT.

The search for alternative axiomatizations of QT is an old topic that goes back to Birkhoff and von Neumann [8], and has been approached in many different ways: extending propositional logic [7, 8], using operational primitives [3-6], [9], searching for informationtheoretic principles [5, 6, 10, 11], [19-21] and building upon the phenomenon of quantum nonlocality [6], [10-13]. Alfsen and Shultz [22] have accomplished a complete characterization of the state spaces of QT from a geometric point of view, but the result does not seem to have an immediate physical meaning. In particular, the fact that the state space of a generalized bit is a three-dimensional (3D) ball is an assumption there, whereas here it is derived from physical requirements.

This work is particularly close to $[4,19]$, from which it takes some material. More concretely, the multiplicativity of capacities and the Simplicity Axiom from [4] are replaced by requirement 5. In comparison to [19], the fact that each state of a generalized bit is mixture of two distinguishable ones, the maximality of the group of reversible transformations and its orthogonality, and the multiplicativity of capacities, is also replaced by requirement 5.

This paper is organized as follows: Section 2 contains an introduction to the framework of generalized probabilistic theories, where some elementary results are stated without proof. In section 3 the five requirements and their significance are explained in detail. Section 4 is the core of this work. It contains the characterization of all theories compatible with the requirements, concluding that the only possibilities are CPT and QT. The conclusion (section 5) recapitulates the results and adds some remarks. The appendix contains all lemmas and their proofs.


Figure 1. General experimental setup. From left to right, there are the preparation, transformation and measurement devices. As soon as the release button is pressed, the preparation device outputs a physical system in the state specified by the knobs. The next device performs the transformation specified by its knobs (which in particular can 'do nothing'). The device on the right carries out the measurement specified by its knobs, and the outcome ( $x$ or $\bar{x}$ ) is indicated by the corresponding light.

## 2. Generalized probabilistic theories

In CPT there can always be a joint probability distribution for all random variables under consideration. The framework of generalized probabilistic theories (GPTs), also called the convex operational framework, generalizes this by allowing the possibility of random variables that cannot have a joint probability distribution or cannot be simultaneously measured (such as noncommuting observables in QT).

This framework assumes that at some level there is a classical reality, where it makes sense to talk about experimentalists performing basic operations such as preparations, mixtures, measurements and counting the relative frequencies of outcomes. These are the primary concepts of this framework. It also provides a unified way for all GPTs to represent states, transformations and measurements. A particular GPT specifies which of these are allowed, but it does not tell their correspondence to actual experimental setups. On its own, a GPT can still make nontrivial predictions such as: the maximal violation of a Bell inequality [1], the complexity-theoretic computational power $[2,18]$ and, in general, all information-theoretic properties of the theory [6].

The framework of GPTs can be stated in different ways, but all lead to the same formalism [3-9]. This formalism is presented in this section at a very basic level, providing some elementary results without proofs.

### 2.1. States

Definition of a system. We associate with a setup like figure 1 a system if, for each configuration of the preparation, transformation and measurement devices, the relative frequencies of the outcomes tend to a unique probability distribution (in the large sample limit).

The probability of a measurement outcome $x$ is denoted by $p(x)$. This outcome can be associated with a binary measurement that tells whether $x$ happens or not (this second event $\bar{x}$ has probability $p(\bar{x})=1-p(x))$. The above definition of a system allows one to associate with each preparation procedure a list of probabilities of the outcomes of all the measurements that can be carried out on a system. As we show in section 4.3, our requirements imply that all these probabilities $p(x)$ are determined by a finite set of them; the smallest such set is used to
represent the state

$$
\psi=\left[\begin{array}{c}
1  \tag{1}\\
p\left(x_{1}\right) \\
\vdots \\
p\left(x_{d}\right)
\end{array}\right]=\left[\begin{array}{c}
\psi^{0} \\
\psi^{1} \\
\vdots \\
\psi^{d}
\end{array}\right] \in \mathcal{S} \subset \mathbb{R}^{d+1} .
$$

The measurement outcomes that characterize the state $x_{1}, \ldots, x_{d}$ are called fiducial, and in general, there is more than one set of them (for example, a $\frac{1}{2}$-spin particle in QT is characterized by the spin in any three linearly independent directions). Note that each of the fiducial outcomes can correspond to a different measurement. The redundant component $\psi^{0}=1$ is reminiscent of QT, where one of the diagonal entries of a density matrix is redundant, since they sum up to 1. In fact $\psi^{0} \neq 1$ is sometimes used to represent unnormalized states, but not in this paper, where only normalized states are considered. The redundant component $\psi^{0}$ allows one to use the tensor-product formalism in composite systems (section 2.4), which simplifies the notation.

The set of all allowed states $\mathcal{S}$ is convex [23], because if $\psi_{1}, \psi_{2} \in \mathcal{S}$, then one can prepare $\psi_{1}$ with probability $q$, and $\psi_{2}$ with probability $1-q$, effectively preparing the state $q \psi_{1}+(1-q) \psi_{2}$. The number of fiducial probabilities $d$ is equal to the (affine) dimension of $\mathcal{S}$; otherwise one fiducial probability would be functionally related to the others and hence redundant.

Suppose there is an $\mathbb{R}^{d+1}$-vector $\psi \notin \mathcal{S}$, which is in the topological closure of $\mathcal{S}$-that is, $\psi$ can be approximated by states $\psi^{\prime} \in \mathcal{S}$ to arbitrary accuracy. Since there is no observable physical difference between perfect preparation and arbitrarily good preparation, we will consider $\psi$ to be a valid state and add it to the state space. This does not change the physical predictions of the theory, but it has the mathematical consequence that state spaces become topologically closed. Since state vectors (1) are bounded, and we are in finite dimensions (shown in section 4.3), state spaces $\mathcal{S}$ are compact convex sets [23].

The pure states of a state space $\mathcal{S}$ are the ones that cannot be written as mixtures: $\psi \neq q \psi_{1}+(1-q) \psi_{2}$ with $\psi_{1} \neq \psi_{2}$ and $0<q<1$. Since $\mathcal{S}$ is compact and convex, all states are mixtures of pure states [23].

### 2.2. Measurements

The probability of measurement outcome $x$ when the system is in state $\psi \in \mathcal{S}$ is given by a function $\Omega_{x}(\psi)$. Suppose the system is prepared in the mixture $q \psi_{1}+(1-q) \psi_{2}$, then the relative frequency of outcome $x$ does not depend on whether the label of the actual preparation $\psi_{k}$ is ignored before or after the measurement; hence

$$
\Omega_{x}\left(q \psi_{1}+(1-q) \psi_{2}\right)=q \Omega_{x}\left(\psi_{1}\right)+(1-q) \Omega_{x}\left(\psi_{2}\right) .
$$

This means that the function $\Omega_{x}$ is affine on $\mathcal{S}$. The redundant component $\psi^{0}$ in (1) allows one to write this function as a linear map $\Omega_{x}: \mathbb{R}^{d+1} \rightarrow \mathbb{R}[3,6]$.

An effect is a linear map $\Omega: \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ such that $\Omega(\psi) \in[0,1]$ for all states $\psi \in \mathcal{S}$. Every function $\Omega_{x}$ associated with an outcome probability $p(x)$ is an effect. The converse is not necessarily true: the framework of GPTs allows one to construct theories where some effects do not represent possible measurement outcomes. These restrictions are analogous to superselection rules, where some (mathematically well-defined) states are not allowed by the physical theory. This is related to requirement 5 . A tight effect $\Omega$ is one for which there are two states $\psi_{0}, \psi_{1} \in \mathcal{S}$ satisfying $\Omega\left(\psi_{0}\right)=0$ and $\Omega\left(\psi_{1}\right)=1$.

An $n$-outcome measurement is specified by $n$ effects $\Omega_{1}, \ldots, \Omega_{n}$ such that $\Omega_{1}(\psi)+\ldots+$ $\Omega_{n}(\psi)=1$ for all $\psi \in \mathcal{S}$. The number $\Omega_{a}(\psi)$ is the probability of outcome $a$ when the measurement is carried out on the state $\psi$. The states $\psi_{1}, \ldots, \psi_{n}$ are distinguishable if there is an $n$-outcome measurement such that $\Omega_{a}\left(\psi_{b}\right)=\delta_{a, b}$, where $\delta_{a, b}=1$ if $a=b$ and $\delta_{a, b}=0$ if $a \neq b$.

The capacity of a state space $\mathcal{S}$ is the size of the largest family of distinguishable states and is denoted by $c$. This is the amount of classical information that can be transmitted by the corresponding type of system, in a single-shot error-free procedure. (In QT the capacity of a system is the dimension of its corresponding Hilbert space, which must not be confused with the dimension of the state space $d=c^{2}-1$, that is, the set of $c \times c$ complex matrices that are positive and have unit trace.) A complete measurement on $\mathcal{S}$ is one capable of distinguishing $c$ states.

### 2.3. Transformations

Each type of system has associated with it the following: a state space, a set of measurements and a set of transformations. A transformation $T$ is a map $T: \mathcal{S} \rightarrow \mathcal{S}$. Similarly to measurements, if a state is prepared as a mixture $q \psi_{1}+(1-q) \psi_{2}$, it does not matter whether the label of the actual preparation $\psi_{k}$ is ignored before or after the transformation. Hence,

$$
T\left(q \psi_{1}+(1-q) \psi_{2}\right)=q T\left(\psi_{1}\right)+(1-q) T\left(\psi_{2}\right)
$$

which implies that $T$ is an affine map. The redundant component $\psi^{0}$ in (1) allows one to extend $T$ to a linear map $T: \mathbb{R}^{d+1} \rightarrow \mathbb{R}^{d+1}[3,6]$.

A transformation $T$ is reversible if its inverse $T^{-1}$ exists and belongs to the set of transformations allowed by the theory. The set of (allowed) reversible transformations of a particular state space $\mathcal{S}$ forms a group $\mathcal{G}$. For the same reason as for the state space itself, we will assume that the group of reversible transformations is topologically closed. Previously, we have seen that a state space $\mathcal{S}$ is bounded and hence the corresponding group of transformations $\mathcal{G}$ is bounded, too. In summary, groups of transformations are compact [24].

### 2.4. Composite systems

Definition of a composite system. Two systems $A$ and $B$ constitute a composite system, denoted $A B$, if a measurement for $A$ together with a measurement for $B$ uniquely specifies a measurement for $A B$. This means that if $x$ and $y$ are measurement outcomes on $A$ and $B$, respectively, the pair $(x, y)$ specifies a unique measurement outcome on $A B$, whose probability distribution $p(x, y)$ does not depend on the temporal order in which the subsystems are measured.

The fact that subsystems are themselves systems implies that each has a well-defined reduced state $\psi_{A}, \psi_{B}$, that does not depend on which transformations and measurements are performed on the other subsystem (see the definition of a system in section 2.1). This is often referred to as no-signaling. Let $x_{1}, \ldots, x_{d_{A}}$ be the fiducial measurements of system $A$, and $y_{1}, \ldots, y_{d_{B}}$ those of $B$. The no-signaling constraints are

$$
\begin{align*}
p\left(x_{i}\right) & =p\left(x_{i}, y_{j}\right)+p\left(x_{i}, \bar{y}_{j}\right)  \tag{2}\\
p\left(y_{i}\right) & =p\left(x_{i}, y_{j}\right)+p\left(\bar{x}_{i}, y_{j}\right)
\end{align*}
$$

for all $i, j$.

An assumption that is often postulated additionally in the GPT context is requirement 2 , which states that the state of a composite system is completely characterized by the statistics of measurements on the subsystems, that is, $p(x, y)$. This and no-signaling (2) imply that states in $A B$ can be represented on the tensor-product vector space [3] as

$$
\psi_{A B}=\left(\begin{array}{c}
1  \tag{3}\\
\vdots \\
p\left(x_{i}\right) \\
\vdots \\
p\left(y_{j}\right) \\
\vdots \\
p\left(x_{i}, y_{j}\right) \\
\vdots
\end{array}\right) \in \mathcal{S}_{A B} \subset \mathbb{R}^{d_{A}+1} \otimes \mathbb{R}^{d_{B}+1} .
$$

The joint probability of two arbitrary local measurement outcomes $x, y$ is given by

$$
\begin{equation*}
p(x, y)=\left(\Omega_{x} \otimes \Omega_{y}\right)\left(\psi_{A B}\right) \tag{4}
\end{equation*}
$$

where $\Omega_{x}$ is the effect representing $x$ in $A$, that is, $p(x)=\Omega_{x}\left(\psi_{A}\right)$, and analogously for $\Omega_{y}$ [3]. (The term 'local' is used when referring to subsystems, and has nothing to do with spatial locations.) In other words, if $\left\{\Omega_{1}^{A}, \ldots, \Omega_{n}^{A}\right\}$ is an $n$-outcome measurement on $A$ and if $\left\{\Omega_{1}^{B}, \ldots, \Omega_{m}^{B}\right\}$ is an $m$-outcome measurement on $B$, then $\left\{\Omega_{a}^{A} \otimes \Omega_{b}^{B} \mid a=1, \ldots, n\right.$; $b=1, \ldots, m\}$ defines a measurement on $A B$ with $n m$ outcomes. Local transformations act on the global state as

$$
\begin{equation*}
\psi_{A B} \rightarrow\left(T_{A} \otimes T_{B}\right)\left(\psi_{A B}\right), \tag{5}
\end{equation*}
$$

where $T_{A}$ is the matrix that represents the transformation in $A$, and analogously for $T_{B}$ [3]. The reduced states

$$
\psi_{A}=\left(\begin{array}{c}
1  \tag{6}\\
\vdots \\
p\left(x_{i}\right) \\
\vdots
\end{array}\right), \quad \psi_{B}=\left(\begin{array}{c}
1 \\
\vdots \\
p\left(y_{j}\right) \\
\vdots
\end{array}\right)
$$

are obtained from $\psi_{A B}$ by picking the right components (3). Alternatively, reduced states can be defined by $\Omega_{A}\left(\psi_{A}\right)=\left(\Omega_{A} \otimes \mathbf{1}\right)\left(\psi_{A B}\right)$ for any effect $\Omega_{A}$ in $A$, where $\mathbf{1}\left(\psi_{B}\right)=\psi_{B}^{0}$ is the unit effect. The reduced state $\psi_{A}$ must belong to the state space of subsystem $A$, denoted $\mathcal{S}_{A}$, and any state in $\mathcal{S}_{A}$ must be the reduction of a state from $\mathcal{S}_{A B}$ (analogously for subsystem B). This implies that all product states

$$
\begin{equation*}
\psi_{A B}=\psi_{A} \otimes \psi_{B} \tag{7}
\end{equation*}
$$

are contained in $\mathcal{S}_{A B}$ [3], and similarly, all tensor products of local measurements and transformations are allowed on $A B$.

Given two fixed state spaces $\mathcal{S}_{A}$ and $\mathcal{S}_{B}$, the previous discussion imposes constraints on the state space of the composite system $\mathcal{S}_{A B}$. However, there are still many different possible joint state spaces $\mathcal{S}_{A B}$, and some of them allow for larger violations of Bell inequalities than do QT. In fact, this has been extensively studied [5, 6], [10-14], and is one of the reasons for the popularity of generalized probabilistic theories.

Nothing prevents Bob's system from being composite itself; hence, one can recursively extend the definition of composite system and formulae (3)-(5) and (7) to more parties.

### 2.5. Equivalent state spaces

Let $\mathcal{L}: \mathcal{S} \rightarrow \mathcal{S}^{\prime}$ be an invertible affine map. If all states are transformed as $\psi \rightarrow \mathcal{L}(\psi)$ and all effects on $\mathcal{S}$ are transformed as $\Omega \rightarrow \Omega \circ \mathcal{L}^{-1}$, then the outcome probabilities $\Omega(\psi)$ are left unchanged. Analogously, if all transformations on $\mathcal{S}$ are mapped as $T \rightarrow \mathcal{L} \circ T \circ \mathcal{L}^{-1}$, then their action on the states is the same. The new state space $\mathcal{S}^{\prime}$, together with the transformed effects and transformations, is then just a different representation of $\mathcal{S}$. In this case, we call $\mathcal{S}$ and $\mathcal{S}^{\prime}$ equivalent. In the new representation, the entries of $\psi$ need not be probabilities as in (1), but it may have other advantages. In this work, several representations are used.

In the standard formalism of QT, states are represented by density matrices; however, they can also be represented as in (1).

Changing the set of fiducial measurements is a particular type of $\mathcal{L}$-transformation. For example, if the components of the Bloch vector (of a quantum spin- $\frac{1}{2}$ particle) correspond to spin measurements in non-orthogonal directions, then the Bloch sphere becomes an ellipsoid.

### 2.6. Instances of generalized probabilistic theories

QT is an instance of GPT and can be specified as follows. The state space $\mathcal{S}_{c}$ with capacity $c$ is equivalent to the set of complex $c \times c$-matrices $\rho$ such that $\rho \geqslant 0$ and $\operatorname{tr} \rho=1$. This set has dimension $d_{c}=c^{2}-1$, and its pure states are rank-one. The effects on $\mathcal{S}_{c}$ have the form $\Omega(\rho)=\operatorname{tr}(M \rho)$, where $M$ is a complex $c \times c$-matrix such that $0 \leqslant M \leqslant \mathbb{I}$. The reversible transformations act as $\rho \rightarrow V \rho V^{\dagger}$ with $V \in \mathrm{SU}(c)$. The capacity of a composite system $A B$ is the product of the capacities for the subsystems $c_{A B}=c_{A} c_{B}$.

CPT is another instance of GPT, and can be specified as follows. The state space $\mathcal{S}_{c}$ with capacity $c$ is equivalent to the set of $c$-outcome probability distributions $[p(1), \ldots, p(c)]$, which has dimension $d_{c}=c-1$ (in geometric terms, each $\mathcal{S}_{c}$ is a simplex). The pure states are the deterministic distributions $p(a)=\delta_{a, b}$ with $b=1, \ldots, c$. The $c$-outcome measurement with effects $\Omega_{a}(\psi)=p(a)$ for $a=1, \ldots, c$ distinguishes the $c$ pure states; hence it is complete. Any other measurement is a function of this one. The reversible transformations act by permuting the entries of the state $[p(1), \ldots, p(c)]$. The capacity of a composite system is also $c_{A B}=c_{A} c_{B}$. Note that CPT can be obtained by restricting the states of QT to diagonal matrices. In other words, CPT is embedded in QT.

An instance of GPT that is not observed in nature is the generalized no-signaling theory [6], colloquially called boxworld. By definition, state spaces contain all correlations (3) satisfying the no-signaling constraints (2). Such state spaces have finitely many pure states, and some of them violate Bell inequalities more strongly than any quantum state [12]. The effects in boxworld are all generated by products of local effects. The group of reversible transformations consists only of relabeling of local measurements and their outcomes, permutations of subsystems and combinations thereof [14].

## 3. The requirements

This section contains the precise statement of the requirements, each followed by an explanation of its significance.

Requirement 1 (Finiteness). A state space with capacity $c=2$ has finite dimension $d$.
If this did not hold, the characterization of a state of a generalized bit would require infinitely many outcome probabilities, making state estimation impossible. It is shown below that this requirement, together with the others, implies that all state spaces with finite capacity $c$ have finite dimension.

Requirement 2 (Local tomography). The state of a composite system $A B$ is completely characterized by the statistics of measurements on the subsystems $A, B$.

In other words, state tomography [3] can be performed locally. This is equivalent to the constraint

$$
\begin{equation*}
\left(d_{A B}+1\right)=\left(d_{A}+1\right)\left(d_{B}+1\right) \tag{8}
\end{equation*}
$$

[3, 4]. This requirement can be recursively extended to more parties by letting subsystems $A, B$ be themselves composite.

Requirement 3 (Equivalence of subspaces). Let $\mathcal{S}_{c}$ and $\mathcal{S}_{c-1}$ be systems with capacities $c$ and $c-1$, respectively. If $\Omega_{1}, \ldots, \Omega_{c}$ is a complete measurement on $\mathcal{S}_{c}$, then the set of states $\psi \in \mathcal{S}_{c}$ with $\Omega_{c}(\psi)=0$ is equivalent to $\mathcal{S}_{c-1}$.

The notions of complete measurements and equivalent state spaces are defined in sections 2.2 and 2.5. In particular, the equivalence of $\mathcal{S}_{c-1}$ and

$$
\begin{equation*}
\mathcal{S}_{c-1}^{\prime}:=\left\{\psi \in \mathcal{S}_{c}: \Omega_{c}(\psi)=0\right\} \subset \mathcal{S}_{c} \tag{9}
\end{equation*}
$$

implies that all measurements and reversible transformations on one of them can be implemented on the other.

This requirement, first introduced in [4], implies that all state spaces with the same capacity are equivalent: if $\mathcal{S}_{c-1}$ and $\tilde{\mathcal{S}}_{c-1}$ are state spaces with capacity $c-1$, then both are equivalent to (9) and hence are equivalent to each other. In other words, the only property that characterizes the type of system is the capacity to carry information. If we start with $\mathcal{S}_{c}$ and apply requirement 3 recursively, we obtain a more general formulation: consider any subset of outcomes $\left\{a_{1}, \ldots, a_{c^{\prime}}\right\} \subseteq\{1, \ldots, c\}$ of the complete measurement $\Omega_{1}, \ldots, \Omega_{c}$; then the set of states $\psi \in \mathcal{S}_{c}$ with

$$
\begin{equation*}
\Omega_{a_{1}}(\psi)+\cdots+\Omega_{a_{c^{\prime}}}(\psi)=1 \tag{10}
\end{equation*}
$$

is equivalent to the state space $\mathcal{S}_{c^{\prime}}$ with capacity $c^{\prime}$. This provides an onion-like structure for all state spaces $\mathcal{S}_{1} \subset \mathcal{S}_{2} \subset \mathcal{S}_{3} \subset \cdots$.

The particular structure of QT simplifies the task of assigning a state space to a physical system or experimental setup. It is not necessary to consider all possible states of the system, but instead, the relevant ones for the context being analyzed need to be considered. For example, an atom is sometimes modeled with a state space having two distinguishable states ( $c=2$ ), even though its constituents have many more degrees of freedom. In particular, if we know that only two energy levels are populated with nonzero probability, we can ignore all others and effectively get a genuine quantum two-level state space. In a theory where this is not true, the effective state space might depend on how many unpopulated energy levels are ignored, or on the detailed internal state of the electron, for example. In order to avoid pathologies like this, we postulate requirement 3 .

Requirement 4 (Symmetry). For every pair of pure states $\psi_{1}, \psi_{2} \in \mathcal{S}$ there is a reversible transformation $G$ mapping one onto the other: $G\left(\psi_{1}\right)=\psi_{2}$.

The set of reversible transformations of a state space $\mathcal{S}_{c}$ forms a group, denoted $\mathcal{G}_{c}$. This group endows $\mathcal{S}_{c}$ with a symmetry, which makes all pure states equivalent. A group $\mathcal{G}_{c}$ is said to be continuous if it is topologically connected: any transformation is the composition of many infinitesimal ones [24]. Hardy invokes the continuity of time evolution in physical systems to justify the continuity of reversible transformations [3, 4]; in this case, state spaces $\mathcal{S}_{c}$ must have infinitely many pure states; this rules out CPT and singles out QT. However, all the analysis in this work is done without imposing continuity, since we find it very interesting that the only theory with state spaces having finitely many pure states, and satisfying the requirements, is CPT.

Requirement 5 (All measurements allowed). All effects on $\mathcal{S}_{2}$ are outcome probabilities of possible measurements.

It is shown below that, in combination with the other requirements, this implies that all effects on all state spaces (with arbitrary $c$ ) appear as outcome probabilities of measurements in the resulting theory. Note that requirement 5 has nontrivial consequences in conjunction with the other requirements: adding effects as allowed measurements to a physical theory extends the applicability of requirement 3 .

For completeness, we would like to mention that requirement 5 can be replaced by the following postulate, which was first put forward in an interesting paper that appeared after the completion of this work [33]. It calls a state 'completely mixed' if it is in the relative interior of state space. See lemma 9 in the appendix for details regarding how the proof of our main result has to be modified in this case.

Requirement $5^{\prime}$ [33]. If a state is not completely mixed, then there exists at least one state that can be perfectly distinguished from it.

## 4. Characterization of all theories satisfying the requirements

### 4.1. The maximally mixed state

We use the following notation: the system with capacity $c$ has a state space $\mathcal{S}_{c}$ with dimension $d_{c}$ and a group of reversible transformations $\mathcal{G}_{c}$. The group $\mathcal{G}_{c}$ is compact (section 2.3) and, hence, has a normalized invariant Haar measure [25]. This allows one to define the maximally mixed state

$$
\begin{equation*}
\mu_{c}=\int_{\mathcal{G}_{c}} G(\psi) \mathrm{d} G \in \mathcal{S}_{c}, \tag{11}
\end{equation*}
$$

where $\psi \in \mathcal{S}_{c}$ is an arbitrary pure state. It follows from requirement 4 that the resulting state $\mu_{c}$ does not depend on the choice of the pure state $\psi$. By construction, the maximally mixed state is invariant:

$$
\begin{equation*}
G\left(\mu_{c}\right)=\mu_{c} \text { for all } G \in \mathcal{G}_{c} . \tag{12}
\end{equation*}
$$

Moreover, lemma 1 shows that it is the only invariant state in $\mathcal{S}_{c}$ (this lemma and all others are stated and proved in the appendix).


Figure 2. The left panel is a state space whose boundary consists of facets (like $\Omega=1$ ). Each facet contains infinitely many states ( $\Omega=1$ contains $\psi_{1}, \psi_{2}$ and all $\left.\psi_{\text {mix }}=q \psi_{1}+(1-q) \psi_{2}\right)$. The right panel is a state space whose boundary has no facets. Any state space has supporting hyperplanes containing a unique state (like $\Omega_{\text {one }}=1$ in both figures).

### 4.2. The generalized bit

A generalized bit is a system with capacity two. For any state $\psi \in \mathcal{S}_{2}$ in the standard representation (1), its Bloch representation is defined by

$$
\hat{\psi}=2\left[\begin{array}{c}
p\left(x_{1}\right)-\mu_{2}^{1}  \tag{13}\\
\vdots \\
p\left(x_{d_{2}}\right)-\mu_{2}^{d_{2}}
\end{array}\right] \in \hat{\mathcal{S}}_{2} \subset \mathbb{R}^{d_{2}} .
$$

States in the Bloch representation do not have the redundant component $\psi^{0}$, so equations (4), (5) and (7) become less simple. The invertible map $\mathcal{L}: \mathcal{S}_{2} \rightarrow \hat{\mathcal{S}}_{2}$ is affine but not linear; hence, effects $\Omega$ in the Bloch representation ( $\hat{\Omega}=\Omega \circ \mathcal{L}^{-1}$ ) are affine but not necessarily linear. The same applies to transformations ( $\hat{G}=\mathcal{L} \circ G \circ \mathcal{L}^{-1}$ ); however, the maximally mixed state in the Bloch representation is the null vector $\hat{\mu}_{2}=\mathbf{0}$; therefore (12) becomes $\hat{G}(\mathbf{0})=\mathbf{0}$, which implies that $\hat{G}$ acts linearly (as a matrix).
Theorem 1. A state in $\hat{\mathcal{S}}_{2}$ is pure if and only if it belongs to the boundary $\partial \hat{\mathcal{S}}_{2}$.

Proof. In any convex set, pure states belong to the boundary [23]. Let us see the converse.
It is shown in [26] that any compact convex set has a supporting hyperplane containing exactly one point of the set. Translated into our language: there is a tight effect $\hat{\Omega}_{\text {one }}$ on $\hat{\mathcal{S}}_{2}$ such that only one state $\hat{\varphi}_{\text {one }} \in \hat{\mathcal{S}}_{2}$ satisfies $\hat{\Omega}_{\text {one }}\left(\hat{\varphi}_{\text {one }}\right)=1$; this is illustrated in figure 2 . According to requirement 5 , the effect $\hat{\Omega}_{\text {one }}$ corresponds to a valid measurement outcome, and so does $\hat{\mathbf{1}}-\hat{\Omega}_{\text {one }}$, where $\hat{\mathbf{1}}(\hat{\psi})=1$ for all $\hat{\psi} \in \hat{\mathcal{S}}_{2}$. Thus, the two effects, $\hat{\Omega}_{\text {one }}$ and $\hat{\mathbf{1}}-\hat{\Omega}_{\text {one }}$, define a complete measurement on $\hat{\mathcal{S}}_{2}$. Imposing requirement 3 on the single outcome $\hat{\Omega}_{\text {one }}$ constrains the state space with unit capacity $\hat{\mathcal{S}}_{1}$ to contain only one state.

Suppose there is a point in the boundary $\hat{\varphi}_{\text {mix }} \in \partial \hat{\mathcal{S}}_{2}$ that is not pure: $\hat{\varphi}_{\text {mix }}=q \hat{\varphi}_{1}+(1-q) \hat{\varphi}_{2}$ with $\hat{\varphi}_{1} \neq \hat{\varphi}_{2}$ and $0<q<1$. Every point in the boundary of a compact convex set has a supporting hyperplane that contains it [23]. In our language: there is a tight effect $\hat{\Omega}$ on $\hat{\mathcal{S}}_{2}$ such that $\hat{\Omega}\left(\hat{\varphi}_{\text {mix }}\right)=1$. The affine function $\hat{\Omega}$ is bounded: $\hat{\Omega}(\hat{\varphi}) \leqslant 1$ for any $\hat{\varphi} \in \hat{\mathcal{S}}_{2}$, which implies $\hat{\Omega}\left(\hat{\varphi}_{1}\right)=\hat{\Omega}\left(\hat{\varphi}_{2}\right)=1$; this is illustrated in figure 2 . Like $\hat{\Omega}_{\text {one }}$, the effect $\hat{\Omega}$ defines a complete measurement, and requirement 3 can be imposed on the single outcome $\hat{\Omega}$, implying that $\hat{\mathcal{S}}_{1}$
contains more than one state. This is in contradiction to the previous paragraph; hence, all points in the boundary are pure.

For the case $d_{2}=1$, the state space $\mathcal{S}_{2}$ is a segment (a 1D ball) and hence the previous and next theorems are trivial. For $d_{2}>1$, the previous theorem implies that $\mathcal{S}_{2}$ contains infinitely many pure states. The next theorem recovers the (quantum-like) Bloch sphere with a yet unknown dimension, $d_{2}$.

Theorem 2. There is a set of fiducial measurements for which $\hat{\mathcal{S}}_{2}$ is a $d_{2}$-dimensional unit ball.

Proof. Lemma 2 shows that there is an invertible real matrix $S$ such that for each $\hat{G} \in \hat{\mathcal{G}_{2}}$ the matrix $S \hat{G} S^{-1}$ is orthogonal. Let us redefine the set $\hat{\mathcal{S}}_{2}$ by transforming the states as $\hat{\varphi} \rightarrow \hat{\varphi}^{\prime}=q S \hat{\varphi}$, where the number $q>0$ is chosen such that all pure states are unit vectors $\left|\hat{\varphi}^{\prime}\right|^{2}=\hat{\varphi}^{\prime \top} \hat{\varphi}^{\prime}=1$. This is possible because in the transformed state space, all pure states are related by orthogonal matrices $\left(S G S^{-1}\right)$ that preserve the norm. Since theorem 1 also applies to the redefined set $\hat{\mathcal{S}}_{2}^{\prime}$, it must be a unit ball. In what follows, we define a new set of fiducial measurements $x_{i}^{\prime}$ such that the Bloch representation (13) associated with the new fiducial probabilities $p\left(x_{i}^{\prime}\right)$ coincides with the redefinition $\hat{\varphi}^{\prime}$.

Requirement 5 reveals that in $\hat{\mathcal{S}}_{2}^{\prime}$, all tight effects are allowed measurements. For each unit vector $\hat{v} \in \mathbb{R}^{d_{2}}$, the function $\left.\hat{\Omega}_{\hat{\hat{v}}} \hat{\varphi}^{\prime}\right)=\left(1+\hat{\nu}^{\mathrm{T}} \hat{\varphi}^{\prime}\right) / 2$ is a tight effect on the unit ball, and conversely, all tight effects on the unit ball are of this form. The new set of fiducial measurements $x_{i}^{\prime}$ has effects $\hat{\Omega}_{x_{i}^{\prime}}=\hat{\Omega}_{\hat{\nu}_{i}}$, where

$$
\hat{v}_{1}=\left[\begin{array}{c}
1  \tag{14}\\
0 \\
\vdots \\
0
\end{array}\right], \quad \hat{v}_{2}=\left[\begin{array}{c}
0 \\
1 \\
\vdots \\
0
\end{array}\right], \quad \ldots, \quad \hat{v}_{d_{2}}=\left[\begin{array}{c}
0 \\
0 \\
\vdots \\
1
\end{array}\right]
$$

is a fixed orthonormal basis for $\mathbb{R}^{d_{2}}$. For any state $\hat{\varphi}^{\prime}$ the new fiducial probabilities are $p\left(x_{i}^{\prime}\right)=$ $\left.\Omega_{x_{i}^{\prime}} \hat{\varphi}^{\prime}\right)=\left(1+\hat{\varphi}^{\prime i}\right) / 2$, which implies $\hat{\varphi}^{\prime i}=2\left[p\left(x_{i}^{\prime}\right)-1 / 2\right]$. This is just (13) with the new fiducial measurements (note that $\hat{\mu}_{2}^{\prime}=\mathbf{0}$ and $\mu_{2}^{\prime i}=\hat{\Omega}_{x_{i}^{\prime}}(\mathbf{0})=1 / 2$ ).

In the rest of the paper, we will use the representation derived in theorem 2 above, where the generalized bit is represented by a unit ball. Moreover, we will drop the prime in $\hat{\mathcal{S}}_{2}^{\prime}, x_{i}^{\prime}, \hat{\varphi}^{\prime}$ used in the proof, and simply write $\hat{\mathcal{S}}_{2}, x_{i}, \hat{\varphi}$.

As argued above, for each pure state $\varphi \in \mathcal{S}_{2}$, there is a binary measurement with associated effect

$$
\begin{equation*}
\Omega_{\varphi}(\psi)=\left(1+\hat{\varphi}^{\mathrm{T}} \hat{\psi}\right) / 2 \tag{15}
\end{equation*}
$$

such that $\hat{\Omega}_{\varphi}(\hat{\varphi})=1$ and $\hat{\Omega}_{\varphi}(-\hat{\varphi})=0$. In summary, there is a correspondence between tight effects and pure states in $\mathcal{S}_{2}$, and each pure state belongs to a distinguishable pair, $\{\hat{\varphi},-\hat{\varphi}\}$.

### 4.3. Capacity and dimension

Requirements 1-3 imply that a state space with finite capacity $c$ has finite dimension $d_{c}$, which generalizes requirement 1 . To show this, consider a system composed of $m$ generalized bits,
with state space denoted by $\mathcal{S}_{2 \times m}$. Since $d_{2}$ is finite, equation (8) implies that $\mathcal{S}_{2 \times m}$ has finite dimension. Owing to the fact that perfectly distinguishable states are linearly independent, its capacity, denoted by $c_{m}$, must be finite, too. Since systems with the same capacity are equivalent, we must have $c_{m} \neq c_{n}$ for $m \neq n$, and the sequence of integers $c_{1}, c_{2}, \ldots$ is unbounded. For any capacity $c$ there is a value of $m$ such that $c \leqslant c_{m}$; hence, by requirement 3 we have $\mathcal{S}_{c} \subset \mathcal{S}_{2 \times m}$, which implies that $\mathcal{S}_{c}$ is finite dimensional.

In QT, the maximally mixed state (11) has two convenient properties. The first property: if $\mu_{A}$ and $\mu_{B}$ are the maximally mixed states of systems $A$ and $B$, then the maximally mixed state of the composite system $A B$ is

$$
\begin{equation*}
\mu_{A B}=\mu_{A} \otimes \mu_{B} . \tag{16}
\end{equation*}
$$

The second property: in the state space $\mathcal{S}_{c}$, there are $c$ pure distinguishable states $\psi_{1}, \ldots, \psi_{c} \in$ $\mathcal{S}_{c}$ such that

$$
\begin{equation*}
\mu_{c}=\frac{1}{c} \sum_{a=1}^{c} \psi_{a} . \tag{17}
\end{equation*}
$$

Lemmas 3 and 5 show that these two properties hold for every theory satisfying our requirements. The following theorem exploits these properties to show that the capacity is multiplicative (one of the axioms in [4]).

Theorem 3. If $c_{A}$ and $c_{B}$ are the capacities of systems $A$ and $B$, then the capacity of the composite system $A B$ is

$$
\begin{equation*}
c_{A B}=c_{A} c_{B} \tag{18}
\end{equation*}
$$

Proof. Equation (17) allows one to write the maximally mixed states of systems $A$ and $B$ as

$$
\mu_{A}=\frac{1}{c_{A}} \sum_{a=1}^{c_{A}} \varphi_{a}^{A} \quad \mu_{B}=\frac{1}{c_{B}} \sum_{b=1}^{c_{B}} \varphi_{b}^{B},
$$

where $\varphi_{1}^{A}, \ldots, \varphi_{c_{A}}^{A} \in \mathcal{S}_{A}$ are pure and distinguishable, and $\varphi_{1}^{B}, \ldots, \varphi_{c_{B}}^{B} \in \mathcal{S}_{B}$ are pure and distinguishable, too. This and equation (16) imply

$$
\begin{equation*}
\mu_{A B}=\mu_{A} \otimes \mu_{B}=\frac{1}{c_{A} c_{B}} \sum_{a=1}^{c_{A}} \sum_{b=1}^{c_{B}} \varphi_{a}^{A} \otimes \varphi_{b}^{B} . \tag{19}
\end{equation*}
$$

All states $\varphi_{a}^{A} \otimes \varphi_{b}^{B} \in \mathcal{S}_{A B}$ can be distinguished with the tensor-product measurement; therefore

$$
\begin{equation*}
c_{A B} \geqslant c_{A} c_{B} . \tag{20}
\end{equation*}
$$

Let $\left(\Omega_{1}, \ldots, \Omega_{c_{A B}}\right)$ be a complete measurement on $A B$ that distinguishes the states $\psi_{1}, \ldots, \psi_{c_{A B}} \in \mathcal{S}_{A B}$; that is, $\Omega_{k}\left(\psi_{k^{\prime}}\right)=\delta_{k, k^{\prime}}$. According to lemma 4 these states can be chosen to be pure. Since $\sum_{k=1}^{c_{A B}} \Omega_{k}\left(\mu_{A B}\right)=1$, there is at least one value of $k$, denoted by $k_{0}$, such that

$$
\begin{equation*}
\Omega_{k_{0}}\left(\mu_{A B}\right) \leqslant 1 / c_{A B} . \tag{21}
\end{equation*}
$$

The product of pure states $\varphi_{1}^{A} \otimes \varphi_{1}^{B}$ is pure [3]; hence, requirement 4 reveals that there is a reversible transformation $G \in \mathcal{G}_{A B}$ such that $G\left(\psi_{k_{0}}\right)=\varphi_{1}^{A} \otimes \varphi_{1}^{B}$. The measurement
$\left(\Omega_{1} \circ G^{-1}, \ldots, \Omega_{c_{A B}} \circ G^{-1}\right)$ distinguishes the states $G\left(\psi_{1}\right), \ldots, G\left(\psi_{c_{A B}}\right)$. Inequality (21), the invariance of $\mu_{A B}$, expansion (19), the positivity of probabilities and $\left(\Omega_{k_{0}} \circ G^{-1}\right)\left(\varphi_{1}^{A} \otimes \varphi_{1}^{B}\right)=1$ imply

$$
\begin{aligned}
\frac{1}{c_{A B}} & \geqslant\left(\Omega_{k_{0}} \circ G^{-1}\right)\left(\mu_{A B}\right) \\
& =\frac{1}{c_{A} c_{B}} \sum_{a, b}\left(\Omega_{k_{0}} \circ G^{-1}\right)\left(\varphi_{a}^{A} \otimes \varphi_{b}^{B}\right) \geqslant \frac{1}{c_{A} c_{B}} .
\end{aligned}
$$

This and (20) imply (18).
It is shown in [4] that the two multiplicativity formulae (8) and (18) imply the existence of a positive integer $r$ such that: for any $c$ the state space $\mathcal{S}_{c}$ has dimension

$$
\begin{equation*}
d_{c}=c^{r}-1 \tag{22}
\end{equation*}
$$

The integer $r$ is a constant of the theory, with values $r=1$ for CPT and $r=2$ for QT.

### 4.4. Recovering classical probability theory

Let us consider all theories with $d_{2}=1$. In this case, equation (22) becomes $d_{c}=c-1$. In [4], it is shown that the only GPT with this relation between capacity and dimension is CPT, as described in section 2.6. We reproduce the proof for completeness.

Theorem 4. The only GPT with $d_{2}=1$ satisfying requirements $1-5$ is classical probability theory.

Proof. Let $\mathcal{S}_{c}$ be a state space and $\left(\Omega_{1}, \ldots, \Omega_{c}\right)$ a complete measurement that distinguishes the states $\psi_{1}, \ldots, \psi_{c} \in \mathcal{S}_{c}$. The vectors $\psi_{1}, \ldots, \psi_{c} \in \mathbb{R}^{c}$ are linearly independent; otherwise $\psi_{a}=\sum_{b \neq a} t_{b} \psi_{b}$ and $1=\Omega_{a}\left(\psi_{a}\right)=\sum_{b \neq a} t_{b} \Omega_{a}\left(\psi_{b}\right)=0$ gives a contradiction. Therefore, any state $\psi \in \mathcal{S}_{c} \subseteq \mathbb{R}^{c}$ can be written in this basis $\psi=\sum_{a} q_{a} \psi_{a}$, where $q_{a}=\Omega_{a}(\psi)$ turns out to be the probability of outcome $a$. The numbers $\left(q_{1}, \ldots, q_{c}\right)$ constitute a probability distribution; hence, there is a one-to-one correspondence between states in $\mathcal{S}_{c}$ and $c$-outcome probability distributions. This kind of set is called a $d_{c}$-simplex. A similar argument shows that the effects $\Omega_{1}, \ldots, \Omega_{c}$ are linearly independent. Hence, any effect $\Omega$ on $\mathcal{S}_{c}$ can be written as $\Omega=\sum_{a} h_{a} \Omega_{a}$, and the constraint $0 \leqslant \Omega\left(\psi_{a}\right) \leqslant 1$ implies $0 \leqslant h_{a} \leqslant 1$. In other words, every measurement on $\mathcal{S}_{c}$ is generated by the complete one.

Every reversible transformation on $\mathcal{S}_{c}$ is a symmetry of the $d_{c}$-simplex, i.e. a permutation of pure states. Due to requirement 4 , there is a reversible transformation on the bit $\mathcal{S}_{2}$ that exchanges the two pure states. Using requirement 3 inductively: if there is a transformation on $\mathcal{S}_{c-1}$ that exchanges two pure states and leaves the rest invariant, this transposition can be implemented on $\mathcal{S}_{c}$, also leaving all other pure states invariant. Therefore, all transpositions can be implemented in $\mathcal{S}_{c}$, and those generate the full group of permutations.

### 4.5. Reversible transformations for the generalized bit

In the rest of the paper, only theories with $d_{2}>1$ are considered. Theorem 2 shows that $\hat{\mathcal{S}}_{2}$ is a $d_{2}$-dimensional unit ball. Equation (22) for $c=2$ implies that $d_{2}$ is odd. The pure states in $\mathcal{S}_{2}$ are
the unit vectors in $\mathbb{R}^{d_{2}}$. A reversible transformation $\hat{G} \in \hat{\mathcal{G}}_{2}$ maps pure states onto pure states; hence it preserves the norm and has to be an orthogonal matrix $\hat{G}^{\mathrm{T}}=\hat{G}^{-1}$. Therefore, $\hat{\mathcal{G}}_{2}$ is a subgroup of the orthogonal group $\mathrm{O}\left(d_{2}\right)$.

Requirement 4 imposes that for any pair of unit vectors $\hat{\varphi}, \hat{\varphi}^{\prime}$ there is $\hat{G} \in \hat{\mathcal{G}}_{2}$ such that $\hat{G}(\hat{\varphi})=\hat{\varphi}^{\prime}$. In other words, $\hat{\mathcal{G}}_{2}$ is transitive on the sphere [27, 28]. According to lemma 6 , if $\hat{\mathcal{G}_{2}}$ is transitive on the sphere, then the largest connected subgroup $\hat{\mathcal{C}}_{2} \subseteq \hat{\mathcal{G}}_{2}$ is also transitive on the sphere. The matrix group $\hat{\mathcal{C}}_{2}$ is compact and connected and hence it is a Lie group (theorem 7.31 in [24]). A classification of all connected compact Lie groups that are transitive on the sphere is presented in [27, 28]. For odd $d_{2}$, the only possibility is $\hat{\mathcal{C}}_{2}=\mathrm{SO}\left(d_{2}\right)$, except for $d_{2}=7$ where there are additional possibilities: $\hat{\mathcal{C}}_{2}=M \mathbb{G}_{2} M^{\mathrm{T}} \subset \mathrm{SO}(7)$ for any $M \in \mathrm{O}(7)$, where $\mathbb{G}_{2}$ is the fundamental representation of the smallest exceptional Lie group [29]. For even $d_{2}$, there are many more possibilities [27, 28], but equation (22) implies that $d_{2}$ must be odd.

The stabilizer of the vector $\hat{v}_{1}$ defined in (14) is the subgroup $\hat{\mathcal{H}}_{2}=\left\{\hat{G} \in \hat{\mathcal{G}}_{2}: \hat{G}\left(\hat{v}_{1}\right)=\hat{v}_{1}\right\}$. Each transformation $\hat{H} \in \hat{\mathcal{H}}_{2}$ has the form

$$
\hat{H}=\left(\begin{array}{cc}
1 & \mathbf{0}^{\mathrm{T}} \\
\mathbf{0} & \bar{H}
\end{array}\right)
$$

where $\bar{H} \in \overline{\mathcal{H}}_{2}$ is the nontrivial part. If $\hat{\mathcal{C}}_{2}=\mathrm{SO}\left(d_{2}\right)$, then $\mathrm{SO}\left(d_{2}-1\right) \subseteq \overline{\mathcal{H}}_{2}$. In the case $d_{2}=7$, if $\hat{\mathcal{C}}_{2}=M \mathbb{G}_{2} M^{\mathrm{T}}$ then $\overline{\mathcal{H}}_{2}$ contains (up to the similarity $M$ ) the real 6 D representation of $\mathrm{SU}(3)$ given by

$$
\bar{H}=\left(\begin{array}{cc}
\operatorname{re} U & \operatorname{im} U  \tag{23}\\
-\operatorname{im} U & \operatorname{re} U
\end{array}\right)
$$

where re $U$ and $\operatorname{im} U$ are the real and imaginary parts of $U \in \mathrm{SU}(3)$ (see exercise 22.27 in [29]).

### 4.6. Two generalized bits

The joint state space of two $\mathcal{S}_{2}$ systems is denoted by $\mathcal{S}_{2,2}$. The multiplicativity of the capacity (18) implies that $\mathcal{S}_{2,2}$ is equivalent to $\mathcal{S}_{4}$. However, we write $\mathcal{S}_{2,2}$ to emphasize the bipartite structure.

In what follows, instead of using the standard representation for bipartite systems (3) we generalize the Bloch representation to two generalized bits. A state $\psi_{A B} \in \mathcal{S}_{2,2}$ has the Bloch representation $\hat{\psi}_{A B}=[\alpha, \beta, C]$ with

$$
\begin{align*}
& \alpha^{i}=2 p\left(x_{i}\right)-1, \\
& \beta^{j}=2 p\left(y_{j}\right)-1,  \tag{24}\\
& C^{i j}=4 p\left(x_{i}, y_{j}\right)-2 p\left(x_{i}\right)-2 p\left(y_{j}\right)+1
\end{align*}
$$

for $i, j=1, \ldots, d_{2}$. Note that $\alpha=\hat{\psi}_{A}$ and $\beta=\hat{\psi}_{B}$ are the reduced states in the Bloch representation (13). The correlation matrix can also be written as $C^{i j}=p\left(x_{i}, y_{j}\right)-p\left(x_{i}, \bar{y}_{j}\right)-$ $p\left(\bar{x}_{i}, y_{j}\right)+p\left(\bar{x}_{i}, \bar{y}_{j}\right)$ and characterizes the correlations between subsystems. Product states have the Bloch representation

$$
\begin{equation*}
\left(\varphi_{A} \otimes \varphi_{B}\right)^{\wedge}=\left[\hat{\varphi}_{A}, \hat{\varphi}_{B}, \hat{\varphi}_{A} \hat{\varphi}_{B}^{\mathrm{T}}\right] \tag{25}
\end{equation*}
$$

with a rank-one correlation matrix. In QT, where $d_{2}=3$, two-qubit density matrices are often represented by $[\alpha, \beta, C]$ through formula (41). Definition (24) implies

$$
\begin{equation*}
-1 \leqslant \alpha^{i}, \beta^{j}, C^{i j} \leqslant 1 \tag{26}
\end{equation*}
$$

The invertible map $\mathcal{L}\left[\psi_{A B}\right]=\hat{\psi}_{A B}$ defined by (24) also determines the Bloch representation of effects $\hat{\Omega}=\Omega \circ \mathcal{L}^{-1}$. In particular, the tensor product of two effects of the form (15) is

$$
\begin{equation*}
\left(\Omega_{\varphi_{A}} \otimes \Omega_{\varphi_{B}}\right)^{\wedge}[\alpha, \beta, C]=\left(1+\hat{\varphi}_{A}^{\mathrm{T}} \alpha+\hat{\varphi}_{B}^{\mathrm{T}} \beta+\hat{\varphi}_{A}^{\mathrm{T}} C \hat{\varphi}_{B}\right) / 4 \tag{27}
\end{equation*}
$$

The map $\mathcal{L}$ also determines the action of reversible transformation in the Bloch representation. Since $\mathcal{L}$ is affine but not linear, the action $\hat{G}=\mathcal{L} \circ G \circ \mathcal{L}^{-1}$ need not be linear. Identities (16, 25) and $\hat{\mu}_{2}=\mathbf{0}$ imply that the maximally mixed state in $\hat{\mathcal{S}}_{2,2}$ is $\hat{\mu}_{2,2}=\left(\hat{\mu}_{2}, \hat{\mu}_{2}, \hat{\mu}_{2} \hat{\mu}_{2}^{\mathrm{T}}\right)=\mathbf{0}$. This and (12) imply that transformations $\hat{\mathcal{G}}_{2,2}$ act on the generic vector $[\alpha, \beta, C]$ as matrices. In particular, local transformations $G_{A}, G_{B} \in \mathcal{G}_{2}$ act as

$$
\begin{equation*}
\left(G_{A} \otimes G_{B}\right)^{\wedge}[\alpha, \beta, C]=\left[\hat{G}_{A} \alpha, \hat{G}_{B} \beta, \hat{G}_{A} C \hat{G}_{B}^{\mathrm{T}}\right] . \tag{28}
\end{equation*}
$$

Section 4.5 concludes that $\hat{\mathcal{G}}_{2}$ consists of orthogonal matrices, and lemma 8 shows that all transformations in $\hat{\mathcal{G}}_{2,2}$ are orthogonal, too. Orthogonal matrices preserve the norm of vectors; therefore all pure states $\psi \in \mathcal{S}_{2,2}$ satisfy

$$
\begin{equation*}
|\hat{\psi}|^{2}=|\alpha|^{2}+|\beta|^{2}+\operatorname{tr}\left(C^{\mathrm{T}} C\right)=3 . \tag{29}
\end{equation*}
$$

The constant in the right-hand side can be obtained by letting $\hat{\psi}=\left[\alpha, \alpha, \alpha \alpha^{\mathrm{T}}\right]$ with $|\alpha|=1$.

### 4.7. Consistency in the subspaces of two generalized bits

In this subsection, we use a trick introduced in [19]: to impose the equivalence between a particular subspace of $\mathcal{S}_{2,2}$ and $\mathcal{S}_{2}$ (requirement 3).

Consider the unit vector $\hat{v}_{1}$ from (14) and the two distinguishable pure states $\hat{\varphi}_{0}=\hat{v}_{1}$ and $\hat{\varphi}_{1}=-\hat{v}_{1}$ from $\hat{\mathcal{S}}_{2}$. The four pure states $\varphi_{a, b}=\varphi_{a} \otimes \varphi_{b} \in \mathcal{S}_{2,2}$ can be distinguished with the complete measurement $\Omega_{a, b}=\Omega_{\varphi_{a}} \otimes \Omega_{\varphi_{b}}$, where $a, b \in\{0,1\}$. Formula (27) implies that

$$
\begin{align*}
& \hat{\Omega}_{0,0}[\alpha, \beta, C]=\left(1+\alpha^{1}+\beta^{1}+C^{1,1}\right) / 4  \tag{30}\\
& \hat{\Omega}_{1,1}[\alpha, \beta, C]=\left(1-\alpha^{1}-\beta^{1}+C^{1,1}\right) / 4 \tag{31}
\end{align*}
$$

Requirement 3 implies that the subspace

$$
\mathcal{S}_{2}^{\prime}=\left\{\psi \in \mathcal{S}_{2,2}:\left(\Omega_{0,0}+\Omega_{1,1}\right)(\psi)=1\right\}
$$

is equivalent to $\mathcal{S}_{2}$. By adding (30) and (31), it becomes clear that a state $\hat{\psi}=[\alpha, \beta, C]$ belongs to $\hat{\mathcal{S}}_{2}^{\prime}$ if and only if $C^{1,1}=1$. Moreover, if $\psi \in \mathcal{S}_{2}^{\prime}$, then it follows from $\hat{\Omega}_{0,1}(\hat{\psi}) \geqslant 0$ and $\hat{\Omega}_{1,0}(\hat{\psi}) \geqslant 0$ that $\alpha^{1}=\beta^{1}$.

Theorem 5. The state space of a generalized bit has dimension three $\left(d_{2}=3\right)$.

Proof. Recall that the case under consideration is odd $d_{2}$ larger than one. The space $\mathcal{S}_{2}^{\prime} \subset \mathcal{S}_{2,2}$ is equivalent to $\mathcal{S}_{2}$, which is a $d_{2}$-dimensional unit ball. If $\varphi_{0,0}$ and $\varphi_{1,1}$ are considered the poles
of this ball, then the equator is the set of states $\psi_{\text {eq }}$ such that $\Omega_{0,0}\left(\psi_{\text {eq }}\right)=\Omega_{1,1}\left(\psi_{\text {eq }}\right)=1 / 2$. Equations (30) and (31) reveal that equator states have $\alpha^{1}=\beta^{1}=0$, and then

$$
\hat{\psi}_{\mathrm{eq}}=\left[\binom{0}{\bar{\alpha}},\binom{0}{\bar{\beta}},\left(\begin{array}{cc}
1 & \bar{\tau}^{\mathrm{T}}  \tag{32}\\
\bar{\gamma} & \bar{C}
\end{array}\right)\right],
$$

where $\bar{\alpha}, \bar{\beta}, \bar{\gamma}, \bar{\tau} \in \mathbb{R}^{d_{2}-1}$ and $\bar{C} \in \mathbb{R}^{\left(d_{2}-1\right) \times\left(d_{2}-1\right)}$. Consider the action of $G_{A} \otimes \mathbb{I}$ for $G_{A} \in \mathcal{G}_{2}$ on an equator state, $\psi_{\text {eq. }}$. Since $\hat{\mathcal{G}}_{2}$ is transitive on the unit sphere, so if $\bar{\gamma} \neq \mathbf{0}$, then there is some $\hat{G}_{A} \in \hat{\mathcal{G}}_{2}$ such that the correlation matrix transforms into

$$
\hat{G}_{A}\left(\begin{array}{cc}
1 & \bar{\tau}^{\mathrm{T}} \\
\bar{\gamma} & \bar{C}
\end{array}\right)=\left(\begin{array}{cc}
\sqrt{1+|\bar{\gamma}|^{2}} & ? \\
\mathbf{0} & ?
\end{array}\right),
$$

which is in contradiction to (26). Therefore $\bar{\gamma}=\mathbf{0}$, and by a similar argument $\bar{\tau}=\mathbf{0}$.
The stabilizer of $\hat{v}_{1}$ is the largest subgroup $\hat{\mathcal{H}}_{2} \subset \hat{\mathcal{G}}_{2}$, which leaves $\hat{v}_{1}$ invariant (section 4.5). For any pair $H_{A}, H_{B} \in \mathcal{H}_{2}$ the identity $\Omega_{a, b} \circ\left(H_{A} \otimes H_{B}\right)=\Omega_{a, b}$ holds, which implies that if $\psi_{\mathrm{eq}}$ belongs to the equator (32), then

$$
\left(H_{A} \otimes H_{B}\right)\left(\psi_{\mathrm{eq}}\right)^{\wedge}=\left[\binom{0}{\bar{H}_{A} \bar{\alpha}},\binom{0}{\bar{H}_{B} \bar{\beta}},\left(\begin{array}{cc}
1 & \mathbf{0}^{\mathrm{T}} \\
\mathbf{0} & \bar{H}_{A} \bar{C} \bar{H}_{B}^{\mathrm{T}}
\end{array}\right)\right]
$$

also belongs to the equator. The equator is a unit ball of dimension $d_{2}-1$. Since the set

$$
\begin{equation*}
\left\{\left(H_{A} \otimes H_{B}\right)\left(\psi_{\mathrm{eq}}\right): H_{A}, H_{B} \in \mathcal{H}_{2}\right\} \tag{33}
\end{equation*}
$$

is a subset of the equator, the dimension of its affine span is at most $d_{2}-1$.
Consider the case $\bar{C}=\mathbf{0}$. The normalization condition (29) implies $|\bar{\alpha}|=|\bar{\beta}|=1$. The set $\left\{\bar{H}_{A} \bar{\alpha}: \bar{H}_{A} \in \overline{\mathcal{H}}_{2}\right\}$ has dimension $d_{2}-1$, and the same for $\left\{\bar{H}_{B} \bar{\beta}: \bar{H}_{B} \in \overline{\mathcal{H}}_{2}\right\}$. Therefore, the set (33) has dimension at least $2\left(d_{2}-1\right)$, generating a contradiction.

Consider the case $\bar{C} \neq \mathbf{0}$. The group action on $\bar{C}$ corresponds to the exterior tensor product $\overline{\mathcal{H}}_{2} \boxtimes \overline{\mathcal{H}}_{2}=\left\{\bar{H}_{A} \otimes \bar{H}_{B}: \bar{H}_{A}, \bar{H}_{B} \in \overline{\mathcal{H}}_{2}\right\}$. If $d_{2}>3$ and $\mathrm{SO}\left(d_{2}-1\right) \subseteq \overline{\mathcal{H}}_{2}$, then $\overline{\mathcal{H}}_{2}$ is irreducible in $\mathbb{C}^{d_{2}-1}$, and a simple character-based argument shows that $\overline{\mathcal{H}}_{2} \boxtimes \overline{\mathcal{H}}_{2}$ is irreducible in $\left(\mathbb{C}^{d_{2}-1}\right)^{\otimes 2}$ (see p 427 in [29]). Hence, the set

$$
\begin{equation*}
\left\{\bar{H}_{A} \bar{C} \bar{H}_{B}^{\mathrm{T}}: \bar{H}_{A}, \bar{H}_{B} \in \overline{\mathcal{H}}_{2}\right\} \tag{34}
\end{equation*}
$$

has dimension $\left(d_{2}-1\right)^{2}$, which conflicts with the dimensionality requirements of (33). If $d_{2}=7$ and $\overline{\mathcal{H}}_{2}$ contains the representation of $\operatorname{SU}(3)$ given in (23), then the subgroup

$$
\bar{H}=\left(\begin{array}{cc}
U & \mathbf{0} \\
\mathbf{0} & U
\end{array}\right)
$$

with $U \in \mathrm{SO}(3) \subset \mathrm{SU}(3)$ has two invariant $\mathbb{C}^{3}$ subspaces. Therefore, the invariant subspaces of $\overline{\mathcal{H}}_{2} \boxtimes \overline{\mathcal{H}}_{2}$ have at least dimension 9 , and independently of $\bar{C}$, the set (34) has at least dimension 9 , which is in conflict with the dimensionality requirements of (33). So the only possibility is $d_{2}=3$.

From now on, only the case $d_{2}=3$ is considered. Section 4.5 reveals that $\mathrm{SO}(3) \subseteq \hat{\mathcal{G}}_{2} \subseteq \mathrm{O}(3)$, which implies that either $\hat{\mathcal{G}}_{2}=\mathrm{O}(3)$ and $\overline{\mathcal{H}}_{2}=\mathrm{O}(2)$, or $\hat{\mathcal{G}}_{2}=\mathrm{SO}(3)$ and $\overline{\mathcal{H}}_{2}=\mathrm{SO}(2)$.

Let us show that the first case is impossible. The group $\overline{\mathcal{H}}_{2}=\mathrm{O}(2)$ is irreducible in $\mathbb{C}^{2}$; therefore, $\overline{\mathcal{H}}_{2} \boxtimes \overline{\mathcal{H}}_{2}$ is irreducible in $\left(\mathbb{C}^{2}\right)^{\otimes 2}$. Three paragraphs above, it is shown that $\bar{C} \neq \mathbf{0}$, hence the set (34) has dimension $\left(d_{2}-1\right)^{2}$, which is a lower bound for the one of (33), which is larger than the allowed one $\left(d_{2}-1=2\right)$.

Let us address the second case. The group $\overline{\mathcal{H}}_{2}=\mathrm{SO}(2)$ is irreducible in $\mathbb{R}^{2}$ but reducible in $\mathbb{C}^{2}$; so the previous argument does not hold. The vector space of $2 \times 2$ real matrices decomposes into the subspace generated by rotations

$$
R_{+}=\left(\begin{array}{cc}
\cos v & \sin v  \tag{35}\\
-\sin v & \cos v
\end{array}\right),
$$

and the one generated by reflections

$$
R_{-}=\left(\begin{array}{cc}
\cos v & \sin v  \tag{36}\\
\sin v & -\cos v
\end{array}\right),
$$

where $\operatorname{det} R_{ \pm}= \pm 1$. For any pair $\bar{H}_{A}, \bar{H}_{B} \in \overline{\mathcal{H}}_{2}$, the matrix $\bar{H}_{A} R_{+} \bar{H}_{B}^{\mathrm{T}}$ is a rotation and the matrix $\bar{H}_{A} R_{-} \bar{H}_{B}^{\mathrm{T}}$ is a reflection; therefore the 2D subspaces (35) and (36) are invariant under $\overline{\mathcal{H}}_{2} \boxtimes \overline{\mathcal{H}}_{2}$. Since the equator has dimension $d_{2}-1=2$, all matrices $\bar{C} \neq \mathbf{0}$ must be fully contained in one of the two subspaces spanned by (35) or (36); otherwise the dimension of the set (33) would be too large again. For the same reason $\bar{\alpha}=\bar{\beta}=\mathbf{0}$.

Depending on whether $\bar{C}$ is in the subspace generated by $R_{+}$from (35) or by $R_{-}$from (36), the states in the equator of $\mathcal{S}_{2}^{\prime}$ are either $\hat{\psi}_{\text {eq }}^{+}$or $\hat{\psi}_{\text {eq }}^{-}$, where

$$
\hat{\psi}_{\mathrm{eq}}^{ \pm}=\left[\left(\begin{array}{l}
0 \\
0 \\
0
\end{array}\right),\left(\begin{array}{l}
0 \\
0 \\
0
\end{array}\right),\left(\begin{array}{ccc}
1 & 0 & 0 \\
0 & \cos v & \sin v \\
0 & \mp \sin v & \pm \cos v
\end{array}\right)\right] .
$$

The proportionality constants in $\bar{C} \propto R_{ \pm}$are fixed by normalization (29). It turns out that both the symmetric case $\hat{\psi}_{\text {eq }}^{-}$and the antisymmetric case $\hat{\psi}_{\text {eq }}^{+}$correspond to different representations of the same physical theory-that is, the corresponding state spaces (together with measurements and transformations) are equivalent in the sense of section 2.5 . To see this, define the linear map $\hat{\tau}: \hat{\mathcal{S}}_{2} \rightarrow \hat{\mathcal{S}}_{2}$ as $\hat{\tau}\left(\alpha_{1}, \alpha_{2}, \alpha_{3}\right)^{\mathrm{T}}:=\left(\alpha_{1}, \alpha_{2},-\alpha_{3}\right)^{\mathrm{T}}$; that is, a reflection in the Bloch ball. The equivalence transformation is defined as $\mathcal{L}:=\tau \otimes \mathbb{I}$ (in quantum information terms, this is a 'partial transposition'). This map respects the tensor product structure, leaves the set of product states invariant and satisfies $\hat{\mathcal{L}}\left(\hat{\psi}_{\text {eq }}^{+}\right)=\hat{\psi}_{\text {eq }}^{-}$[19]. In other words, we have reduced the discussion of the antisymmetric theory to that of the symmetric theory ${ }^{6}$, which will be considered in the rest of the paper.

The orthogonality of the matrices in $\hat{\mathcal{G}}_{2,2}$ implies that $\hat{\mathcal{S}}_{2}^{\prime}$ is a 3 D ball, and not just affinely related to it. Hence, all states on the surface of the ball $\hat{\mathcal{S}}_{2}^{\prime} \subset \hat{\mathcal{S}}_{2,2}$ can be parameterized in polar
${ }^{6}$ As a physical interpretation of the antisymmetric case, consider two observers who have never met before, but who have independently built devices to measure spin- $\frac{1}{2}$ particles in three orthogonal directions. If they never had the chance to agree on a common 'handedness' of spatial coordinate systems, and happen to have chosen two different orientations, they will measure antisymmetric correlation matrices on shared quantum states. The 'threebit no-go result' from [19] can be interpreted as follows: if there is a third observer, then it is impossible that every pair of parties measures antisymmetric correlation matrices.
coordinates $u \in[0, \pi)$ and $v \in[0,2 \pi)$ as

$$
\hat{\psi}(u, v)=\left[\left(\begin{array}{c}
\cos u  \tag{37}\\
0 \\
0
\end{array}\right),\left(\begin{array}{c}
\cos u \\
0 \\
0
\end{array}\right),\left(\begin{array}{ccc}
1 & 0 & 0 \\
0 & \sin u \cos v & \sin u \sin v \\
0 & \sin u \sin v & -\sin u \cos v
\end{array}\right)\right] .
$$

These states cannot be written as proper mixtures of other states from $\hat{\mathcal{S}}_{2}^{\prime}$. It is easy to see that this implies that they are pure states in $\hat{\mathcal{S}}_{2,2}$.

### 4.8. The Hermitian representation

In this subsection, a new (more familiar) representation is introduced, where states in $\mathcal{S}_{2}$ are represented by $2 \times 2$ Hermitian matrices. For any state $\psi \in \mathcal{S}_{2}$ in the standard representation (1), define the linear map

$$
\begin{equation*}
\mathcal{L}[\psi]=\psi^{0} \frac{\mathbb{I}-\sigma^{1}-\sigma^{2}-\sigma^{3}}{2}+\sum_{i=1}^{3} \psi^{i} \sigma^{i} \tag{38}
\end{equation*}
$$

The Pauli matrices

$$
\sigma^{1}=\left(\begin{array}{ll}
0 & 1 \\
1 & 0
\end{array}\right), \quad \sigma^{2}=\left(\begin{array}{cc}
0 & -\mathrm{i} \\
\mathrm{i} & 0
\end{array}\right), \quad \sigma^{3}=\left(\begin{array}{cc}
1 & 0 \\
0 & -1
\end{array}\right),
$$

together with the identity $\mathbb{I}$, constitute an orthogonal basis for the real vector space of Hermitian matrices. In terms of the Bloch representation, the map (38) has the familiar form

$$
\mathcal{L}[\psi]=\frac{1}{2}\left(\mathbb{I}+\sum_{i=1}^{3} \hat{\psi}^{i} \sigma^{i}\right) .
$$

All positive unit-trace $2 \times 2$ Hermitian matrices can be written in this way with $\hat{\psi}$ in the unit sphere. Since $\hat{\mathcal{S}}_{2}$ is a 3 D unit sphere, the set $\mathcal{L}\left[\mathcal{S}_{2}\right]$ is the set of quantum states. The extreme points of $\mathcal{L}\left[\mathcal{S}_{2}\right]$ are the rank-one projectors: each pure state $\psi \in \mathcal{S}_{2}$ satisfies $\mathcal{L}[\psi]=|\psi\rangle\langle\psi|$, where the vector $|\psi\rangle \in \mathbb{C}^{2}$ is defined up to a global phase. The effect (15) associated with the pure state $\varphi \in \mathcal{S}_{2}$ is

$$
\begin{equation*}
\Omega_{\varphi}(\psi)=\left(\Omega_{\varphi} \circ \mathcal{L}^{-1}\right)(\mathcal{L}[\psi])=\operatorname{tr}(|\varphi\rangle\langle\varphi| \mathcal{L}[\psi]) . \tag{39}
\end{equation*}
$$

Note that the state $\varphi$ and its associated effect $\Omega_{\varphi}$ are both represented by $|\varphi\rangle\langle\varphi|$. The action of a reversible transformation $\hat{G} \in \hat{\mathcal{G}_{2}}=\mathrm{SO}(3)$ in the Hermitian representation is

$$
\mathcal{L}[G(\psi)]=U \mathcal{L}[\psi] U^{\dagger}
$$

where $U \in \mathrm{SU}(2)$ is related to $\hat{G}$ via

$$
\begin{equation*}
\sum_{j=1}^{3} \hat{G}^{j i} \sigma^{j}=U \sigma^{i} U^{\dagger} \tag{40}
\end{equation*}
$$

and $\hat{G}^{j i}$ are the matrix components (equation VII.5.12 in [25]). In summary, the generalized bit in all theories satisfying $d_{2}>1$ and the requirements is equivalent to the qubit in QT.

### 4.9. Reconstructing quantum theory

In this subsection, the main result of this work is proved. But before that, let us introduce some notation.

In QT, the state space with capacity $c$ and the corresponding group of reversible transformations are

$$
\begin{aligned}
& \mathcal{S}_{c}^{\mathrm{Q}}=\left\{\rho \in \mathbb{C}^{c \times c}: \rho \geqslant 0, \operatorname{tr} \rho=1\right\}, \\
& \mathcal{G}_{c}^{\mathrm{Q}}=\left\{U \otimes U^{*}: U \in \operatorname{SU}(c)\right\} .
\end{aligned}
$$

The joint state space of $m$ generalized bits is denoted by $\mathcal{S}_{2 \times m}$ and the corresponding group of reversible transformations by $\mathcal{G}_{2 \times m}$. The Hermitian representation of a state $\psi \in \mathcal{S}_{2 \times m}$ is defined as $\mathcal{L}^{\otimes m}[\psi]$, where $\mathcal{L}^{\otimes m}:=\mathcal{L} \otimes \cdots \otimes \mathcal{L}$, and $\mathcal{L}$ has been defined in (38). The map $\mathcal{L}^{\otimes m}$ acts independently on each tensor factor; hence it translates the tensor product structure from the standard representation (4), (5) and (7) into the Hermitian one. For example, if $\varphi \in \mathcal{S}_{2}$ is a pure state, then $\mathcal{L}^{\otimes m}\left[\varphi^{\otimes m}\right]=|\varphi\rangle\left\langle\left.\varphi\right|^{\otimes m}\right.$. The notation

$$
\begin{aligned}
& \mathcal{S}_{2 \times m}^{\mathrm{H}}=\mathcal{L}^{\otimes m}\left[\mathcal{S}_{2 \times m}\right], \\
& \mathcal{G}_{2 \times m}^{\mathrm{H}}=\mathcal{L}^{\otimes m} \circ \mathcal{G}_{2 \times m} \circ\left(\mathcal{L}^{\otimes m}\right)^{-1}
\end{aligned}
$$

will be useful. The Hermitian representation of a state $\hat{\psi}_{A B}=[\alpha, \beta, C] \in \hat{\mathcal{S}}_{2,2}$ is

$$
\begin{equation*}
\mathcal{L}^{\otimes 2}\left[\psi_{A B}\right]=\frac{1}{4}\left[\mathbb{I} \otimes \mathbb{I}+\sum_{i=1}^{3} \alpha^{i} \sigma^{i} \otimes \mathbb{I}+\sum_{j=1}^{3} \beta^{j} \mathbb{I} \otimes \sigma^{j}+\sum_{i, j=1}^{3} C^{i j} \sigma^{i} \otimes \sigma^{j}\right] . \tag{41}
\end{equation*}
$$

The action of local transformations $G_{A}, G_{B} \in \mathcal{G}_{2}$ on $\psi_{A B} \in \mathcal{S}_{2,2}$ is

$$
\begin{equation*}
\mathcal{L}^{\otimes 2}\left[\left(G_{A} \otimes G_{B}\right)\left(\psi_{A B}\right)\right]=\left(U_{A} \otimes U_{B}\right) \rho_{A B}\left(U_{A} \otimes U_{B}\right)^{\dagger} \tag{42}
\end{equation*}
$$

where $\rho_{A B}=\mathcal{L}^{\otimes 2}\left[\psi_{A B}\right]$ and $U_{A}, U_{B} \in \mathrm{SU}(2)$ are related to $G_{A}, G_{B}$ via (40). Now, we are ready to prove theorem 6 .

Theorem 6. The only GPT with $d_{2}>1$ satisfying requirements $1-5$ is $Q T$.

Proof. We start by reproducing an argument from [19] that shows that $\mathcal{S}_{4}^{\mathrm{Q}} \subseteq \mathcal{S}_{2,2}^{\mathrm{H}}$. A particular family of pure states in $\mathcal{S}_{2,2}$ is $\psi(u)=\psi(u, 0)$, defined in (37). The Hermitian representation of $\psi(u)$ is the projector $\mathcal{L}^{\otimes 2}[\psi(u)]=|\psi(u)\rangle \psi(u) \mid$ onto the $\mathbb{C}^{2} \otimes \mathbb{C}^{2}$-vector

$$
|\psi(u)\rangle=\cos \frac{u}{2}|+\rangle \otimes|+\rangle+\sin \frac{u}{2}|-\rangle \otimes|-\rangle,
$$

where $|+\rangle=(1,1)^{\mathrm{T}} / \sqrt{2}$ and $|-\rangle=(-1,1)^{\mathrm{T}} / \sqrt{2}$. From the Schmidt decomposition, it follows that all rank-one projectors in $\mathbb{C}^{4 \times 4}$ can be written as $\left(U_{A} \otimes U_{B}\right)|\psi(u)\rangle\langle\psi(u)|\left(U_{A} \otimes U_{B}\right)^{\dagger}$ for some value of $u$ and some local unitaries $U_{A}, U_{B} \in \mathrm{SU}(2)$. Thus, all rank-one projectors are pure states in $\mathcal{S}_{2,2}^{\mathrm{H}}$. Their mixtures generate all of $\mathcal{S}_{4}^{\mathrm{Q}}$; therefore $\mathcal{S}_{4}^{\mathrm{Q}} \subseteq \mathcal{S}_{2,2}^{\mathrm{H}}$.

Direct calculation shows that

$$
\begin{equation*}
\operatorname{tr}\left(\mathcal{L}^{\otimes 2}[\psi] \mathcal{L}^{\otimes 2}\left[\psi^{\prime}\right]\right)=\frac{1}{4}+\frac{1}{4}\left[\alpha^{\mathrm{T}} \alpha^{\prime}+\beta^{\mathrm{T}} \beta^{\prime}+\operatorname{tr}\left(C^{\mathrm{T}} C^{\prime}\right)\right] \tag{43}
\end{equation*}
$$

for any pair of states $\hat{\psi}=[\alpha, \beta, C]$ and $\hat{\psi}^{\prime}=\left[\alpha^{\prime}, \beta^{\prime}, C^{\prime}\right]$ from $\hat{\mathcal{S}}_{2,2}$. Lemma 8 shows that all $\hat{G} \in \hat{\mathcal{G}}_{2,2}$ are orthogonal matrices. Therefore, the Euclidean inner product between states, as in the right-hand side of (43), is preserved by the action of any $\hat{G} \in \hat{\mathcal{G}}_{2,2}$. Equality (43) maps this property to the Hermitian representation: any $H \in \mathcal{G}_{2,2}^{\mathrm{H}}$ preserves the Hilbert-Schmidt inner product between states:

$$
\begin{equation*}
\operatorname{tr}\left[H(\rho) H\left(\rho^{\prime}\right)\right]=\operatorname{tr}\left(\rho \rho^{\prime}\right) \tag{44}
\end{equation*}
$$

for all $\rho, \rho^{\prime} \in \mathcal{S}_{2,2}^{\mathrm{H}}$.
For any pure state $\varphi \in \mathcal{S}_{2}$, the rank-one projector $|\varphi\rangle\langle\varphi| \otimes|\varphi\rangle\langle\varphi|=\mathcal{L}^{\otimes 2}[\varphi \otimes \varphi]$ is a pure state in $\mathcal{S}_{2,2}^{\mathrm{H}}$, and $\operatorname{tr}(|\varphi\rangle\langle\varphi| \otimes|\varphi\rangle\langle\varphi| \rho)=\left(\Omega_{\varphi} \otimes \Omega_{\varphi}\right) \circ\left(\mathcal{L}^{\otimes 2}\right)^{-1}(\rho)$ is a measurement on $\mathcal{S}_{2,2}^{\mathrm{H}}$. Any rank-one projector $|\psi\rangle\langle\psi| \in \mathbb{C}^{4 \times 4}$ is a pure state in $\mathcal{S}_{2,2}^{\mathrm{H}}$; hence there is $H \in \mathcal{G}_{2,2}^{\mathrm{H}}$ such that $H(|\varphi\rangle\langle\varphi| \otimes|\varphi\rangle\langle\varphi|)=|\psi\rangle\langle\psi|$. Composing the transformation $H$ with the effect $\Omega_{\varphi} \otimes \Omega_{\varphi}$ generates the effect $\left(\Omega_{\varphi} \otimes \Omega_{\varphi}\right) \circ\left(\mathcal{L}^{\otimes 2}\right)^{-1} \circ H^{-1}$, which maps any $\rho \in \mathcal{S}_{2,2}^{\mathrm{H}}$ to

$$
\begin{equation*}
\operatorname{tr}\left[|\varphi\rangle\langle\varphi| \otimes|\varphi\rangle\langle\varphi| H^{-1}(\rho)\right]=\operatorname{tr}[|\psi\rangle\langle\psi| \rho], \tag{45}
\end{equation*}
$$

where (44) has been used. In summary, every rank-one projector $|\psi\rangle\langle\psi| \in \mathbb{C}^{4 \times 4}$ has an associated effect (45) which is an allowed measurement on $\mathcal{S}_{2,2}^{\mathrm{H}}$, and these generate all quantum effects.

We have seen that all quantum states $\mathcal{S}_{4}^{\mathrm{Q}}$ are contained in $\mathcal{S}_{4}^{\mathrm{H}}$. But can there be other states? If so, the associated Hermitian matrices should have a negative eigenvalue (note that all states in the Hermitian representation (41) have unit trace). If $\rho$ has a negative eigenvalue and $|\psi\rangle$ is the corresponding eigenvector, then the associated measurement outcome (45) has negative probability. Hence, we conclude that $\mathcal{S}_{4}^{\mathrm{Q}}=\mathcal{S}_{4}^{\mathrm{H}}$, and similarly for the measurements.

All reversible transformations $H \in \mathcal{G}_{4}^{\mathrm{H}}$ map pure states to pure states, i.e. rank-one projectors to rank-one projectors. According to Wigner's theorem [30], every map of this kind can be written as $H(|\psi\rangle\langle\psi|)=(U|\psi\rangle)(U|\psi\rangle)^{\dagger}$, where $U$ is either unitary or anti-unitary. If $U$ is anti-unitary, it follows from Wigner's normal form [31] that there is a 2D $U$-invariant subspace spanned by two orthonormal vectors $\left|\theta_{0}\right\rangle,\left|\theta_{1}\right\rangle \in \mathbb{C}^{4}$ such that $U\left(t_{0}\left|\theta_{0}\right\rangle+t_{1}\left|\theta_{1}\right\rangle\right)$ equals either $\bar{t}_{0}\left|\theta_{0}\right\rangle+\bar{t}_{1}\left|\theta_{1}\right\rangle$ or $\bar{t}_{1} \mathrm{e}^{\mathrm{i} s}\left|\theta_{0}\right\rangle+\bar{t}_{0} \mathrm{e}^{-\mathrm{i} s}\left|\theta_{1}\right\rangle$ for some $s \in \mathbb{R}$. In both cases, $U$ acts as a reflection in the corresponding Bloch ball, which contradicts requirement 3 , because we know that $\mathcal{G}_{2}=\mathrm{SO}$ (3). Therefore $\mathcal{G}_{4}^{\mathrm{H}} \subseteq \mathcal{G}_{4}^{\mathrm{Q}}$.

We know that $\mathcal{G}_{2,2}^{\mathrm{H}}$ contains all local unitaries. Since this group is transitive on the pure states, it contains at least one unitary that maps a product state to an entangled state. It is well known [32] that this implies that the corresponding group of unitaries constitutes a universal gate set for quantum computation; that is, it generates every unitary operation on two qubits. This proves that $\mathcal{G}_{4}^{\mathrm{H}}=\mathcal{G}_{4}^{\mathrm{Q}}$.

Consider $m$ generalized bits as a composite system. From the previously discussed case of $\mathcal{S}_{2,2}$, we know that every unitary operation on every pair of generalized bits is an allowed transformation on $\mathcal{S}_{2^{\times m}}^{\mathrm{H}}$. But two-qubit unitaries generate all unitary transformations [32]; hence $\mathcal{G}_{2^{m}}^{\mathrm{Q}} \subseteq \mathcal{G}_{2^{\times m}}^{\mathrm{H}}$. By applying all these unitaries to $|\varphi\rangle\left\langle\left.\varphi\right|^{\otimes m}\right.$, all pure quantum states are generated; hence $\mathcal{S}_{2^{m}}^{\mathrm{Q}} \subseteq \mathcal{S}_{2^{\times m}}^{\mathrm{H}}$. Reasoning as in the $\mathcal{S}_{2,2}$ case, for every rank-one projector $|\psi\rangle\langle\psi|$ acting on $\mathbb{C}^{2^{m}}$, the associated effect that maps $\rho \in \mathcal{S}_{2^{\times m}}^{\mathrm{H}}$ to $\operatorname{tr}(|\psi\rangle\langle\psi| \rho)$ is an allowed measurement outcome on $\mathcal{S}_{2^{\times m}}^{\mathrm{H}}$. This implies that all matrices in $\mathcal{S}_{2^{\times m}}^{\mathrm{H}}$ have positive eigenvalues; therefore, $\mathcal{S}_{2^{\times m}}^{\mathrm{H}}=\mathcal{S}_{2^{m}}^{\mathrm{Q}}$ and $\mathcal{G}_{2^{\times m}}^{\mathrm{H}}=\mathcal{G}_{2^{m}}^{\mathrm{Q}}$.

The remaining cases of capacities $c$ that are not powers of two are treated by applying requirement 3 , assuming that $\mathcal{S}_{c} \subset \mathcal{S}_{2^{m}}$ for large enough $m$.

## 5. Conclusion

We have imposed five physical requirements on the framework of generalized probabilistic theories. These requirements are simple and have a clear physical meaning in terms of basic operational procedures. It is shown that the only theories compatible with them are CPT and QT. If requirement 4 is strengthened by imposing the continuity of reversible transformations, then the only theory that survives is QT. Any other theory violates at least one of the requirements; hence, the relaxation of each constitutes a different way to go beyond QT.

The standard formulation of QT includes two postulates that do not follow from our requirements: (i) the update rule for the state after a measurement and (ii) the Schrödinger equation. If desired, these can be incorporated into our derivation of QT by imposing the following two extra requirements: (i) if a system is measured twice 'in rapid succession' with the same measurement, the same outcome is obtained both times [4] and (ii) closed systems evolve reversibly and continuously in time.

This derivation of QT contains two steps that deserve a special mention. Firstly, a direct consequence of requirement 3 is that $\mathcal{S}_{2}$ is fully surrounded by pure states, which together with requirement 4 implies that $\mathcal{S}_{2}$ is a ball. Secondly, this ball has dimension three, since $d=3$ is the only value for which $\mathrm{SO}(d-1)$ is reducible in $\mathbb{C}^{d}$.

Modifications and generalizations of QT are of interest in themselves and could be essential to construct a QT of gravity. Some well-known attempts [15, 16] have shown that straightforward modifications of QT's mathematical formalism quickly lead to inconsistencies, such as superluminal signaling [17]. This work provides an alternative way to proceed. We have shown that the Hilbert space formalism of QT follows from five simple physical requirements. This gives five different consistent ways to go beyond QT, each obtained by relaxing one of our requirements.

## Acknowledgments

The authors are grateful to Anne Beyreuther, Jens Eisert, Volkher Scholz, Tony Short and Christopher Witte for discussions. Special thanks to Lucien Hardy for pointing out the multiplicity of groups that are transitive on the sphere and, correspondingly, the need to address the 7D ball as a special case. Lluís Masanes is financially supported by Caixa Manresa and benefits from the Spanish MEC projects TOQATA (FIS2008-00784) and QOIT (Consolider Ingenio 2010), the EU Integrated Project SCALA and the STREP project NAMEQUAM. Markus Müller was supported by the EU (QESSENCE). Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research and Innovation.

## Appendix. Lemmas

Lemma 1. In any state space $\mathcal{S}_{c}$, the only state $\psi \in \mathcal{S}_{c}$ that is invariant under all reversible transformations

$$
\begin{equation*}
G(\psi)=\psi \text { for all } G \in \mathcal{G}_{c} \tag{A.1}
\end{equation*}
$$

is the maximally mixed state $\mu_{c}$, defined in (11).

Proof. Suppose $\psi \in \mathcal{S}_{c}$ satisfies (A.1). Any state can be written as a mixture of pure states: $\psi=\sum_{k} q_{k} \psi_{k}$. The normalization $\int_{\mathcal{G}_{c}} \mathrm{~d} G=1$, condition (A.1), the linearity of $G$, the purity of all $\psi_{k}$, the definition of $\mu_{c}$ and $\sum_{k} q_{k}=1$ imply that

$$
\begin{aligned}
\psi & =\int_{\mathcal{G}_{c}} \psi \mathrm{~d} G=\int_{\mathcal{G}_{c}} G(\psi) \mathrm{d} G \\
& =\sum_{k} q_{k} \int_{\mathcal{G}_{c}} G\left(\psi_{k}\right) \mathrm{d} G=\sum_{k} q_{k} \mu_{c}=\mu_{c},
\end{aligned}
$$

which proves the claim.
Lemma 2. If $\mathcal{G}$ is a compact real matrix group, then there is a real matrix $S>0$ such that for each $G \in \mathcal{G}$ the matrix $S G S^{-1}$ is orthogonal.

Proof. Since the group $\mathcal{G}$ is compact, there is an invariant Haar measure [25], which allows us to define

$$
P=\int_{\mathcal{G}} G^{\mathrm{T}} G \mathrm{~d} G .
$$

Since each $G$ is invertible, the matrix $G^{\mathrm{T}} G$ is strictly positive, and $P$ is positive too. Define $S=\sqrt{P}>0$, where both $S, S^{-1}$ are real and symmetric. For any $G \in \mathcal{G}$, we have $\left(S G S^{-1}\right)^{\mathrm{T}}\left(S G S^{-1}\right)=\mathbb{I}$, which implies orthogonality.

Lemma 3. If $\mu_{A}$ and $\mu_{B}$ are the maximally mixed states of the state spaces $\mathcal{S}_{A}$ and $\mathcal{S}_{B}$, then the maximally mixed state of the composite system $\mathcal{S}_{A B}$ is

$$
\mu_{A B}=\mu_{A} \otimes \mu_{B} .
$$

Proof. The pure states $\psi^{A}$ in $\mathcal{S}_{A}$ linearly span $\mathbb{R}^{d_{A}+1}$, and the pure states $\psi^{B}$ in $\mathcal{S}_{B}$ linearly span $\mathbb{R}^{d_{B}+1}$. Therefore, pure product states $\psi^{A} \otimes \psi^{B}$ span $\mathbb{R}^{d_{A}+1} \otimes \mathbb{R}^{d_{B}+1}$. In particular, the maximally mixed state (11) of $\mathcal{S}_{A B}$ can be written as

$$
\begin{equation*}
\mu_{A B}=\sum_{a, b} t_{a, b} \psi_{a}^{A} \otimes \psi_{b}^{B} \tag{A.2}
\end{equation*}
$$

where $t_{a, b} \in \mathbb{R}$ are not necessarily positive coefficients, and all $\psi_{a}^{A}, \psi_{b}^{B}$ are pure. From definition (1), the first component of the vector equality (A.2) implies $\sum_{a, b} t_{a, b}=1$. The maximally mixed state is invariant under all reversible transformations, in particular under the local ones

$$
\begin{aligned}
\mu_{A B} & =\int_{\mathcal{G}_{A}} \mathrm{~d} G_{A} \int_{\mathcal{G}_{B}} \mathrm{~d} G_{B}\left(G_{A} \otimes G_{B}\right)\left(\mu_{A B}\right) \\
& =\sum_{a, b} t_{a, b}\left[\int_{\mathcal{G}_{A}} \mathrm{~d} G_{A} G_{A}\left(\psi_{a}^{A}\right)\right] \otimes\left[\int_{\mathcal{G}_{B}} \mathrm{~d} G_{B} G_{B}\left(\psi_{b}^{B}\right)\right] \\
& =\sum_{a, b} t_{a, b} \mu_{A} \otimes \mu_{B}=\mu_{A} \otimes \mu_{B},
\end{aligned}
$$

where the same tricks as those of lemma 1 have been used.

Lemma 4. For every tight effect $\Omega$, there is a pure state $\psi$ such that $\Omega(\psi)=1$. Also, if a measurement $\Omega_{1}, \ldots, \Omega_{n}$ distinguishes $n$ states $\psi_{1}, \ldots, \psi_{n}$, then these states can be chosen to be pure.

Proof. By definition, for each tight effect $\Omega$ there is a (not necessarily pure) state $\psi^{\prime}$ such that $\Omega\left(\psi^{\prime}\right)=1$. Every $\psi^{\prime}$ can be written as a mixture of pure states $\psi_{k}$, i.e. $\psi^{\prime}=\sum_{k} q_{k} \psi_{k}$ with $q_{k}>0$ and $\sum_{k} q_{k}=1$. Effects are linear functions such that $\Omega(\psi) \leqslant 1$ for any state $\psi$. Therefore, it must happen that all pure states $\psi_{k}$ in the above decomposition satisfy $\Omega\left(\psi_{k}\right)=1$.

To prove the second part, let $\psi_{1}^{\prime}, \ldots, \psi_{n}^{\prime}$ be the states that are distinguished by the measurement, i.e. $\Omega_{a}\left(\psi_{b}^{\prime}\right)=\delta_{a, b}$. Every $\psi_{b}^{\prime}$ can be written as a convex combination of pure states $\psi_{b}^{\prime}=\sum_{k} q_{k} \psi_{b, k}$. But effects are linear functions such that $0 \leqslant \Omega(\psi) \leqslant 1$ for any state $\psi$. Hence, $\Omega\left(\psi_{b}^{\prime}\right)=0$ is only possible if $\Omega\left(\psi_{b, k}\right)=0$ for all $k$, and similarly for the case $\Omega\left(\psi_{b}^{\prime}\right)=1$. It follows that $\Omega_{a}\left(\psi_{b, 1}\right)=\delta_{a, b}$.

Lemma 5. If $\mathcal{S}_{c}$ is a state space with capacity $c \geqslant 1$ and $\mu_{c}$ the corresponding maximally mixed state, then there are c pure distinguishable states $\psi_{1}, \ldots, \psi_{c} \in \mathcal{S}_{c}$ such that

$$
\mu_{c}=\frac{1}{c} \sum_{a=1}^{c} \psi_{a}
$$

Proof. Since $\mathcal{S}_{1}$ contains a single state the claim is trivially true for $c=1$. Since $\mathcal{S}_{2}$ is the $d_{2}{ }^{-}$ dimensional unit ball, the two antipodal points $\hat{\varphi}_{1}$ and $\hat{\varphi}_{2}=-\hat{\varphi}_{1}$ are pure and distinguishable and satisfy

$$
\begin{equation*}
\mu_{2}=\frac{1}{2}\left(\varphi_{1}+\varphi_{2}\right) . \tag{A.3}
\end{equation*}
$$

Now, consider the joint state space of $n$ generalized bits, denoted by $\mathcal{S}_{2 \times n}$. Lemma 3 and (A.3) imply that the maximally mixed state of $\mathcal{S}_{2 \times n}$ is

$$
\begin{equation*}
\mu_{(n)}=\left(\mu_{2}\right)^{\otimes n}=\frac{1}{2^{n}} \sum_{a_{i} \in\{1,2\}} \varphi_{a_{1}} \otimes \cdots \otimes \varphi_{a_{n}} . \tag{A.4}
\end{equation*}
$$

The states $\varphi_{a_{1}} \otimes \cdots \otimes \varphi_{a_{n}} \in \mathcal{S}_{2 \times n}$ for all $a_{1}, \ldots, a_{n} \in\{1,2\}$ can be perfectly distinguished by the corresponding product measurement; hence, the capacity of $\mathcal{S}_{2^{\times n}}$, denoted by $c_{n}$, satisfies

$$
\begin{equation*}
c_{n} \geqslant 2^{n} . \tag{A.5}
\end{equation*}
$$

Let $\left(\Omega_{1}, \ldots, \Omega_{c_{n}}\right)$ be a complete measurement that distinguishes the states $\psi_{1}, \ldots, \psi_{c_{n}} \in \mathcal{S}_{2 \times n}$. According to lemma 4 these states can be chosen to be pure. Since $\sum_{k=1}^{c_{n}} \Omega_{k}\left(\mu_{(n)}\right)=1$, there is at least one value of $k$, denoted by $k_{0}$, such that

$$
\begin{equation*}
\Omega_{k_{0}}\left(\mu_{(n)}\right) \leqslant 1 / c_{n} \tag{A.6}
\end{equation*}
$$

The state $\varphi_{1} \in \mathcal{S}_{2}$ from (A.3) is pure; hence $\varphi_{1}^{\otimes n} \in \mathcal{S}_{2^{\times n}}$ is pure too. Requirement 4 reveals that there is a reversible transformation $G$ acting on $\mathcal{S}_{2 \times n}$ such that $G\left(\psi_{k_{0}}\right)=\varphi_{1}^{\otimes n}$. The measurement $\left(\Omega_{1} \circ G^{-1}, \ldots, \Omega_{c_{n}} \circ G^{-1}\right)$ distinguishes the states $G\left(\psi_{1}\right), \ldots, G\left(\psi_{c_{n}}\right)$. Inequality (A.6), the
invariance of $\mu_{(n)}$ under $G$, expansion (A.4), the positivity of probabilities and ( $\Omega_{k_{0}} \circ$ $\left.G^{-1}\right)\left(\varphi_{1}^{\otimes n}\right)=1$ imply that

$$
\begin{aligned}
\frac{1}{c_{n}} & \geqslant\left(\Omega_{k_{0}} \circ G^{-1}\right)\left(\mu_{(n)}\right) \\
& =\frac{1}{2^{n}} \sum_{a_{i} \in\{1,2\}}\left(\Omega_{k_{0}} \circ G^{-1}\right)\left(\varphi_{a_{1}} \otimes \cdots \otimes \varphi_{a_{n}}\right) \geqslant \frac{1}{2^{n}} .
\end{aligned}
$$

This and (A.5) imply that $c_{n}=2^{n}$. This together with (A.4) shows the assertion of the lemma for state spaces whose capacity is a power of two. The rest of the cases are shown by induction.

Let us prove that if the claim of the lemma holds for a state space with capacity $c$, with $c>1$, then it holds for a state space with capacity $c-1$, too. The induction hypothesis reveals that there is a complete measurement $\left(\Omega_{1}, \ldots, \Omega_{c}\right)$ that distinguishes the pure states $\psi_{1}, \ldots, \psi_{c} \in \mathcal{S}_{c}$, and $\mu_{c}=\frac{1}{c} \sum_{k=1}^{c} \psi_{k} \in \mathcal{S}_{c}$ is the corresponding maximally mixed state. Requirement 3 reveals that the state space $\mathcal{S}_{c-1}$ is equivalent to

$$
\mathcal{S}_{c-1}^{\prime}=\left\{\psi \in \mathcal{S}_{c} \mid \Omega_{1}(\psi)+\cdots+\Omega_{c-1}(\psi)=1\right\} .
$$

According to requirement 3 , for each $G \in \mathcal{G}_{c-1}$ there is $G^{\prime} \in \mathcal{G}_{c}$ which implements $G$ on $\mathcal{S}_{c-1}^{\prime}$. Hence $G^{\prime}\left(\psi_{k}\right) \in \mathcal{S}_{c-1}^{\prime}$ for $k=1, \ldots, c-1$, which implies $\left(\Omega_{c} \circ G^{\prime}\right)\left(\psi_{k}\right)=0$ for those $k$ values and

$$
\begin{aligned}
\frac{1}{c} & =\Omega_{c}\left(\mu_{c}\right)=\left(\Omega_{c} \circ G^{\prime}\right)\left(\mu_{c}\right), \\
& =\frac{1}{c} \sum_{k=1}^{c}\left(\Omega_{c} \circ G^{\prime}\right)\left(\psi_{k}\right)=\frac{1}{c}\left(\Omega_{c} \circ G^{\prime}\right)\left(\psi_{c}\right),
\end{aligned}
$$

and $\left(\Omega_{c} \circ G^{\prime}\right)\left(\psi_{c}\right)=1$. Requirement 3 reveals that the set

$$
\mathcal{S}_{1}^{\prime}=\left\{\psi \in \mathcal{S}_{c} \mid \Omega_{c}(\psi)=1\right\}
$$

is equivalent to $\mathcal{S}_{1}$, which contains a single state. This and $\Omega_{c}\left(\psi_{c}\right)=1 \mathrm{imply}$ that $G^{\prime}\left(\psi_{c}\right)=\psi_{c}$ and then $G^{\prime}\left(\mu_{c-1}^{\prime}\right)=\mu_{c-1}^{\prime}$, where we define

$$
\mu_{c-1}^{\prime}=\frac{1}{c-1} \sum_{k=1}^{c-1} \psi_{k}=\frac{c}{c-1} \mu_{c}-\frac{1}{c-1} \psi_{c} \in \mathcal{S}_{c-1}^{\prime}
$$

For any $G \in \mathcal{G}_{c-1}$ the corresponding $G^{\prime}$ satisfies $G^{\prime}\left(\mu_{c-1}^{\prime}\right)=\mu_{c-1}^{\prime}$. Due to lemma 1, the invariant state $\mu_{c-1}^{\prime}$ must be the maximally mixed state in $\mathcal{S}_{c-1}^{\prime}$, which has the claimed form, and requirement 3 extends this to $\mathcal{S}_{c-1}$.

Lemma 6. Let $\mathcal{S}$ be a state space such that the subset of pure states $\mathcal{P}$ is a connected topological manifold. If the corresponding group of transformations $\mathcal{G}$ is transitive on $\mathcal{P}$, then the largest connected subgroup $\mathcal{C} \subseteq G$ is transitive on $\mathcal{P}$, too.

Proof. This proof involves basic notions of point set topology.
Since $\mathcal{G}$ is compact, it is the union of a finite number of (disjoint) connected components $\mathcal{G}=\mathcal{C}_{1} \cup \cdots \cup \mathcal{C}_{n}$. If $n=1$ the lemma is trivial. Let $\mathcal{C}$ be the connected component $\mathcal{C}_{i}$ containing the identity matrix $\mathbb{I}$, which is the largest connected subgroup of $\mathcal{G}$. Each connected component $\mathcal{C}_{i}$ is clopen (open and closed) and compact and is a coset of the group: $\mathcal{C}_{i}=G_{i} \circ \mathcal{C}$ for some $G_{i} \in \mathcal{G}$ [29].

Pick $\psi \in \mathcal{P}$, and consider the continuous surjective map $f: \mathcal{G} \rightarrow \mathcal{P}$, defined by $f(G)=$ $G(\psi) \in \mathcal{P}$. Since $\mathcal{C}$ is compact, $f(\mathcal{C}) \subseteq \mathcal{P}$ is compact too. Since the manifold $\mathcal{P}$ is, in particular, a Hausdorff space, $f(\mathcal{C})$ is closed. Consider the set $\mathcal{D}=f^{-1}(f(\mathcal{C}))$. If two group elements $G, H \in \mathcal{G}$ are in the same component, i.e. $G^{-1} H \in \mathcal{C}$, then $G \in \mathcal{D}$ implies $H \in \mathcal{D}$, assuming that $\mathcal{C}$ is a normal subgroup of $\mathcal{G}$. This implies that $\mathcal{D}$ is the union of some connected components $\mathcal{C}_{i}$, and so is $\mathcal{G} \backslash \mathcal{D}$. In particular, $\mathcal{G} \backslash \mathcal{D}$ is compact; thus, $f(\mathcal{G} \backslash \mathcal{D})$ is compact and hence closed. Therefore, $f(\mathcal{C})=\mathcal{P} \backslash f(\mathcal{G} \backslash \mathcal{D})$ is open. We have thus proved that $f(\mathcal{C}) \neq \emptyset$ is clopen. Since $\mathcal{P}$ is connected, it follows that $f(\mathcal{C})=\mathcal{P}$.

The following lemma shows that there are transformations for two generalized bits which perform the 'classical' swap and the 'classical' controlled-NOT in a particular basis. Note that these transformations do not necessarily swap other states that are not in the given basis, as in QT. However, they implement a minimal amount of reversible computational power which exceeds, for example, that of boxworld, where no controlled-NOT operation is possible [14].

Lemma 7. For each pair of distinguishable states $\varphi_{0}, \varphi_{1} \in \mathcal{S}_{2}$, there are transformations $G_{\text {swap }}, G_{\text {cnot }} \in \mathcal{G}_{2,2}$ such that

$$
\begin{align*}
& G_{\text {swap }}\left(\varphi_{a} \otimes \varphi_{b}\right)=\varphi_{b} \otimes \varphi_{a},  \tag{A.7}\\
& G_{\text {cnot }}\left(\varphi_{a} \otimes \varphi_{b}\right)=\varphi_{a} \otimes \varphi_{a \oplus b}, \tag{A.8}
\end{align*}
$$

for all $a, b \in\{0,1\}$, where $\oplus$ is addition modulo 2 .

Proof. Let $\left(\Omega_{0}, \Omega_{1}\right)$ be the measurement that distinguishes $\left(\varphi_{0}, \varphi_{1}\right)$, i.e. $\Omega_{a}\left(\varphi_{b}\right)=\delta_{a, b}$. Define $\psi_{a, b}=\varphi_{a} \otimes \varphi_{b}$ and $\Omega_{a, b}=\Omega_{a} \otimes \Omega_{b}$ for $a, b \in\{0,1\}$. Define

$$
\begin{aligned}
& \mathcal{S}_{3}^{\prime}=\left\{\psi \in \mathcal{S}_{2,2} \mid\left(\Omega_{0,1}+\Omega_{1,0}+\Omega_{1,1}\right)(\psi)=1\right\}, \\
& \mathcal{S}_{2}^{\prime}=\left\{\psi \in \mathcal{S}_{2,2} \mid\left(\Omega_{0,1}+\Omega_{1,0}\right)(\psi)=1\right\},
\end{aligned}
$$

and note that $\mathcal{S}_{2}^{\prime} \subset \mathcal{S}_{3}^{\prime} \subset \mathcal{S}_{2,2}$. At the end of section 4.2, it is shown that two distinguishable states $\varphi_{0}, \varphi_{1} \in \mathcal{S}_{2}$ have Bloch representation satisfying $\hat{\varphi}_{0}=-\hat{\varphi}_{1}$. According to requirement 4 , there is $G \in \mathcal{G}_{2}$ such that $G\left(\varphi_{0}\right)=\varphi_{1}$, and by linearity, $\hat{G}\left(\hat{\varphi}_{1}\right)=-\hat{G}\left(\hat{\varphi}_{0}\right)=-\hat{\varphi}_{1}=\hat{\varphi}_{0}$. Requirement 3 implies that there is a transformation $G_{\text {swap }}^{\prime}$ for $\mathcal{S}_{3}^{\prime}$ such that $G_{\text {swap }}^{\prime}\left(\psi_{0,1}\right)=\psi_{1,0}$ and $G_{\text {swap }}^{\prime}\left(\psi_{1,0}\right)=\psi_{0,1}$. According to lemma 5 , the maximally mixed state in $\mathcal{S}_{3}^{\prime}$ can be written as $\mu_{3}^{\prime}=\left(\psi_{0,1}+\psi_{1,0}+\psi_{1,1}\right) / 3$. Equalities $G_{\text {swap }}^{\prime}\left(\mu_{3}^{\prime}\right)=\mu_{3}^{\prime}$ and $G_{\text {swap }}^{\prime}\left(\psi_{0,1}+\psi_{1,0}\right)=\psi_{0,1}+\psi_{1,0}$ imply that $G_{\text {swap }}^{\prime}\left(\psi_{1,1}\right)=\psi_{1,1}$. Using requirement 3 again, there is a reversible transformation $G_{\text {swap }} \in \mathcal{G}_{2,2}$ that implements $G_{\text {swap }}^{\prime}$ in the subspace $\mathcal{S}_{3}^{\prime}$. Repeating the argument with the maximally mixed state (now in $\mathcal{S}_{2,2}$ ), we conclude that $G_{\text {swap }}\left(\psi_{1,1}\right)=\psi_{1,1}$; hence $G_{\text {swap }}$ satisfies (A.7).

The existence of $G_{\text {cnot }}$ is shown similarly, by exchanging the roles of $\psi_{0,1}$ and $\psi_{1,1}$.

Lemma 8. Reversible transformations for two generalized bits in the Bloch representation (24) are orthogonal:

$$
\hat{\mathcal{G}}_{2,2} \subseteq \mathrm{O}\left(d_{4}\right) .
$$

Proof. In section 4.6, the Bloch representation for two generalized bits is defined, and it is argued that reversible transformations $\hat{\mathcal{G}}_{2,2}$ act on $[\alpha, \beta, C]$ as matrices. In particular, local transformations (28) are

$$
\left(G_{A} \otimes G_{B}\right)^{\wedge}=\left(\begin{array}{ccc}
\hat{G}_{A} & \mathbf{0} & \mathbf{0}  \tag{A.9}\\
\mathbf{0} & \hat{G}_{B} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \hat{G}_{A} \otimes \hat{G}_{B}
\end{array}\right)
$$

where each diagonal block acts on an entry of $[\alpha, \beta, C]$, and $\hat{G}_{A}, \hat{G}_{B} \in \hat{\mathcal{G}}_{2}$. In section 4.5 , it is argued that $\hat{\mathcal{G}}_{2} \subseteq \mathrm{O}\left(d_{2}\right)$; hence local transformations (A.9) are orthogonal.

Lemma 2 shows the existence of a real matrix $S>0$ such that for any $\hat{G} \in \hat{\mathcal{G}}_{2,2}$ the matrix $S \hat{G} S^{-1}$ is orthogonal. In particular,

$$
\left[S \cdot\left(G_{A} \otimes G_{B}\right)^{\wedge} \cdot S^{-1}\right]^{\mathrm{T}}\left[S \cdot\left(G_{A} \otimes G_{B}\right)^{\wedge} \cdot S^{-1}\right]=\mathbb{I},
$$

which implies the commutation relation

$$
\begin{equation*}
S \cdot\left(G_{A} \otimes G_{B}\right)^{\wedge}=\left(G_{A} \otimes G_{B}\right)^{\wedge} \cdot S . \tag{A.10}
\end{equation*}
$$

Section 4.5 concludes that $d_{2}$ is odd, and that $\mathrm{SO}\left(d_{2}\right) \subseteq \hat{\mathcal{G}}_{2}$ except when $d_{2}=7$, where $M \mathbb{G}_{2} M^{-1} \subseteq \hat{\mathcal{G}}_{2}$ and $\mathbb{G}_{2}$ is the fundamental representation of the smallest exceptional Lie group [29]. For $d_{2} \geqslant 3$ these groups act irreducibly in $\mathbb{C}^{d_{2}}$; hence, $\hat{\mathcal{G}}_{2}$ acts irreducibly in $\mathbb{C}^{d_{2}}$, too [29]. The first two diagonal blocks in (A.9) are irreducible. The exterior tensor product of two irreducible representations (in $\mathbb{C}^{d}$ ) is also an irreducible representation; hence, the third diagonal block in (A.9) is also irreducible. This, together with (A.10), implies that

$$
S=\left(\begin{array}{ccc}
a \mathbb{I} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & b \mathbb{I} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & s \mathbb{I}
\end{array}\right)
$$

for some $a, b, s>0$ (Schur's lemma [29]).
According to lemma 7, for each unit vector $\alpha \in \hat{\mathcal{S}}_{2}$ there is a transformation $G_{\text {swap }} \in \mathcal{G}_{2,2}$ such that

$$
\begin{aligned}
\hat{G}_{\text {swap }}[\alpha, 0,0] & =\hat{G}_{\text {swap }}\left(\left[\alpha, \alpha, \alpha \alpha^{\mathrm{T}}\right]+\left[\alpha,-\alpha,-\alpha \alpha^{\mathrm{T}}\right]\right) / 2 \\
& =\left(\left[\alpha, \alpha, \alpha \alpha^{\mathrm{T}}\right]+\left[-\alpha, \alpha,-\alpha \alpha^{\mathrm{T}}\right]\right) / 2 \\
& =[0, \alpha, 0] .
\end{aligned}
$$

Since $S \hat{G}_{\text {swap }} S^{-1}$ is orthogonal, the vectors $[\alpha, 0,0]$ and $\left[0, b a^{-1} \alpha, 0\right]$ have the same modulus; hence $a=b$. Also, there is a transformation $G_{\text {cnot }} \in \mathcal{G}_{2,2}$ such that

$$
\begin{aligned}
\hat{G}_{\text {cnot }}\left[0,0, \alpha \alpha^{\mathrm{T}}\right] & =\hat{G}_{\text {cnot }}\left(\left[\alpha, \alpha, \alpha \alpha^{\mathrm{T}}\right]+\left[-\alpha,-\alpha, \alpha \alpha^{\mathrm{T}}\right]\right) / 2 \\
& =\left(\left[\alpha, \alpha, \alpha \alpha^{\mathrm{T}}\right]+\left[-\alpha, \alpha,-\alpha \alpha^{\mathrm{T}}\right]\right) / 2 \\
& =[0, \alpha, 0] .
\end{aligned}
$$

Since $S \hat{G}_{\text {cnot }} S^{-1}$ is orthogonal, the vectors $\left[0,0, \alpha \alpha^{\mathrm{T}}\right]$ and $\left[0, b s^{-1} \alpha, 0\right]$ have the same modulus; hence $s=b$. Consequently, $S=a \mathbb{I}$, and the claim follows.

## Lemma 9. The results of this paper also hold if requirement 5 is replaced by requirement $5^{\prime}$.

Proof. Assume requirement $5^{\prime}$, but not requirement 5. It follows that $\mathcal{S}_{1}$ contains a single state only: if it contained more than one state, there would exist some $\psi \in \mathcal{S}_{1}$ which is not completely mixed, which would then be distinguishable from some other state, contradicting that $\mathcal{S}_{1}$ has capacity 1 . Requirement 5 is used in the proof of theorem 1 . This proof is easily modified to comply with requirement $5^{\prime}$ instead: adopting the notation from the proof, the state $\varphi_{\text {mix }}$ is not completely mixed and thus it is distinguishable from some other state. This proves the existence of some $\hat{\Omega}$ with the claimed properties, and the other arguments remain unchanged, proving that $\hat{\mathcal{S}}_{2}$ can be represented as a unit ball. All pure states in $\hat{\mathcal{S}}_{2}$ are not completely mixed and hence they have a corresponding tight effect which is physically allowed. But in the case of every state on the surface of the ball, there exists only one unique tight effect that gives probability one for that state. Hence, all these effects must be allowed, and since they generate the set of all effects, this proves requirement 5 for use in (14) and the rest of the paper.

## References

[1] Bell J S 1964 Physics 1195
[2] Shor P W 1997 Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer SIAM J. Comput. 26 1484-509 (arXiv:quant-ph/9508027v2)
[3] Hardy L 2009 Foliable operational structures for general probabilistic theories arXiv:0912.4740v1
[4] Hardy L 2001 Quantum theory from five reasonable axioms arXiv:quant-ph/0101012v4
[5] Barnum H and Wilce A 2008 Information processing in convex operational theories DCM/QPL (Oxford University, 2008) arXiv:0908.2352v1
[6] Barrett J 2007 Information processing in generalized probabilistic theories Phys. Rev. A 75032304 (arXiv:quant-ph/0508211v3)
[7] Mackey G W 1963 The Mathematical Foundations of Quantum Mechanics (New York: Benjamin)
[8] Birkhoff G and von Neumann J 1936 The logic of quantum mechanics Ann. Math. 37823
[9] Chiribella G, D'Ariano G M and Perinotti P 2010 Probabilistic theories with purification Phys. Rev. A 81062348 (arXiv:0908.1583v5)
[10] van Dam W 2005 Implausible consequences of superstrong nonlocality arXiv:quant-ph/0501159v1
[11] Pawlowski M, Paterek T, Kaszlikowski D, Scarani V, Winter A and Zukowski M 2009 A new physical principle: information causality Nature 4611101 (arXiv:0905.2292v3)
[12] Popescu S and Rohrlich D 1997 Causality and nonlocality as axioms for quantum mechanics Proc. Symp. Causality and Locality in Modern Physics and Astronomy (Toronto, York University, 1997) arXiv:quantph/9709026v2
[13] Navascues M and Wunderlich H 2009 A glance beyond the quantum model Proc. R. Soc. Lond. A 466 881-90 (arXiv:0907.0372v1)
[14] Gross D, Müller M, Colbeck R and Dahlsten O C O 2009 All reversible dynamics in maximally non-local theories are trivial Phys. Rev. Lett. 104080402 (arXiv:0910.1840v2)
[15] Gleason A 1957 J. Math. Mech. 6885
[16] Weinberg S 1989 Ann. Phys., NY 194336
[17] Gisin N 1990 Weinberg's nonlinear quantum mechanics and supraluminal communications Phys. Lett. A 143 1-2
[18] Aaronson S 2004 Is quantum mechanics an Island in theory space? arXiv:quant-ph/0401062v2
[19] Dakić B and Brukner C 2009 Quantum theory and beyond: is entanglement special? arXiv:0911.0695v1
[20] Fuchs C A 2002 Quantum mechanics as quantum information (and only a little more) arXiv:quantph/0205039v1
[21] Brassard G 2005 Is information the key? Nat. Phys. 12
[22] Alfsen E M and Shultz F W 2003 Geometry of State Spaces of Operator Algebras (Boston, MA: Birkhäuser)
[23] Rockafellar R T 1970 Convex Analysis (Princeton, NJ: Princeton University Press)
[24] Baker A 2006 Matrix Groups. An Introduction to Lie Group Theory (Berlin: Springer)
[25] Simon B 1996 Representations of Finite and Compact Groups (Graduate Studies in Mathematics vol 10) (Providence, RI: American Mathematical Society)
[26] Straszewicz S 1935 Über exponierte Punkte abgeschlossener Punktmengen Fund. Math. 24 139-43
[27] Onishchik A L and Gorbatsevich V V 1993 Lie groups and Lie algebras I Encyclopedia of Mathematical Sciences vol 20 (Berlin: Springer)
[28] Onishchik A L 1963 Transitive compact transformation groups Mat. Sb. (N.S.) 60 447-85 Onishchik A L 1966 Am. Math. Soc. Transl. 55 153-94 (Engl. Transl.)
[29] Fulton W and Harris J 2004 Representation Theory (Graduate Texts in Mathematics) (Berlin: Springer)
[30] Bargmann V 1964 Note on Wigner's theorem on symmetry operations J. Math. Phys. 5 862-8
[31] Wigner E P 1960 Normal form of antiunitary operators J. Math. Phys. 1409-13
[32] Bremner M J, Dawson C M, Dodd J L, Gilchrist A, Harrow A W, Mortimer D, Nielsen M A and Osborne T J 2002 Practical scheme for quantum computation with any two-qubit entangling gate Phys. Rev. Lett. 89247902 (arXiv:quant-ph/0207072v1)
[33] Chiribella G, D'Ariano G M and Perinotti P 2010 Informational derivation of quantum theory arXiv:1011.6451v2


[^0]:    ${ }^{5}$ Author to whom any correspondence should be addressed.

