

## Informational derivation of quantum theory

Giulio Chiribella\*

*Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Ontario, Canada N2L 2Y5<sup>†</sup>*

Giacomo Mauro D’Ariano<sup>‡</sup> and Paolo Perinotti<sup>§</sup>

*QUIT Group, Dipartimento di Fisica “A. Volta” and INFN Sezione di Pavia, via Bassi 6, I-27100 Pavia, Italy<sup>||</sup>*

(Received 29 November 2010; published 11 July 2011)

We derive quantum theory from purely informational principles. Five elementary axioms—causality, perfect distinguishability, ideal compression, local distinguishability, and pure conditioning—define a broad class of theories of information processing that can be regarded as standard. One postulate—purification—singles out quantum theory within this class.

DOI: [10.1103/PhysRevA.84.012311](https://doi.org/10.1103/PhysRevA.84.012311)

PACS number(s): 03.67.Ac, 03.65.Ta

### I. INTRODUCTION

More than 80 years after its formulation, quantum theory is still mysterious. The theory has a solid mathematical foundation, addressed by Hilbert, von Neumann, and Nordheim in 1928 [1] and brought to completion in the monumental work by von Neumann [2]. However, this formulation is based on the abstract framework of Hilbert spaces and self-adjoint operators, which, to say the least, are far from having an intuitive physical meaning. For example, the postulate stating that the pure states of a physical system are represented by unit vectors in a suitable Hilbert space appears as rather artificial: which are the physical laws that lead to this very specific choice of mathematical representation? The problem with the standard textbook formulations of quantum theory is that the postulates therein impose particular mathematical structures without providing any fundamental reason for this choice: the mathematics of Hilbert spaces is adopted without further questioning as a prescription that “works well” when used as a black box to produce experimental predictions. In a satisfactory axiomatization of quantum theory, instead, the mathematical structures of Hilbert spaces (or  $C^*$  algebras) should emerge as consequences of physically meaningful postulates, that is, postulates formulated exclusively in the language of physics: this language refers to notions like physical system, experiment, or physical process and not to notions like Hilbert space, self-adjoint operator, or unitary operator. Note that any serious axiomatization has to be based on postulates that can be precisely translated in mathematical terms. However, the point with the present status of quantum theory is that there are postulates that have a precise mathematical statement, but cannot be translated back into language of physics. Those are the postulates that one would like to avoid.

The need for a deeper understanding of quantum theory in terms of fundamental principles was clear since

the very beginning. Von Neumann himself expressed his dissatisfaction with his mathematical formulation of quantum theory with the surprising words “I don’t believe in Hilbert space anymore,” reported by Birkhoff in [3]. Realizing the physical relevance of the axiomatization problem, Birkhoff and von Neumann made an attempt to understand quantum theory as a new form of logic [4]: the key idea was that propositions about the physical world must be treated in a suitable logical framework, different from classical logics, where the operations AND and OR are no longer distributive. This work inaugurated the tradition of quantum logics, which led to several attempts to axiomatize quantum theory, notably by Mackey [5] and Jauch and Piron [6] (see Ref. [7] for a review on the more recent progresses of quantum logics). In general, a certain degree of technicality, mainly related to the emphasis on infinite-dimensional systems, makes these results far from providing a clear-cut description of quantum theory in terms of fundamental principles. Later Ludwig initiated an axiomatization program [8] adopting an operational approach, where the basic notions are those of preparation devices and measuring devices and the postulates specify how preparations and measurements combine to give the probabilities of experimental outcomes. However, despite the original intent, Ludwig’s axiomatization did not succeed in deriving Hilbert spaces from purely operational notions, as some of the postulates still contained mathematical notions with no operational interpretation.

More recently, the rise of quantum information science moved the emphasis from logics to information processing. The new field clearly showed that the mathematical principles of quantum theory imply an enormous amount of information-theoretic consequences, such as the no-cloning theorem [9,10], the possibility of teleportation [11], secure key distribution [12–14], or of factoring numbers in polynomial time [15]. The natural question is whether the implication can be reversed: is it possible to retrieve quantum theory from a set of purely informational principles? Another contribution of quantum information has been to shift the emphasis to finite dimensional systems, which allow for a simpler treatment but still possess all the remarkable quantum features. In a sense, the study of finite dimensional systems allows one to decouple the

\*gchiribella@perimeterinstitute.ca

<sup>†</sup><http://www.perimeterinstitute.ca>

<sup>‡</sup>dariano@unipv.it

<sup>§</sup>paolo.perinotti@unipv.it

<sup>||</sup><http://www.qubit.it>

conceptual difficulties in our understanding of quantum theory from the technical difficulties of infinite dimensional systems.

In this scenario, Hardy's 2001 work [16] re-opened the debate about the axiomatizations of quantum theory with fresh ideas. Hardy's proposal was based on five main assumptions about the relation between dimension of the state space and the number of perfectly distinguishable states of a given system, about the structure of composite systems, and about the possibility of connecting any two pure states of a physical system through a continuous path of reversible transformations. However, some of these assumptions directly refer to the mathematical properties of the state space (in particular, the "simplicity axiom" 2, which is an abstract statement about the functional dependence of the state space dimension on the number of perfectly distinguishable states). Very recently, building on Hardy's work there have been two new attempts of axiomatization by Dakic and Bruckner [17] and Masanes and Müller [18]. Although these works succeeded in removing the "simplicity axiom," they still contain mathematical assumptions that cannot be understood in elementary physical terms (see, e.g., requirement 5 of Ref. [18], which assumes that "all mathematically well-defined measurements are allowed by the theory").

Another approach to the axiomatization of quantum theory was pursued by one of the authors in a series of works [19] culminated in Ref. [20]. These works tackled the problem using operational principles related to tomography and calibration of physical devices, experimental complexity, and to the composition of elementary transformations. In particular this research introduced the concept of dynamically faithful states, namely states that can be used for the complete tomography of physical processes. Although this approach went very close to deriving quantum theory, in this case one mathematical assumption without operational interpretation was needed (see the CJ postulate of Ref. [20]).

In this paper we provide a complete derivation of finite dimensional quantum theory based on purely operational principles. Our principles do not refer to abstract properties of the mathematical structures that we use to represent states, transformations, or measurements, but only to the way in which states, transformations, and measurements combine with each other. More specifically, our principles are of *informational* nature: they assert basic properties of information processing, such as the possibility or impossibility to carry out certain tasks by manipulating physical systems. In this approach the rules by which information can be processed determine the physical theory, in accordance with Wheeler's program "it from bit," for which he argued that "all things physical are information-theoretic in origin" [21]. Note that, however, our axiomatization of quantum theory is relevant, as a rigorous result, also for those who do not share Wheeler's ideas on the informational origin of physics. In particular, in the process of deriving quantum theory we provide alternative proofs for many key features of the Hilbert space formalism, such as the spectral decomposition of self-adjoint operators or the existence of projections. The interesting feature of these proofs is that they are obtained by manipulation of the principles, without assuming Hilbert spaces from the start.

The main message of our work is simple: within a standard class of theories of information processing, quantum theory

is uniquely identified by a single postulate: *purification*. The purification postulate, introduced in Ref. [22], expresses a distinctive feature of quantum theory, namely that the ignorance about a part is always compatible with the maximal knowledge of the whole. The key role of this feature was noticed already in 1935 by Schrödinger in his discussion about entanglement [23], of which he famously wrote "I would not call that *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought." In a sense, our work can be viewed as the concrete realization of Schrödinger's claim: the fact that every physical state can be viewed as the marginal of some pure state of a compound system is indeed the key to single out quantum theory within a standard set of possible theories. It is worth stressing, however, that the purification principle assumed in this paper includes a requirement that was not explicitly mentioned in Schrödinger's discussion: if two pure states of a composite system AB have the same marginal on system A, then they are connected by some reversible transformation on system B. In other words, we assume that all purifications of a given mixed state are equivalent under local reversible operations [24].

The purification principle expresses a law of *conservation of information*, stating that at least in principle, irreversibility can always be reduced to the lack of control over an environment. More precisely, the purification principle is equivalent to the statement that every irreversible process can be simulated in an essentially unique way by a reversible interaction of the system with an environment, which is initially in a pure state [22]. This statement can also be extended to include the case of measurement processes, and in that case it implies the possibility of arbitrarily shifting the cut between the observer and the observed system [22]. The possibility of such a shift was considered by von Neumann as a "fundamental requirement of the scientific viewpoint" (see p. 418 of [2]) and his discussion of the measurement process was exactly aimed to show that quantum theory fulfils this requirement.

Besides Schrödinger's discussion on entanglement and von Neumann's discussion of the measurement process, the purification principle is deeply rooted in the structure of quantum theory. At the purely mathematical level it plays a crucial role in the theory of  $C^*$  algebras of operators on separable Hilbert spaces, where the purification principle is equivalent to the Gelfand-Naimark-Segal (GNS) construction [25] and implies the celebrated Stinespring's theorem [26]. On the other hand, purification is a cornerstone of quantum information, lying at the origin of most quantum protocols. As it was shown in Ref. [22], the purification principle directly implies crucial features like no-cloning, teleportation, no-information without disturbance, error correction, the impossibility of bit commitment, and the "no-programming" theorem of Ref. [27].

In addition to the purification postulate, our derivation of quantum theory is based on five informational axioms. The reason why we call them "axioms," as opposed to the the purification "postulate," is that they are not at all specific of quantum theory. These axioms represent standard features of information processing that everyone would, more or

less implicitly, assume. They define a class of theories of information processing that includes, for example, classical information theory, quantum information theory, and quantum theory with superselection rules. The question whether there are other theories satisfying our five axioms and, in case of a positive answer, the full classification of these theories is currently an open problem.

Here we informally illustrate the five axioms, leaving the detailed description to the remaining part of the paper:

(1) *Causality*: the probability of a measurement outcome at a certain time does not depend on the choice of measurements that will be performed later.

(2) *Perfect distinguishability*: if a state is not completely mixed (i.e., if it cannot be obtained as a mixture from any other state), then there exists at least one state that can be perfectly distinguished from it.

(3) *Ideal compression*: every source of information can be encoded in a suitable physical system in a lossless and maximally efficient fashion. Here *lossless* means that the information can be decoded without errors and *maximally efficient* means that every state of the encoding system represents a state in the information source.

(4) *Local distinguishability*: if two states of a composite system are different, then we can distinguish between them from the statistics of local measurements on the component systems.

(5) *Pure conditioning*: if a pure state of system AB undergoes an atomic measurement on system A, then each outcome of the measurement induces a pure state on system B. (Here *atomic measurement* means a measurement that cannot be obtained as a coarse graining of another measurement.)

All these axioms are satisfied by classical information theory. Axiom 5 is even trivial for classical theory, because the only pure states of a composite system AB are the product of pure states of the component systems A and B, and hence the state of system B will be pure irrespectively of what we do on system A.

A stronger version of axiom 5, introduced in Ref. [20], is the following:

(5') *Atomicity of composition*: the sequential composition of two atomic operations is atomic. (Here *atomic transformation* means a transformation that cannot be obtained from coarse graining.)

However, it turns out that axiom 5 is enough for our derivation: thanks to the purification postulate we will be able to show the nontrivial implication: axiom 5  $\Rightarrow$  axiom 5' (see lemma 16).

The paper is organized as follows. In Sec. II we review the framework of *operational-probabilistic theories* introduced in Ref. [22]. This framework will provide the basic notions needed for the formulation of our principles. In Sec. III we introduce the principles from which we will derive quantum theory. In Sec. IV we prove some direct consequences of the principles that will be used later in the paper. In Sec. V we discuss the properties of perfectly distinguishable states, while in Sec. VI we prove the existence of a duality between pure states and atomic effects.

The results about distinguishability and duality of pure states and atomic effects allow us to show in Sec. VII that every system has a well defined *informational dimension*—the operational counterpart of the Hilbert space dimension.

Section VIII contains the proof that every state can be decomposed as a convex combination of perfectly distinguishable pure states. Similarly, any element of the vector space spanned by the states can be written as a linear combination of perfectly distinguishable states. This result corresponds to the spectral theorem for self-adjoint operators on complex Hilbert spaces. In Sec. IX we prove some results about the maximum teleportation probability, which allow us to derive a functional relation between the dimension of the state space and the number of perfectly distinguishable states of the system. The mathematical representation of systems with two perfectly distinguishable states is derived in Sec. X, where we prove that such systems are indeed two-dimensional quantum systems—also known as qubits. In Sec. XI we construct projections on the faces of the state space of any system and prove their main properties. These results lead to the derivation of the operational analog of the superposition principle in Sec. XII which allows us to prove that systems with the same number of perfectly distinguishable states are operationally equivalent (Sec. XII B). The properties of the projections and the superposition principle are then exploited in Sec. XIII, where we extend the density matrix representation from qubits to higher dimensional systems, thus proving that a system with  $d$  perfectly distinguishable states is indeed a quantum system with  $d$ -dimensional Hilbert space. We conclude the paper with Sec. XIV, where we review our results, discussing future directions for this research.

## II. THE FRAMEWORK

This section provides a brief summary of the framework of *operational-probabilistic theories*, which was formulated in Ref. [22]. We refer to Ref. [22] for an exhaustive presentation of the details of the framework and of the ideas behind it. The operational-probabilistic framework combines the operational language of circuits with the toolbox of probability theory: on the one hand experiments are described by circuits resulting from the connection of physical devices, on the other hand each device in the circuit can have classical outcomes and the theory provides the probability distribution of outcomes when the devices are connected to form closed circuits (that is, circuits that start with a preparation and end with a measurement).

The notions discussed in this section will allow us to draw a precise distinction between principles with an operational content and exclusively mathematical principles: with the expression “operational principle” we will mean a principle that can be expressed using only the basic notions of the the operational-probabilistic framework.

### A. Circuits with outcomes

A *test* represents one use of a physical device, like a Stern-Gerlach magnet, a beamsplitter, or a photon counter. The device will have an *input system* and an *output system*, labeled by capital letters. The corresponding test can have different classical outcomes, represented by different values of an index  $i \in X$ :

$$\text{---} \text{A} \text{---} \boxed{\{\mathcal{C}_i\}_{i \in X}} \text{---} \text{B} \text{---}$$



The same construction holds for observation tests: every event  $a_j \in \text{Eff}(A)$  induces a function  $\hat{a}_j : \text{St}(A) \rightarrow \mathbb{R}$ , given by  $\hat{a}_j(\rho_i) := (a_j | \rho_i)$ . If two events  $a_j, a'_j \in \text{Eff}(A)$  induce to the same function, then it is impossible to distinguish between them from the statistics of the experiments allowed in our theory. This means that for our purposes the two events are the same: accordingly, we will take equivalence classes with respect to the relation  $a_j \simeq a'_j$  if  $\hat{a}_j = \hat{a}'_j$ . To avoid introducing new notation, from now on we will identify the event  $a_j \in \text{Eff}(A)$  with the corresponding function  $\hat{a}_j$  and we will call it *effect*. Accordingly, we will refer to observation tests as collection of effects  $\{a_j\}_{j \in Y}$ . The effects in  $\text{Eff}(A)$  generate a real vector space, denoted by  $\text{Eff}_{\mathbb{R}}(A)$ .

A vector in  $\text{St}_{\mathbb{R}}(A)$  [or  $\text{Eff}_{\mathbb{R}}(A)$ ] can be extended to a linear function on  $\text{Eff}_{\mathbb{R}}(A)$  [or  $\text{St}_{\mathbb{R}}(A)$ ]. In this way, states and effects can be thought as elements of two real vector spaces, one dual to the other. In this paper we will restrict our attention to *finite dimensional* vector spaces: operationally this means that the state of a given physical system is completely determined by the statistics of a finite number of finite-outcome measurements. The dimension of the vector space  $\text{St}_{\mathbb{R}}(A)$ , which by construction is equal to the dimension of its dual  $\text{Eff}_{\mathbb{R}}(A)$ , will be denoted by  $D_A$ . We will refer to  $D_A$  as the *size* of system  $A$ .

Finally, the vector spaces  $\text{St}_{\mathbb{R}}(A)$  and  $\text{Eff}_{\mathbb{R}}(A)$  can be equipped with suitable norms, which have an operational meaning related to optimal discrimination schemes [22]. The norm of an element  $\delta \in \text{St}_{\mathbb{R}}(A)$  is given by [22]

$$\|\delta\| = \sup_{a_0 \in \text{Eff}(A)} (a_0 | \delta) - \inf_{a_1 \in \text{Eff}(A)} (a_1 | \delta),$$

while the norm of an element  $\xi \in \text{Eff}_{\mathbb{R}}(A)$  is given by

$$\|\xi\| = \sup_{\rho \in \text{St}(A)} |(\xi | \rho)|.$$

We will always take the set of states  $\text{St}(A)$  to be closed in the operational norm. The convenience of this choice is the convenience of using real numbers instead of rational ones: dealing with a single real number is much easier than dealing with a Cauchy sequence of rational numbers. Operationally taking  $\text{St}(A)$  to be closed is very natural: the fact that there is a sequence of states  $\{\rho_n\}_{n=1}^{\infty}$  that converges to  $\rho \in \text{St}_{\mathbb{R}}(A)$  means that there is a procedure to prepare  $\rho$  with arbitrary precision and hence that  $\rho$  deserves the name of “state”.

We conclude this subsection by noting that every event  $\mathcal{C}_k$  from  $A$  to  $B$  induces a linear map  $\hat{\mathcal{C}}_k$  from  $\text{St}_{\mathbb{R}}(A)$  to  $\text{St}_{\mathbb{R}}(B)$ , uniquely defined by

$$\hat{\mathcal{C}}_k : |\rho\rangle \in \text{St}(A) \mapsto \mathcal{C}_k |\rho\rangle \in \text{St}(B).$$

Likewise, for every system  $C$  the event  $\mathcal{C}_k \otimes \mathcal{I}_C$  induces a linear map  $\hat{\mathcal{C}}_k \otimes \mathcal{I}_C$  from  $\text{St}_{\mathbb{R}}(AC)$  to  $\text{St}_{\mathbb{R}}(BC)$ . If two events  $\mathcal{C}_k$  and  $\mathcal{C}'_k$  induce the same maps for every possible system  $C$ , then there is no experiment in the theory that is able to distinguish between them. This means that for our purposes the two events are the same: accordingly, we will take equivalence classes with respect to the relation  $\mathcal{C}_k \simeq \mathcal{C}'_k$  if  $\hat{\mathcal{C}}_k \otimes \mathcal{I}_C = \hat{\mathcal{C}'_k} \otimes \mathcal{I}_C$  for every system  $C$ . In this case, we will say that two events represent the same *transformation*. Accordingly, we will refer to tests  $\{\mathcal{C}_i\}_{i \in X}$  as collections of transformations.

The deterministic transformations (corresponding to single-outcome tests) will be called *channels*.

### C. Basic definitions in the operational-probabilistic framework

Here we summarize few elementary definitions that will be used later in the paper. The meaning of the definitions in the case of quantum theory is also discussed.

#### 1. Coarse graining, refinement, atomic transformations, pure, mixed and completely mixed states

First, we start from the notions of *coarse graining* and *refinement*. Coarse graining arises when we join together some outcomes of a test: we say that the test  $\{\mathcal{D}_j\}_{j \in Y}$  is a coarse graining of the test  $\{\mathcal{C}_i\}_{i \in X}$  if there is a disjoint partition  $\{X_j\}_{j \in Y}$  of  $X$  such that

$$\mathcal{D}_j = \sum_{i \in X_j} \mathcal{C}_i.$$

Conversely, if  $\{\mathcal{D}_j\}_{j \in Y}$  is a coarse graining of  $\{\mathcal{C}_i\}_{i \in X}$ , we say that  $\{\mathcal{C}_i\}_{i \in X}$  is a refinement of  $\{\mathcal{D}_j\}_{j \in Y}$ . Intuitively, a test that refines another is a test that extracts information in a more precise way: it is a test with better “resolving power.”

The notion of refinement also applies to a single transformation: a refinement of the transformation  $\mathcal{C}$  is given by a test  $\{\mathcal{C}_i\}_{i \in X}$  and a subset  $X_0$  such that

$$\mathcal{C} = \sum_{i \in X_0} \mathcal{C}_i.$$

Accordingly, we say that each transformation  $\mathcal{C}_i, i \in X_0$  is a *refinement* of  $\mathcal{C}$ . A transformation  $\mathcal{C}$  is *atomic* if it has only trivial refinements: if  $\mathcal{C}_i$  refines  $\mathcal{C}$ , then  $\mathcal{C}_i = p\mathcal{C}$  for some probability  $p \geq 0$ . A test that consists of atomic transformations is a test whose “resolving power” cannot be further improved.

When discussing states (i.e., transformations with trivial input) we will use the word *pure* as a synonym of atomic. A pure state describes a situation of maximal knowledge about the system’s preparation, a knowledge that cannot be further refined.

As usual, a state that is not pure will be called *mixed*. An important notion is that of *completely mixed state*.

*Definition 1 (Completely mixed state).* A state is completely mixed if any other state can refine it: precisely,  $\omega \in \text{St}(A)$  is completely mixed if for every  $\rho \in \text{St}(A)$  there is a nonzero probability  $p > 0$  such that  $p\rho$  is a refinement of  $\omega$ .

Intuitively, a completely mixed state describes a situation of complete ignorance about the system’s preparation: if a system is described by a completely mixed state, then it means that we know so little about its preparation that, in fact, every preparation is possible.

We conclude this paragraph with a couple of definitions that will be used throughout the paper.

*Definition 2 (Reversible transformation).* A transformation  $\mathcal{U} \in \text{Transf}(A, B)$  is reversible if there exists another transformation  $\mathcal{U}^{-1} \in \text{Transf}(B, A)$  such that  $\mathcal{U}^{-1}\mathcal{U} = \mathcal{I}_A$  and  $\mathcal{U}\mathcal{U}^{-1} = \mathcal{I}_B$ . When  $A = B$  the reversible transformations form a group, indicated as  $\mathbf{G}_A$ .

*Definition 3 (Operationally equivalent systems).* Two systems A and B are operationally equivalent if there exists a reversible transformation  $\mathcal{U}$  from A to B.

When two systems are operationally equivalent one can convert one into the other in a reversible fashion.

**2. Examples in quantum theory**

Consider a quantum system with Hilbert space  $\mathcal{H} = \mathbb{C}^d$ ,  $d < \infty$ . In this case a preparation test is a collection of unnormalized density matrices  $\{\rho_i\}_{i \in X}$  (i.e., of nonnegative  $d \times d$  complex matrices with trace bounded by 1) such that

$$\sum_{i \in X} \text{Tr}[\rho_i] = 1.$$

Preparation tests are often called *quantum information sources* in quantum information theory. A generic state  $\rho$  is an unnormalized density matrix. A deterministic state, corresponding to a single-outcome preparation test, is a normalized density matrix  $\rho$ , with  $\text{Tr}[\rho] = 1$ .

Diagonalizing  $\rho = \sum_i \alpha_i |\psi_i\rangle\langle\psi_i|$  we then obtain that each matrix  $\alpha_i |\psi_i\rangle\langle\psi_i|$  is a refinement of  $\rho$ . More generally, every matrix  $\sigma$  such that  $\sigma \leq \rho$  is a refinement of  $\rho$ . Up to a positive rescaling, all matrices with support contained in the support of  $\rho$  are refinements of  $\rho$ . A quantum state  $\rho$  is atomic (pure) if and only if it is proportional to a rank-one projection. A quantum state is completely mixed if and only if its density matrix has full rank. Note that the quantum state  $\chi = \frac{I_d}{d}$ , where  $I_d$  is the identity  $d \times d$  matrix, is a particular example of completely mixed state, but not the only example. Precisely,  $\chi = \frac{I_d}{d}$  is the unique unitarily invariant state in dimension  $d$ .

Let us now consider the case of observation tests: in quantum theory an observation test is given by a *POVM (positive operator-valued measure)*, namely by a collection  $\{P_j\}_{j \in Y}$  of nonnegative  $d \times d$  matrices such that

$$\sum_{j \in Y} P_j = I_d.$$

An effect is then a nonnegative matrix  $P \geq 0$  upper bounded by the identity. In quantum theory there is only one deterministic effect, corresponding to a single-outcome observation test: the unique deterministic effect given by the identity matrix. As we will see in the following section, the fact that the deterministic effect is unique is equivalent to the fact that quantum theory is a *causal* theory.

An effect  $P$  is atomic if and only if  $P$  is proportional to a rank-one projector. An observation test is atomic if it is a POVM with rank-one elements.

Finally, a general test from an input system with Hilbert space  $\mathcal{H}_1 = \mathbb{C}^{d_1}$  to an output system with Hilbert space  $\mathcal{H}_2 = \mathbb{C}^{d_2}$  is given by a *quantum instrument*, namely by a collection  $\{\mathcal{C}_k\}_{k \in Z}$  of completely positive trace nonincreasing maps sending linear operators on  $\mathcal{H}_1$  to linear operators on  $\mathcal{H}_2$ , with the property that

$$\mathcal{C}_Z := \sum_{k \in Z} \mathcal{C}_k$$

is trace preserving. A general transformation is then given by a trace nonincreasing map, called *quantum operation*, whereas a deterministic transformation, corresponding to a

single-outcome test, is given by a trace-preserving map, called *quantum channel*.

Any quantum operation  $\mathcal{C}$  can be written in the Kraus form  $\mathcal{C}(\rho) = \sum_i C_i \rho C_i^\dagger$ , where  $C_i : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  are the Kraus operators. Up to a positive scaling, every quantum operation  $\mathcal{D}$  such that the Kraus operators of  $\mathcal{D}$  belong to the linear span of the Kraus operators of  $\mathcal{C}$  is a refinement of  $\mathcal{C}$ . A map  $\mathcal{C}$  is atomic if and only if there is only one Kraus operator in its Kraus form. A reversible transformation in quantum theory is a unitary map  $\mathcal{U}(\rho) = U \rho U^\dagger$ , where  $U : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  is a unitary operator, that is  $U^\dagger U = I_1$  and  $U U^\dagger = I_2$  where  $I_1$  ( $I_2$ ) is the identity operator on  $\mathcal{H}_1$  ( $\mathcal{H}_2$ ). Two quantum systems are operationally equivalent if and only if the corresponding Hilbert spaces have the same dimension.

**D. Operational principles**

We are now in position to make precise the usage of the expression “operational principle” in the context of this paper. By operational principle we mean here a principle that can be stated using only the operational-probabilistic language, i.e., using only

(1) the notions of system, test, outcome, probability, state, effect, transformation;

(2) their specifications: atomic, pure, mixed, completely mixed; and

(3) more complex notions constructed from the above terms (e.g., the notion of “reversible transformation”).

The distinction between operational principles and principles referring to abstract mathematical properties, mentioned in the Introduction, should now be clear: for example, a statement like “the pure states of a system cannot be cloned” is a valid operational principle, because it can be analyzed in basic operational-probabilistic terms as “for every system A there exists no transformation  $\mathcal{C}$  with input system A and output system AA such that  $\mathcal{C}|\varphi\rangle = |\varphi\rangle|\varphi\rangle$  for every pure state  $\varphi$  of A.” On the contrary, a statement like “the state space of a system with two perfectly distinguishable states is a three-dimensional sphere” is not a valid operational principle, because there is no way to express what it means for a state space to be a three-dimensional sphere in terms of basic operational notions. The fact that a state space is a sphere may be eventually derived from operational principles, but cannot be assumed as a starting point.

**III. THE PRINCIPLES**

We now state the principles used in our derivation. The first five principles express generic features that are shared by both classical and quantum theory. They could be even included in the definition of the background framework: they define the simple model of information processing in which we try to single out quantum theory. For this reason we will call them *axioms*. The sixth principle in our derivation has a different status: it expresses the genuinely quantum features. A major message of our work is that, within a broad class of theories of information processing, quantum theory is completely described by the purification principle. To emphasize the special role of the sixth principle we will call it *postulate*, in analogy with the parallel postulate of Euclidean geometry.

## A. Axioms

### 1. Causality

The first axiom of our list, *causality* [22], is so basic that could be considered as part of the background framework. We decided to explicitly present it as an axiom for two reasons: The first reason is that the framework of operational-probabilistic theories can be developed even without this requirement (see Ref. [22] for the general framework and Refs. [31,32] for two explicit examples of noncausal theories). The second reason is that we want to stress that causality is an essential ingredient in our derivation. This observation is important in view of possible extensions of quantum theory to quantum gravity scenarios where the causal structure is not defined from the start (see, e.g., Hardy in Ref. [33]).

*Axiom 1 (Causality)*. The probability of preparations is independent of the choice of observations.

In technical terms: if  $\{\rho_i\}_{i \in X} \subset \text{St}(A)$  is a preparation test, then the conditional probability of the preparation  $\rho_i$  given the choice of the observation-test  $\{a_j\}_{j \in Y}$  is the marginal

$$p(i|\{a_j\}) := \sum_{j \in Y} (a_j|\rho_i).$$

The axiom states that the marginal probability  $p(i|\{a_j\})$  is independent of the choice of the observation-test  $\{a_j\}$ : if  $\{a_j\}_{j \in Y}$  and  $\{b_k\}_{k \in Z}$  are two different observation tests, then one has  $p(i|\{a_j\}) = p(i|\{b_k\})$ . Loosely speaking, one may refer to causality as a requirement of *no signaling from the future*: indeed, causality is equivalent to the fact that the probability of an outcome at a certain time does not depend on the choice of operations that will be done at later times [20].

An operational-probabilistic theory that satisfies the causality axiom 1 will be called *causal*. As we already mentioned, causality is a very basic requirement and could be considered as part of the framework: it provides the notions used to state the other axioms and it implies several facts that will be used frequently in the paper. In fact, in our derivation we do not use the causality axiom directly, but only through its consequences. In the following we briefly summarize the facts and the notations that characterize the framework of *causal operational-probabilistic theories*, introduced and discussed in detail in Ref. [22]. Similar structures have been subsequently considered in Refs. [34,35] within a formal description of circuits in foliable space-time regions.

First, causality is equivalent to the existence of an effect  $e_A$  such that  $e_A = \sum_{j \in X} a_j$  for every observation-test  $\{a_j\}_{j \in Y}$ . We call the effect  $e_A$  the *deterministic effect* for system A. By definition, the effect  $e_A$  is unique. The subindex A in  $e_A$  will be dropped when no confusion can arise.

In a causal theory every test  $\{\mathcal{E}_i\}_{i \in X} \subset \text{Transf}(A, B)$  satisfies the condition

$$\sum_{i \in X} (e_B|\mathcal{E}_i = (e_A|).$$

As a consequence, a transformation  $\mathcal{C} \in \text{Transf}(A, B)$  satisfies the condition

$$(e_B|\mathcal{C} \leq (e_A|, \quad (4)$$

with the equality if and only if  $\mathcal{C}$  is a channel (i.e., a deterministic transformation, corresponding to a single-outcome test). In

Eq. (4) we used the notation  $(a| \leq (a'|$  to mean  $(a|\rho) \leq (a'|\rho)$  for every  $\rho \in \text{St}(A)$ .

In a causal theory the norm of a state  $\rho_i \in \text{St}(A)$  is given by  $\|\rho_i\| = (e|\rho_i)$ . Accordingly, one can define the normalized state

$$\bar{\rho}_i := \frac{\rho_i}{(e|\rho_i)}.$$

In a causal theory one can always allow for *rescaled preparations*: conditionally to the outcome  $i \in X$  in the preparation-test  $\{\rho_i\}_{i \in X}$  we can say that we prepared the normalized state  $\bar{\rho}_i$ . For this reason, every state in a causal theory is proportional to a normalized state.

The set of normalized states will be denoted by  $\text{St}_1(A)$ . Since the set of all states  $\text{St}(A)$  is closed in the operational norm, also the set of normalized states  $\text{St}_1(A)$  is closed. Moreover, the set  $\text{St}_1(A)$  is convex [22]: this means that for every pair of normalized states  $\rho_1, \rho_2 \in \text{St}_1(A)$  and for every probability  $p \in [0, 1]$  the convex combination  $\rho_p = p\rho_1 + (1-p)\rho_2$  is a normalized state. Operationally, the state  $\rho_p$  is obtained by

- (1) performing a binary test with outcomes  $\{1, 2\}$  and outcome probabilities  $p_1 = p$  and  $p_2 = 1 - p$ ;
- (2) for outcome  $i$  preparing  $\rho_i$ , thus realizing the preparation-test  $\{p_i \rho_i\}_{i=1,2}$ ;
- (3) coarse graining over the outcomes, thus obtaining  $\rho_p = p\rho_1 + (1-p)\rho_2$ .

The step 2 (preparation of a state conditionally on the outcome of a previous test) is possible because the theory is causal [22].

The pure normalized states are the extreme points of the convex set  $\text{St}_1(A)$ . For a normalized state  $\rho \in \text{St}_1(A)$  we define the *face identified by  $\rho$*  as follows.

*Definition 4 (Face identified by a state)*. The face identified by  $\rho \in \text{St}_1(A)$  is the set  $F_\rho$  of all normalized states  $\sigma \in \text{St}_1(A)$  such that  $\rho = p\sigma + (1-p)\tau$ , for some nonzero probability  $p > 0$  and some normalized state  $\tau \in \text{St}_1(A)$ .

In other words,  $F_\rho$  is the set of all normalized states that show up in the convex decompositions of  $\rho$ . Clearly, if  $\varphi$  is a pure state, then one has  $F_\varphi = \{\varphi\}$ . The opposite situation is that of completely mixed states: by definition 1, a state  $\omega \in \text{St}_1(A)$  is completely mixed if every state  $\sigma \in \text{St}_1(A)$  can stay in its convex decomposition, that is, if  $F_\omega = \text{St}_1(A)$ . An equivalent condition for a state to be completely mixed is the following.

*Lemma 1*. A state  $\omega \in \text{St}_1(A)$  is completely mixed if and only if  $\text{Span}(F_\omega) = \text{St}_\mathbb{R}(A)$ .

*Proof*. The condition is clearly necessary. It is also sufficient because for a state  $\sigma \in \text{St}_1(A)$  the relation  $\sigma \in \text{Span}(F_\omega)$  implies  $\sigma \in F_\omega$  (see lemma 16 of Ref. [22]). ■

A completely mixed state can never be distinguished from another state with zero error probability.

*Proposition 1*. Let  $\rho \in \text{St}_1(A)$  be a completely mixed state and  $\sigma \in \text{St}_1(A)$  be an arbitrary state. Then, the probability of error in distinguishing  $\rho$  from  $\sigma$  is strictly greater than zero.

*Proof*. By contradiction, suppose that one can distinguish between  $\rho$  and  $\sigma$  with zero error probability. This means that there exists a binary test  $\{a_\rho, a_\sigma\}$  such that  $(a_\rho|\sigma) = (a_\sigma|\rho) = 0$ . Since  $\rho$  is completely mixed there exists a probability  $p > 0$  and a state  $\tau \in \text{St}_1(A)$  such that  $\rho = p\sigma + (1-p)\tau$ . Hence, the condition  $(a_\sigma|\rho) = 0$  implies  $(a_\sigma|\sigma) = 0$ . Therefore, we have  $(a_\rho|\sigma) + (a_\sigma|\sigma) = 0$ . This is in contradiction with the

normalization of the probabilities in the test  $\{a_\rho, a_\sigma\}$ , which would require  $(a_\rho|\sigma) + (a_\sigma|\rho) = 1$ . ■

**2. Perfect distinguishability**

Our second axiom regards the task of state discrimination. As we saw in proposition 1, if a state is completely mixed, then it is impossible to distinguish it perfectly from any other state. Axiom 2 states the converse.

*Axiom 2 (Perfect distinguishability).* Every state that is not completely mixed can be perfectly distinguished from some other state.

Note that the statement of axiom 2 holds for quantum and for classical information theory. In quantum theory a completely mixed state is a density matrix with full rank. If a density matrix  $\rho$  has not full rank, then it must have a kernel: hence, every density matrix  $\sigma$  with support in the kernel of  $\rho$  will be perfectly distinguishable from  $\rho$ , as stated in axiom 2. Applying the same reasoning for density matrices that are diagonal in a given basis, one can easily see that axiom 2 is satisfied also by classical information theory.

To the best of our knowledge, the perfect distinguishability property has never been considered in the literature as an axiom, probably because in most works it came for free as a consequence of stronger mathematical assumptions. For example, one can obtain the perfect distinguishability property from the *no-restriction hypothesis* of Ref. [22], stating that for every system A any binary probability rule [i.e., any pair of positive functionals  $a_0, a_1 \in \text{Eff}_{\mathbb{R}}(A)$  such that  $a_0 + a_1 = e_A$ ] actually describes a measurement allowed by the theory. This assumption was made, for example, in Ref. [18] in the case of systems with at most two distinguishable states (see requirement 5 of Ref. [18]). Note that the difference between the perfect distinguishability axiom and the no-restriction hypothesis is that the former can be expressed in purely operational terms, whereas the latter requires the notion of “positive functional” which is not part of the basic operational language.

**3. Ideal compression**

The third axiom is about information compression. An *information source* for system A is a preparation-test  $\{\rho_i\}_{i \in X}$ , where each  $\rho_i \in \text{St}(A)$  is an unnormalized state and  $\sum_{i \in X} (e|\rho_i) = 1$ . A compression scheme is given by an *encoding operation*  $\mathcal{E}$  from A to a *smaller system* C, that is, to a system C such that  $D_C \leq D_A$ . The compression scheme is *lossless for the source*  $\{\rho_i\}_{i \in X}$  if there exists a *decoding operation*  $\mathcal{D}$  from C to A such that  $\mathcal{D}\mathcal{E}|\rho_i) = |\rho_i)$  for every value of the index  $i \in X$ . This means that the decoding allows one to perfectly retrieve the states  $\{\rho_i\}_{i \in X}$ . We say that a compression scheme is *lossless* for the state  $\rho$ , if it is lossless for every source  $\{\rho_i\}_{i \in X}$  such that  $\rho = \sum_{i \in X} \rho_i$ . Equivalently, this means that the restriction of  $\mathcal{D}\mathcal{E}$  to the face identified by  $\rho$  is equal to the identity channel:  $\mathcal{D}\mathcal{E}|\sigma) = \sigma$  for every  $\sigma \in F_\rho$ .

A lossless compression scheme is *maximally efficient* if the encoding system C has the smallest possible size, that is, if the system C has no more states than exactly those needed to compress  $\rho$ . This happens when every normalized state  $\tau \in \text{St}_1(C)$  comes from the encoding of some normalized state  $\sigma \in F_\rho$ , namely  $|\tau) = \mathcal{E}|\sigma)$ .

We say that a compression scheme that is lossless and maximally efficient is *ideal*. Our second axiom states that ideal compression is always possible.

*Axiom 3 (Ideal compression).* For every state there exists an ideal compression scheme.

It is easy to see that this statement holds in quantum theory and in classical probability theory. For example, if  $\rho$  is a density matrix on a  $d$ -dimensional Hilbert space and  $\text{rank}(\rho) = r$ , then the ideal compression is obtained by just encoding  $\rho$  in an  $r$ -dimensional Hilbert space. As long as we do not tolerate losses, this is the most efficient one-shot compression we can devise in quantum theory. Similar observations hold for classical information theory.

It is important to emphasize the difference between our “ideal compression” axiom and the “subspace” axiom of Refs. [16–18]: differently from the subspace axiom, the compression axiom is not an axiom about the structure of perfectly distinguishable states available for a given system. For example, here we do not assume that all systems with the same number of perfectly distinguishable states are equivalent. This fact will be proved from the principles in Sec. XII B.

**4. Local distinguishability**

The fourth axiom consists in the assumption of local distinguishability, here presented in the formulation of Ref. [22].

*Axiom 4 (Local distinguishability).* If two bipartite states are different, then they give different probabilities for at least one product experiment.

In more technical terms: if  $\rho, \sigma \in \text{St}_1(AB)$  are states and  $\rho \neq \sigma$ , then there are two effects  $a \in \text{Eff}(A)$  and  $b \in \text{Eff}(B)$  such that

$$\left( \begin{array}{c} \text{A} \\ \text{B} \end{array} \middle| \rho \right) \begin{array}{c} \text{---} a \\ \text{---} b \end{array} \neq \left( \begin{array}{c} \text{A} \\ \text{B} \end{array} \middle| \sigma \right) \begin{array}{c} \text{---} a \\ \text{---} b \end{array}$$

Local distinguishability is equivalent to the fact that two distant parties, holding systems A and B, respectively, can distinguish between the two states  $\rho, \sigma \in \text{St}_1(AB)$  using only local operations and classical communication and achieving an error probability strictly larger than  $p_{\text{ran}} = 1/2$ , the probability of error in random guess [22]. Again, this statement holds in ordinary quantum theory (on complex Hilbert spaces) and in classical information theory.

Another equivalent condition to local distinguishability is the *local tomography axiom*, introduced in Refs. [19,36]. The local tomography axiom imposes that every bipartite state can be reconstructed from the statistics of local measurements on the component systems. Technically, local tomography is in turn equivalent to the relation  $D_{AB} = D_A D_B$  [16] and to the fact that every state  $\rho \in \text{St}(AB)$  can be written as

$$\rho = \sum_{i=1}^{D_A} \sum_{j=1}^{D_B} \rho_{ij} \alpha_i \otimes \beta_j,$$

where  $\{\alpha_i\}_{i=1}^{D_A}$  ( $\{\beta_j\}_{j=1}^{D_B}$ ) is a basis for the vector space  $\text{St}_{\mathbb{R}}(A)$  [ $\text{St}_{\mathbb{R}}(B)$ ]. The analog condition also holds for effects: every bipartite effect  $E \in \text{Eff}(AB)$  can be written as

$$E = \sum_{i=1}^{D_A} \sum_{j=1}^{D_B} E_{ij} a_i \otimes b_j,$$

where  $\{a_i\}_{i=1}^{D_A}$  ( $\{b_j\}_{j=1}^{D_B}$ ) is a basis for the vector space  $\text{Eff}_{\mathbb{R}}(A)$  [ $\text{Eff}_{\mathbb{R}}(B)$ ].

An important consequence of local distinguishability, observed in Ref. [22], is that a transformation  $\mathcal{C} \in \text{Transf}(AB)$  is completely specified by its action on  $\text{St}(A)$ : thanks to local distinguishability we have the implication

$$\mathcal{C}|\rho\rangle = \mathcal{C}'|\rho\rangle \quad \forall \rho \in \text{St}(A) \implies \mathcal{C} = \mathcal{C}'. \quad (5)$$

(see lemma 14 of Ref. [22] for the proof). Note that Eq. (5) does not hold for quantum theory on real Hilbert spaces [22].

### 5. Pure conditioning

The fourth axiom states how the outcomes of a measurement on one side of a pure bipartite state can induce pure states on the other side. In this case we consider *atomic measurements*, that is, measurements described by observation-tests  $\{a_i\}_{i \in X}$  where each effect  $a_i$  is atomic. Intuitively, atomic measurements are those with maximum “resolving power.”

*Axiom 5 (Pure conditioning)*. If a bipartite system is in a pure state, then each outcome of an atomic measurement on one side induces a pure state on the other.

The pure conditioning property holds in quantum theory and in classical information theory as well. In fact, the statement is trivial in classical information theory, because the only pure bipartite states are the product of pure states: no matter which measurement is performed on one side, the remaining state on the other side will necessarily be pure.

The pure conditioning property, as formulated above, has been recently introduced in Ref. [37]. A stronger version of axiom 5 is the *atomicity of composition* introduced in Ref. [20]:

*5' Atomicity of composition*: the sequential composition of two atomic operations is atomic.

Since pure states and atomic effects are a particular case of atomic transformations, axiom 5' implies axiom 5. In our derivation, however, also the converse implication holds: indeed, thanks to the purification postulate we will be able to show that axiom 5 implies axiom 5' (see lemma 16).

### B. The purification postulate

The last postulate in our list is the purification postulate, which was introduced and explored in detail in Ref. [22]. While the previous axioms were also satisfied by classical probability theory, the purification axiom introduces in our derivation the genuinely quantum features. A *purification* of the state  $\rho \in \text{St}_1(A)$  is a pure state  $\Psi_\rho$  of some composite system AB, with the property that  $\rho$  is the marginal of  $\Psi_\rho$ , that is,

$$\rho \text{---} A = \left( \Psi_\rho \begin{array}{c} \text{---} A \\ \text{---} B \\ \text{---} e_B \end{array} \right).$$

Here we refer to the system B as the *purifying system*. The purification axiom states that every state can be obtained as the marginal of a pure bipartite state in an essentially unique way.

*Postulate 1 (Purification)*. Every state has a purification. For fixed purifying system, every two purifications of the

same state are connected by a reversible transformation on the purifying system.

Informally speaking, our postulate states that the ignorance about a part is always compatible with a maximal knowledge of the whole. The existence of pure bipartite states with mixed marginal was already recognized by Schrödinger as the characteristic trait of quantum theory [23]. Here, however, we also emphasize the importance of the uniqueness of purification up to reversible transformations: this property sets up a relation between pure states and reversible transformations that generates most of the structure of quantum theory. As shown in Ref. [22], an impressive number of quantum features are actually direct consequences of purification. In particular, purification implies the possibility of simulating any irreversible process through a reversible interaction of the system with an environment that is finally discarded.

## IV. FIRST CONSEQUENCES OF THE PRINCIPLES

### A. Results about ideal compression

Let  $\rho \in \text{St}_1(A)$  be a state and let  $\mathcal{E} \in \text{Transf}(A,C)$  [or  $\mathcal{D} \in \text{Transf}(C,A)$ ] be its encoding (or decoding) in the ideal compression scheme of axiom 3.

Essentially the encoding operation  $\mathcal{E} \in \text{Transf}(A,C)$  identifies the face  $F_\rho$  with the state space  $\text{St}_1(C)$ . In the following we provide a list of elementary lemmas showing that all statements about  $F_\rho$  can be translated into statements about  $\text{St}_1(C)$  and vice versa.

*Lemma 2*. The composition of decoding and encoding is the identity on C, namely  $\mathcal{E}\mathcal{D} = \mathcal{I}_C$ .

*Proof*. Since the compression is maximally efficient, for every state  $\tau \in \text{St}_1(C)$  there is a state  $\sigma \in F_\rho$  such that  $\mathcal{E}\sigma = \tau$ . Using the fact that  $\mathcal{D}\mathcal{E}\sigma = \sigma$  (the compression is lossless) we then obtain  $\mathcal{E}\mathcal{D}\tau = \mathcal{E}\mathcal{D}\mathcal{E}\sigma = \mathcal{E}\sigma = \tau$ . By local distinguishability [see Eq. (5)], this implies  $\mathcal{E}\mathcal{D} = \mathcal{I}_C$ . ■

*Lemma 3*. The image of  $\text{St}_1(C)$  under the decoding operation  $\mathcal{D}$  is  $F_\rho$ .

*Proof*. Since the compression is maximally efficient, for all  $\tau \in \text{St}_1(C)$  there exists  $\sigma \in F_\rho$  such that  $\tau = \mathcal{E}\sigma$ . Then,  $\mathcal{D}\tau = \mathcal{D}\mathcal{E}\sigma = \sigma$ . This implies that  $\mathcal{D}[\text{St}_1(C)] \subseteq F_\rho$ . On the other hand, since the compression is lossless, for every state  $\sigma \in F_\rho$  one has  $\mathcal{D}\mathcal{E}\sigma = \sigma$ . This implies the inclusion  $F_\rho \subseteq \mathcal{D}[\text{St}_1(C)]$ . ■

*Lemma 4*. If the state  $\varphi \in F_\rho$  is pure, then the state  $\mathcal{E}\varphi \in \text{St}_1(C)$  is pure. If the state  $\psi \in \text{St}_1(C)$  is pure, then the state  $\mathcal{D}\psi \in F_\rho$  is pure.

*Proof*. Suppose that  $\varphi \in F_\rho$  is pure and that  $\mathcal{E}\varphi$  can be written as  $\mathcal{E}\varphi = p\sigma + (1-p)\tau$  for some  $p > 0$  and some  $\sigma, \tau \in \text{St}_1(C)$ . Applying  $\mathcal{D}$  on both sides we obtain  $\varphi = p\mathcal{D}\sigma + (1-p)\mathcal{D}\tau$ . Since  $\varphi$  is pure we must have  $\mathcal{D}\sigma = \mathcal{D}\tau = \varphi$ . Now, applying  $\mathcal{E}$  on all terms of the equality and using lemma 2 we obtain  $\sigma = \tau = \mathcal{E}\varphi$ . This proves that  $\mathcal{E}\varphi$  is pure. Conversely, suppose that  $\psi \in \text{St}_1(C)$  is pure and  $\mathcal{D}\psi = p\sigma + (1-p)\tau$  for some  $p > 0$  and some  $\sigma, \tau \in \text{St}_1(A)$ . Since  $\mathcal{D}\psi$  is in the face  $F_\rho$  (lemma 3), also  $\sigma$  and  $\tau$  are in the same face. Applying  $\mathcal{E}$  on both sides of the equality  $\mathcal{D}\psi = p\sigma + (1-p)\tau$  and using lemma 2 we obtain  $\psi = \mathcal{E}\mathcal{D}\psi = p\mathcal{E}\sigma + (1-p)\mathcal{E}\tau$ . Since  $\psi$  is pure we must

have  $\mathcal{E}\sigma = \mathcal{E}\tau = \psi$ . Applying  $\mathcal{D}$  on all terms of the equality we then have  $\sigma = \tau = \mathcal{D}\psi$ , thus proving that  $\mathcal{D}\psi$  is pure. ■

We say that a state  $\sigma \in F_\rho$  is *completely mixed relative to the face*  $F_\rho$  if every state  $\tau \in F_\rho$  can stay in the convex decomposition of  $\sigma$ . In other words,  $\sigma$  is completely mixed relative to  $F_\rho$  if one has  $F_\sigma = F_\rho$ . Note that in general  $\sigma \in F_\rho$  implies  $F_\sigma \subseteq F_\rho$ .

We then have the following.

*Lemma 5.* If the state  $\omega \in F_\rho$  is completely mixed relative to  $F_\rho$ , then the state  $\mathcal{E}\omega \in \text{St}_1(\mathbf{C})$  is completely mixed. If the state  $\nu \in \text{St}_1(\mathbf{C})$  is completely mixed, then the state  $\mathcal{D}\nu \in F_\rho$  is completely mixed relative to  $F_\rho$ .

*Proof.* Suppose that  $\omega$  is completely mixed relative to  $F_\rho$ . Then every state  $\sigma \in F_\rho$  can stay in its convex decomposition, say  $\omega = p\sigma + (1-p)\sigma'$  with  $p > 0$  and  $\sigma' \in F_\rho$ . Applying  $\mathcal{E}$  we have

$$\mathcal{E}\omega = p\mathcal{E}\sigma + (1-p)\mathcal{E}\sigma'. \quad (6)$$

Since the compression is maximally efficient, for every state  $\tau \in \text{St}_1(\mathbf{C})$  there exists a state  $\sigma \in F_\rho$  such that  $\tau = \mathcal{E}\sigma$ . Choosing the suitable  $\sigma \in F_\rho$  and substituting  $\tau$  to  $\mathcal{E}\sigma$  in Eq. (6) we then obtain that for every state  $\tau \in \text{St}_1(\mathbf{C})$  there exists probability  $p > 0$  and a state  $\sigma' \in F_\rho$  such that

$$\mathcal{E}\omega = p\tau + (1-p)\mathcal{E}\sigma'.$$

This implies that  $\mathcal{E}\omega$  is completely mixed. Suppose now that  $\nu \in \text{St}_1(\mathbf{C})$  is completely mixed. Then every state  $\tau \in \text{St}_1(\mathbf{C})$  can stay in its convex decomposition, say  $\nu = p\tau + (1-p)\tau'$  with  $p > 0$  and  $\tau' \in \text{St}_1(\mathbf{C})$ . Applying  $\mathcal{D}$  on both sides we have

$$\mathcal{D}\nu = p\mathcal{D}\tau + (1-p)\mathcal{D}\tau'. \quad (7)$$

Now, using lemma 3 we have that every state  $\sigma \in F_\rho$  can be written as  $\sigma = \mathcal{D}\tau$  for some  $\tau \in \text{St}_1(\mathbf{C})$ . Choosing the suitable  $\tau \in \text{St}_1(\mathbf{C})$  and substituting  $\sigma$  to  $\mathcal{D}\tau$  in Eq. (7) we then obtain that for every state  $\sigma \in F_\rho$  there exists a probability  $p > 0$  and a state  $\tau' \in \text{St}_1(\mathbf{C})$  such that  $\mathcal{D}\nu = p\sigma + (1-p)\mathcal{D}\tau'$ . Therefore,  $\mathcal{D}\nu$  is completely mixed relative to  $F_\rho$ . ■

We now show that the system  $\mathbf{C}$  used for ideal compression of the state  $\rho$  is unique up to operational equivalence.

*Lemma 6.* If two systems  $\mathbf{C}$  and  $\mathbf{C}'$  allow for ideal compression of a state  $\rho \in \text{St}_1(\mathbf{A})$ , then  $\mathbf{C}$  and  $\mathbf{C}'$  are operationally equivalent.

*Proof.* Let  $\mathcal{E}, \mathcal{D}$  and  $\mathcal{E}', \mathcal{D}'$  denote the encoding and decoding schemes for systems  $\mathbf{C}$  and  $\mathbf{C}'$ , respectively. Define the transformations  $\mathcal{U} := \mathcal{E}'\mathcal{D} \in \text{Transf}(\mathbf{C}, \mathbf{C}')$  and  $\mathcal{V} = \mathcal{E}\mathcal{D}' \in \text{Transf}(\mathbf{C}', \mathbf{C})$ . It is easy to see that  $\mathcal{U}$  is reversible and  $\mathcal{U}^{-1} = \mathcal{V}$ . Indeed, since the restriction of  $\mathcal{D}'\mathcal{E}'$  and  $\mathcal{D}\mathcal{E}$  to the face  $F_\rho$  is the identity, using lemma 3 one has  $\mathcal{D}'\mathcal{E}'\mathcal{D} = \mathcal{D}$  and similarly  $\mathcal{D}\mathcal{E}\mathcal{D}' = \mathcal{D}'$ . Hence we have  $\mathcal{U}\mathcal{V} = \mathcal{E}'\mathcal{D}\mathcal{E}\mathcal{D}' = \mathcal{E}'\mathcal{D}' = \mathcal{I}_{\mathbf{C}'}$ , and  $\mathcal{V}\mathcal{U} = \mathcal{E}\mathcal{D}'\mathcal{E}'\mathcal{D} = \mathcal{E}\mathcal{D} = \mathcal{I}_{\mathbf{C}}$ . ■

It is useful to introduce the notion of *equality upon input of  $\rho$* . We say that two transformations  $\mathcal{A}, \mathcal{A}' \in \text{Transf}(\mathbf{A}, \mathbf{B})$  are *equal upon input of  $\rho \in \text{St}(\mathbf{A})$*  if their restrictions to the face identified by  $\rho$  are equal, that is, if  $\mathcal{A}\sigma = \mathcal{A}'\sigma$  for every  $\sigma \in F_\rho$ . If  $\mathcal{A}$  and  $\mathcal{A}'$  are equal upon input of  $\rho$  we write  $\mathcal{A} =_\rho \mathcal{A}'$ .

Using the notion of equality upon input of  $\rho$  we can rephrase the fact that the compression is lossless for  $\rho$  as  $\mathcal{D}\mathcal{E} =_\rho \mathcal{I}_{\mathbf{A}}$ . Similarly, we can state the following.

*Lemma 7.* The encoding  $\mathcal{E}$  is deterministic upon input of  $\rho$ , that is  $(e_{\mathbf{C}}|\mathcal{E} =_\rho (e_{\mathbf{A}}|$ .

*Proof.* For every  $\sigma \in F_\rho$  we have  $(e_{\mathbf{C}}|\mathcal{E}|\sigma) \geq (e_{\mathbf{A}}|\mathcal{E}|\sigma) = (e_{\mathbf{A}}|\sigma) = 1$ , having used Eq. (4) and the fact that the compression is lossless. Since probabilities are bounded by 1, this implies  $(e_{\mathbf{C}}|\mathcal{E}|\sigma) = (e_{\mathbf{A}}|\sigma)$  for every  $\sigma \in F_\rho$ , that is,  $(e_{\mathbf{C}}|\mathcal{E} =_\rho (e_{\mathbf{A}}|$ . ■

A similar result holds for the decoding.

*Lemma 8.* The decoding  $\mathcal{D}$  is deterministic, that is  $(e_{\mathbf{A}}|\mathcal{D} = (e_{\mathbf{C}}|$ .

*Proof.* For every  $\tau \in \text{St}_1(\mathbf{A})$  we have  $(e_{\mathbf{A}}|\mathcal{D}|\tau) \geq (e_{\mathbf{C}}|\mathcal{E}\mathcal{D}|\tau) = (e_{\mathbf{C}}|\tau)$ , having used Eq. (4) and lemma 2. Hence  $(e_{\mathbf{A}}|\mathcal{D} = (e_{\mathbf{C}}|$ . ■

## B. Results about purification

The purification postulate 1 implies a large number of quantum features, as it was shown in Ref. [22]. Here we review only the facts that are useful for our derivation, referring to Ref. [22] for the proofs.

An elementary consequence of the uniqueness of purification is that the group  $\mathbf{G}_{\mathbf{A}}$  of reversible transformations on  $\mathbf{A}$  acts transitively on the set of pure states.

*Lemma 9 (Transitivity on pure states).* For every couple of pure states  $\varphi, \varphi' \in \text{St}_1(\mathbf{A})$  there is a reversible transformation  $\mathcal{U} \in \mathbf{G}_{\mathbf{A}}$  such that  $\varphi' = \mathcal{U}\varphi$ .

*Proof.* See lemma 20 of Ref. [22]. ■

Transitivity implies that for every system  $\mathbf{A}$  there is a unique state  $\chi_{\mathbf{A}} \in \text{St}_1(\mathbf{A})$  that is invariant under reversible transformations, that is, a unique state such that  $\mathcal{U}\chi_{\mathbf{A}} = \chi_{\mathbf{A}}$  for every  $\mathcal{U} \in \mathbf{G}_{\mathbf{A}}$ .

*Lemma 10 (Uniqueness of the invariant state).* For every system  $\mathbf{A}$ , there is a unique state  $\chi_{\mathbf{A}}$  invariant under all reversible transformations in  $\mathbf{G}_{\mathbf{A}}$ . The invariant state has the following properties:

- (1)  $\chi_{\mathbf{A}}$  is completely mixed
- (2)  $\chi_{\mathbf{AB}} = \chi_{\mathbf{A}} \otimes \chi_{\mathbf{B}}$ .

*Proof.* See corollary 34 and theorem 4 of Ref. [22]. The proof of item 2 uses the local distinguishability axiom. ■

When there is no ambiguity we will drop the subindex  $\mathbf{A}$  and simply write  $\chi$ .

The uniqueness of purification in postulate 1 requires that if  $\Psi_\rho, \Psi'_\rho \in \text{St}_1(\mathbf{AB})$  are two purifications of  $\rho \in \text{St}_1(\mathbf{A})$ , then there exists a reversible transformation  $\mathcal{U} \in \mathbf{G}_{\mathbf{B}}$  such that  $\Psi'_\rho = (\mathcal{I}_{\mathbf{A}} \otimes \mathcal{U})\Psi_\rho$ . The following lemma extends the uniqueness property to purifications with different purifying systems.

*Lemma 11 (Uniqueness of the purification up to channels on the purifying systems).* Let  $\Psi \in \text{St}_1(\mathbf{AB})$  and  $\Psi' \in \text{St}_1(\mathbf{AC})$  be two purifications of  $\rho \in \text{St}_1(\mathbf{A})$ . Then there exists a channel  $\mathcal{C} \in \text{Transf}(\mathbf{B}, \mathbf{C})$  such that

*Proof.* See lemma 21 of Ref. [22]. ■

Another consequence of the uniqueness of purification is the fact that any ensemble decomposition of a given mixed state can be obtained by performing a measurement on the purifying system.

*Lemma 12 (Purification of preparation-tests).* Let  $\rho \in \text{St}_1(A)$  be a state and  $\Psi_\rho \in \text{St}_1(AB)$  be a purification of  $\rho$ . If  $\{\rho_i\}_{i \in X}$  be a preparation test such that  $\sum_{i \in X} \rho_i = \rho$ , then there exists an observation-test  $\{a_i\}_{i \in X}$  on the purifying system such that

$$\rho_i \text{---} A = \left( \Psi_\rho \begin{array}{c} A \\ B \end{array} \right) b_i .$$

*Proof.* See lemma 8 of Ref. [22]. ■

An easy consequence is the following.

*Corollary 1.* If  $\Psi_\rho \in \text{St}_1(AB)$  is a purification of  $\rho \in \text{St}_1(A)$  and  $\sigma$  belongs to the face  $F_\rho$ , then there exists an effect  $b$  and a nonzero probability  $p > 0$  such that

$$p \sigma \text{---} A = \left( \Psi_\rho \begin{array}{c} A \\ B \end{array} \right) b .$$

An important consequence of purification and local distinguishability is the relation between equality upon input of  $\rho$  and equality on the purifications of  $\rho$ .

*Theorem 1 (Equality upon input of  $\rho$  vs equality on purifications of  $\rho$ ).* Let  $\Psi \in \text{St}_1(AC)$  be a purification of  $\rho \in \text{St}_1(A)$ , and let  $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, B)$  be two transformations. Then one has

$$(\mathcal{A} \otimes \mathcal{I}_C) \Psi_\rho = (\mathcal{A}' \otimes \mathcal{I}_C) \Psi_\rho \iff \mathcal{A} =_\rho \mathcal{A}' .$$

*Proof.* See theorem 1 of Ref. [22]. The proof of the direction  $\iff$  uses the local distinguishability axiom. ■

As a consequence, the purification of a completely mixed state allows for the tomography of transformations:

*Corollary 2.* Let  $\omega \in \text{St}_1(A)$  be completely mixed and  $\Psi_\omega \in \text{St}_1(AC)$  is a purification of  $\omega$ . Then, for all transformations  $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, B)$  one has

$$(\mathcal{A} \otimes \mathcal{I}_C) \Psi_\omega = (\mathcal{A}' \otimes \mathcal{I}_C) \Psi_\omega \iff \mathcal{A} = \mathcal{A}' .$$

*Proof.* By theorem 1 the first condition is equivalent to  $\mathcal{A} =_\omega \mathcal{A}'$ . Since  $\omega$  is completely mixed, this means  $\mathcal{A}\sigma = \mathcal{A}'\sigma$  for every  $\sigma \in \text{St}_1(A)$ . By local distinguishability [see Eq. (5)] this implies  $\mathcal{A} = \mathcal{A}'$ . ■

Corollary shows that the state  $(\mathcal{A} \otimes \mathcal{I}_C) \Psi_\omega$  characterizes the transformation  $\mathcal{A}$  completely. We will express this fact by saying that the state  $\Psi_\omega$  is *dynamically faithful* [20], or just *faithful*, for short. Using this notion we can rephrase corollary 2.

*Corollary 3.* If  $\Psi \in \text{St}_1(AC)$  is pure and its marginal on system A is completely mixed, then  $\Psi$  is dynamically faithful for system A.

Let us choose a fixed faithful state for system A, say  $\Psi \in \text{St}_1(AC)$ . Then for every transformation  $\mathcal{C} \in \text{Transf}(A, B)$  we can define the *Choi state*  $R_\mathcal{C} \in \text{St}(BC)$  as

$$R_\mathcal{C} \begin{array}{c} B \\ C \end{array} := \left( \Psi \begin{array}{c} A \\ C \end{array} \right) \mathcal{C} \begin{array}{c} B \\ \end{array} .$$

We then have the following.

*Theorem 2 (Choi isomorphism).* For a given faithful state  $\Psi \in \text{St}_1(AC)$  the map  $\mathcal{C} \mapsto R_\mathcal{C} := (\mathcal{C} \otimes \mathcal{I}_C) \Psi$  has the following properties:

(1) It defines a bijective correspondence between tests  $\{\mathcal{C}_i\}_{i \in X}$  from A to B and collections of states  $\{R_i\}_{i \in X}$  for BC satisfying

$$\sum_{i \in X} (e|_B |R_i)_{BC} = (e|_A | \Psi)_{AC} .$$

(2) The transformation  $\mathcal{C}$  is atomic if and only if the corresponding state  $R_\mathcal{C}$  is pure.

*Proof.* See theorem 17 of Ref. [22]. ■

A simple consequence of the Choi isomorphism is the following.

*Corollary 4.* Let  $\{\mathcal{C}_i\}_{i \in X} \subset \text{Transf}(A, B)$  be a collection of transformations. Then,  $\{\mathcal{C}_i\}_{i \in X}$  is a test if and only if

$$\sum_{i \in X} (e|_B \mathcal{C}_i = (e|_A .$$

In particular, let  $\{a_i\}_{i \in X} \subset \text{Eff}(A)$  be a collection of effects. Then,  $\{a_i\}_{i \in X}$  is an observation test if and only if

$$\sum_{i \in X} (a_i| = (e| . \quad (8)$$

*Proof.* Apply item 1 of theorem 2 to the collection of states  $\{R_i\}_{i \in X}$  defined by  $R_i := (\mathcal{C}_i \otimes \mathcal{I}_C) \Psi$ . ■

A much deeper consequence of the Choi isomorphism is the following theorem.

*Theorem 3 (States specify the theory)*

Let  $\Theta, \Theta'$  be two theories satisfying the purification postulate. If  $\Theta$  and  $\Theta'$  have the same sets of normalized states, then  $\Theta' = \Theta$ .

*Proof.* See theorem 19 of Ref. [22]. ■

Thanks to theorem 3 to derive quantum theory we will only need to prove that our principles imply that for every system A the normalized states  $\text{St}_1(A)$  can be described as positive Hermitian matrices with unit trace. Once this is proved, theorem 3 automatically ensures that all the dynamics and all the measurements allowed by the theory are exactly the dynamics and the measurements allowed in quantum theory.

Note that in the definition of the Choi state we left the freedom to choose the faithful state  $\Psi \in \text{St}_1(AC)$ . Among many possibilities, one convenient choice is to take a faithful state  $\Phi \in \text{St}_1(AC)$  obtained as a purification of the invariant state  $\chi \in \text{St}_1(A)$ . Moreover, as we will see in the next paragraph, we can always choose the purifying system C in such a way that the marginal on C is completely mixed.

### C. Results about the combination of compression and purification

An important consequence of the combination of the purification postulate with the compression axiom is the fact that one can always choose a purification of  $\rho$  such that the marginal state on the purifying system is completely mixed. To prove this result we need the following lemma.

*Lemma 1.* Let  $\rho \in \text{St}_1(A)$  be a state and let  $\Psi_\rho \in \text{St}_1(AB)$  be a purification of  $\rho$ . If  $\mathcal{E} \in \text{Transf}(A, C)$  is the encoding operation in the compression scheme of axiom 3, then the state  $\Psi'_\rho := (\mathcal{E} \otimes \mathcal{I}_B) \Psi_\rho$  is pure.

*Proof.* Let  $\mathcal{D} \in \text{Transf}(C, A)$  be the decoding operation. Since the compression is lossless for  $\rho$  we know that  $\mathcal{D}\mathcal{E} =_\rho \mathcal{I}_A$ . By theorem 1 this is equivalent to the condition  $(\mathcal{D}\mathcal{E} \otimes \mathcal{I}_B) \Psi'_\rho = \Psi_\rho$ .

$\mathcal{S}_B)\Psi_\rho = \Psi_\rho$ . Now, suppose that  $(\mathcal{E} \otimes \mathcal{S}_B)\Psi_\rho = \sum_{i \in X} \Gamma_i$ . Applying  $\mathcal{D}$  on both sides we then obtain  $\Psi_\rho = \sum_{i \in X} (\mathcal{D} \otimes \mathcal{S}_B)\Gamma_i$ , and, since  $\Psi_\rho$  is pure, for every  $i \in X$  we must have  $(\mathcal{D} \otimes \mathcal{S}_B)\Gamma_i = p_i \Psi_\rho$ , where  $p_i \geq 0$  is some probability. Finally, since  $\mathcal{E}\mathcal{D} = \mathcal{I}_C$  (lemma 2), one has  $\Gamma_i = p_i(\mathcal{E} \otimes \mathcal{S}_B)\Psi_\rho$ . Hence,  $(\mathcal{E} \otimes \mathcal{S}_B)\Psi_\rho$  admits only decompositions with  $\Gamma_i = p_i(\mathcal{E} \otimes \mathcal{S}_B)\Psi_\rho$ , that is,  $(\mathcal{E} \otimes \mathcal{S}_B)\Psi_\rho$  is pure. ■

We are now in position to prove the desired result.

**Theorem 4.** For every state  $\rho \in \text{St}_1(A)$  there exists a system C and a purification  $\Psi_\rho \in \text{St}_1(AC)$  of  $\rho$  such that the marginal state on system C is completely mixed. Moreover, the system C is unique up to operational equivalence.

*Proof.* Take an arbitrary purification of  $\rho$ , say  $\Phi_\rho \in \text{St}_1(AB)$  for some purifying system B. Define the marginal state on system B as  $|\theta\rangle_B := (e|_A|\Phi_\rho)_{AB}$  and define the state  $\Psi_\rho := (\mathcal{I}_A \otimes \mathcal{E})\Phi_\rho$ , where  $\mathcal{E} \in \text{Transf}(B,C)$  the encoding operation for state  $\theta$ . By lemma 13 we know that  $\Psi_\rho \in \text{St}(AC)$  is pure. Using lemma 7 and theorem 1 we obtain  $(e_C||\Psi_\rho) = [(e_C|\mathcal{E}||\Phi_\rho) = (e_B||\Phi_\rho) = |\rho\rangle)$ , that is,  $\Psi_\rho$  is a purification of  $\rho$ . Finally, the marginal on system C is given by  $\tilde{\rho} = \mathcal{E}\theta$ , which by lemma 5 is completely mixed. This proves the first part of the thesis. It remains to show that the system C is uniquely defined up to operational equivalence. Suppose that  $\Psi'_\rho \in \text{St}(AC')$  is another purification of  $\rho$  with the property that the marginal on system C' is completely mixed. Since  $\Psi_\rho$  and  $\Psi'_\rho$  are two purifications of the same state, there must be two channels  $\mathcal{C} \in \text{Transf}(C,C')$  and  $\mathcal{R} \in \text{Transf}(C',C)$  such that  $\Psi'_\rho = (\mathcal{I}_A \otimes \mathcal{C})\Psi_\rho$  and  $\Psi_\rho = (\mathcal{I}_A \otimes \mathcal{R})\Psi'_\rho$  (lemma 11). Combining the two equalities one obtains  $\Psi_\rho = (\mathcal{I}_A \otimes \mathcal{R}\mathcal{C})\Psi_\rho$ . Now, the marginal of  $\Psi_\rho$  on system C is completely mixed, and this implies that  $\Psi_\rho$  is faithful for system C (corollary 3). Hence we have  $\mathcal{R}\mathcal{C} = \mathcal{I}_C$ . Repeating the same argument for  $\Psi'_\rho$  we obtain  $\mathcal{C}\mathcal{R} = \mathcal{I}_{C'}$ . Therefore,  $\mathcal{C}$  is reversible and  $\mathcal{R} = \mathcal{C}^{-1}$ . This proves that C and C' are operationally equivalent. ■

The following facts will also be useful.

**Corollary 5.** Let  $\Psi_\rho \in \text{St}_1(AB)$  be a purification of  $\rho \in \text{St}_1(A)$  and let  $\mathcal{E} \in \text{Transf}(A,C)$  be the encoding for  $\rho$ . Then, the state  $(\mathcal{E} \otimes \mathcal{S}_B)\Psi_\rho \in \text{St}_1(CB)$  is dynamically faithful for C.

*Proof.* The marginal of  $(\mathcal{E} \otimes \mathcal{S}_B)\Psi_\rho$  on system C is  $\mathcal{E}\rho$ , which is completely mixed by lemma 5. Hence,  $(\mathcal{E} \otimes \mathcal{S}_B)\Psi_\rho$  is dynamically faithful by corollary 3. ■

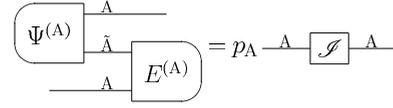
**Lemma 14.** The decoding transformation  $\mathcal{D} \in \text{Transf}(C,A)$  in the ideal compression for  $\rho \in \text{St}_1(A)$  is atomic.

*Proof.* Let  $\Psi_\rho \in \text{St}_1(AB)$  be a purification of  $\rho$ , for some purifying system B. Since  $\mathcal{D}\mathcal{E} =_\rho \mathcal{I}_A$  (the compression is lossless), we have  $(\mathcal{D}\mathcal{E} \otimes \mathcal{S}_B)|\Psi_\rho\rangle = |\Psi_\rho\rangle$  (theorem 1). Now, by corollary 5  $(\mathcal{E} \otimes \mathcal{S}_B)|\Psi_\rho\rangle$  is faithful for C and by lemma 13  $(\mathcal{E} \otimes \mathcal{S}_B)|\Psi_\rho\rangle$  is pure. Using the Choi isomorphism with the faithful state  $\Psi := (\mathcal{E} \otimes \mathcal{S}_B)\Psi_\rho$  we then obtain that  $\mathcal{D}$  is atomic. ■

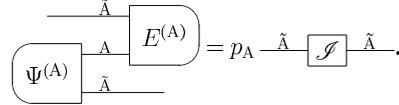
#### D. Teleportation and the link product

For every system A one can choose a completely mixed state  $\omega_A$  and a purification  $\Psi^{(A)} \in \text{St}(A\tilde{A})$  such that the marginal on system  $\tilde{A}$  is completely mixed (cf. theorem 4). Any such purification allows for a probabilistic teleportation scheme:

**Lemma 15 (Probabilistic teleportation).** There exists an atomic effect  $E^{(A)} \in \text{Eff}(\tilde{A}A)$  and a nonzero probability  $p_A$  such that



and



*Proof.* See corollary 19 of Ref. [22]. ■

Let us choose  $\Psi^{(A)}$  to be the faithful state in the definition of the Choi isomorphism. Then the sequential composition of transformation induces a composition of Choi states in following way.

**Corollary 6 (Link product).** For two transformations  $\mathcal{C} \in \text{Transf}(A,B)$  and  $\mathcal{D} \in \text{Transf}(B,C)$  the Choi state of  $\mathcal{D}\mathcal{C} \in \text{Transf}(A,C)$  is given by the link product

$$R_{\mathcal{D}\mathcal{C}} = \frac{1}{p_B} \left( R_{\mathcal{D}} \begin{array}{c} C \\ \tilde{B} \\ B \\ \tilde{A} \end{array} R_{\mathcal{C}} \right) E^{(B)} \quad (9)$$

*Proof.* See corollary 22 of Ref. [22]. ■

We conclude this paragraph with an important result that follows from the combination of the link product structure with the pure conditioning axiom.

**Lemma 16 (Atomicity of composition).** The composition of two atomic transformations is atomic.

*Proof.* Let  $\mathcal{C} \in \text{Transf}(A,B)$  and  $\mathcal{D} \in \text{Transf}(B,C)$  be two atomic transformations. By the Choi isomorphism, the (unnormalized) states  $R_{\mathcal{C}}$  and  $R_{\mathcal{D}}$  are pure. Since the teleportation effect  $E^{(B)}$  in Eq. (9) is atomic (lemma 15), the pure conditioning axiom 5 implies the state  $R_{\mathcal{D}\mathcal{C}}$  is pure. By the Choi isomorphism this means that  $\mathcal{D}\mathcal{C}$  is atomic. ■

#### E. No information without disturbance

We say that a test  $\{\mathcal{C}_i\}_{i \in X} \subset \text{Transf}(A)$  is *nondisturbing upon input of  $\rho$*  if  $\sum_{i \in X} \mathcal{C}_i =_\rho \mathcal{I}_A$ . If  $\rho$  is completely mixed, we simply say that the test is *nondisturbing*.

A consequence of the purification postulate is the following “no-information without disturbance” result.

**Lemma 17 (No information without disturbance).** A test  $\{\mathcal{C}_i\}_{i \in X} \subset \text{Transf}(A)$  is nondisturbing upon input of  $\rho$  if and only if there is a set of probabilities  $\{p_i\}_{i \in X}$  such that  $\mathcal{C}_i =_\rho p_i \mathcal{I}_A$  for every  $i \in X$ .

*Proof.* See theorem 10 of Ref. [22]. ■

The no-information without disturbance result implies the following geometrical limitation.

**Corollary 7.** For every system A the convex set of states  $\text{St}_1(A)$  is not a segment.

*Proof.* The proof is by contradiction. Suppose that for some system A the set  $\text{St}_1(A)$  is a segment. The segment

has only two pure states, say  $\varphi_1$  and  $\varphi_2$ , and every other state  $\rho \in \text{St}_1(A)$  is completely mixed. Then the distinguishability axiom 2 imposes that  $\varphi_1$  and  $\varphi_2$  are perfectly distinguishable. Take the binary test  $\{a_1, a_2\}$  such that  $(a_i|\varphi_j) = \delta_{ij}$  and define the “measure-and-prepare” test  $\{\mathcal{C}_1, \mathcal{C}_2\}$  as  $\mathcal{C}_i = |\varphi_i\rangle\langle a_i|$ ,  $i = 1, 2$  (the possibility of preparing a state depending on the outcome of a previous measurement is guaranteed by causality [22]). Since every state  $\rho$  in the segment can be written as convex combination of the two extreme points, we have that the test  $\{\mathcal{C}_1, \mathcal{C}_2\}$  is nondisturbing:  $(\mathcal{C}_1 + \mathcal{C}_2)\rho = \rho$  for every  $\rho$ . This is in contradiction with lemma 17 because  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are not proportional to the identity. ■

We know that no information can be extracted without disturbance. In the following we will prove a result in the converse direction: if a measurement extracts no information, than it can be realized in a nondisturbing fashion. To show this result we first need the following.

*Lemma 18.* For every observation test  $\{a_i\}_{i \in X} \subset \text{Eff}(A)$  with finite outcome set  $X$  there is a system  $C$  and a test  $\{\mathcal{A}_i\}_{i \in X} \subset \text{Transf}(A, C)$  consisting of atomic transformations such that  $(a_i| = (e_C|\mathcal{A}_i$ .

*Proof.* Let  $|\Psi\rangle_{AB}$  be a pure faithful state for system  $A$  and let  $|R_i\rangle_B = (a_i|_A|\Psi\rangle_{AB}$  the Choi state of  $a_i$ . Take a purification of  $R_i$ , say  $|\Psi_i\rangle_{BC}$  for some purifying system  $C$  [38]. Then, by the Choi isomorphism there is a test  $\{\mathcal{A}_i\}_{i \in X}$ , with input  $A$  and output  $C$ , such that

$$\begin{array}{c} \text{C} \\ \text{---} \\ \Psi_i \\ \text{---} \\ \text{B} \end{array} = \begin{array}{c} \text{A} \quad \text{C} \\ \text{---} \quad \text{---} \\ \Psi \quad \mathcal{A}_i \\ \text{---} \quad \text{---} \\ \text{B} \end{array}$$

(see item 1 of theorem 2). Moreover, each transformation  $\mathcal{A}_i : A \rightarrow C$  is atomic (item 2 of theorem 2). Applying the deterministic effect  $(e_C|$  on both sides we then obtain  $|R_i\rangle_B = (e_C|\Psi_i)_{CA} = (e_C|\mathcal{A}_i|\Psi)_{AB}$ . By definition of  $R_i$ , this implies  $(a_i|_A|\Psi)_{AB} = (e_C|\mathcal{A}_i|\Psi)_{AB}$ , and, since  $\Psi$  is dynamically faithful,  $(a_i|_A = (e_C|\mathcal{A}_i$ . ■

*Theorem 5.* Let  $\rho \in \text{St}_1(A)$  be a state,  $a \in \text{Eff}(A)$  be an effect, and  $\mathcal{A} \in \text{Transf}(A, B)$  be an atomic transformation such that  $(a|_A = (e|_B\mathcal{A}$ . If  $(a| = p(e|$  for some  $p \geq 0$ , then there exists a channel  $\mathcal{C} \in \text{Transf}(B, A)$  such that  $\mathcal{C}\mathcal{A} = p\mathcal{I}_A$ .

*Proof.* Consider a purification of  $\rho$ , say  $\Psi_\rho \in \text{St}(AC)$ , and define the state  $\Sigma \in \text{St}_1(BC)$  by  $|\Sigma\rangle := \frac{1}{p}(\mathcal{A} \otimes \mathcal{I}_C)|\Psi_\rho\rangle$ . By the atomicity of composition 16 the state  $\Sigma$  is pure. Moreover, we have

$$\begin{aligned} (e_B|\Sigma)_{BC} &= \frac{1}{p}(a|_A|\Psi_\rho)_{AB} \\ &= (e_A|\Psi_\rho)_{AC}, \end{aligned}$$

having used theorem 1 in the last equality. This implies that  $\Psi_\rho$  and  $\Sigma$  are different purifications of the same mixed state on system  $C$ . Then, by lemma 11 there exists a channel  $\mathcal{C} \in \text{Transf}(B, A)$  such that  $|\Psi_\rho\rangle = (\mathcal{C} \otimes \mathcal{I}_C)|\Sigma\rangle = \frac{1}{p}(\mathcal{C}\mathcal{A} \otimes \mathcal{I}_C)|\Psi_\rho\rangle$ . By theorem 1, the last equality implies  $\mathcal{C}\mathcal{A} = p\mathcal{I}_A$ . ■

We now make a simple observation that combined with theorem 5 will lead to some interesting consequences.

*Lemma 19.* If  $(a|\rho) = \|a\|$ , then  $a = p\|a\|e$ . Similarly, if  $(a|\rho) = 0$ , then  $a = p\|a\|0$ .

*Proof.* By definition,  $\sigma \in F_\rho$  iff there exists  $p > 0$  and  $\tau \in \text{St}_1(A)$  such that  $\rho = p\sigma + (1-p)\tau$ . If  $(a|\rho) = \|a\|$ , then we have  $\|a\| = p(a|\sigma) + (1-p)(a|\tau)$ . Since  $(a|\sigma)$  and  $(a|\tau)$  cannot be larger than  $\|a\|$ , the only way to have the equality is to have  $(a|\sigma) = (a|\tau) = \|a\|$ . By definition, this amounts to say  $a = p\|a\|e$ . Similarly, if  $(a|\rho) = 0$ , one has  $0 = p(a|\sigma) + (1-p)(a|\tau)$ , which is satisfied only if  $(a|\sigma) = (a|\tau) = 0$ , that is, if  $a = p\|a\|0$ . ■

As consequence, we have the following.

*Corollary 8.* Let  $\rho \in \text{St}_1(A)$  be a state,  $a \in \text{Eff}(A)$  be an effect, and  $\mathcal{A} \in \text{Transf}(A, B)$  be an atomic transformation such that  $(a|_A = (e|_B\mathcal{A}$ . If  $(a|\rho) = 1$ , then  $\mathcal{A}$  is correctable upon input of  $\rho$ , that is, there exists a correction operation  $\mathcal{C} \in \text{Transf}(B, A)$  such that  $\mathcal{C}\mathcal{A} = \rho\mathcal{I}_A$ .

*Proof.* If  $(a|\rho) = 1$ , then clearly  $\|a\| = 1$ . Lemma 19 then implies  $(a| = \rho(e|$ . Applying theorem 5 we finally obtain the thesis. ■

*Corollary 9.* Let  $\rho \in \text{St}_1(A)$  be a state,  $a \in \text{Eff}(A)$  be an effect such that  $(a|\rho) = 1$ . Then there exists a transformation  $\mathcal{C} \in \text{Transf}(A)$  such that  $(a| = (e|\mathcal{C}$  and  $\mathcal{C} = \rho\mathcal{I}$ .

*Proof.* Straightforward consequence of lemma 18 and of corollary 8. ■

Finally, we say that an observation-test  $\{a_i\}_{i \in X}$  is *noninformative upon input of  $\rho$*  if we have  $(a_i| = p_i(e|$  for every  $i \in X$ . This means that the test  $\{a_i\}_{i \in X}$  is unable to distinguish the states in the face  $F_\rho$ . As a consequence of theorem 5 we have the following “no disturbance without information” result.

*Corollary 10 (No disturbance without information).* If the test  $\{a_i\}_{i \in X}$  is noninformative upon input of  $\rho$  then there is a test  $\{\mathcal{D}_i\}_{i \in X} \subset \text{Transf}(A)$  that is nondisturbing upon input of  $\rho$  and satisfies  $(e|\mathcal{D}_i = (a_i|$  for every  $i \in X$ .

*Proof.* By lemma 18 there exists a test  $\{\mathcal{A}_i\} \subset \text{Transf}(A, B)$  such that each transformation  $\mathcal{A}_i$  is atomic and  $(e|\mathcal{A}_i = (a_i|$ . By theorem 5, for each  $\mathcal{A}_i$  there is a correction channel  $\mathcal{C}_i$  such that  $\mathcal{C}_i\mathcal{A}_i = p_i\mathcal{I}_A$ . Defining  $\mathcal{D}_i := \mathcal{C}_i\mathcal{A}_i$  we then obtain the thesis. ■

## V. PERFECTLY DISTINGUISHABLE STATES

In this section we prove some basic facts about perfectly distinguishable states. Let us start from the definition.

*Definition 5 (Perfectly distinguishable states).* The normalized states  $\{\rho_i\}_{i=1}^N \subset \text{St}_1(A)$  are perfectly distinguishable if there exists an observation-test  $\{a_i\}_{i=1}^N$  such that  $(a_j|\rho_i) = \delta_{ij}$ . The observation-test  $\{a_i\}_{i=1}^N$  is called perfectly distinguishing.

From the distinguishability axiom 2 it is clear that every nontrivial system has at least two perfectly distinguishable states.

*Lemma 20.* For every nontrivial system  $A$  there are at least two perfectly distinguishable states.

*Proof.* Let  $\varphi$  be a pure state of  $A$ . Obviously,  $\varphi$  is not completely mixed (unless the system  $A$  has only one state, that is, unless  $A$  is trivial). Hence, by axiom 2 there exists at least a state  $\sigma$  that is perfectly distinguishable from  $\varphi$ . ■

An equivalent condition for perfect distinguishability is the following.

*Lemma 21.* The states  $\{\rho_i\}_{i=1}^N \subset \text{St}_1(A)$  are perfectly distinguishable if and only if there exists an observation-test  $\{a_i\}_{i=1}^N$  such that  $(a_i|\rho_i) = 1$  for every  $i$ .

*Proof.* The condition  $(a_i|\rho_i) = 1, \forall i = 1, \dots, N$  is clearly necessary. On the other hand, the condition  $(a_i|\rho_i) = 1, \forall i = 1, \dots, N$  implies

$$(a_i|\rho_i) = 1 = \sum_{j=1}^N (a_j|\rho_i) = (a_i|\rho_i) + \sum_{i \neq j} (a_j|\rho_i).$$

Since all probabilities are nonnegative, we must have  $(a_j|\rho_i) = 0$  for  $i \neq j$ , and therefore,  $(a_j|\rho_i) = \delta_{ij}$ . ■

A very general fact about state discrimination is expressed by the following.

*Lemma 22.* If  $\rho$  is perfectly distinguishable from  $\sigma$  and  $\rho'$  (or  $\sigma'$ ) belongs to the face identified by  $\rho$  (or  $\sigma$ ), then  $\rho'$  is perfectly distinguishable from  $\sigma'$ .

*Proof.* Let  $\{a, e - a\}$  be the binary observation test that distinguishes perfectly between  $\rho$  and  $\sigma$ . By definition,  $a \in \text{Eff}(A)$  is such that  $(a|\rho) = 1$  and  $(a|\sigma) = 0$ . Now, by lemma 19,  $(a|\rho') = 1$  and  $(a|\sigma') = 0$  for all  $\rho' \in F_\rho$  and  $\sigma' \in F_\sigma$ . ■

Thanks to purification and to the local distinguishability axiom 4, we are also in position to show a much stronger result.

*Lemma 23.* Let  $\{\rho_i\}_{i=1}^N \subset F_\rho$  and  $\{\rho_j\}_{j=N+1}^{N+M} \subset F_\sigma$  be two sets of perfectly distinguishable states.

If  $\rho$  is perfectly distinguishable from  $\sigma$ , then the states  $\{\rho_i\}_{i=1}^{N+M}$  are perfectly distinguishable.

*Proof.* Let  $\{a, e_A - a\}$  be the observation test such that  $(a|\rho) = 1$  and  $(a|\sigma) = 0$ . Now, by corollary 9 there is a transformation  $\mathcal{C} \in \text{Transf}(A)$  such that  $(e_A|\mathcal{C} = |a|$  and  $\mathcal{C} =_\rho \mathcal{I}_A$ . Similarly, there exists a transformation  $\mathcal{C}' \in \text{Transf}(A)$  such that  $(e_A|\mathcal{C}' = |e_A| - |a|$  and  $\mathcal{C}' =_\sigma \mathcal{I}_A$ . We can then define the following observation test:

$$(c_i| = \begin{cases} (a_i|\mathcal{C} & i \leq N \\ (b_i|\mathcal{C}' & N + 1 \leq i \leq N + M \end{cases}$$

where  $\{a_i\}_{i=1}^N$  (or  $\{b_j\}_{j=N+1}^{N+M}$ ) is the observation test that perfectly distinguishes among the states  $\{\rho_i\}_{i=1}^N$  (or  $\{\rho_j\}_{j=N+1}^{N+M}$ ). By corollary 4 [see in particular Eq. (8)],  $\{c_i\}_{i=1}^{N+M}$  is indeed an observation test: each  $c_i$  is an effect and one has the normalization

$$\begin{aligned} \sum_{i=1}^{N+M} (c_i| &= \sum_{i=1}^N (a_i|\mathcal{C} + \sum_{i=N+1}^{N+M} (b_i|\mathcal{C}' \\ &= (e_A|\mathcal{C} + (e_A|\mathcal{C}' \\ &= |a| + (e_A| - |a|) = (e_A|. \end{aligned}$$

Moreover, since  $\mathcal{C} =_\rho \mathcal{I}_A$  and  $\mathcal{C}' =_\sigma \mathcal{I}_A$ , one has  $(c_i|\rho_i) = 1$  for every  $i = 1, \dots, N + M$ . By lemma 21, this implies that the states  $\{\rho_i\}_{i=1}^{N+M}$  are perfectly distinguishable. ■

*Definition 6.* A set of perfectly distinguishable states  $\{\rho_i\}_{i=1}^N$  is *maximal* if there is no state  $\rho_{N+1} \in \text{St}_1(A)$  such that the states  $\{\rho_i\}_{i=1}^{N+1}$  are perfectly distinguishable.

*Theorem 6.* A set of perfectly distinguishable states  $\{\rho_i\}_{i=1}^N$  is maximal if and only if the state  $\omega = \sum_{i=1}^N \rho_i / N$  is completely mixed.

*Proof.* We first prove that if  $\omega$  is completely mixed, then the set  $\{\rho_i\}_{i=1}^N$  must be maximal. Indeed, if there existed a state  $\rho_{N+1}$  such that  $\{\rho_i\}_{i=1}^{N+1}$  are perfectly distinguishable, then clearly  $\rho_{N+1}$  would be distinguishable from

$\omega$ . This is absurd because by proposition 1 no state can be perfectly distinguished from a completely mixed state. Conversely, if  $\{\rho_i\}_{i=1}^N$  is maximal, then  $\omega$  is completely mixed. If it were not, by the distinguishability axiom 2,  $\omega$  would be perfectly distinguishable from some state  $\rho_{N+1}$ . By lemma 23, this would imply that the states  $\{\rho_i\}_{i=1}^{N+1}$  are perfectly distinguishable, in contradiction with the hypothesis that the set  $\{\rho_i\}_{i=1}^N$  is maximal. ■

*Lemma 24.* Every set of perfectly distinguishable pure states can be extended to a maximal set of perfectly distinguishable pure states.

*Proof.* Let  $\{\varphi_i\}_{i=1}^N$  be a nonmaximal set of perfectly distinguishable pure states. By definition, there exists a state  $\sigma$  such that  $\{\varphi_i\}_{i=1}^N \cup \{\sigma\}$  is perfectly distinguishable. Let  $\varphi_{N+1}$  be a pure state in  $F_\sigma$ . By lemma 19 the states  $\{\varphi_i\}_{i=1}^{N+1}$  will be perfectly distinguishable. Since the dimension of  $\text{St}_{\mathbb{R}}(A)$  is finite and distinguishable states are linearly independent, iterating this procedure one finally obtains a maximal set of pure states in a finite number of steps. ■

*Corollary 11.* Any pure state belongs to a maximal set of perfectly distinguishable pure states.

We conclude this section with a few elementary facts about how the ideal compression of axiom 3 preserves the distinguishability properties. In the following we will choose a state  $\rho \in \text{St}_1(A)$  and  $\mathcal{E} \in \text{Transf}(A, C)$  [or  $\mathcal{D} \in \text{Transf}(C, A)$ ] will be the encoding (or decoding) in the ideal compression scheme for  $\rho$ .

*Lemma 25.* If the states  $\{\rho_i\}_{i=1}^k \subset F_\rho$  are perfectly distinguishable, then the states  $\{\mathcal{E}\rho_i\}_{i=1}^k \subset \text{St}_1(C)$  are perfectly distinguishable. Conversely, if the states  $\{\sigma_i\}_{i=1}^k \subset \text{St}_1(C)$  are perfectly distinguishable, then the states  $\{\mathcal{D}\sigma_i\}_{i=1}^k \subset F_\rho$  are perfectly distinguishable.

*Proof.* Let  $\{a_i\}_{i=1}^k$  be the observation test such that  $(a_i|\rho_i) = 1$  for every  $i = 1, \dots, k$ . Since the compression is lossless, we have  $\mathcal{D}\mathcal{E}|\rho_i) = |\rho_i)$  and  $(a_i|\mathcal{D}\mathcal{E}|\rho_i) = 1$ . Now, consider the test  $\{c_i\}_{i=1}^k$  defined by  $(c_i| = (a_i|\mathcal{D}$ . Clearly we have  $(c_i|\mathcal{E}|\rho_i) = 1$  for every  $i = 1, \dots, k$ . By lemma 21 this means that the states  $\{\mathcal{E}\rho_i\}_{i=1}^k$  are perfectly distinguishable. Similarly, let  $\{b_i\}_{i=1}^k$  be the observation test that distinguishes the set  $\{\sigma_i\}_{i=1}^k$ . Since  $\mathcal{E}\mathcal{D} = \mathcal{I}_C$  (lemma 2), we can conclude by the same argument that the states  $\{\mathcal{D}\sigma_i\}_{i=1}^k$  are perfectly distinguishable. ■

We say that a set of perfectly distinguishable states  $\{\rho_i\}_{i=1}^k \subset F_\rho$  is *maximal in the face*  $F_\rho$  if there is no state  $\rho_{k+1} \in F_\rho$  such that the states  $\{\rho_i\}_{i=1}^{k+1}$  are perfectly distinguishable. We then have the following.

*Corollary 12.* If  $\{\rho_i\}_{i=1}^k \subset F_\rho$  is a maximal set of perfectly distinguishable states in the face  $F_\rho$ , then  $\{\mathcal{E}\rho_i\}_{i=1}^k \in \text{St}_1(C)$  is a maximal set of perfectly distinguishable states. Conversely, if  $\{\sigma_i\}_{i=1}^k \in \text{St}_1(C)$  is a maximal set of perfectly distinguishable states, then  $\{\mathcal{D}\sigma_i\}_{i=1}^k$  is a maximal set of perfectly distinguishable states in the face  $F_\rho$ .

*Proof.* Distinguishability of the states  $\{\mathcal{E}\rho_i\}_{i=1}^k$  and  $\{\mathcal{D}\sigma_i\}_{i=1}^k$  is proved by lemma 25. Let us now prove maximality. By contradiction, suppose that the set  $\{\rho_i\}_{i=1}^k$  is maximal in the face  $F_\rho$  while the set  $\{\sigma_i\}_{i=1}^k, \sigma_i := \mathcal{E}\rho_i$  is not maximal. This means that there exists a state  $\sigma_{k+1} \in \text{St}_1(C)$  such that the states  $\{\sigma_i\}_{i=1}^{k+1}$  are perfectly distinguishable. By lemma 25 the states  $\{\mathcal{D}\sigma_i\}_{i=1}^{k+1}$  are perfectly distinguishable.





*Lemma 35.* Let  $\{\rho_i\}_{i=1}^N \subset \text{St}_1(A)$  be a set of states. If there exists a set of effects  $\{b_i\}_{i=1}^N \subset \text{Eff}(A)$  (not necessarily an observation test) such that  $(b_i|\rho_j) = \delta_{ij}$ , then the states  $\{\rho_i\}_{i=1}^N$  are perfectly distinguishable.

*Proof.* For each  $i = 1, \dots, N$  consider the binary test  $\{b_i, e - b_i\}$ . Since by hypothesis  $(b_i|\rho_j) = \delta_{ij}$ , the test  $\{b_i, e - b_i\}$  can perfectly distinguish  $\rho_i$  from any mixture of the states  $\{\rho_j\}_{j \neq i}$ . In particular, this means that, for every  $M < N$ ,  $\rho_{M+1}$  can be perfectly distinguished from the mixture  $\omega_M = \sum_{j=1}^M \rho_j / M$ . Note that, by definition, the states  $\{\rho_i\}_{i=1}^M$  belong to the face  $F_{\omega_M}$ . We now prove by induction on  $M$  that the states  $\{\rho_i\}_{i=1}^M$  are perfectly distinguishable. This is true for  $M = 1$ . Now, suppose that the states  $\{\rho_i\}_{i=1}^M$  are perfectly distinguishable. Since the state  $\rho_{M+1}$  is perfectly distinguishable from  $\omega_M$ , by lemma 23 we have that the states  $\{\rho_i\}_{i=1}^{M+1}$  are perfectly distinguishable. Taking  $M = N - 1$  the thesis follows. ■

We now show that the invariant state  $\chi$  is a mixture of perfectly distinguishable pure states.

*Theorem 9.* For every maximal set of perfectly distinguishable pure states  $\{\varphi_i\}_{i=1}^{d_A} \subset \text{St}_1(A)$  one has

$$\chi = \frac{1}{d_A} \sum_{i=1}^{d_A} \varphi_i.$$

*Proof.* Let  $\{a_i\}_{i=1}^{d_A}$  be the observation test such that  $(a_i|\varphi_j) = \delta_{ij}$ , and  $\Phi \in \text{St}_1(AB)$  be a purification of  $\chi$ , chosen in such a way that the marginal on system B is completely mixed (theorem 4). Let  $\{\psi_i\}_{i=1}^{d_A} \subset \text{St}_1(B)$  be the pure states defined by

$$\begin{array}{c} \text{A} \\ \text{---} \\ \Phi \\ \text{---} \\ \text{B} \end{array} \begin{array}{c} a_i \\ \text{---} \\ \text{---} \\ \text{---} \end{array} = \frac{1}{d_A} \begin{array}{c} \psi_i \\ \text{---} \\ \text{---} \\ \text{---} \end{array} \text{---} \text{B}$$

and, for each  $i$ , let  $b_i$  be the atomic effect such that

$$\begin{array}{c} \text{A} \\ \text{---} \\ \Phi \\ \text{---} \\ \text{B} \end{array} \begin{array}{c} \text{---} \\ \text{---} \\ b_i \\ \text{---} \end{array} = \frac{1}{d_A} \begin{array}{c} \varphi_i \\ \text{---} \\ \text{---} \\ \text{---} \end{array} \text{---} \text{A} \quad (16)$$

(here we used lemma 30 and the fact that  $p_{\max} = 1/d_A$ ). Then we have

$$\begin{array}{c} \psi_i \\ \text{---} \\ \text{---} \\ \text{---} \end{array} \text{---} \text{B} \begin{array}{c} b_j \\ \text{---} \\ \text{---} \\ \text{---} \end{array} = d_A \begin{array}{c} \text{A} \\ \text{---} \\ \Phi \\ \text{---} \\ \text{B} \end{array} \begin{array}{c} a_i \\ \text{---} \\ \text{---} \\ b_j \\ \text{---} \end{array} \quad (17) \\ = \begin{array}{c} \varphi_j \\ \text{---} \\ \text{---} \\ \text{---} \end{array} \text{---} \text{A} \begin{array}{c} a_i \\ \text{---} \\ \text{---} \\ \text{---} \end{array} = \delta_{ij}.$$

By lemma 35 this implies that the states  $\{\psi_i\}_{i=1}^{d_A}$  are perfectly distinguishable. Now, since the marginal of  $|\Phi\rangle_{AB}$  on system B is completely mixed, theorem 6 states that the set  $\{\psi_i\}_{i=1}^{d_A}$  is maximal. Let  $\{b'_i\}_{i=1}^{d_A}$  be the observation test such that  $(b'_i|\psi_j) = \delta_{ij}$ . By lemma 27 each  $b'_i$  must be atomic. On the other hand, there is a unique atomic effect  $b_i$  such that  $(b_i|\psi_i) = 1$  (theorem 8). Therefore,  $b'_i = b_i$ . This means that the effects  $\{b_i\}_{i=1}^{d_A}$  form an observation test. Once this fact has been proved, using Eq. (16) we obtain

$$\begin{aligned} |\chi\rangle_A &= (e_B|\Phi)_{AB} \\ &= \sum_i (b_i|\Phi)_{AB} \end{aligned}$$

$$= 1/d_A \sum_i |\varphi_i\rangle. \quad \blacksquare$$

As a consequence, we have the following.

*Corollary 16 (Existence of conjugate systems).* For every system A there exists a system  $\tilde{A}$ , called the *conjugate system*, and a purification  $\Phi \in \text{St}_1(A\tilde{A})$  of the invariant state  $\chi_A$  such that  $d_{\tilde{A}} = d_A$  and the marginal on  $\tilde{A}$  is the invariant state  $\chi_{\tilde{A}}$ . The conjugate system  $\tilde{A}$  is unique up to operational equivalence.

*Proof.* We first prove that  $\tilde{A}$  is unique up to operational equivalence. The defining property of the conjugate system  $\tilde{A}$  is that the marginal of  $\Phi$  on  $\tilde{A}$  is the invariant state  $\chi_{\tilde{A}}$ , which is completely mixed. Theorem 4 then implies that  $\tilde{A}$  is unique up to operational equivalence. Let us now show the existence of  $\tilde{A}$ . Take a purification of  $\chi_A$ , with purifying system  $\tilde{A}$  chosen so that the marginal of  $\Phi$  on  $\tilde{A}$  is completely mixed (this is possible thanks to theorem 4). Now, the states  $\{\psi_i\}_{i=1}^{d_A} \subset \text{St}(B)$ , defined by  $\frac{1}{d_A}|\psi_i\rangle := [(a_i|\otimes \mathcal{I}_{\tilde{A}})|\Phi\rangle]$ , are perfectly distinguishable [see Eq. (17) in the proof of theorem 9]. Hence, by theorem 6 they are a maximal set of perfectly distinguishable pure states. This implies  $d_{\tilde{A}} = d_A$ . Finally, by theorem 9 one has  $1/d_{\tilde{A}} \sum_{i=1}^{d_{\tilde{A}}} \psi_i = \chi_{\tilde{A}}$ . ■

*Corollary 17.* The distance between the invariant state  $\chi_A$  and an arbitrary pure state  $\varphi \in \text{St}_1(A)$  is

$$\|\chi - \varphi\| = \frac{2(d_A - 1)}{d_A}.$$

*Proof.* Take a maximal set of perfectly distinguishable pure states  $\{\varphi_i\}_{i=1}^{d_A}$  such that  $\varphi_1 = \varphi$  (corollary 11). Since  $\chi = \sum_{i=1}^{d_A} \varphi_i / d_A$  one has  $\chi - \varphi = \frac{(d_A-1)}{d_A}(\sigma - \varphi_1)$ , where  $\sigma = \sum_{i=2}^{d_A} \varphi_i / (d_A - 1)$ . Hence, one has  $\|\chi - \varphi\| = \frac{(d_A-1)}{d_A} \|\sigma - \varphi_1\| = \frac{2(d_A-1)}{d_A}$ , having used that  $\sigma$  and  $\varphi_1$  are perfectly distinguishable and therefore  $\|\sigma - \varphi_1\| = 2$  (see subsection II-I in Ref. [22]). ■

We can now prove the following strong result.

*Theorem 10 (Spectral decomposition).* For every system A, every mixed state can be written as a convex combination of perfectly distinguishable pure states.

*Proof.* The proof is by induction on the dimension of the system. If  $d_A = 1$ , the thesis trivially holds. Now suppose that the thesis holds for any system B with dimension  $d_B \leq N$ , and take a mixed state  $\rho \in \text{St}_1(A)$  where  $d_A = N + 1$ . There are two possibilities: either (1)  $\rho$  is not completely mixed or (2)  $\rho$  is completely mixed. Suppose that (1)  $\rho$  is not completely mixed. Then by the compression axiom 3 one can encode it in a system C, using an encoding operation  $\mathcal{E} \in \text{Transf}(A, C)$ . Now, the maximum number of perfectly distinguishable states in C is equal to the maximum number of perfectly distinguishable states in the face  $F_\rho$  (corollary 12). Since  $\rho$  is not completely mixed, we must have  $d_C \leq N$ . Using the induction hypothesis we then obtain that the state  $\mathcal{E}\rho \in \text{St}_1(C)$  is a mixture of perfectly distinguishable pure states, say  $\mathcal{E}\rho = \sum_i p_i \psi_i$ . Applying the decoding operation  $\mathcal{D} \in \text{Transf}(C, A)$  we get  $\rho = \mathcal{D}\mathcal{E}\rho = \sum_i p_i \mathcal{D}\psi_i$ . Since by lemmas 4 and 25 we know that the states  $\{\mathcal{D}\psi_i\}_{i=1}^{d_C}$  are pure and perfectly distinguishable, this is the desired decomposition for  $\rho$ . Now suppose that  $\rho$  is completely mixed (2). Consider the half-line in  $\text{St}_{\mathbb{R}}(A)$  defined by  $\sigma_t = (1 + t)\rho - t\chi$ ,  $t \geq 0$ .

Since the set of normalized states  $\text{St}_1(A)$  is compact, the line will cross its border at some point  $t_0$ . Therefore, one will have

$$\rho = \frac{1}{1+t_0}\sigma_{t_0} + \frac{t_0}{1+t_0}\chi$$

for some state  $\sigma_{t_0}$  on the border of  $\text{St}_1(A)$ , that is, for some state that is not completely mixed. But we know from the discussion of point (1) that the state  $\sigma_{t_0}$  is a mixture of perfectly distinguishable pure states, say  $\sigma_{t_0} = \sum_{i=1}^k p_i \varphi_i$ . By lemma 24 this set can be extended to a maximal set of perfectly distinguishable pure states  $\{\varphi_i\}_{i=1}^{d_A}$ . On the other hand, theorem 9 states that  $\chi = \sum_{i=1}^{d_A} \varphi_i/d_A$ . This implies the desired decomposition

$$\rho = \sum_{i=1}^{d_A} \left[ \frac{q_i}{1+t_0} + \frac{t_0}{d_A(1+t_0)} \right] \varphi_i,$$

where  $q_i = p_i$  for  $1 \leq i \leq k$ , and  $q_i = 0$  otherwise. ■

It is easy to show that the marginals of a pure bipartite state have the same spectral decomposition.

*Corollary 18.* Let  $\Psi \in \text{St}_1(AB)$  be a pure state, and let  $\rho$  and  $\tilde{\rho}$  be the marginals of  $\Psi$  on systems A and B, respectively. If  $\rho$  has spectral decomposition  $\rho = \sum_{i=1}^{d_A} p_i \varphi_i$ , with  $p_i > 0$  for every  $i = 1, \dots, r$ ,  $r \leq d_A$ , then  $\tilde{\rho}$  has spectral decomposition  $\tilde{\rho} = \sum_{i=1}^r p_i \psi_i$ .

*Proof.* Let  $\{a_i\}_{i=1}^{d_A}$  be the observation test such that  $(a_i|\varphi_j) = \delta_{ij}$ ,  $\{b_i\}_{i=1}^r$  be the observation test such that  $(b_i|_B|\Psi)_{AB} = p_i|\varphi_i)_A$  for every  $i \leq r$ . For  $i \leq r$  define the pure state  $\psi_i \in \text{St}_1(B)$  and the probability  $q_i$  via the relation

$$q_i |\psi_i)_B := (a_i|_A|\Psi)_{AB}.$$

[Note that  $\psi_i$  is pure due to the pure conditioning axiom.] By definition we have

$$\begin{aligned} q_i(b_j|\psi_i) &= (a_i \otimes b_j|\Psi) \\ &= (a_i|\varphi_j) \\ &= p_i \delta_{ij} \quad \forall i \leq r, \quad \forall j \leq r. \end{aligned}$$

The above relation implies  $q_i = \sum_{j=1}^{d_A} q_i(b_j|\psi_i) = \sum_j p_i \delta_{ij} = p_i$  and  $(b_j|\psi_i) = \delta_{ij}$ . Hence the states  $\{\psi_i\}_{i=1}^r$  are perfectly distinguishable. On the other hand, we have  $(a_i \otimes e_B|\Psi) = (a_i|\rho) = 0 \quad \forall i > r$ , which implies  $(a_i|_A|\Psi)_{AB} = 0, \forall i > r$ . Therefore, we obtained

$$\begin{aligned} |\tilde{\rho})_B &= (e|_A|\Psi)_{AB} \\ &= \sum_{i=1}^{d_A} (a_i|_A|\Psi)_{AB} \\ &= \sum_{i=1}^r (a_i|_A|\Psi)_{AB} \\ &= \sum_{i=1}^r p_i |\psi_i)_A, \end{aligned}$$

which is the desired spectral decomposition. ■

The spectral decomposition of states has many consequences. Here we just discuss the simplest ones, which are needed for the purpose of the derivation of quantum theory.

A first consequence is the following lemma.

*Lemma 36.* Let  $\varphi \in \text{St}_1(A)$  be a pure state and let  $a \in \text{Eff}(A)$  be the unique atomic effect such that  $(a|\varphi) = 1$ . If  $\varphi$  is perfectly distinguishable from  $\rho$ , then  $(a|\rho) = 0$ .

*Proof.* Let us write  $\rho = \sum_{i=1}^k p_i \varphi_i$ , with  $\{\varphi_i\}_{i=1}^k$  perfectly distinguishable pure states and  $p_i > 0$  for each  $i$ . Now, by lemma 23 the states  $\{\varphi_1, \dots, \varphi_k, \varphi\}$  are perfectly distinguishable, and by lemma 24 this set can be extended to a maximal set of perfectly distinguishable pure states  $\{\gamma_m\}_{m=1}^{d_A}$ , with  $\gamma_i = \varphi_i$  for  $i \leq k$  and  $\gamma_{k+1} = \varphi$ . Denote by  $\{c_m\}_{m=1}^{d_A}$  the observation test that perfectly distinguishes between the states  $\{\gamma_m\}$ . Note that, by definition,  $(c_{k+1}|\varphi) = 1$  and  $(c_{k+1}|\varphi_j) = 0$  for every  $j \neq k+1$ . Also, recall that  $c_{k+1}$  is atomic (lemma 27). By the duality of theorem 8 we have  $a = c_{k+1}$  and, therefore,  $(a|\rho) = \sum_{i=1}^k p_i (c_{k+1}|\varphi_i) = 0$ . ■

Another consequence of theorem 10 is the following characterization of the completely mixed states as *full rank* states.

*Corollary 19 (Characterization of completely mixed states).*

A state  $\rho \in \text{St}_1(A)$ , written as a mixture  $\rho = \sum_{i=1}^{d_A} p_i \varphi_i$  of a maximal set of perfectly distinguishable pure states  $\{\varphi_i\}_{i=1}^{d_A}$ , is completely mixed if and only if  $p_i > 0$  for every  $i = 1, \dots, d_A$ .

*Proof.* Necessity: If  $p_i = 0$  for some  $i$ , then  $\rho$  is perfectly distinguishable from  $\varphi_i$ . Hence, it cannot be completely mixed. Sufficiency: let  $p_{\min} = \min\{p_i, i = 1, \dots, d_A\}$ . Then we have  $\rho = p_{\min}\chi + (1 - p_{\min})\sigma$ , where  $\sigma$  is the state defined by  $\sigma = 1/(1 - p_{\min}) \sum_{i=1}^{d_A} (p_i - p_{\min}/d_A)\varphi_i$ . Since  $\rho$  contains  $\chi$  in its convex decomposition, and since  $\chi$  is completely mixed, we conclude that  $\rho$  is completely mixed. ■

In particular, for two-dimensional systems we have the result.

*Corollary 20.* For  $d_A = 2$  any state on the border of  $\text{St}_1(A)$  is pure.

Another consequence of theorem 10 is that every element in the vector space  $\text{St}_{\mathbb{R}}(A)$  can be written as a linear combination of perfectly distinguishable states.

*Corollary 21.* For every  $\xi \in \text{St}_{\mathbb{R}}(A)$  there exists a maximal set of perfectly distinguishable pure states  $\{\varphi_i\}_{i=1}^{d_A}$  and a set of real numbers  $\{c_i\}_{i=1}^{d_A}$  such that  $(\xi) = \sum_i c_i |\varphi_i)$ .

*Proof.* Write  $\xi$  as  $\xi = c_+\rho - c_-\sigma$ , where  $c_+, c_- \geq 0$  and  $\rho$  and  $\sigma$  are normalized states. If  $c_- = 0$  there is nothing to prove, because  $\xi$  is proportional to a state. Then, suppose that  $c_- > 0$ . Write  $\sigma$  as  $\sigma = \sum_i p_i \psi_i$  where  $\{\psi_i\}$  are perfectly distinguishable and define  $k = \max\{p_i\}$ . Then one has  $\chi + 1/(c_-kd_A)\xi = (\chi - 1/(kd_A)\sigma) + c_+/(c_-kd_A)\rho$ . Now, by definition  $\chi - 1/(kd_A)\sigma$  is proportional to a state: indeed we have  $[\chi - 1/(kd_A)\sigma] = 1/d_A \sum_i (1 - p_i/k)\psi_i$ , and, by definition  $1 - p_i/k \geq 0$ . Therefore  $\chi + 1/(c_-kd_A)\xi$  is proportional to a state, say  $\chi + 1/(c_-kd_A)\xi = t\tau$ , with  $t > 0$ . Writing  $\tau$  as  $\tau = \sum_i q_i \varphi_i$ , where  $\{\varphi_i\}_{i=1}^{d_A}$  is a maximal set of perfectly distinguishable pure states, we then obtain  $\xi = (c_-kd_A)(t\tau - \chi) = (c_-kd_A) \sum_i (tq_i - 1/d_A)\varphi_i$ , which is the desired decomposition. ■

In quantum theory corollary 21 is equivalent to the fact that every Hermitian matrix is diagonal in a suitable orthonormal basis. A simple consequence of corollary 21 is the following.

*Corollary 22.* For every system A with  $d_A = 2$  there is a continuous set of pure states.

*Proof.* Let  $\xi \in \text{St}_{\mathbb{R}}(A)$  be an arbitrary vector such that  $(e|\xi) = 0$ . Note that since the convex set  $\text{St}_1(A)$  cannot be a

segment (corollary 7), we must have  $D_A = \dim[\text{St}_{\mathbb{R}}(A)] > 2$  and, therefore, the space of vectors  $\xi$  such that  $(e|\xi) = 0$  is at least two dimensional. By corollary 21 we have  $\xi = c(\varphi_1 - \varphi_2) = 2c(\varphi_1 - \chi)$ , where  $c \geq 0$ ,  $\{\varphi_1, \varphi_2\}$  are two perfectly distinguishable pure states and we used the fact that  $\chi = \frac{1}{2}(\varphi_1 + \varphi_2)$ . Let us define  $\varphi_{\xi} := \varphi_1$ . With this definition, if  $\varphi_{\xi_1} = \varphi_{\xi_2}$  then one has  $\xi_2 = t\xi_1$  for some  $t \geq 0$ . Now, since there is a continuous infinity of vectors  $\xi$  (up to scaling), there must be a continuous set of pure states. ■

We conclude this section with the dual result to the ‘‘spectral decomposition’’ of corollary 21.

*Corollary 23.* For every  $x \in \text{Eff}_{\mathbb{R}}(A)$  there exists a perfectly distinguishing observation-test  $\{a_i\}_{i=1}^{d_A}$  and a set of real numbers  $\{d_i\}_{i=1}^{d_A}$  such that  $(x| = \sum_i d_i (a_i|$ .

*Proof.* Let  $\Phi \in \text{St}_1(A\tilde{A})$  be a purification of the invariant state  $\chi_A$ , where  $\tilde{A}$  is the conjugate system defined in corollary 16. Take the Choi vector  $|R_x\rangle_{\tilde{A}} := (x|_A |\Phi\rangle_{A\tilde{A}}$ . By corollary 21 there exists a maximal set of perfectly distinguishable pure states  $\{|\psi_i\rangle_{\tilde{A}}\}_{i=1}^{d_A}$  and a set of real numbers  $\{c_i\}_{i=1}^{d_A}$  such that  $|R_x\rangle = \sum_i c_i |\psi_i\rangle_{\tilde{A}}$ . Let  $\{a_i\}_{i=1}^{d_A} \subset \text{Eff}(A)$  be the observation test such that  $1/d_A |\psi_i\rangle_{\tilde{A}} = (a_i|_A |\Phi\rangle_{A\tilde{A}}$  for every  $i = 1, \dots, d_A$  (recall that by corollary 16 the marginal of  $\Phi$  on system  $\tilde{A}$  is the invariant state  $\chi_{\tilde{A}}$  and  $d_{\tilde{A}} = d_A$ ). The test  $\{a_i\}_{i=1}^{d_A}$  is perfectly distinguishing: if  $\{b_i\}_{i=1}^{d_A}$  is the observation test such that  $(b_i|\psi_j) = \delta_{ij}$  and  $\varphi_i \in \text{St}_1(A)$  is the state defined by  $(\varphi_i)_A := d_A (b_i|_A |\Phi\rangle_{A\tilde{A}}$ , then we have

$$\begin{aligned} (a_i|\varphi_j) &= d_A (a_i \otimes b_j | \Phi) \\ &= (b_j|\psi_i) \\ &= \delta_{ij}. \end{aligned}$$

Moreover, we have

$$\begin{aligned} (x|_A |\Phi\rangle_{A\tilde{A}} &= |R_x\rangle_{\tilde{A}} \\ &= \sum_i c_i |\psi_i\rangle_{\tilde{A}} \\ &= \sum_i c_i d_A (a_i|_A |\Phi\rangle_{A\tilde{A}}. \end{aligned}$$

Since  $\Phi$  is dynamically faithful, this implies  $(x| = \sum_i d_i (a_i|$ , where  $d_i := c_i d_A$ . ■

## IX. TELEPORTATION REVISITED

In this section we revisit probabilistic teleportation using the results about informational dimension. The key point is the section will be the proof the equality  $D_A = d_A^2$ , which relates the dimension of the vector space  $\text{St}_{\mathbb{R}}(A)$  with the informational dimension  $d_A$ .

### A. Probability of teleportation

We start by showing a probabilistic teleportation scheme that achieves success probability  $p_A = 1/d_A$  for every system  $A$ .

*Theorem 11 (Probability of teleportation).* For every system  $A$ , probabilistic teleportation can be achieved with probability  $p_A = 1/d_A^2$ .

*Proof.* Let  $\tilde{A}$  and  $|\Phi\rangle_{A\tilde{A}}$  be the conjugate system and the pure state defined in corollary 16. Then, the state  $|\Phi\rangle_{A\tilde{A}}|\Phi\rangle_{A\tilde{A}}$  satisfies the identity

$$\begin{aligned} \begin{array}{c} \Lambda \\ \Phi \\ \tilde{\Lambda} \\ e \end{array} &= \begin{array}{c} \Lambda \\ \chi \\ \tilde{\Lambda} \end{array} = \begin{array}{c} \Lambda \\ \chi \\ \tilde{\Lambda} \end{array}. \end{aligned}$$

On the other hand, by lemma 33 the maximum probability of a pure state in the convex decomposition of  $\chi_{A\tilde{A}}$  is  $p_{\max} = 1/d_{A\tilde{A}}$ , and by corollaries 15 and 16 one has  $p_{\max} = 1/(d_A d_{\tilde{A}}) = 1/d_A^2$ . Therefore, by lemma 12 there exists an atomic effect  $E$  such that

$$\begin{array}{c} \Lambda \\ \Phi \\ \tilde{\Lambda} \\ E \end{array} = \frac{1}{d_A^2} \begin{array}{c} \Lambda \\ \Phi \\ \tilde{\Lambda} \end{array} \quad (18)$$

and, since  $\Phi$  is dynamically faithful,

$$\begin{array}{c} \Lambda \\ \Phi \\ \tilde{\Lambda} \\ E \end{array} = \frac{1}{d_A^2} \begin{array}{c} \Lambda \\ \mathcal{J} \\ \Lambda \end{array} \quad (19)$$

as can be verified applying both members of Eq. (19) to  $\Phi$ , thus obtaining Eq. (18). ■

### B. Isotropic states and effects

Here we define two maps that send reversible transformations of  $A$  to reversible transformations of  $\tilde{A}$ : the transpose and the conjugate. Using these maps we will also define the notions of isotropic states and effects and we will prove some properties of them.

Let us start from the definition of the transpose.

*Lemma 37 (Transpose of a reversible transformation).* Let  $\Phi \in \text{St}(A\tilde{A})$  be a purification of the invariant state  $\chi_A$ . The reversible transformations of system  $\tilde{A}$  are in one-to-one correspondence with the reversible transformations of system  $A$  via the transposition  $\tau$  defined as follows:

$$\begin{array}{c} \Lambda \\ \Phi \\ \tilde{\Lambda} \\ \mathcal{U} \end{array} = \begin{array}{c} \Lambda \\ \Phi \\ \tilde{\Lambda} \\ \mathcal{U}^{\tau} \end{array} \quad (20)$$

[note that the transposition is defined with respect to the given state  $\Phi$ ].

*Proof.* Since  $(\mathcal{U} \otimes \mathcal{J}_{\tilde{A}})|\Phi\rangle$  and  $|\Phi\rangle$  are purifications of the same state  $\chi_A$ , there exists a reversible transformation  $\mathcal{U}^{\tau} \in \mathbf{G}_{\tilde{A}}$  such that Eq. (20) holds. Since  $\Phi$  is dynamically faithful on  $A$ , the map  $\mathcal{U} \mapsto \mathcal{U}^{\tau}$  is injective. Furthermore, the map is surjective: for every reversible  $\mathcal{V} \in \mathbf{G}_{\tilde{A}}$  the states  $(\mathcal{J}_A \otimes \mathcal{V})|\Phi\rangle$  and  $|\Phi\rangle$  are two purifications of the same state  $\chi_{\tilde{A}}$ , and, by the uniqueness of purification stated in postulate 1, there exists a reversible  $\mathcal{U} \in \mathbf{G}_A$  such that

$$\begin{array}{c} \Lambda \\ \Phi \\ \tilde{\Lambda} \\ \mathcal{V} \end{array} = \begin{array}{c} \Lambda \\ \Phi \\ \tilde{\Lambda} \\ \mathcal{U} \end{array}, \quad (21)$$

namely  $\mathcal{V} = \mathcal{U}^{\tau}$ . ■

The conjugate is just defined as the inverse of the transpose.

*Definition 7.* Let  $\tau$  be the transpose defined with respect to the state  $\Phi \in \text{St}_1(A\tilde{A})$ . The conjugate of the reversible channel  $\mathcal{U} \in \mathbf{G}_A$  is the reversible channel  $\mathcal{U}^* \in \mathbf{G}_{\tilde{A}}$  defined by  $\mathcal{U}^* := (\mathcal{U}^\tau)^{-1}$ .

We can now give the definition of isotropic pure state (isotropic atomic effect).

*Definition 8.* A pure state  $\Psi \in \text{St}(A\tilde{A})$  [an atomic effect  $F \in \text{Eff}(\tilde{A}A)$ ] is isotropic if it is invariant under the  $\mathcal{U} \otimes \mathcal{U}^*$  (under  $\mathcal{U}^* \otimes \mathcal{U}$ ). Diagrammatically

$$\begin{array}{c}
 \begin{array}{c} \Psi \\ \hline \tilde{A} \end{array} \begin{array}{c} A \\ \hline \tilde{A} \end{array} \begin{array}{c} \mathcal{U} \\ \hline \mathcal{U}^* \end{array} \begin{array}{c} A \\ \hline \tilde{A} \end{array} = \begin{array}{c} \Psi \\ \hline \tilde{A} \end{array} \begin{array}{c} A \\ \hline \tilde{A} \end{array} \quad \forall \mathcal{U} \in \mathbf{G}_A \\
 \\
 \begin{array}{c} \tilde{A} \\ \hline \tilde{A} \end{array} \begin{array}{c} \mathcal{U}^* \\ \hline \mathcal{U} \end{array} \begin{array}{c} \tilde{A} \\ \hline \tilde{A} \end{array} \begin{array}{c} F \\ \hline F \end{array} = \begin{array}{c} \tilde{A} \\ \hline \tilde{A} \end{array} \begin{array}{c} F \\ \hline F \end{array} \quad \forall \mathcal{U} \in \mathbf{G}_A. \quad (22)
 \end{array}$$

An example of isotropic state is  $\Phi$ : indeed, by definition of conjugate we have, for every  $\mathcal{U} \in \mathbf{G}_A$ ,

$$\begin{aligned}
 (\mathcal{U} \otimes \mathcal{U}^*)|\Phi\rangle &= (\mathcal{U} \otimes (\mathcal{U}^\tau)^{-1})|\Phi\rangle \\
 &= (\mathcal{I}_A \otimes (\mathcal{U}^\tau)^{-1}\mathcal{U}^\tau)|\Phi\rangle = |\Phi\rangle.
 \end{aligned}$$

As a consequence, the teleportation effect  $E$  is isotropic: indeed one has

$$\begin{aligned}
 \begin{array}{c} \Phi \\ \hline \tilde{A} \end{array} \begin{array}{c} A \\ \hline \tilde{A} \end{array} \begin{array}{c} \mathcal{U}^* \\ \hline \mathcal{U} \end{array} \begin{array}{c} A \\ \hline \tilde{A} \end{array} \begin{array}{c} E \\ \hline E \end{array} &= \begin{array}{c} \Phi \\ \hline \tilde{A} \end{array} \begin{array}{c} A \\ \hline \tilde{A} \end{array} \begin{array}{c} \mathcal{U}^{-1} \\ \hline \mathcal{U}^\tau \end{array} \begin{array}{c} A \\ \hline \tilde{A} \end{array} \begin{array}{c} E \\ \hline E \end{array} \\
 &= \frac{1}{d_A^2} \begin{array}{c} \Phi \\ \hline \tilde{A} \end{array} \begin{array}{c} A \\ \hline \tilde{A} \end{array} \begin{array}{c} \mathcal{U}^{-1} \\ \hline \mathcal{U}^\tau \end{array} \begin{array}{c} A \\ \hline \tilde{A} \end{array} \\
 &= \frac{1}{d_A^2} \begin{array}{c} \Phi \\ \hline \tilde{A} \end{array} \begin{array}{c} A \\ \hline \tilde{A} \end{array} \\
 &= \begin{array}{c} \Phi \\ \hline \tilde{A} \end{array} \begin{array}{c} A \\ \hline \tilde{A} \end{array} \begin{array}{c} E \\ \hline E \end{array}
 \end{aligned}$$

which implies  $(E|(\mathcal{U}^* \otimes \mathcal{U}) = (E|$ , since the state  $\Phi \otimes \Phi$  is dynamically faithful.

We now show that all isotropic pure states (isotropic atomic effects) are connected to the state  $\Phi$  (to the effect  $E$ ) through a local reversible transformation.

*Lemma 38.* If a pure state  $\Psi \in \text{St}_1(A\tilde{A})$  is isotropic then  $|\Psi\rangle = (\mathcal{V} \otimes \mathcal{I}_{\tilde{A}})|\Phi\rangle$  for some reversible transformation  $\mathcal{V} \in \mathbf{G}_A$  such that  $\mathcal{V}\mathcal{U} = \mathcal{U}\mathcal{V}$  for every  $\mathcal{U} \in \mathbf{G}_A$ .

*Proof.* Since  $\Psi$  satisfies Eq. (22), its marginal on system  $\tilde{A}$  is the invariant state  $|\chi_{\tilde{A}}\rangle$ . Since  $\Psi$  and  $\Phi$  are purifications of the same state, there must exist a reversible channel  $\mathcal{V} \in \mathbf{G}_A$  such that  $|\Psi\rangle = (\mathcal{V} \otimes \mathcal{I}_{\tilde{A}})|\Phi\rangle$ . Moreover, we have for every

$$\mathcal{U} \in \mathbf{G}_A$$

$$\begin{aligned}
 (\mathcal{U}\mathcal{V}\mathcal{U}^{-1} \otimes \mathcal{I}_{\tilde{A}})|\Phi\rangle &= (\mathcal{U}\mathcal{V} \otimes \mathcal{U}^*)|\Phi\rangle \\
 &= (\mathcal{U} \otimes \mathcal{U}^*)|\Psi\rangle \\
 &= |\Psi\rangle \\
 &= (\mathcal{V} \otimes \mathcal{I}_{\tilde{A}})|\Phi\rangle.
 \end{aligned}$$

Since  $\Phi$  is dynamically faithful, the above equation implies  $\mathcal{U}\mathcal{V}\mathcal{U}^{-1} = \mathcal{V}$  for every  $\mathcal{U} \in \mathbf{G}_A$ . ■

By the duality between states and effects, it is easy to obtain the following.

*Lemma 39.* Let  $A \in \text{Eff}(A\tilde{A})$  be the atomic effect such that  $(A|\Phi) = 1$ . If an atomic effect  $F \in \text{Eff}(\tilde{A}A)$  is isotropic then  $(F|_{\tilde{A}A} = (A|_{\tilde{A}A}(\mathcal{I}_{\tilde{A}} \otimes \mathcal{V}))$  for some reversible transformation  $\mathcal{V} \in \mathbf{G}_A$  such that  $\mathcal{V}\mathcal{U} = \mathcal{U}\mathcal{V}$  for every  $\mathcal{U} \in \mathbf{G}_A$ .

*Proof.* Let  $\Psi$  be the pure state such that  $(F|\Psi) = 1$ . Clearly  $\Psi$  is isotropic: one has  $(F|(\mathcal{U} \otimes \mathcal{U}^*)|\Psi) = (F|\Psi) = 1$ , and, therefore,  $(\mathcal{U} \otimes \mathcal{U}^*)|\Psi) = |\Psi)$ . By lemma 38, there exists a reversible transformation  $\mathcal{V}$  such that  $|\Psi) = (\mathcal{V}^{-1} \otimes \mathcal{I}_{\tilde{A}})|\Phi)$  and  $\mathcal{V}^{-1}\mathcal{U} = \mathcal{U}\mathcal{V}^{-1}$  for every  $\mathcal{U} \in \mathbf{G}_A$ . Now, this implies  $(F|(\mathcal{V}^{-1} \otimes \mathcal{I}_{\tilde{A}})|\Phi) = (F|\Psi) = 1$ , which by theorem 8 implies  $(F| = (A|(\mathcal{V} \otimes \mathcal{I}_{\tilde{A}})$ . ■

As a consequence, every isotropic effect is connected to the teleportation effect by a local reversible transformation:

*Corollary 24.* If an atomic effect  $F \in \text{Eff}(\tilde{A}A)$  is isotropic then  $(F|_{\tilde{A}A} = (E|_{\tilde{A}A}(\mathcal{I}_{\tilde{A}} \otimes \mathcal{V}))$  for some reversible transformation  $\mathcal{V} \in \mathbf{G}_A$  such that  $\mathcal{V}\mathcal{U} = \mathcal{U}\mathcal{V}$  for every  $\mathcal{U} \in \mathbf{G}_A$ .

*Proof.* Since  $(E|$  and  $(F|$  are both isotropic, lemma 39 implies that they are both connected to  $(A|$  through a local reversible transformation, say  $\mathcal{V}$  and  $\mathcal{W}$ , respectively. Therefore, they are connected to each other through the transformation  $\mathcal{W}\mathcal{V}^{-1}$ . ■

### C. Dimension of the state space

In this subsection we use the local distinguishability axiom to prove the equality  $D_A = d_A^2$  (see theorem 12). As a consequence, we will be able to represent the states of a system  $A$  as square  $d_A \times d_A$  Hermitian complex matrices, that is, Hermitian operators on the complex Hilbert space  $\mathbb{C}^{d_A}$ . Theorem 12 is thus the point where the complex field (as opposed to the real field) enters in our derivation. Notice that, even if the local distinguishability excludes quantum theory on real Hilbert spaces since the very beginning, to prove the emergence of complex Hilbert spaces we need to use all the six principles.

Due to local distinguishability, any bipartite state  $\Psi \in \text{St}(AB)$  can be written as

$$|\Psi\rangle = \sum_{i=1}^{D_A} \sum_{j=1}^{D_B} \Psi_{ij} |\alpha_i\rangle |\beta_j\rangle,$$

where  $\{\alpha_i\}$  ( $\{\beta_j\}$ ) is a basis for the vector space  $\text{St}_{\mathbb{R}}(A)$  [ $\text{St}_{\mathbb{R}}(B)$ ]. Similarly, a bipartite effect  $F \in \text{Eff}(BA)$  can be written as

$$(F| = \sum_{k=1}^{D_B} \sum_{l=1}^{D_A} F_{kl} (\beta_k^* | (\alpha_l^* |$$

with  $(\alpha_i^*|\alpha_i) = \delta_{ii}$  and  $(\beta_k^*|\beta_j) = \delta_{jk}$ . Finally, a transformation  $\mathcal{C}$  from A to B can be written as

$$\mathcal{C} = \sum_{j=1}^{D_B} \sum_{i=1}^{D_A} C_{ji} |\beta_j\rangle (\alpha_i^*|.$$

In this matrix representation the teleportation diagram of Eq. (19) becomes

$$\Phi E = \frac{I_{D_A}}{d_A^2}, \quad (23)$$

where  $I_{D_A}$  is the identity matrix in dimension  $D_A$ . On the other hand, we also have

$$1 \geq (E|\Phi) = \text{Tr}[\Phi E] = \frac{D_A}{d_A^2}$$

and, therefore,

$$D_A \leq d_A^2.$$

We now show that one has the equality, using the following standard lemma.

*Lemma 40.* With a suitable choice of basis for the vector space  $\text{St}_{\mathbb{R}}(A)$ , every reversible transformation  $\mathcal{U} \in \mathbf{G}_A$  is represented by a matrix  $M_{\mathcal{U}}$  of the form

$$M_{\mathcal{U}} = \begin{pmatrix} 1 & 0 \\ 0 & O_{\mathcal{U}} \end{pmatrix}, \quad (24)$$

where  $O_{\mathcal{U}}$  is an orthogonal  $(D_A - 1) \times (D_A - 1)$  matrix.

*Proof.* Let  $\{\xi_i\}$  be a basis for  $\text{St}_{\mathbb{R}}(A)$ , chosen in such a way that the first basis vector is  $\chi$ , while the remaining vectors satisfy  $(e|\xi_i) = 0, \forall i = 2, \dots, D_A$ . Such a choice is always possible since every vector  $v \in \text{St}_{\mathbb{R}}(A)$  can be written as  $v = (e|v)\chi + \xi$ , where  $\xi$  satisfies  $(e|\xi) = 0$ . Now, since  $\mathcal{U}\chi = \chi$ , the first column of  $M_{\mathcal{U}}$  must be  $(1, 0, \dots, 0)^T$ . Moreover, since for every normalized state  $\rho$ ,  $\mathcal{U}\rho$  is a normalized state, one must have  $(e|\mathcal{U}|\xi) = 0$  for every  $\xi$  such that  $(e|\xi) = 0$ . Hence the first row of  $M_{\mathcal{U}}$  must be  $(1, 0, \dots, 0)$ , namely  $M_{\mathcal{U}}$  has the block form of Eq. (24). It remains to show that, with a suitable choice of basis, the matrix  $O_{\mathcal{U}}$  in the second block can be chosen to be orthogonal. Observe that by definition the matrices  $\{M_{\mathcal{U}}\}_{\mathcal{U} \in \mathbf{G}_A}$  form a representation of the group  $\mathbf{G}_A$ : indeed, one has  $M_{\mathcal{I}} = I_{D_A}$  and  $M_{\mathcal{U}\mathcal{V}} = M_{\mathcal{U}}M_{\mathcal{V}}$  for every  $\mathcal{U}, \mathcal{V} \in \mathbf{G}$ . Consider the positive definite matrix  $P$  defined by the integral

$$P := \int d\mathcal{U} O_{\mathcal{U}}^T O_{\mathcal{U}},$$

where  $d\mathcal{U}$  is the Haar measure on the compact group  $\mathbf{G}_A$  (see corollary 30 of Ref. [22] for the proof of compactness) and  $A^T$  denotes the transpose of  $A$ . By definition, one has  $P^T = P$  and  $O_{\mathcal{U}}^T P O_{\mathcal{U}} = P$  for every  $\mathcal{U} \in \mathbf{G}_A$ . Let us now define the new representation

$$O'_{\mathcal{U}} := P^{\frac{1}{2}} O_{\mathcal{U}} P^{-\frac{1}{2}},$$

obtained from  $O_{\mathcal{U}}$  by a change of basis in the subspace spanned by  $\{\xi_i\}_{i=2}^{D_A}$ . With this choice, each matrix  $O'_{\mathcal{U}}$  is orthogonal:

$$\begin{aligned} O'^T_{\mathcal{U}} O'_{\mathcal{U}} &= (P^{\frac{1}{2}} O_{\mathcal{U}} P^{-\frac{1}{2}})^T (P^{\frac{1}{2}} O_{\mathcal{U}} P^{-\frac{1}{2}}) \\ &= P^{-\frac{1}{2}} (O_{\mathcal{U}}^T P O_{\mathcal{U}}) P^{-\frac{1}{2}} = I_{D_A-1}. \end{aligned}$$

As a consequence, we have the following. ■

*Corollary 25.* For every system A, the group of reversible transformations  $\mathbf{G}_A$  is (isomorphic to) a compact subgroup of  $O(D_A - 1)$ .

*Lemma 41.* Let  $E \in \text{Eff}(A\bar{A})$  be the teleportation effect of Eq. (19). Then one has  $(E|\Phi) = 1$ .

*Proof.* Let  $A \in \text{Eff}(A\bar{A})$  be the atomic effect such that  $(A|\Phi) = 1$ . We now prove that  $A = E$ . Indeed, by corollary 24 there exists a reversible transformation  $\mathcal{V} \in \mathbf{G}_A$  such that  $(A| = (E|(\mathcal{V} \otimes \mathcal{I}_{\bar{A}})$ . Using a basis for  $\text{St}_{\mathbb{R}}(A)$  such that the transformations in  $\mathbf{G}_A$  are represented by orthogonal matrices as in Eq. (24), one has

$$\begin{aligned} 1 &= (A|\Phi) \\ &= (E|(\mathcal{V} \otimes \mathcal{I}_{\bar{A}})|\Phi) \\ &= \text{Tr}[EM_{\mathcal{V}}\Phi] \\ &= \text{Tr}[\Phi EM_{\mathcal{V}}] \\ &= \frac{\text{Tr}[M_{\mathcal{V}}]}{d_A^2}, \end{aligned}$$

having used Eq. (23) for the last equality. Using the inequality  $\text{Tr}[M_{\mathcal{V}}] \leq \text{Tr}[I_{D_A}]$ , that holds for every orthogonal  $D_A \times D_A$  matrix, we then obtain

$$\begin{aligned} 1 &= \frac{\text{Tr}[M_{\mathcal{V}}]}{d_A^2} \\ &\leq \frac{\text{Tr}[I_{D_A}]}{d_A^2} \\ &= \text{Tr}[E\Phi] \\ &= (E|\Phi) \\ &\leq 1, \end{aligned}$$

and, therefore  $(E|\Phi) = 1$ . ■

*Theorem 12 (Dimension of the state space).* The dimension  $D_A$  of the vector space generated by the states in  $\text{St}(A)$  is  $D_A = d_A^2$ .

*Proof.* Using lemma 41 and Eq. (23) we obtain  $1 = (E|\Phi) = \text{Tr}[E\Phi] = \text{Tr}[I_{D_A}]/d_A^2 = D_A/d_A^2$ . Hence,  $D_A = d_A^2$ . ■

An interesting consequence of the relation  $(E|\Phi) = 1$  is the following.

*Corollary 26 (No inversion).* Let us write an arbitrary state  $\rho \in \text{St}_1(A)$  as  $\rho = \chi_A + \xi$ , with  $(e|\xi) = 0$ . Then, the linear map  $\mathcal{N}$  defined by  $\mathcal{N}(\rho) = \chi_A - \xi$  is not a physical transformation.

*Proof.* Write the state  $\Phi$  as  $\Phi = \chi_A \otimes \chi_{\bar{A}} + \Xi$ . Since  $(e|_A|\Phi)_{A\bar{A}} = |\chi|_{\bar{A}}$  one must have  $(e|_A|\Xi)_{A\bar{A}} = 0$ . Therefore,  $\Xi$  must be of the form  $\Xi = \sum_i \alpha_i \otimes \beta_i$  with  $(e|\alpha_i) = 0$  for all  $i$ . Applying the transformation  $\mathcal{N}$  one then obtains  $(\mathcal{N} \otimes \mathcal{I}_{\bar{A}})\Phi = \chi_A \otimes \chi_{\bar{A}} - \Xi$ . We now prove that this is not a state, and therefore,  $\mathcal{N}$  cannot be a physical transformation.

Let  $E$  be the teleportation effect. Since  $(E|\Phi) = 1$ , we have  $1 = (E|\chi_A \otimes \chi_A) + (E|\Xi) = 1/d_A^2 + (E|\Xi)$ . Now, we have

$$(E|(\mathcal{N} \otimes \mathcal{I}_A)|\Phi) = \frac{1}{d_A^2} - (E|\Xi) = \frac{2}{d_A^2} - 1.$$

Since this quantity is negative for every  $d_A > 1$ , the map  $\mathcal{N}$  cannot be a physical transformation. ■

*Corollary 27.* The matrix  $M_{\mathcal{N}}$  defined as

$$M_{\mathcal{N}} = \begin{pmatrix} 1 & 0 \\ 0 & -I_{D_A-1} \end{pmatrix} \quad (25)$$

cannot represent a physical transformation of system A.

## X. DERIVATION OF THE QUBIT

In this section we show that every two-dimensional system in our theory is a *qubit*. With this expression we mean that the normalized states in  $\text{St}_1(A)$  can be represented as density matrices for a quantum system with two-dimensional Hilbert space. With this choice of representation we also show that the effects in  $\text{Eff}(A)$  are all the positive Hermitian matrices bounded by the identity, and that the reversible transformations  $\mathbf{G}_A$  act on the states by conjugation with unitary matrices in  $\text{SU}(2)$ .

The first step is to prove that the set of normalized states  $\text{St}_1(A)$  is a sphere. The idea of the proof is a simple geometric observation: in the ordinary three-dimensional space the sphere is the only compact convex set that has an infinite number of pure states connected by orthogonal transformations. The complete proof is given in the following.

*Theorem 13 (The Bloch sphere).* The normalized states of a system A with  $d_A = 2$  form a sphere and the group  $\mathbf{G}_A$  is  $\text{SO}(3)$ .

*Proof.* According to corollary 25, the group of reversible transformations  $\mathbf{G}_A$  is a compact subgroup of the orthogonal group  $\text{O}(3)$ . It cannot be the whole  $\text{O}(3)$  because, as we saw in corollary 27, the inversion  $-I$  cannot represent a physical transformation. We now show that  $\mathbf{G}_A$  must be  $\text{SO}(3)$  by excluding all the other possibilities. From corollary 22 we know that the system A has a continuum of pure states. Therefore, the group  $\mathbf{G}_A$  must contain a continuous set of transformations. Now, from the classification of the closed subgroups of  $\text{O}(3)$  we know that there are only two possibilities: (i)  $\mathbf{G}_A$  is  $\text{SO}(3)$  and (ii)  $\mathbf{G}_A$  is the subgroup generated by  $\text{SO}(2)$ , the group of rotations around a fixed axis, say the  $z$  axis, and possibly by the reflections with respect to planes containing the  $z$  axis. Note that the reflection in the  $xy$  plane is forbidden, because the composition of this reflection with the rotation of  $\pi$  around the  $z$  axis would give the inversion, which is forbidden by corollary 26. Case (ii) is excluded because in this case the action of the group  $\mathbf{G}_A$  cannot be transitive. The detailed proof is as follows: because of the  $\text{SO}(2)$  symmetry, the set of pure states must contain at least a circle in the  $xy$  plane. This circle will be necessarily invariant under all operations in the group. However, since the convex set of states is three dimensional, there is at least a pure state outside the circle. Clearly there is no way to transform a state on the circle into a state outside the circle by means of an operation in  $\mathbf{G}_A$ . This is in contradiction with the fact that every two pure states are connected by a reversible transformation.

Hence, case (ii) is ruled out. The only remaining alternative is (i), namely that  $\mathbf{G}_A = \text{SO}(3)$  and, hence, the set of pure states generated by its action on a fixed pure state is a sphere. ■

Since the convex set of density matrices on a two-dimensional Hilbert space is a sphere, we can represent the states in  $\text{St}_1(A)$  as density matrices. Precisely, we can choose three orthogonal axes passing through the center of the sphere and call them  $x, y, z$  axes, take  $\varphi_{+,k}, \varphi_{-,k}$ ,  $k = x, y, z$  to be the two perfectly distinguishable pure states in the direction of the  $k$  axis and define  $\sigma_k := \varphi_{k,+} - \varphi_{k,-}$ . From the geometry of the sphere we know that any state  $\rho \in \text{St}_1(A)$  can be written as

$$|\rho) = |\chi) + \frac{1}{2} \sum_{k=x,y,z} n_k |\sigma_k), \quad \sum_{k=x,y,z} n_k^2 \leq 1, \quad (26)$$

where the pure states are those for which  $\sum_{k=x,y,z} n_k^2 = 1$ . The Bloch representation  $S_\rho$  of quantum state  $\rho$  is then obtained by associating the basis vectors  $\chi, \sigma_x, \sigma_y, \sigma_z$  to the matrices

$$S_\chi = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad S_{\sigma_x} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ S_{\sigma_y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad S_{\sigma_z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and by defining  $S_\rho$  by linearity from Eq. (26). Clearly, in this way we obtain

$$S_\rho = \frac{1}{2} \begin{pmatrix} 1 + n_z & n_x - in_y \\ n_x + in_y & 1 - n_z \end{pmatrix},$$

which is the expression of a generic density matrix. Denoting by  $M_2(\mathbb{C})$  the set of complex two-by-two matrices we have the following.

*Corollary 28 (Qubit density matrices).* For  $d_A = 2$  the set of states  $\text{St}_1(A)$  is isomorphic to the set of density matrices in  $M_2(\mathbb{C})$  through the isomorphism  $\rho \mapsto S_\rho$ .

Once we decide to represent the states in  $\text{St}_1(A)$  as matrices, the effects in  $\text{Eff}(A)$  are necessarily represented by matrices too. The matrix representation of an effect, given by the map  $a \in \text{Eff}(A) \mapsto E_a \in M_2(\mathbb{C})$  is defined uniquely by the relation

$$\text{Tr}[E_a S_\rho] = (a|\rho) \quad \forall \rho \in \text{St}(A).$$

We then have the following.

*Corollary 29.* For  $d_A = 2$  the set of effects  $\text{Eff}(A)$  is isomorphic to the set of positive Hermitian matrices  $P \in M_2(\mathbb{C})$  such that  $P \leq I$ .

*Proof.* Clearly the matrix  $E_a$  must be positive for every effect  $a$ , since we have  $\text{Tr}[E_a S_\rho] = (a|\rho) \geq 0$  for every density matrix  $S_\rho$ . Moreover, since we have  $\text{Tr}[E_a S_\rho] = (a|\rho) \leq 1$  for every density matrix  $S_\rho$ , we must have  $E_a \leq I$ . Finally, we know that for every couple of perfectly distinguishable pure states  $\varphi, \varphi_\perp$  there exists an atomic effect  $a$  such that  $(a|\varphi) = 1$  and  $(a|\varphi_\perp) = 0$ . Since the two pure states  $\varphi, \varphi_\perp$  are represented by orthogonal rank-one projectors  $S_\varphi$  and  $S_{\varphi_\perp}$ , we must have  $E_a = S_\varphi$ . This proves that the atomic effects are the whole set of positive rank-one projectors. As a consequence, also every positive matrix  $P$  with  $P \leq I$  must represent some effect  $a$ . ■

Finally, the reversible transformations are represented as conjugations by unitary matrices in  $\text{SU}(2)$ .

*Corollary 30.* For every reversible transformation  $\mathcal{U} \in \mathbf{G}_A$  with  $d_A = 2$  there exists a unitary matrix  $U \in \text{SU}(2)$  such that

$$S_{\mathcal{U}\rho} = US_\rho U^\dagger, \quad \rho \in \text{St}(A). \quad (27)$$

Conversely, for every  $U \in \text{SU}(2)$  there exists a reversible transformation  $\mathcal{U} \in \mathbf{G}_A$  such that Eq. (27) holds.

*Proof.* Every rotation of the Bloch sphere is represented by conjugation by some  $\text{SU}(2)$  matrix. Conversely, every conjugation by an  $\text{SU}(2)$  matrix represents some rotation on the Bloch sphere. On the other hand, we know that  $\mathbf{G}_A$  is the group of all rotations on the Bloch sphere (theorem 13). ■

Note that we proved that all two-dimensional systems A and B in our theory have the same states [ $\text{St}_1(A) \simeq \text{St}_1(B)$ ], the same effects [ $\text{Eff}(A) \simeq \text{Eff}(B)$ ], and the same reversible transformations ( $\mathbf{G}_A \simeq \mathbf{G}_B$ ), but we did not show that A and B are operationally equivalent. For example, A and B could be different when we compose them with a third system C: the set of states  $\text{St}_1(AC)$  and  $\text{St}_1(BC)$  could be nonisomorphic. The fact that every couple of two-dimensional systems A and B are operationally equivalent will be proved later (cf. corollary 40).

We conclude this section with a simple fact that will be very useful later.

*Corollary 31 (Superposition principle for qubits).* Let  $\{\varphi_1, \varphi_2\} \subset \text{St}_1(A)$  be two perfectly distinguishable pure states of a system A with  $d_A = 2$ . Let  $\{a_1, a_2\}$  be the observation test such that  $(a_i|\varphi_j) = \delta_{ij}$ . Then, for every probability  $0 \leq p \leq 1$  there exists a pure state  $\psi_p \in \text{St}_1(A)$  such that

$$(a_1|\psi_p) = p, \quad (a_2|\psi_p) = 1 - p. \quad (28)$$

Precisely, the set of pure states  $\psi_p \in \text{St}_1(A)$  satisfying Eq. (28) is a circle in the Bloch sphere.

*Proof.* Elementary property of density matrices. ■

## XI. PROJECTIONS

In this section we define the projection on a face  $F$  of the convex set  $\text{St}_1(A)$  and we prove several properties of projections. The projection on the face  $F$  will be defined as an atomic operation  $\Pi_F \in \text{Transf}(A)$  that acts as the identity on states in the face  $F$  and that annihilates the states on the *orthogonal face*  $F^\perp$ . In the following we first introduce the concept of orthogonal face, then prove the existence and uniqueness of projections, and finally give some useful results on the projection of a pure state on two orthogonal faces.

### A. Orthogonal faces and orthogonal complements

In order to introduce the notion of orthogonal face we need first a few elementary results. We start by showing that there is a canonical way to associate a state  $\omega_F$  to a face  $F$ .

*Lemma 42 (State associated to a face).* Let  $F$  be a face of the convex set  $\text{St}_1(A)$  and let  $\{\varphi_i\}_{i=1}^{|F|}$  be a maximal set of perfectly distinguishable pure states in  $F$ . Then the state  $\omega_F := \frac{1}{|F|} \sum_{i=1}^{|F|} \varphi_i$  depends only on the face  $F$  and not on the particular set  $\{\varphi_i\}_{i=1}^{|F|}$ . Moreover,  $F$  is the face identified by  $\omega_F$ .

*Proof.* Suppose that  $F$  is the face identified by  $\rho$  and let  $\mathcal{E} \in \text{Transf}(A, C)$  [or  $\mathcal{D} \in \text{Transf}(C, A)$ ] be the encoding (or decoding) in the ideal compression for  $\rho$ . By lemma 4

and corollary 12,  $\{\mathcal{E}\varphi_i\}_{i=1}^{|F|}$  is a maximal set of perfectly distinguishable pure states of C and by theorem 9 one has  $\chi_C = \frac{1}{|F|} \sum_{i=1}^{|F|} \mathcal{E}\varphi_i$ . Hence,  $\omega_F = \frac{1}{|F|} \sum_{i=1}^{|F|} \varphi_i = \frac{1}{|F|} \sum_{i=1}^{|F|} \mathcal{D}\mathcal{E}\varphi_i = \mathcal{D}\chi_C$ . Since the right-hand side of the equality is independent of the particular set  $\{\varphi_i\}_{i=1}^{|F|}$ , the state  $\omega_F$  in the left-hand side is independent too. To prove that  $F$  is the face identified by  $\omega_F$  it is enough to observe that  $\omega_F$  is completely mixed relative to  $F$ : this fact follows from the relation  $\omega_F = \mathcal{D}\chi_C$  and from lemma 5. ■

We now define the *orthogonal complement* of the state  $\omega_F$ .

*Definition 9.* The orthogonal complement of the state  $\omega_F$  is the state  $\omega_F^\perp \in \text{St}_1(A) \cup \{0\}$  defined as follows:

- (1) if  $|F| = d_A$ , then  $\omega_F^\perp = 0$ ;
- (2) if  $F < d_A$ , then  $\omega_F^\perp$  is defined by the relation

$$\chi_A = \frac{|F|}{d_A} \omega_F + \frac{d_A - |F|}{d_A} \omega_F^\perp. \quad (29)$$

An easy way to write the orthogonal complement.

*Lemma 43.* Take a maximal set  $\{\varphi_i\}_{i=1}^{|F|}$  of perfectly distinguishable pure states in  $F$  and extend it to a maximal set  $\{\varphi_i\}_{i=1}^{d_A}$  of perfectly distinguishable pure states in  $\text{St}_1(A)$ , then for  $|F| < d_A$  we have

$$\omega_F^\perp = \frac{1}{d_A - |F|} \sum_{i=|F|+1}^{d_A} \varphi_i.$$

*Proof.* By definition, for  $|F| < d_A$  we have  $\omega_F^\perp = \frac{1}{d_A - |F|} (d_A \chi_A - |F| \omega_F)$ . Substituting the expressions  $\chi_A = \frac{1}{d_A} \sum_{i=1}^{d_A} \varphi_i$  and  $\omega_F = \frac{1}{|F|} \sum_{i=1}^{|F|} \varphi_i$  we then obtain the thesis. ■

Note, however, that by definition the orthogonal complement  $\omega_F^\perp$  depends only on the face  $F$  and not on the choice of the maximal set in lemma 43.

An obvious consequence of lemma 43.

*Corollary 32.* The states  $\omega_F$  and  $\omega_F^\perp$  are perfectly distinguishable.

*Proof.* Take a maximal set  $\{\varphi_i\}_{i=1}^{|F|}$  of perfectly distinguishable pure states in  $F$ , extend it to a maximal set  $\{\varphi_i\}_{i=1}^{d_A}$ , and take the observation test such that  $(a_i|\varphi_j) = \delta_{ij}$ . Then the binary test  $\{a_F, e - a_F\}$ , defined by  $a_F := \sum_{i=1}^{|F|} a_i$  distinguishes perfectly between  $\omega_F$  and  $\omega_F^\perp$ . ■

We say that a state  $\tau \in \text{St}_1(A)$  is *perfectly distinguishable from the face  $F$*  if  $\tau$  is perfectly distinguishable from every state  $\sigma$  in the face  $F$ . With this definition we have the following.

*Lemma 44.* The following are equivalent:

- (1)  $\tau$  is perfectly distinguishable from the face  $F$ ,
- (2)  $\tau$  is perfectly distinguishable from  $\omega_F$ ,
- (3)  $\tau$  belongs to the face identified by  $\omega_F^\perp$ , that is,  $\tau \in F_{\omega_F^\perp}$ .

*Proof.* (1  $\Leftrightarrow$  2)  $\tau$  is perfectly distinguishable from  $\omega_F$  if and only if then there exists a binary test  $\{a, e - a\}$  such that  $(a|\tau) = 1$  and  $(a|\omega_F) = 0$ . By lemma 19 this is equivalent to the condition  $(a|\tau) = 1$  and  $a =_{\omega_F} 0$ , that is,  $\tau$  is distinguishable from any state  $\sigma$  in the face identified by  $\omega_F$ , which by definition is  $F$ . (2  $\rightarrow$  3) Let  $\{\varphi_i\}_{i=1}^{|F|}$  be a maximal set of perfectly distinguishable states in  $F$ ,  $\omega_F = \frac{1}{|F|} \sum_{i=1}^{|F|} \varphi_i$ , and let  $\{\varphi_i\}_{i=|F|+1}^k$  be the maximal set of perfectly distinguishable pure states in the spectral decomposition  $\tau = \sum_{i=|F|+1}^k p_i \varphi_i$ ,

with  $p_i > 0$  for every  $i = |F| + 1, \dots, k$ . Since  $\tau$  is perfectly distinguishable from  $\omega_F$ , by lemma 23 we have that the states  $\{\varphi_i\}_{i=1}^k$  are all perfectly distinguishable. Let us extend this set to a maximal set  $\{\varphi_i\}_{i=1}^{d_A}$ . By lemma 43 have  $\omega_F^\perp = \frac{1}{d_A - |F|} \sum_{i=|F|+1}^{d_A} \varphi_i$ . Hence, all the states  $\{\varphi_i\}_{i=|F|+1}^{d_A}$  are in the face  $F_{\omega_F^\perp}$ . Since  $\tau$  is a mixture of these states, it also belongs to the face  $F_{\omega_F^\perp}$ . (3  $\Rightarrow$  2). Since  $\omega_F$  and  $\omega_F^\perp$  are perfectly distinguishable, if  $\tau$  belongs to the face identified by  $\omega_F^\perp$ , then by lemma 22  $\tau$  is perfectly distinguishable from  $\omega_F$ . ■

*Corollary 33.* If  $\rho$  is perfectly distinguishable from  $\sigma$  and from  $\tau$ , then  $\rho$  is perfectly distinguishable from any convex mixture of  $\sigma$  and  $\tau$ .

*Proof.* Let  $F$  be the face identified by  $\rho$ . Then by lemma 44 we have  $\sigma, \tau \in F_{\omega_F^\perp}$ . Since  $F_{\omega_F^\perp}$  is a convex set, any mixture of  $\sigma$  and  $\tau$  belongs to it. By lemma 44, this means that any mixture of  $\sigma$  and  $\tau$  is perfectly distinguishable from  $\rho$ . ■

We are now ready to give the definition of orthogonal face.

*Definition 10 (Orthogonal face).* The *orthogonal face*  $F^\perp$  is the set of all states that are perfectly distinguishable from the face  $F$ .

By lemma 44 it is clear that  $F^\perp$  is the face identified by  $\omega_F^\perp$ , that is  $F^\perp = F_{\omega_F^\perp}$ .

In the following we list few elementary facts about orthogonal faces.

*Lemma 45.* The following properties hold

- (1)  $|F^\perp| = d_A - |F|$ ,
- (2)  $\chi_A = \frac{|F|}{d_A} \omega_F + \frac{|F^\perp|}{d_A} \omega_{F^\perp}$ ,
- (3)  $\omega_{F^\perp} = \omega_F^\perp$ ,
- (4)  $\omega_{F^\perp}^\perp = \omega_F$ ,
- (5)  $(F^\perp)^\perp = F$ .

*Proof. Item 1.* If  $|F| = d_A$  the thesis is obvious. If  $|F| < d_A$ , take a maximal set  $\{\varphi_i\}_{i=1}^{|F|}$  (or  $\{\varphi_j\}_{j=|F|+1}^{|F|+|F^\perp|}$ ) of perfectly distinguishable pure states in  $F$  (or  $F^\perp$ ). Hence we have

$$\omega_F = \frac{1}{|F|} \sum_{i=1}^{|F|} \varphi_i \quad \left( \text{or } \omega_{F^\perp} = \frac{1}{|F^\perp|} \sum_{j=|F|+1}^{|F|+|F^\perp|} \varphi_j \right).$$

By corollary 32 the states  $\omega_F$  and  $\omega_{F^\perp}$  are perfectly distinguishable. Hence the states  $\{\varphi_i\}_{i=1}^{|F|+|F^\perp|}$  are perfectly distinguishable jointly (lemma 23). Now we must have  $|F| + |F^\perp| = d_A$ , otherwise there would be a pure state  $\psi$  that is perfectly distinguishable from the states  $\{\varphi_i\}_{i=1}^{|F|+|F^\perp|}$ . This implies that  $\psi$  belongs to  $F^\perp$  and that states  $\{\psi\} \cup \{\varphi_j\}_{j=|F|+1}^{|F|+|F^\perp|}$  are perfectly distinguishable in  $F^\perp$ , in contradiction with the hypotheses that the set  $\{\varphi_j\}_{j=|F|+1}^{|F|+|F^\perp|}$  is maximal in  $F^\perp$ . *Item 2* Immediate from item 1 and definition 9. *Items 3 and 4* Both items follow by comparison of item 2 with Eq. (29). *Item 5* By condition 3 of lemma 44,  $(F^\perp)^\perp$  is the face identified by the state  $\omega_{F^\perp}^\perp$ , which, by item 4, is  $\omega_F$ . Since the face identified by  $\omega_F$  is  $F$ , we have  $(F^\perp)^\perp = F$ . ■

We now show that there is a canonical way to associate an effect  $a_F$  to a face  $F$ .

*Definition 11 (Effect associated to a face).* We say that  $a_F \in \text{Eff}(A)$  is the *effect associated to the face*  $F \subseteq \text{St}_1(A)$  if and only if  $a_F = \omega_F e$  and  $a_F = \omega_F^\perp 0$ .

In other words, the definition imposes that  $(a_F|\rho) = 1$  for every  $\rho \in F$  and  $(a_F|\sigma) = 0$  for every  $\sigma \in F^\perp$ .

*Lemma 46.* A state  $\rho \in \text{St}_1(A)$  belongs to the face  $F$  if and only if  $(a_F|\rho) = 1$ .

*Proof.* By definition, if  $\rho$  belongs to  $F$ , then  $(a_F|\rho) = 1$ . Conversely, if  $(a_F|\rho) = 1$ , then  $\rho$  is perfectly distinguishable from  $\omega_F^\perp$ , because  $(a_F|\omega_F^\perp) = 0$ . Now, we know that  $\omega_F^\perp$  is equal to  $\omega_{F^\perp}$  (item 4 of lemma 45). By item 2 of lemma 44 the fact that  $\rho$  is perfectly distinguishable from  $\omega_{F^\perp}$  implies that  $\rho$  belongs to  $(F^\perp)^\perp$ , which is just  $F$  (item 5 of lemma 45). ■

We now show that the effect  $a_F$  associated to the face  $F$  exists and is unique. A preliminary result needed to this purpose is the following.

*Lemma 47.* The effect  $a_F$  must have the form  $a_F = \sum_{i=1}^{|F|} a_i$ , where  $a_i$  is the atomic effect such that  $(a_i|\varphi_i) = 1$  and  $\{\varphi_i\}_{i=1}^{|F|}$  is a maximal set of perfectly distinguishable pure states in  $F$ .

*Proof.* By corollary 23 we have that  $a_F$  can be written as  $(a_F| = \sum_i d_i (a_i|$  where  $\{a_i\}_{i=1}^{d_A}$  is a perfectly distinguishing test. Moreover, since  $a_F$  is an effect, we must have  $d_i \geq 0$  for all  $i = 1, \dots, d_A$ . Now, by definition we have  $(a_F|\omega_F^\perp) = 0$ , which implies  $d_i (a_i|\omega_F^\perp) = 0$  for every  $i = 1, \dots, d_A$ , that is,  $(a_i|\omega_F^\perp) = 0$  whenever  $d_i \neq 0$ . Let us focus on the values of  $i$  for which  $d_i \neq 0$ . Let  $\varphi_i$  be the pure state such that  $(a_i|\varphi_i) = 1$ . The condition  $(a_i|\omega_F^\perp) = 0$  implies that  $\varphi_i$  is perfectly distinguishable from  $\omega_F^\perp$ . Therefore,  $\varphi_i$  belongs to  $(F^\perp)^\perp$ , which is  $F$ . Since by definition we must have  $(a_F|\varphi_i) = 1$ , this also implies that  $d_i = 1$ . In summary, we proved that  $a_F = \sum' a_i$  where the prime means that the sum is restricted to those values of  $i$  such that  $\varphi_i \in F$ . The condition  $a_F = \omega_F e$  also implies that the number of terms in the sum must be exactly  $|F|$ . The thesis is then proved by suitably relabelling the effects  $\{a_i\}_{i=1}^{d_A}$ , in such a way that  $\varphi_i$  belongs to  $F$  for every  $i = 1, \dots, |F|$ . ■

*Lemma 48.* The effect  $a_F$  associated to the face  $F$  is unique.

*Proof.* Suppose that  $a_F = \sum_{i=1}^{|F|} a_i$  and  $a'_F = \sum_{i=1}^{|F|} a'_i$  are two effects associated to the face  $F$ , both written as in lemma 47. Let  $\{\varphi_i\}_{i=1}^{|F|}$  (or  $\{\varphi'_i\}_{i=1}^{|F|}$ ) be the maximal set of perfectly distinguishable pure states in  $F$  such that  $(a_i|\varphi_i) = 1$  for every  $i = 1, \dots, |F|$  [or  $(a'_i|\varphi'_i) = 1$  for every  $i = 1, \dots, |F|$ ], and let  $\{\psi_j\}_{j=1}^{|F^\perp|}$  be a maximal set of perfectly distinguishable pure states in  $F^\perp$ . Since  $\omega_F$  and  $\omega_F^\perp$  are perfectly distinguishable, the states  $\{\varphi_i\}_{i=1}^{|F|} \cup \{\psi_j\}_{j=1}^{|F^\perp|}$  (or  $\{\varphi'_i\}_{i=1}^{|F|} \cup \{\psi_j\}_{j=1}^{|F^\perp|}$ ) are perfectly distinguishable (lemma 23). Moreover, the set is maximal since  $|F| + |F^\perp| = d_A$ . Let  $b_j$  be the atomic effect such that  $(b_j|\psi_j) = 1$ . Then, the test that distinguishes the states  $\{\varphi_i\}_{i=1}^{|F|} \cup \{\psi_j\}_{j=1}^{|F^\perp|}$  (or  $\{\varphi'_i\}_{i=1}^{|F|} \cup \{\psi_j\}_{j=1}^{|F^\perp|}$ ) is given by  $\{a_i\}_{i=1}^{|F|} \cup \{b_j\}_{j=1}^{|F^\perp|}$  (or  $\{a'_i\}_{i=1}^{|F|} \cup \{b_j\}_{j=1}^{|F^\perp|}$ ) and its normalization reads

$$e = \sum_{i=1}^{|F|} a_i + \sum_{j=1}^{|F^\perp|} b_j = a_F + \sum_{j=1}^{|F^\perp|} b_j,$$

$$e = \sum_{i=1}^{|F|} a'_i + \sum_{j=1}^{|F^\perp|} b_j = a'_F + \sum_{j=1}^{|F^\perp|} b_j.$$

By comparison we obtain  $a_F = a'_F$ . ■

### B. Projections

We are now in position to define the projection on a face.

*Definition 12 (Projection).* Let  $F$  be a face of  $\text{St}_1(\mathbf{A})$ . A *projection* on the face  $F$  is an atomic transformation  $\Pi_F$  such that

- (1)  $\Pi_F =_{\omega_F} \mathcal{S}_A$ ,
- (2)  $\Pi_F =_{\omega_F^\perp} 0$ .

When  $F$  is the face identified by a pure state  $\varphi \in \text{St}_1(\mathbf{A})$ , we have  $F = \{\varphi\}$  and call  $\Pi_{\{\varphi\}}$  a *projection on the pure state*  $\varphi$ .

The first condition in definition 12 means that the projection  $\Pi_F$  does not disturb the states in the face  $F$ . The second condition means that  $\Pi_F$  annihilates all states in the orthogonal face  $F^\perp$ . As a notation, we will indicate with  $\Pi_F^\perp$  the projection on the face  $F^\perp$ , that is, we will use the definition  $\Pi_F^\perp := \Pi_{F^\perp}$ .

An equivalent condition for  $\Pi_F$  to be a projection on the face  $F$  is the following.

*Lemma 49.* Let  $\{\varphi_i\}_{i=1}^{d_A}$  be a maximal set of perfectly distinguishable pure states for system  $\mathbf{A}$ . The transformation  $\Pi_F$  in  $\text{Transf}(\mathbf{A})$  is a projection on the face generated by the subset  $\{\varphi_i\}_{i=1}^{|F|}$  if and only if

- (1)  $\Pi_F =_{\omega_F} \mathcal{S}_A$ ,
- (2)  $\Pi_F|\varphi_l\rangle = 0$  for all  $l > |F|$ .

*Proof.* The condition is clearly necessary, since by definition  $12 \Pi_F|\varphi_l\rangle = 0$  for  $l > |F|$ . On the other hand, if  $\Pi_F|\varphi_l\rangle = 0$  for  $l > |F|$  then by definition of  $\omega_F^\perp$  we have  $\Pi_F|\omega_F^\perp\rangle = 0$  and, therefore,  $\Pi_F =_{\omega_F^\perp} 0$ . ■

A result that will be useful later.

*Lemma 50.* The transformation  $\Pi_F \otimes \mathcal{S}_B$  is a projection on the face  $\tilde{F}$  identified by the state  $\omega_F \otimes \chi_B$ .

*Proof.*  $\Pi_F \otimes \mathcal{S}_B$  is atomic, being the product of two atomic transformations. We now show that  $\Pi_F \otimes \mathcal{S}_B =_{\omega_F \otimes \chi_B} \mathcal{S}_A \otimes \mathcal{S}_B$ : Indeed, by the local tomography axiom it is easy to see that every state  $\sigma \in F_{\omega_F \otimes \chi_B}$  can be written as  $|\sigma\rangle = \sum_{i=1}^r \sum_{j=1}^{d_B} \sigma_{ij} |\alpha_i\rangle |\beta_j\rangle$ , where  $\{\alpha_i\}_{i=1}^r$  is a basis for  $\text{Span}(F)$  and  $\{\beta_j\}_{j=1}^{d_B}$  is a basis for  $\text{St}_1(\mathbf{B})$ . Since  $\Pi_F =_{\omega_F} \mathcal{S}_A$ , we have

$$\begin{aligned} |\sigma\rangle &= (\Pi_F \otimes \mathcal{S}_B) |\sigma\rangle \\ &= \sum_{i=1}^r \sum_{j=1}^{d_B} \sigma_{ij} \Pi_F |\alpha_i\rangle |\beta_j\rangle \\ &= \sum_{i=1}^r \sum_{j=1}^{d_B} \sigma_{ij} |\alpha_i\rangle |\beta_j\rangle \\ &= |\sigma\rangle, \end{aligned}$$

which implies  $\Pi_F \otimes \mathcal{S}_B =_{\omega_F \otimes \chi_B} \mathcal{S}_A \otimes \mathcal{S}_B$ . Finally, note that  $\omega_{\tilde{F}} = \omega_F \otimes \chi_B$ , while  $\omega_{\tilde{F}}^\perp = \omega_F^\perp \otimes \chi_B$ . Since we have  $(\Pi_F \otimes \mathcal{S}_B)|\omega_{\tilde{F}}^\perp\rangle = \Pi_F|\omega_F^\perp\rangle \otimes |\chi_B\rangle = 0$ , we can conclude  $\Pi_F \otimes \mathcal{S}_B =_{\omega_{\tilde{F}}^\perp} 0$ . Hence  $\Pi_F \otimes \mathcal{S}_B$  is a projection on  $\tilde{F}$ . ■

In the following we will show that for every face  $F$  there exists a unique projection  $\Pi_F$  and we will prove several properties of projections. Let us start from an elementary observation.

*Lemma 51.* Let  $\varphi$  be a pure state in the face  $F \subseteq \text{St}_1(\mathbf{A})$  and let  $a \in \text{Eff}(\mathbf{A})$  be the atomic effect such that  $(a|\varphi) = 1$ . If

$\mathcal{A} \in \text{Transf}(\mathbf{A})$  is an atomic transformation such that  $\mathcal{A} =_{\omega_F} \mathcal{S}_A$ , then  $(a|\mathcal{A} = (a|$ . Moreover, if  $a_F$  is the effect associated to the face  $F$ , then we have  $(a_F|\mathcal{A} = (a_F|$ .

*Proof.* By lemma 16, the effect  $(a|\mathcal{A}$  is atomic. Now, since  $\mathcal{A}|\varphi\rangle = |\varphi\rangle$ , we have  $(a|\mathcal{A}|\varphi\rangle = (a|\varphi) = 1$ . However, by theorem 8  $(a|$  is the unique atomic effect such that  $(a|\varphi) = 1$ . Hence,  $(a|\mathcal{A} = (a|$ . Moreover, writing  $a_F$  as  $a_F = \sum_{i=1}^{|F|} a_i$  with  $(a_i|\varphi_i) = 1$ ,  $\varphi_i \in F$  (lemma 47), we obtain  $(a_F|\mathcal{A} = \sum_{i=1}^{|F|} (a_i|\mathcal{A} = \sum_{i=1}^{|F|} (a_i| = (a_F|$ . ■

When applied to the case of projections, the above lemma gives the following.

*Corollary 34.* Let  $\varphi$  be a pure state in the face  $F \subseteq \text{St}_1(\mathbf{A})$  and let  $a \in \text{Eff}(\mathbf{A})$  be the atomic effect such that  $(a|\varphi) = 1$ . Then we have  $(a|\Pi_F = (a|$ . Moreover, if  $a_F$  is the effect associated to the face  $F$ , then we have  $(a_F| = (a_F|\Pi_F$ .

The counterpart of corollary 34 is given as follows.

*Lemma 52.* Let  $\psi$  be a pure state in the face  $F^\perp$  and let  $b$  be the atomic effect such that  $(b|\psi) = 1$ . Then, we have  $(b|\Pi_F = 0$ . Moreover, if  $a_F^\perp$  is the effect associated to the face  $F^\perp$ , then we have  $(a_F^\perp|\Pi_F = 0$ .

*Proof.* By lemma 16, the effect  $(b|\Pi_F$  is atomic. Hence  $(b|\Pi_F$  must be proportional to an atomic effect  $b'$  with  $\|b'\| = 1$ , for some proportionality constant  $\lambda \in [0, 1]$ , that is  $(b|\Pi_F = \lambda(b'|$ . We want to prove that  $\lambda$  is zero. By contradiction, suppose that  $\lambda \neq 0$ . Let  $\psi'$  be the pure state such that  $(b'|\psi') = 1$ . Now, since  $\Pi_F|\omega_F^\perp\rangle = 0$ , we have  $0 = (b|\Pi_F|\omega_F^\perp\rangle = \lambda(b'|\omega_F^\perp)$ , which implies  $(b'|\omega_F^\perp) = 0$ . Hence,  $\psi'$  is perfectly distinguishable from  $\omega_F^\perp$ , which in turn implies that  $\psi'$  belongs to  $(F^\perp)^\perp = F$ . We then have  $\lambda = (b|\Pi_F|\psi') = (b|\psi') = 0$  (the last equality follows from the fact that  $\psi$  and  $\psi'$  belong to  $F^\perp$  and  $F$ , respectively, and hence are perfectly distinguishable). This is in contradiction with the assumption  $\lambda \neq 0$ , thus concluding the proof that  $(b|\Pi_F = 0$ .

Moreover, writing  $a_F^\perp$  as  $a_F^\perp = \sum_{i=1}^{|F^\perp|} b_i$  with  $(b_i|\psi_i) = 1$ ,  $\psi_i \in F^\perp$ , we obtain  $(a_F^\perp|\Pi_F = \sum_{i=1}^{|F^\perp|} (b_i|\Pi_F = 0$ . ■

Combining corollary 34 and lemma 52 we obtain an important property of projections, expressed by the following.

*Corollary 35.* If  $\Pi_F$  is a projection on the face  $F$ , then one has  $(e_A|\Pi_F = (a_F|$ .

*Proof.* The thesis follows from corollary 34 and lemma 52 and from the fact that  $a_F + a_F^\perp = e$ . ■

In the following we will see that for every face  $F$  there exists a unique projection. To prove that, let us start from the existence.

*Lemma 53 (Existence of projections).* For every face  $F$  of  $\text{St}_1(\mathbf{A})$  there exists a projection  $\Pi_F$ .

*Proof.* By lemma 18, there exists a system  $\mathbf{B}$  and an atomic transformation  $\mathcal{A} \in \text{Transf}(\mathbf{A}, \mathbf{B})$  with  $(e_B|\mathcal{A} = (a_F|$ . Then, if  $\Psi_{\omega_F} \in \text{St}(\mathbf{AC})$  is a purification of  $\omega_F$ , we can define the state  $|\Sigma\rangle_{\text{BC}} := (\mathcal{A} \otimes \mathcal{I}_C)|\Psi_{\omega_F}\rangle_{\text{AC}}$ . By lemma 16  $\Sigma$  is a pure state. Moreover, the pure states  $\Sigma$  and  $\Psi_{\omega_F}$  have the same marginal on system  $\mathbf{C}$ : indeed, we have  $(e_B|\Sigma) = [(e_B|\mathcal{A})|\Psi_{\omega_F}] = (a_F|\Psi_{\omega_F})$  and, by definition,  $a_F =_{\omega_F} e_A$ , which by theorem 1 implies  $(a_F|\Psi_{\omega_F}) = (e_A|\Psi_{\omega_F})$ . If  $\varphi_0$  and  $\psi_0$  are two arbitrary pure states of  $\mathbf{A}$  and  $\mathbf{B}$ , respectively, the uniqueness of purification stated by postulate 1 implies that there exists a reversible channel  $\mathcal{U} \in \mathbf{G}_{\text{AB}}$

such that

$$(30)$$

Now, take the atomic effect  $b \in \text{Eff}(B)$  such that  $(b|\psi_0) = 1$ , and define the transformation  $\Pi_F \in \text{Transf}(A)$  as

Applying  $b$  on both sides of Eq. (30) we then obtain

$$(\Pi_F \otimes \mathcal{I}_C)|\Psi_{\omega_F}\rangle = |\Psi_{\omega_F}\rangle$$

and, therefore,  $\Pi_F =_{\omega_F} \mathcal{I}_A$ . Moreover, the transformation  $\Pi_F$  is atomic, being the composition of atomic transformations (lemma 16). Finally, we have  $\Pi_F =_{\omega_F^\perp} 0$ : indeed, by construction of  $\Pi_F$  we have

$$\begin{aligned} (e_A|\Pi_F|\rho) &= (e_A \otimes b|\mathcal{U}(\mathcal{S} \otimes \mathcal{I}_A)|\rho \otimes \varphi_0) \\ &\leq (e_A \otimes e_B|\mathcal{U}(\mathcal{S} \otimes \mathcal{I}_A)|\rho \otimes \varphi_0) \\ &= (e_A|\mathcal{S}|\rho) \\ &= (a_F|\rho). \end{aligned}$$

This implies  $(e_A|\Pi_F|\omega_F^\perp) = (a_F|\omega_F^\perp)0$  and, therefore,  $\Pi_F =_{\omega_F^\perp} 0$ . In conclusion,  $\Pi_F$  is the desired projection. ■

To prove the uniqueness of the projection  $\Pi_F$  we need two auxiliary lemmas, given in the following.

*Lemma 54.* Let  $\Phi \in \text{St}_1(A\bar{A})$  be a purification of the invariant state  $\chi_A$ , and let  $\Pi_F \in \text{Transf}(A)$  be a projection on the face  $F \subseteq \text{St}_1(A)$ . Then, the pure state  $\Phi_F \in \text{St}_1(A\bar{A})$  defined by

$$|\Phi_F\rangle := \frac{d_A}{|F|}(\Pi_F \otimes \mathcal{I}_{\bar{A}})|\Phi\rangle \quad (31)$$

is a purification of  $\omega_F$ .

*Proof.* The state  $\Phi_F$  is pure by lemma 16. Let us choose a maximal set of perfectly distinguishable pure states  $\{\varphi_i\}_{i=1}^{d_A}$  such that  $\{\varphi_i\}_{i=1}^{|F|}$  is maximal in  $F$ . Now, we have

$$\begin{aligned} (e_{\bar{A}}|\Phi_F)_{A\bar{A}} &= \frac{d_A}{|F|}[\Pi_F \otimes (e_{\bar{A}}|)|\Phi\rangle_{A\bar{A}}] \\ &= \frac{d_A}{|F|}\Pi_F|\chi_A\rangle \end{aligned}$$

having used the relation  $(e_{\bar{A}}|\Phi)_{A\bar{A}} = |\chi_A\rangle$  (corollary 16). We then obtain

$$\begin{aligned} (e_{\bar{A}}|\Phi_F)_{A\bar{A}} &= \frac{d_A}{|F|}\Pi_F|\chi_A\rangle \\ &= \frac{1}{|F|}\sum_{i=1}^{d_A}\Pi_F|\varphi_i\rangle \end{aligned}$$

$$\begin{aligned} &= \frac{1}{|F|}\sum_{i=1}^{|F|}|\varphi_i\rangle \\ &= |\omega_F\rangle \end{aligned}$$

having used that  $\chi_A = \sum_{i=1}^{d_A}\varphi_i/d_A$  (theorem 9), and the definition of  $\Pi_F$ . ■

*Lemma 55.* Let  $\Pi_F \in \text{Transf}(A)$  be a projection. A transformation  $\mathcal{C} \in \text{Transf}(A)$  satisfies  $\mathcal{C} =_{\omega_F} \mathcal{I}_A$  if and only if

$$\mathcal{C}\Pi_F = \Pi_F. \quad (32)$$

*Proof.* Let  $\Phi_F$  be the purification of  $\omega_F$  defined in lemma 54. Since  $\mathcal{C} =_{\omega_F} \mathcal{I}_A$ , we have  $(\mathcal{C} \otimes \mathcal{I})|\Phi_F\rangle = |\Phi_F\rangle$ . In other words, we have  $(\mathcal{C}\Pi_F \otimes \mathcal{I})|\Phi\rangle = (\Pi_F \otimes \mathcal{I})|\Phi\rangle$ . Since  $\Phi$  is dynamically faithful, this implies that  $\mathcal{C}\Pi_F = \Pi_F$ . Conversely, Eq. (32) implies that for  $\sigma \in F_{\omega_F}$ ,  $\mathcal{C}|\sigma\rangle = \mathcal{C}\Pi_F|\sigma\rangle = \Pi_F|\sigma\rangle = |\sigma\rangle$ , namely  $\mathcal{C} =_{\omega_F} \mathcal{I}_A$ . ■

*Theorem 14 (Uniqueness of projections).* The projection  $\Pi_F$  satisfying definition 12 is unique.

*Proof.* Let  $\Pi_F$  and  $\Pi'_F$  be two projections on the same face  $F$ , and define the pure states  $\Phi_F$  and  $\Phi'_F$  as in lemma 54. Now,  $\Phi_F$  and  $\Phi'_F$  are both purifications of the same state  $\tilde{\omega}_F \in \bar{A}$ : indeed, one has

$$\begin{aligned} (e_A|\Phi_F)_{A\bar{A}} &= \frac{d_A}{|F|}(e_A|\Pi_F)|\Phi\rangle_{A\bar{A}} \\ &= \frac{d_A}{|F|}(e_F|\Phi)_{A\bar{A}} \\ &= \frac{d_A}{|F|}(e_A|\Pi'_F)|\Phi\rangle_{A\bar{A}} \\ &= (e_A|\Phi'_F)_{A\bar{A}} \end{aligned}$$

having used the relation  $(e_A|\Pi_F = (a_F| = (e_A|\Pi'_F$ , which comes from corollary 34 and from the uniqueness of the effect  $a_F$  (lemma 48). By the uniqueness of purification, we have  $|\Phi'_F\rangle = (\mathcal{U} \otimes \mathcal{I}_{\bar{A}})|\Phi_F\rangle$  for some reversible transformation  $\mathcal{U} \in \mathbf{G}_A$ . This implies  $(\Pi'_F \otimes \mathcal{I}_{\bar{A}})|\Phi\rangle = (\mathcal{U}\Pi_F \otimes \mathcal{I}_{\bar{A}})|\Phi\rangle$ , and, since  $\Phi$  is dynamically faithful,  $\Pi'_F = \mathcal{U}\Pi_F$ . Since by definition 12 we have  $\Pi'_F =_{\omega_F} \mathcal{I}_A$  and  $\Pi_F =_{\omega_F} \mathcal{I}_A$ , we can conclude that  $\mathcal{U} =_{\omega_F} \mathcal{I}_A$ . Finally, using lemma 55 with  $\mathcal{C} = \mathcal{U}$  we obtain  $\Pi'_F = \mathcal{U}\Pi_F = \Pi_F$ . ■

We now show a few simple properties of projections. In the following, given a maximal set of perfectly distinguishable pure states  $\{\varphi_i\}_{i=1}^{d_A}$  and any subset  $V \subseteq \{1, \dots, d_A\}$  we define (with a slight abuse of notation)  $\omega_V := \sum_{i \in V} \varphi_i/|V|$ , and  $\Pi_V$  as the projection on the face  $F_V := F_{\omega_V}$ . We will refer to  $F_V$  as *the face generated by V*.

*Lemma 56.* For two arbitrary subsets  $V, W \subseteq \{1, \dots, d_A\}$  one has

$$\Pi_V \Pi_W = \Pi_{V \cap W}.$$

In particular, if  $V \cap W = \emptyset$  one has  $\Pi_V \Pi_W = 0$ .

*Proof.* First of all,  $\Pi_V \Pi_W$  is atomic, being the product of two atomic transformations. Moreover, since the face  $F_{V \cap W}$  is contained in the faces  $F_V$  and  $F_W$ , we have  $\Pi_V \Pi_W|\rho\rangle = \Pi_V|\rho\rangle = |\rho\rangle$  for every  $\rho \in F_{V \cap W}$ . In other words,  $\Pi_V \Pi_W =_{\omega_{V \cap W}} \mathcal{I}_A$ . Moreover, if  $l \notin V \cap W$  we have  $\Pi_V \Pi_W|\varphi_l\rangle = 0$ . By lemma 49 and by the uniqueness of projections (theorem 14) we then obtain that  $\Pi_V \Pi_W$  is the projection on the face generated by  $V \cap W$ . ■

*Corollary 36 (Idempotence).* Every projection  $\Pi_F$  satisfies the identity  $\Pi_F^2 = \Pi_F$ .

*Proof.* Consider a maximal set of perfectly distinguishable pure states  $\{\varphi_i\}_{i=1}^{d_A}$  such that  $\{\varphi_i\}_{i \in V}$  is maximal in  $F$ . In this way  $F$  is the face generated by  $V$ , and, therefore  $\Pi_F = \Pi_V$ . The thesis follows by taking  $V = W$  in lemma 56. ■

*Corollary 37.* For every state  $\rho \in \text{St}_1(\mathbb{A})$  such that  $\rho \notin F^\perp$ , the normalized state  $\rho'$  defined by

$$|\rho'\rangle = \frac{\Pi_F |\rho\rangle}{(e|\Pi_F|\rho)} \quad (33)$$

belongs to the face  $F$ .

*Proof.* By corollary 35, we have  $(e|\Pi_F = (a_F|$ . Since  $\rho \notin F^\perp$ , we must have  $(e|\Pi_F|\rho) = (a_F|\rho) > 0$ , and, therefore, the state  $\rho'$  in Eq. (33) is well defined. Moreover, using the definition of  $\rho'$  we obtain

$$\begin{aligned} (a_F|\rho') &= \frac{(a_F|\Pi_F|\rho)}{(e|\Pi_F|\rho)} \\ &= 1 \end{aligned}$$

having used corollaries 34 and 35 for the last equality. Finally, lemma 46 implies that  $\rho'$  belongs to the face  $F$ . ■

*Corollary 38.* Let  $\Pi_{\{\varphi\}}$  be the projection on the pure state  $\varphi \in \text{St}_1(\mathbb{A})$  and  $a$  be the atomic effect such that  $(a|\varphi) = 1$ . Then for every state  $\rho \in \text{St}_1(\mathbb{A})$  one has  $\Pi_{\{\varphi\}}|\rho\rangle = p|\varphi\rangle$  where  $p = (a|\rho)$ .

*Proof.* Recall that, by corollary 35, we have  $(a| = (e|\Pi_{\{\varphi\}}$ . If  $(a|\rho) = 0$  then clearly  $\Pi_{\{\varphi\}}|\rho\rangle = 0$ . Otherwise, the proof is a straightforward application of corollary 37. ■

We conclude the present subsection with a result that will be useful in the next subsection.

*Lemma 57.* An atomic transformation  $\mathcal{A} \in \text{Transf}(\mathbb{A})$  satisfies  $\mathcal{A} =_{\omega_F} \mathcal{I}_A$  if and only if

$$\Pi_F \mathcal{A} = \Pi_F. \quad (34)$$

*Proof.* Suppose that  $\mathcal{A} =_{\omega_F} \mathcal{I}_A$ . Let  $\Phi \in \text{St}_1(\mathbb{A}\tilde{\mathbb{A}})$  be a purification of the invariant state  $\chi_A$  and define the two pure states

$$\begin{aligned} |\Phi_F\rangle &:= \frac{d_A}{|F|} (\Pi_F \otimes \mathcal{I}_{\tilde{A}})|\Phi\rangle, \\ |\Phi'_F\rangle &:= \frac{d_A}{|F|} (\Pi_F \mathcal{A} \otimes \mathcal{I}_{\tilde{A}})|\Phi\rangle. \end{aligned}$$

Then we have

$$\begin{aligned} (e_A||\Phi'_F) &= [(a_F|\mathcal{A}||\Phi) \\ &= (a_F||\Phi) \\ &= (e_A||\Phi_F) \end{aligned}$$

having used the condition  $(a_F|\mathcal{A} = (a_F|$  (lemma 51). Now we proved that  $\Phi_F$  and  $\Phi'_F$  have the same marginal on system  $\tilde{\mathbb{A}}$ . By the uniqueness of purification, there exists a reversible transformation  $\mathcal{V} \in \mathbf{G}_A$  such that  $|\Phi'_F\rangle = (\mathcal{V} \otimes \mathcal{I}_{\tilde{A}})|\Phi_F\rangle$ . Since  $\Phi$  is dynamically faithful, this implies  $\Pi_F \mathcal{A} = \mathcal{V} \Pi_F$ .

Now, for every  $\rho$  in  $F$  one has  $\mathcal{V}|\rho\rangle = \mathcal{V}\Pi_F|\rho\rangle = \Pi_F \mathcal{A}|\rho\rangle = |\rho\rangle$ , namely  $\mathcal{V} =_{\omega_F} \mathcal{I}_A$ . Applying lemma 55 with

$\mathcal{C} = \mathcal{V} \Pi_F$  and using the idempotence of projections we then obtain

$$\begin{aligned} \Pi_F \mathcal{A} &= \mathcal{V} \Pi_F \\ &= (\mathcal{V} \Pi_F) \Pi_F \\ &= \Pi_F \Pi_F \\ &= \Pi_F. \end{aligned}$$

Conversely, suppose that Eq. (34) is satisfied. Let  $\varphi \in F$  be a pure state in  $F$  and  $a$  be the atomic effect such that  $(a|\varphi) = 1$ . Then, we have

$$(a|\mathcal{A}|\varphi) = (a|\Pi_F \mathcal{A}|\varphi) = (a|\Pi_F|\varphi) = (a|\varphi) = 1$$

having used the relation  $(a|\Pi_F = (a|$  (corollary 34). Then, by theorem 7  $\mathcal{A}\varphi = \varphi$ . Since  $\varphi \in F$  is arbitrary this implies  $\mathcal{A} =_{\omega_F} \mathcal{I}_A$ . ■

### C. Projection of a pure state on two orthogonal faces

In Sec. X we proved a number of results concerning two-dimensional systems. Some properties of two-dimensional systems will be extended to the case of generic systems using the following lemma.

*Lemma 58.* Consider a pure state  $\varphi \in \text{St}_1(\mathbb{A})$  and two complementary projections  $\Pi_F$  and  $\Pi_{F^\perp}$ . Then  $\varphi$  belongs to the face identified by the state  $|\theta\rangle := (\Pi_F + \Pi_{F^\perp})|\varphi\rangle$ .

*Proof.* If  $\Pi_F|\varphi\rangle = 0$  (or  $\Pi_{F^\perp}|\varphi\rangle = 0$ ), then there is nothing to prove: this means that  $\Pi_{F^\perp}|\varphi\rangle = |\varphi\rangle$  (or  $\Pi_F|\varphi\rangle = |\varphi\rangle$ ) and the thesis is trivially true. Suppose now that  $\Pi_F|\varphi\rangle \neq 0$  and  $\Pi_{F^\perp}|\varphi\rangle \neq 0$ . Using the notation  $\Pi_1 := \Pi_F$ ,  $\Pi_2 := \Pi_{F^\perp}$ , we can define the two pure states  $|\varphi_i\rangle := \Pi_i|\varphi\rangle / (e|\Pi_i|\varphi)$ ,  $i = 1, 2$ , and the probabilities  $p_i = (e|\Pi_i|\varphi)$ . In this way we have  $\Pi_i|\varphi\rangle = p_i|\varphi_i\rangle$  for  $i = 1, 2$  and  $\theta = p_1\varphi_1 + p_2\varphi_2$ . Taking the atomic effect  $(a_i|$  such that  $(a_i|\varphi_i) = 1$  we have  $a_{F_\theta} = a_1 + a_2$ , where  $a_{F_\theta}$  is the effect associated to the face  $F_\theta$ . Recalling that  $(a_i|\Pi_i = (a_i|$  for  $i = 1, 2$  (corollary 34), we then conclude the following:

$$\begin{aligned} (a_{F_\theta}|\varphi) &= [(a_1| + (a_2|)|\varphi) \\ &= (a_1|\Pi_1|\varphi) + (a_2|\Pi_2|\varphi) \\ &= \sum_{i=1,2} p_i (a_i|\varphi_i) = 1. \end{aligned}$$

Finally, lemma 46 yields  $\varphi \in F_\theta$ . ■

A consequence of lemma 58 is the following.

*Lemma 59.* Let  $\varphi \in \text{St}_1(\mathbb{A})$  be a pure state,  $a \in \text{Eff}(\mathbb{A})$  be the unique atomic effect such that  $(a|\varphi) = 1$ , and  $F$  be a face in  $\text{St}_1(\mathbb{A})$ . If  $\rho$  is perfectly distinguishable from  $\Pi_F|\varphi\rangle$  and from  $\Pi_{F^\perp}|\varphi\rangle$  then  $\rho$  is perfectly distinguishable from  $|\varphi\rangle$ . In particular, one has  $(a|\rho) = 0$ .

*Proof.* Since  $\rho$  is perfectly distinguishable from  $\Pi_F|\varphi\rangle$  and  $\Pi_{F^\perp}|\varphi\rangle$ , it is also perfectly distinguishable from any convex combination of them (corollary 33). Equivalently,  $\rho$  is perfectly distinguishable from the face  $F_\theta$  identified by  $|\theta\rangle := \Pi_F|\varphi\rangle + \Pi_{F^\perp}|\varphi\rangle$ . In particular, it must be perfectly distinguishable from  $\varphi$ , which belongs to  $F_\theta$  by virtue of lemma 58. If  $a$  is the atomic effect such that  $(a|\varphi) = 1$ , then by lemma 36 we have  $(a|\rho) = 0$ . ■

A technical result that will be useful in the following.

*Lemma 60.* Let  $\varphi \in \text{St}_1(\mathbb{A})$  be a pure state such that  $\Pi_F|\varphi\rangle \neq 0$  and  $\Pi_{\tilde{F}}^\perp|\varphi\rangle \neq 0$ . Define the pure states  $|\varphi_1\rangle := \Pi_F|\varphi\rangle/(e|\Pi_F|\varphi\rangle)$  and  $|\varphi_2\rangle := \Pi_{\tilde{F}}^\perp|\varphi\rangle/(e|\Pi_{\tilde{F}}^\perp|\varphi\rangle)$  and the mixed state  $|\theta\rangle := (\Pi_F + \Pi_{\tilde{F}}^\perp)|\varphi\rangle$ . Then, we have

$$\begin{aligned}\Pi_F \Pi_{F_\theta} &= \Pi_{\{\varphi_1\}}, \\ \Pi_{\tilde{F}}^\perp \Pi_{F_\theta} &= \Pi_{\{\varphi_2\}}.\end{aligned}$$

*Proof.* Let  $\{\psi_i\}_{i=1}^{|F|}$  be a maximal set of perfectly distinguishable pure states in  $F$ , chosen in such a way that  $\psi_1 = \varphi_1$ , and let  $\{\psi_i\}_{i=|F|+1}^{d_A}$  be a maximal set of perfectly distinguishable pure states in  $F^\perp$ , chosen in such a way that  $\psi_{|F|+1} = \varphi_2$ . Defining the sets  $V := \{1, \dots, |F|\}$ ,  $W := \{|F| + 1, \dots, d_A\}$ , and  $U := \{1, |F| + 1\}$  we then have  $\Pi_V = \Pi_F$ ,  $\Pi_W = \Pi_{\tilde{F}}^\perp$ , and  $\Pi_U = \Pi_{F_\theta}$ . Using lemma 56 we obtain

$$\begin{aligned}\Pi_F \Pi_\theta &= \Pi_V \Pi_U \\ &= \Pi_{V \cap U} \\ &= \Pi_{\{\psi_1\}} \\ &= \Pi_{\{\varphi_1\}}\end{aligned}$$

and

$$\begin{aligned}\Pi_{\tilde{F}}^\perp \Pi_\theta &= \Pi_W \Pi_U \\ &= \Pi_{W \cap U} \\ &= \Pi_{\{\psi_{|F|+1}\}} \\ &= \Pi_{\{\varphi_2\}}\end{aligned} \quad \blacksquare$$

We conclude this subsection with an important observation about the group of reversible transformations that act as the identity on two orthogonal faces  $F$  and  $F^\perp$ . If  $F$  is a face of  $\text{St}_1(\mathbb{A})$ , let us define  $\mathbf{G}_{F, F^\perp}$  as the group of all reversible transformations  $\mathcal{U} \in \mathbf{G}_A$  such that

$$\mathcal{U} =_{\omega_F} \mathcal{S}_A, \quad \mathcal{U} =_{\omega_{\tilde{F}}^\perp} \mathcal{S}_A.$$

Then we have the following.

*Theorem 15.* For every face  $F \subset \text{St}_1(\mathbb{A})$  such that  $F \neq \{0\}$  and  $F \neq \text{St}_1(\mathbb{A})$ , the group  $\mathbf{G}_{F, F^\perp}$  is topologically equivalent to a circle.

*Proof.* Let  $\mathcal{U}$  be a transformation in  $\mathbf{G}_{F, F^\perp}$ ,  $\Phi \in \text{St}(\mathbb{A}\tilde{\mathbb{A}})$  be a purification of the invariant state  $\chi_A$  and  $|\Phi_{\mathcal{U}}\rangle := (\mathcal{U} \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Phi\rangle$  be the Choi state of  $\mathcal{U}$ . Define the orthogonal faces  $\tilde{F} := F_{\omega_F \otimes \chi_A}$  and  $\tilde{F}^\perp := F_{\omega_{\tilde{F}}^\perp \otimes \chi_A}$ , and the projections  $\Pi_{\tilde{F}} := \Pi_F \otimes \mathcal{S}_{\tilde{\mathbb{A}}}$  and  $\Pi_{\tilde{F}}^\perp := \Pi_{\tilde{F}}^\perp \otimes \mathcal{S}_{\tilde{\mathbb{A}}}$  (see lemma 50). Using lemma 57 we then obtain

$$\begin{aligned}\Pi_{\tilde{F}}|\Phi_{\mathcal{U}}\rangle &= (\Pi_F \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Phi_{\mathcal{U}}\rangle \\ &= (\Pi_F \mathcal{U} \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Phi\rangle \\ &= (\Pi_F \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Phi\rangle \\ &= \frac{|F|}{d_A} |\Phi_F\rangle\end{aligned}$$

and, similarly,

$$\begin{aligned}\Pi_{\tilde{F}}^\perp|\Phi_{\mathcal{U}}\rangle &= (\Pi_{\tilde{F}}^\perp \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Phi_{\mathcal{U}}\rangle \\ &= (\Pi_{\tilde{F}}^\perp \mathcal{U} \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Phi\rangle \\ &= (\Pi_{\tilde{F}}^\perp \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Phi\rangle \\ &= \frac{|F^\perp|}{d_A} |\Phi_{F^\perp}\rangle.\end{aligned}$$

This means that the projections of  $\Phi_{\mathcal{U}}$  on the faces  $\tilde{F}$  and  $\tilde{F}^\perp$  are independent of  $\mathcal{U}$ . Also, it means that  $\Phi_{\mathcal{U}}$  belongs to the face  $F_\theta$  identified by the state  $|\theta\rangle := \frac{|F|}{d_A} |\Phi_F\rangle + \frac{|F^\perp|}{d_A} |\Phi_{F^\perp}\rangle$  (lemma 58). Now, by the compression axiom,  $F_\theta$  is isomorphic to the state space of a qubit, say with  $\Phi_F$  and  $\Phi_{F^\perp}$  indicating the north and south poles of the Bloch sphere, respectively, and we know that all the Choi states  $\{\Phi_{\mathcal{U}}\}_{\mathcal{U} \in \mathbf{G}_{F, F^\perp}}$  are at the same latitude [precisely, the latitude is the angle  $\zeta$  given by  $\cos \zeta = (|F| - |F^\perp|)/d_A$ ]. This implies that the states  $\{\Phi_{\mathcal{U}}\}_{\mathcal{U} \in \mathbf{G}_{F, F^\perp}}$  are a subset of a circle  $\mathbf{C}_\zeta$  in the Bloch sphere describing the face  $F_\theta$ . Precisely, the circle  $\mathbf{C}_\zeta$  is given by

$$\begin{aligned}\mathbf{C}_\zeta := \left\{ \Psi \in F_\theta \mid \Pi_{\{\Phi_F\}}|\Psi\rangle = \frac{|F|}{d_A} |\Phi_F\rangle, \right. \\ \left. \Pi_{\{\Phi_{F^\perp}\}}|\Psi\rangle = \frac{|F^\perp|}{d_A} |\Phi_{F^\perp}\rangle \right\}.\end{aligned}$$

We now prove that in fact they are the whole circle. Let  $\Psi$  be a state in  $\mathbf{C}_\zeta$ . Since  $|\Psi\rangle$  belongs to the face  $F_\theta$ , we obtain

$$\begin{aligned}(\Pi_F \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Psi\rangle &= \Pi_{\tilde{F}}|\Psi\rangle \\ &= \Pi_{\tilde{F}} \Pi_{F_\theta} |\Psi\rangle \\ &= \Pi_{\{\Phi_F\}} |\Psi\rangle \\ &= \frac{|F|}{d_A} |\Phi_F\rangle\end{aligned}$$

(the third equality comes from lemma 60 with the substitutions  $F \rightarrow \tilde{F}$ ,  $\varphi \rightarrow \Psi$ ,  $\varphi_1 \rightarrow \Phi_F$ , and  $\varphi_2 \rightarrow \Phi_{F^\perp}$ ) and, similarly,

$$\begin{aligned}(\Pi_{\tilde{F}}^\perp \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Psi\rangle &= \Pi_{\tilde{F}}^\perp |\Psi\rangle \\ &= \Pi_{\tilde{F}}^\perp \Pi_{F_\theta} |\Psi\rangle \\ &= \Pi_{\{\Phi_{F^\perp}\}} |\Psi\rangle \\ &= \frac{|F^\perp|}{d_A} |\Phi_{F^\perp}\rangle.\end{aligned}$$

Therefore, we have

$$\begin{aligned}(e_A|\Psi\rangle) &= [(a_F| + (a_{\tilde{F}}^\perp|)]|\Psi\rangle \\ &= [(e_A| \Pi_F \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Psi\rangle + [(e_A| \Pi_{\tilde{F}}^\perp \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Psi\rangle] \\ &= \frac{|F|}{d_A} (e_A|\Phi_F\rangle) + \frac{|F^\perp|}{d_A} (e_A|\Phi_{F^\perp}\rangle) \\ &= [(e_A| \Pi_F \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Phi\rangle + [(e_A| \Pi_{\tilde{F}}^\perp \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Phi\rangle] \\ &= [(a_F| + (a_{\tilde{F}}^\perp|)]|\Phi\rangle \\ &= (e_A|\Phi\rangle) \\ &= |\chi_{\tilde{\mathbb{A}}}\rangle.\end{aligned}$$

Since  $\Psi$  and  $\Phi$  are both purifications of the invariant state  $\chi_{\tilde{\mathbb{A}}}$ , by the uniqueness of purification there must be a reversible transformation  $\mathcal{U} \in \mathbf{G}_A$  such that  $|\Psi\rangle = (\mathcal{U} \otimes \mathcal{S}_{\tilde{\mathbb{A}}})|\Phi\rangle$ . Finally, it is easy to check that  $\Pi_F \mathcal{U} = \Pi_F$  and  $\Pi_{\tilde{F}}^\perp \mathcal{U} = \Pi_{\tilde{F}}^\perp$ , which, by lemma 57 implies  $\mathcal{U} =_{\omega_F} \mathcal{S}_A$  and  $\mathcal{U} =_{\omega_{\tilde{F}}^\perp} \mathcal{S}_A$ . This proves that the Choi states  $\{\Phi_{\mathcal{U}}\}_{\mathcal{U} \in \mathbf{G}_{F, F^\perp}}$  are the whole circle  $\mathbf{C}_\zeta$ . Since the Choi isomorphism is continuous in the operational norm (see theorem 14 of [22]), the group  $\mathbf{G}_{F, F^\perp}$  is topologically equivalent to a circle.  $\blacksquare$

## XII. THE SUPERPOSITION PRINCIPLE

The validity of the superposition principle, proved for two-dimensional systems using the geometry of the Bloch sphere (corollary 31), can be now extended to arbitrary systems thanks to lemma 58.

*Theorem 16 (Superposition principle for general systems).* Let  $\{\varphi_i\}_{i=1}^{d_A} \subseteq \text{St}_1(A)$  be a maximal set of perfectly distinguishable pure states and  $\{a_i\}_{i=1}^{d_A}$  be the observation test such that  $(a_i|\varphi_j) = \delta_{ij}$ . Then, for every choice of probabilities  $\{p_i\}_{i=1}^{d_A}$ ,  $p_i \geq 0$ ,  $\sum_{i=1}^{d_A} p_i = 1$  there exists at least one pure state  $\varphi_p \in \text{St}_1(A)$  such that

$$p_i = (a_i|\varphi_p) \quad \forall i = 1, \dots, d_A \quad (35)$$

or, equivalently,

$$\Pi_{\{\varphi_i\}}|\varphi_p) = p_i |\varphi_i) \quad \forall i = 1, \dots, d_A, \quad (36)$$

where  $\Pi_{\{\varphi_i\}}$  is the projection on  $\varphi_i$ .

*Proof.* Let us first prove the equivalence between Eqs. (35) and (36). From Eq. (36) we obtain Eq. (35) using the relation  $(e|\Pi_{\{i\}} = (a_i|$ , which follows from corollary 35. Conversely, from Eq. (35) we obtain Eq. (36) using corollary 38. Now, we will prove Eq. (35) by induction. The statement for  $N = 2$  is proved by corollary 31. Assume that the statement holds for every system B of dimension  $d_B = N$  and suppose that  $d_A = N + 1$ . Let  $F$  be the face identified by  $\omega_F = 1/N \sum_{i=1}^N \varphi_i$  and  $F^\perp$  be the orthogonal face, identified by the state  $\varphi_{N+1}$ . Now there are two cases: either  $p_{N+1} = 1$  or  $p_{N+1} \neq 1$ . If  $p_{N+1} = 1$ , then there is nothing to prove: the desired state is  $\varphi_{N+1}$ . Then, suppose that  $p_{N+1} \neq 1$ . Using the induction hypothesis and the compression axiom 3 we can find a state  $\psi_q \in F$  such that  $(a_i|\psi_q) = q_i$ , with  $q_i = p_i/(1 - p_{N+1})$ ,  $i = 1, \dots, N$ . Let us then define a new maximal set of perfectly distinguishable pure states  $\{\varphi'_i\}_{i=1}^{N+1}$ , with  $\varphi'_1 = \psi_q$  and  $\varphi'_{N+1} = \varphi_{N+1}$ . Note that one has  $\omega_F = 1/N \sum_{i=1}^N \varphi'_i$ , that is,  $F$  is the face generated by the states  $\{\varphi'_i\}_{i=1}^N$ . Now consider the two-dimensional face  $F'$  identified by  $\theta = 1/2(\varphi'_1 + \varphi'_{N+1})$ . By corollary 31 (superposition principle for qubits) we know that there exists a pure state  $\varphi \in F'$  with  $(a'_1|\varphi) = 1 - p_{N+1}$  and  $(a'_{N+1}|\varphi) = p_{N+1}$ . Let us define  $V := \{1, \dots, N\}$  and  $W := \{1, N+1\}$ . Then, we have  $\Pi_F = \Pi_V$  and  $\Pi_{F'} = \Pi_W$ , and by lemma 56,

$$\begin{aligned} \Pi_F|\varphi) &= \Pi_F\Pi_{F'}|\varphi) \\ &= \Pi_{V \cap W}|\varphi) \\ &= \Pi_{\{\varphi'_i\}}|\varphi) \\ &= \Pi_{\{\psi_q\}}|\varphi) \\ &= (1 - p_{N+1})|\psi_q) \end{aligned}$$

having used corollary 38 for the last equality. Finally, for  $i = 1, \dots, N$  we have

$$\begin{aligned} (a_i|\varphi) &= (a_i|\Pi_F|\varphi) \\ &= (1 - p_{N+1})(a_i|\psi_q) \\ &= (1 - p_{N+1})q_i \\ &= p_i. \end{aligned}$$

On the other hand we have  $(a_{N+1}|\varphi) = (a'_{N+1}|\varphi) = p_{N+1}$ . ■

## A. Completeness for purification

Using the superposition principle and the spectral decomposition of theorem 10 we can now show that every state of system A has a purification in AB provided  $d_B \geq d_A$ :

*Lemma 61.* For every state  $\rho \in \text{St}_1(A)$  and for every system B with  $d_B \geq d_A$  there exists a purification of  $\rho$  in  $\text{St}_1(AB)$ .

*Proof.* Take the spectral decomposition of  $\rho$ , given by  $\rho = \sum_{i=1}^{d_A} p_i \varphi_i$ , where  $\{p_i\}$  are probabilities and  $\{\varphi_i\}_{i=1}^{d_A} \subseteq \text{St}_1(A)$  is a maximal set of perfectly distinguishable pure states. Let  $\{\psi_i\}_{i=1}^{d_B}$  be a maximal set of perfectly distinguishable pure states and  $\{a_i\}_{i=1}^{d_A} \subseteq \text{Eff}(A)$  [or  $\{b_i\}_{i=1}^{d_B} \subseteq \text{Eff}(B)$ ] be the test such that  $(a_i|\varphi_j) = \delta_{ij}$  [or  $(b_i|\psi_j) = \delta_{ij}$ ]. Clearly  $\{\varphi_i \otimes \psi_j\}$  is a maximal set of perfectly distinguishable pure states for AB. Then, by the superposition principle (theorem 16) there exists a pure state  $\Psi_\rho$  such that  $(a_i \otimes b_j|\Psi_\rho) = p_i \delta_{ij}$ . Equivalently, we have  $(b_i|_B|\Psi_\rho)_{AB} = p_i |\varphi_i)_A$  for every  $i = 1, \dots, d_A$  and  $(b_i|_B|\Psi_\rho)_{AB} = 0$  for  $i > d_A$ . Summing over  $i$  we then obtain  $(e|_B|\Psi_\rho)_{AB} = \sum_{i=1}^{d_B} (b_i|_B|\Psi_\rho)_{AB} = \sum_{i=1}^{d_A} p_i |\varphi_i)_A = |\rho)_A$ . ■

In the terminology of Ref. [22], lemma 61 states that a system B with  $d_B \geq d_A$  is complete for the purification of system A.

As a consequence of lemma 61 we have the following.

*Corollary 39.* Every system B with  $d_B = d_A$  is operationally equivalent to the conjugate system  $\tilde{A}$ .

*Proof.* By corollary 61, the invariant state  $\chi_A \in \text{St}_1(A)$  has a purification  $\Psi$  in  $\text{St}_1(AB)$ . By corollary 18, the marginal of  $\Psi$  on B is the invariant state  $\chi_B$ . By definition, this means that B is a conjugate system of A. Since the conjugate system  $\tilde{A}$  is unique up to operational equivalence (corollary 16), this implies the thesis. ■

## B. Equivalence of systems with equal dimension

We are now in position to prove that two systems A and B with the same dimension are operationally equivalent, namely that there is a reversible transformation from A to B. In other words, we prove that the informational dimension classifies the systems of our theory up to operational equivalence. The fact that this property is derived from the principles, rather than being assumed from the start, is one of the important differences of our work with respect to Refs. [16–18]. Another difference is that here the equivalence of systems with the same dimension is proved after the derivation of the qubit, whereas in Refs. [16–18] the derivation of the qubit requires the equivalence of systems with the same dimension.

*Corollary 40 (Operational equivalence of systems with equal dimension).* Every two systems A and B with  $d_A = d_B$  are operationally equivalent.

*Proof.* By corollary 39, A and B are both operationally equivalent to the conjugate system  $\tilde{A}$ . Hence they are operationally equivalent to each other. ■

## C. Reversible operations of perfectly distinguishable pure states

An important consequence of the superposition principle is the possibility of transforming an arbitrary maximal set of perfectly distinguishable pure states into another via a reversible transformation:

*Corollary 41.* Let A and B be two systems with  $d_A = d_B =: d$  and let  $\{\varphi_i\}_{i=1}^d$  (or  $\{\psi_i\}_{i=1}^d$ ) be a maximal set of perfectly distinguishable pure states in A (or B). Then, there exists a reversible transformation  $\mathcal{U} \in \text{Transf}(A, B)$  such that  $\mathcal{U}|\varphi_i\rangle = |\psi_i\rangle$ .

*Proof.* Let  $\Phi \in \text{St}(A\tilde{A})$  be a purification of the invariant state  $\chi_A$ . Although we know that A and  $\tilde{A}$  are operationally equivalent (corollary 39) we use the notation A and  $\tilde{A}$  to distinguish between the two subsystems of  $A\tilde{A}$ . Define the pure state  $\tilde{\varphi}_i$  via the relation  $(a_i|_A|\Phi)_{A\tilde{A}} = \frac{1}{d}|\tilde{\varphi}_i\rangle_{\tilde{A}}$ , where  $\{a_i\}_{i=1}^d$  is the observation test such that  $(a_i|\varphi_i) = \delta_{ij}$ . Let  $\{\tilde{a}_i\}_{i=1}^d$  be the observation test such that  $(\tilde{a}_i|\tilde{\varphi}_j) = \delta_{ij}$ . Then, by lemma 30 we have

$$(\tilde{a}_i|_{\tilde{A}}|\Phi)_{A\tilde{A}} = \frac{1}{d}|\varphi_i\rangle_A. \quad (37)$$

On the other hand, if  $\{b_i\}_{i=1}^d$  is the observation test such that  $(b_i|\psi_j) = \delta_{ij}$ , then using the superposition principle (theorem 16) we can construct a state  $\Psi \in \text{St}_1(B\tilde{A})$  such that  $(b_i|\Psi) = \delta_{ij}/d$ , or, equivalently,

$$(\tilde{a}_i|_{\tilde{A}}|\Psi)_{B\tilde{A}} = \frac{1}{d}|\psi_i\rangle_B. \quad (38)$$

Now,  $\Phi$  and  $\Psi$  have the same marginal on system  $\tilde{A}$ : they are both purifications of the invariant state  $\chi_{\tilde{A}}$ . Moreover, A and B are operationally equivalent because they have the same dimension (corollary 40). Hence, by the uniqueness of purification, there must be a reversible transformation  $\mathcal{U} \in \text{Transf}(A, B)$  such that

$$|\Psi\rangle_{B\tilde{A}} = (\mathcal{U} \otimes \mathcal{I}_{\tilde{A}})|\Phi\rangle_{A\tilde{A}}. \quad (39)$$

Combining Eqs. (37), (38), and (39) we finally obtain

$$\begin{aligned} \frac{1}{d}\mathcal{U}|\varphi_i\rangle_A &= [\mathcal{U} \otimes (\tilde{a}_i|_{\tilde{A}})|\Phi]_{B\tilde{A}} \\ &= (\tilde{a}_i|_{\tilde{A}}|\Psi)_{B\tilde{A}} \\ &= \frac{1}{d}|\psi_i\rangle_B, \end{aligned}$$

that is,  $\mathcal{U}|\varphi_i\rangle = |\psi_i\rangle$  for every  $i = 1, \dots, d$ . ■

### XIII. DERIVATION OF THE DENSITY MATRIX FORMALISM

The goal of this section is to show that our set of axioms implies that

- (1) the set of states for a system A of dimension  $d_A$  is the set of density matrices on the Hilbert space  $\mathbb{C}^{d_A}$ ,
- (2) the set of effects is the set of positive matrices bounded by the identity, and
- (3) the pairing between a state and an effect is given by the trace of the product of the corresponding matrices.

Using the result of theorem 3, we will then obtain that all the physical transformations in our theory are exactly the physical transformations allowed in quantum mechanics. This will conclude our derivation of quantum theory.

#### A. The basis

In order to specify the correspondence between states and matrices we choose a particular basis for the vector space

$\text{St}_{\mathbb{R}}(A)$ . For this purpose, we adopt the choice of basis used in Ref. [16]. The basis is constructed as follows: Let us first choose a maximal set of  $d_A$  perfectly distinguishable states  $\{\varphi_m\}_{m=1}^{d_A}$ , and declare that they are the first  $d_A$  basis vectors. Then, for every  $m < n$  the face  $F_{mn}$  generated by  $\{\varphi_m, \varphi_n\}$  defines a ‘‘two-dimensional subsystem’’: precisely, the face  $F_{mn} := F_{\omega_{mn}}$  with  $\omega_{mn} := \frac{\varphi_m + \varphi_n}{2}$  can be ideally encoded in a two-dimensional system. Now, the convex set of states of a two-dimensional system is the Bloch sphere, and we can choose the  $z$  axis to be the line joining the two states  $\{\varphi_m, \varphi_n\}$ , for example, with the positive direction of the  $z$  axis being the direction from  $\varphi_m$  to  $\varphi_n$ . Once the direction of the  $z$  axis has been specified, we can choose the  $x$  and  $y$  axes. Note that any couple of orthogonal directions in the plane orthogonal to  $z$  axis is a valid choice for the  $x$  and  $y$  axes (here we do not restrict ourselves to the choice of a right-handed coordinate system). At the moment there is no relation among the different choices of axes made for different values of  $m$  and  $n$ . However, to prove that the states are represented by positive matrices, later we will have to find a suitable way of connecting all these choices of axes.

Let  $\varphi_{x,+}^{mn}, \varphi_{x,-}^{mn} \in F_{mn}$  ( $\varphi_{y,+}^{mn}, \varphi_{y,-}^{mn} \in F_{mn}$ ) be the two perfectly distinguishable states in the direction of the  $x$  axis ( $y$  axis) and define

$$\sigma_k^{mn} := \varphi_{k,+}^{mn} - \varphi_{k,-}^{mn}, \quad k = x, y. \quad (40)$$

An immediate observation is the following.

*Lemma 62.* The four vectors  $\{\varphi_m, \varphi_n, \sigma_x^{mn}, \sigma_y^{mn}\} \subseteq \text{St}_{\mathbb{R}}(A)$  are linearly independent.

*Proof.* Linear independence is evident from the geometry of the Bloch sphere. ■

We now show that the collection of all vectors obtained in this way is a basis for  $\text{St}_{\mathbb{R}}(A)$ . To this purpose we use the following.

*Lemma 63.* Let  $V \subset \{1, \dots, d_A\}$ , and consider the projection  $\Pi_V$ . Then, for  $m \in V$  and  $n \notin V$ , one has  $\Pi_V|\sigma_k^{mn}\rangle = 0$  for  $k = x, y$ .

*Proof.* Using lemma 56 and corollary 38 we obtain

$$\begin{aligned} \Pi_V|\varphi_{k,\pm}^{mn}\rangle &= \Pi_V\Pi_{\{m,n\}}|\varphi_{k,\pm}^{mn}\rangle \\ &= \Pi_{\{m\}}|\varphi_{k,\pm}^{mn}\rangle \\ &= |\varphi_m\rangle(a_m|\varphi_{k,\pm}^{mn}). \end{aligned}$$

Since the face  $F_{mn}$  is isomorphic to the Bloch sphere and the state since  $\varphi_{k,\pm}^{mn}, k = x, y$  lie on the equator of the Bloch sphere, we know that  $(a_m|\varphi_{k,\pm}^{mn}) = \frac{1}{2}$ . This implies

$$\begin{aligned} \Pi_V|\sigma_k^{mn}\rangle &= \Pi_V(|\varphi_{k,+}^{mn}\rangle - |\varphi_{k,-}^{mn}\rangle) \\ &= |\varphi_m\rangle\left(\frac{1}{2} - \frac{1}{2}\right) = 0. \end{aligned}$$

■  
*Lemma 64.* The vectors  $\{\varphi_n\}_{n=1}^{d_A} \cup \{\sigma_k^{mn}\}_{n>m=1, \dots, d_A, k=x, y}$  form a basis for  $\text{St}_{\mathbb{R}}(A)$ .

*Proof.* Since the number of vectors is exactly  $d_A^2$ , to prove that they form a basis it is enough to show that they are linearly independent. Suppose that there exists a vector of coefficients

$\{c_m\} \cup \{c_k^{mn}\}$  such that

$$\sum_m c_m \varphi_m + \sum_{n>m, k=x,y} c_k^{mn} \sigma_k^{mn} = 0.$$

Applying the projection  $\Pi_{\{m,n\}}$  on both sides and using lemma 63 we obtain

$$c_m |\varphi_m\rangle + c_n |\varphi_n\rangle + c_{mn}^x |\sigma_x^{mn}\rangle + c_{mn}^y |\sigma_y^{mn}\rangle = 0.$$

However, from lemma 62 we know that the vectors  $\{\varphi_m, \varphi_n, \sigma_x^{mn}, \sigma_y^{mn}\}$  are linearly independent. Consequently,  $c_m = c_n = c_{mn}^k = 0$  for all  $m, n, k$ . ■

### B. The matrices

Since the state space  $\text{St}(A)$  for system  $A$  spans a real vector space of dimension  $D_A = d_A^2$ , we can decide to represent the vectors  $\{\varphi_m\}_{m=1}^{d_A} \cup \{\sigma_k^{mn}\}_{n>m=1, \dots, d_A, k=x,y}$  as Hermitian  $d_A \times d_A$  matrices. Precisely, we associate the vector  $\varphi_m$  to the matrix  $S_{\varphi_m}$  defined by

$$[S_{\varphi_m}]_{rs} = \delta_{rm} \delta_{sm}, \quad (41)$$

the vector  $\sigma_x^{mn}$  to the matrix

$$[S_{\sigma_x^{mn}}]_{rs} = \delta_{rm} \delta_{sn} + \delta_{rn} \delta_{sm} \quad (42)$$

and the vector  $\sigma_y^{mn}$  to the matrix

$$[S_{\sigma_y^{mn}}]_{rs} = i\lambda (\delta_{rm} \delta_{sn} - \delta_{rn} \delta_{sm}), \quad (43)$$

where  $\lambda$  can take the values  $+1$  or  $-1$ . The freedom in the choice of  $\lambda$  will be useful in Sec. [XIII C](#), where we will introduce the representation of composite systems of two qubits. However, this choice of sign plays no role in the present subsection, and for simplicity we will take the positive sign.

Recall that in principle any orthogonal direction in the plane orthogonal to the  $z$  axis can be chosen to be the  $x$  axis. In general, the other possible choices for the  $x$  axis will lead to matrices of the form

$$[S_{\sigma_{x,\theta}^{mn}}]_{rs} = \delta_{rm} \delta_{sn} e^{i\theta} + \delta_{rn} \delta_{sm} e^{-i\theta}, \quad \theta \in [0, 2\pi), \quad (44)$$

and the corresponding choice for the  $y$  axis will lead to a matrices of the form

$$[S_{\sigma_{y,\theta}^{mn}}]_{rs} = i\lambda (\delta_{rm} \delta_{sn} e^{i\theta} - \delta_{rn} \delta_{sm} e^{-i\theta}), \quad \theta \in [0, 2\pi). \quad (45)$$

Since the vectors  $\{\varphi_m\}_{m=1}^{d_A} \cup \{\sigma_k^{mn}\}_{n>m=1, \dots, d_A; k=x,y}$  are a basis for the real vector space  $\text{St}_{\mathbb{R}}(A)$ , we can expand any state  $\rho \in \text{St}(A)$  on them:

$$|\rho\rangle = \sum_m \rho_m |\varphi_m\rangle + \sum_{n>m, k=x,y} \rho_k^{mn} |\sigma_k^{mn}\rangle \quad (46)$$

and the expansion coefficients  $\{\rho_m\}_{m=1}^{d_A} \cup \{\rho_k^{mn}\}_{n>m=1, \dots, d_A; k=x,y}$  are all real. Hence each state  $\rho$  is in one-to-one correspondence with a Hermitian matrix, given by

$$S_\rho = \sum_m \rho_m S_{\varphi_m} + \sum_{n>m, k=x,y} \rho_k^{mn} S_{\sigma_k^{mn}}. \quad (47)$$

Since effects are linear functionals on states, they are also represented by Hermitian matrices. We will indicate with  $E_a$

the Hermitian matrix associated to the effect  $a \in \text{Eff}(A)$ . The matrix  $E_a$  is uniquely defined by the relation

$$(a|\rho) = \text{Tr}[E_a S_\rho].$$

In the rest of the section we show that the set of matrices  $\{S_\rho | \rho \in \text{St}_1(A)\}$  is the whole set of positive Hermitian matrices with unit trace and that the set of matrices  $\{E_a | a \in \text{Eff}(A)\}$  is the set of positive Hermitian matrices bounded by the identity.

Let us start from some simple facts:

*Lemma 65.* The invariant state  $\chi_A$  has matrix representation  $S_{\chi_A} = \frac{I_{d_A}}{d_A}$ , where  $I_{d_A}$  is the identity matrix in dimension  $d_A$ .

*Proof.* Obvious from the expression  $\chi_A = \frac{1}{d} \sum_m \varphi_m$  and from the matrix representation of the states  $\{\varphi_m\}_{m=1}^{d_A}$  in Eq. (41). ■

*Lemma 66.* Let  $a_m \in \text{Eff}(A)$  be the atomic effect such that  $(a_m|\varphi_m) = 1$ . Then, the effect  $a_m$  has matrix representation  $E_{a_m}$  such that  $E_{a_m} = S_{\varphi_m}$ .

*Proof.* Let  $\rho \in \text{St}_1(A)$  be an arbitrary state. Expanding  $\rho$  as in Eq. (46) and using lemma 62 we obtain  $(a_m|\rho) = \rho_m$ . On the other hand, by Eq. (47) we have that  $\rho_m$  is the  $m$ th diagonal element of the matrix  $S_\rho$ : by definition of  $S_{\varphi_m}$  [Eq. (41)], this implies  $\rho_m = \text{Tr}[S_{\varphi_m} S_\rho]$ . Now, by construction we have  $\text{Tr}[E_{a_m} S_\rho] = (a_m|\rho) = \rho_m = \text{Tr}[S_{\varphi_m} S_\rho]$  for every  $\rho \in \text{St}_1(A)$ . Hence  $E_{a_m} = S_{\varphi_m}$ . ■

*Lemma 67.* The deterministic effect  $e \in \text{Eff}(A)$  has matrix representation  $E_e = I_{d_A}$ .

*Proof.* Obvious from the expression  $e = \sum_m a_m$ , combined with lemma 66 and Eq. (41). ■

*Corollary 42.* For every state  $\rho \in \text{St}_1(A)$  one has

$$\text{Tr}[S_\rho] = 1.$$

*Proof.*  $\text{Tr}[S_\rho] = \text{Tr}[E_e S_\rho] = (e|\rho) = 1$ . ■

*Theorem 17.* The matrix elements of  $S_\varphi$  for a pure state  $\varphi \in \text{St}_1(A)$  are  $(S_\varphi)_{mn} = \sqrt{p_m p_n} e^{i\theta_{mn}}$ , with  $\sum_{m=1}^{d_A} p_m = 1$ ,  $\theta_{mn} \in [0, 2\pi)$ ,  $\theta_{mn} = 0$  and  $\theta_{mn} = -\theta_{nm}$ .

*Proof.* First of all, the diagonal elements of  $S_\varphi$  are given by  $[S_\varphi]_{mm} = (a_m|\varphi)$  [cf. Eqs. (46) and (47)]. Denoting the  $m$ th element by  $p_m$ , we clearly have  $\sum_{m=1}^{d_A} p_m = (e|\varphi) = 1$ . Now, the projection  $\Pi_{\{m,n\}}|\varphi\rangle$  is a state in the face  $F_{mn}$ , and, by our choice of representation, the corresponding matrix  $S_{\Pi_{\{m,n\}}|\varphi\rangle}$  is proportional to a pure qubit state (nonnegative rank-one matrix). On the other hand, it is easy to see from Eqs. (46) and (47) that  $S_{\Pi_{\{m,n\}}|\varphi\rangle}$  is the matrix with the same elements as  $S_\varphi$  in the block corresponding to the qubit  $(m, n)$  and 0 elsewhere. In order to be positive and rank-one the corresponding  $2 \times 2$  submatrix must have the off-diagonal elements  $(S_\varphi)_{mn} = \sqrt{p_m p_n} e^{i\theta_{mn}}$  for some  $\theta_{mn} \in [0, 2\pi)$  with  $\theta_{nm} = -\theta_{mn}$ . Repeating the same argument for all choices of indices  $m, n$ , the thesis follows. ■

*Theorem 18.* For a pure state  $\varphi \in \text{St}_1(A)$ , the corresponding atomic effect  $a_\varphi$  such that  $(a_\varphi|\varphi) = 1$  has a matrix representation  $E_\varphi$  with the property that  $E_\varphi = S_\varphi$ .

*Proof.* We already know that the statement holds for  $d_A = 2$ , where we proved the Bloch sphere representation, equivalent to the fact that states and effects are represented as  $2 \times 2$  positive complex matrices, with the set of pure states identified with the set of all rank-one projectors. Let us now consider a generic system  $A$ . For every  $m < n$ , the face  $F_{mn}$  generated

by  $\{\varphi_m, \varphi_n\}$  can be encoded in a two-dimensional system. Therefore, the matrices  $S_{\Pi_{\{m,n\}}|\varphi}$  and  $E_{(a|\Pi_{\{m,n\}})}$  are positive [also, recall that all matrix elements outside the  $(m, n)$  block are zero]. Let  $\varphi_{\perp}^{(mn)}$  be the pure state in the face  $F_{mn}$  that is perfectly distinguishable from  $\Pi_{\{m,n\}}|\varphi$ . Note that, since  $\varphi_{\perp}^{(mn)}$  belongs to the face  $F_{mn}$ , it is also perfectly distinguishable from  $\Pi_{\{1, \dots, d_A\} \setminus \{m, n\}}|\varphi$ . Hence  $\varphi_{\perp}^{(mn)}$  is perfectly distinguishable from  $\varphi$  and, in particular,  $(a|\varphi_{\perp}^{(mn)}) = 0$  (lemma 59). This implies the relation

$$\begin{aligned} \text{Tr}[E_{(a|\Pi_{\{m,n\}}} S_{\varphi_{\perp}^{(mn)}})] &= (a|\Pi_{\{m,n\}}|\varphi_{\perp}^{(mn)}) \\ &= (a|\varphi_{\perp}^{(mn)}) = 0. \end{aligned}$$

Now, since the matrix  $E_{(a|\Pi_{\{m,n\}})}$  is positive, the above relation implies  $E_{(a|\Pi_{\{m,n\}}} = c_{mn} S_{\Pi_{\{m,n\}}|\varphi}$ , where  $c_{mn} \geq 0$ . Finally, repeating the argument for all possible values of  $(m, n)$ , we obtain that  $c_{mn} = c$  for every  $m, n$ , that is,  $E_a = c S_{\varphi}$ . Taking the trace on both sides we obtain  $\text{Tr}[E_a] = c$ . To prove that  $c = 1$ , we use the relation  $\text{Tr}[E_a]/d_A = (a|\chi_A) = 1/d_A$ . ■

We conclude with a simple corollary that will be used in the next subsection.

*Corollary 43.* Let  $\varphi \in \text{St}_1(A)$  be a pure state and let  $\{\gamma_i\}_{i=1}^r \subset \text{St}_1(A)$  be a set of pure states. If the state  $\varphi$  can be written as

$$|\varphi\rangle = \sum_i x_i |\gamma_i\rangle$$

for some real coefficients  $\{x_i\}_{i=1}^r$ , then the atomic effect  $a$  such that  $(a|\varphi) = 1$  is given by

$$(a| = \sum_i x_i (c_i|,$$

where  $c_i$  is the atomic effect such that  $(c_i|\gamma_i) = 1$ .

*Proof.* For every  $\rho \in \text{St}(A)$  by theorem 18 one has

$$\begin{aligned} (a|\rho) &= \text{Tr}[E_a S_{\rho}] = \text{Tr}[S_{\varphi} S_{\rho}] = \sum_i x_i \text{Tr}[S_{\gamma_i} S_{\rho}] \\ &= \sum_i x_i \text{Tr}[E_{c_i} S_{\rho}] = \sum_i x_i (c_i|\rho), \end{aligned}$$

thus implying the thesis. ■

### C. Choice of axes for a two-qubit system

If A and B are two systems with  $d_A = d_B = 2$ , then we can use two different types of matrix representations for the states of the composite system AB.

The first type of representation is the representation  $S_{\varphi}$  introduced through lemma 64: here we will refer to it as the *standard representation*. Note that there are many different representations of this type because for every pair  $(m, n)$  there is freedom in choice of the  $x$  and  $y$  axis [cf. Eqs. (44) and (45)].

The second type of representation is the *tensor product representation*  $T_{\varphi}$ , defined by the tensor product of matrices representing states of systems A and B: for a state  $|\rho\rangle = \sum_{i,j} \rho_{ij} |\alpha_i\rangle|\beta_j\rangle$ , with  $\alpha_i \in \text{St}(A), \beta_j \in \text{St}(B)$ , we have

$$T_{\rho} := \sum_{i,j} \rho_{ij} S_{\alpha_i}^A \otimes S_{\beta_j}^B, \quad (48)$$

where  $S^A$  (or  $S^B$ ) is the matrix representation for system A (or B). Here the freedom is in the choice of the axes for the Bloch spheres of qubits A and B. Since A and B are operationally equivalent, we will indicate the elements of the bases for  $\text{St}_{\mathbb{R}}(A)$  and  $\text{St}_{\mathbb{R}}(B)$  with the same letters:  $\{\varphi_m\}_{m=1}^2$  for the two perfectly distinguishable pure states and  $\{\sigma_k\}_{k=x,y}$  for the remaining basis vectors.

We now show a few properties of the tensor representation. Let  $F_A$  denote the matrix corresponding to the effect  $A \in \text{Eff}(AB)$  in the tensor representation, that is, the matrix defined by

$$(A|\rho) := \text{Tr}[F_A T_{\rho}] \quad \forall \rho \in \text{St}(AB). \quad (49)$$

It is easy to show that the matrix representation for effects must satisfy the analog of Eq. (48).

*Lemma 68.* Let  $A \in \text{Eff}(AB)$  be a bipartite effect, written as  $(A| = \sum_{i,j} A_{ij} (a_i| (b_j|$ . Then one has

$$F_A = \sum_{i,j} A_{ij} E_{a_i}^A \otimes E_{b_j}^B,$$

where  $E_{a_i}^A$  (or  $E_{b_j}^B$ ) is the matrix representing the single-qubit effect  $a_i$  (or  $b_j$ ) in the standard representation for qubit A (or B).

*Proof.* For every bipartite state  $|\rho\rangle = \sum_{k,l} \rho_{kl} |\alpha_k\rangle|\beta_l\rangle$  one has

$$\begin{aligned} \text{Tr}[F_A T_{\rho}] &= (A|\rho) \\ &= \sum_{i,j,k,l} A_{ij} \rho_{kl} (a_i|\alpha_k) (b_j|\beta_l) \\ &= \sum_{i,j,k,l} A_{ij} \rho_{kl} \text{Tr}[E_{a_i}^A S_{\alpha_k}^A] \text{Tr}[E_{b_j}^B S_{\beta_l}^B] \\ &= \sum_{i,j,k,l} A_{ij} \rho_{kl} \text{Tr}[(E_{a_i}^A \otimes E_{b_j}^B) T_{\alpha_k \otimes \beta_l}] \\ &= \sum_{i,j} A_{ij} \text{Tr}[(E_{a_i}^A \otimes E_{b_j}^B) T_{\rho}] \end{aligned}$$

which implies the thesis. ■

*Corollary 44.* Let  $\Psi \in \text{St}_1(AB)$  be a pure state and let  $A \in \text{Eff}(AB)$  be the atomic effect such that  $(A|\Psi) = 1$ . Then one has  $F_A = T_{\Psi}$ .

*Proof.* Let  $\{a_i\}_{i=1}^4$  (or  $\{\beta_j\}_{j=1}^4$ ) be a set of pure states that span  $\text{St}_{\mathbb{R}}(A)$  [or  $\text{St}_{\mathbb{R}}(B)$ ] and expand  $\Psi$  as  $|\Psi\rangle = \sum_{i,j} c_{ij} |\alpha_i\rangle|\beta_j\rangle$ . Then, corollary 43 yields  $(A| = \sum_{i,j} c_{ij} (a_i| (b_j|$  where  $a_i$  and  $b_j$  are the atomic effects such that  $(a_i|\alpha_i) = (b_j|\beta_j) = 1$ . Therefore, we have

$$F_A = \sum_{i,j} c_{ij} E_{a_i}^A \otimes E_{b_j}^B = \sum_{i,j} c_{ij} S_{\alpha_i}^A \otimes S_{\beta_j}^B = T_{\Psi}. \quad \blacksquare$$

*Corollary 45.* For every bipartite state  $\rho \in \text{St}_1(AB)$ ,  $d_A = d_B = 2$  one has  $\text{Tr}[T_{\rho}] = 1$ .

*Proof.* For each qubit we have

$$E_{a_1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad E_{a_2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \quad (50)$$

Hence  $E_{e_A}^A = E_{e_B}^B = I$ , where  $I$  is the  $2 \times 2$  identity matrix. By lemma 68 we then have  $F_{e_A \otimes e_B} = I \otimes I$  and, therefore,  $\text{Tr}[T_\rho] = \text{Tr}[F_{e_A \otimes e_B} T_\rho] = (e_A \otimes e_B | \rho) = 1$ . ■

Finally, an immediate consequence of local distinguishability is the following.

*Lemma 69.* Suppose that  $\mathcal{U} \in \mathbf{G}_A$  and  $\mathcal{V} \in \mathbf{G}_B$  are two reversible transformations for qubits A and B, respectively, and that  $U, V \in \text{SU}(2)$  are such that

$$\begin{aligned} S_{\mathcal{U}\rho}^A &= U S_\rho^A U^\dagger \quad \forall \rho \in \text{St}_1(A), \\ S_{\mathcal{V}\sigma}^B &= V S_\sigma^B V^\dagger \quad \forall \sigma \in \text{St}_1(B). \end{aligned}$$

Then, we have  $T_{(\mathcal{U} \otimes \mathcal{V})\tau} = (U \otimes V) T_\tau (U^\dagger \otimes V^\dagger)$  for every  $\tau \in \text{St}_1(AB)$ .

*Proof.* The thesis follows by linearity expanding  $\tau$  as  $\tau = \sum_{i,j=1}^4 \tau_{ij} \alpha_i \otimes \beta_j$ , where  $\{\alpha_i\}_{i=1}^4$  and  $\{\beta_j\}_{j=1}^4$  are bases for the  $\text{St}_{\mathbb{R}}(A)$  and  $\text{St}_{\mathbb{R}}(B)$ . ■

The rest of this subsection is aimed at showing that, with a suitable choice of matrix representation for system B, the standard representation coincides with the tensor representation, that is,  $S_\rho = T_\rho$  for every  $\rho \in \text{St}(AB)$ . This technical result is important because some properties used in our derivation are easily proved in the standard representation, while the property expressed by lemma 69 is easily proved in the tensor representation: it is then essential to show that we can construct a representation that enjoys both properties.

The four states  $\{\varphi_m \otimes \varphi_n\}_{m,n=1}^2$  are clearly a maximal set of perfectly distinguishable pure states in AB. In the following we will construct the standard representation starting from this set.

*Lemma 70.* For a composite system AB with  $d_A = d_B = 2$  one can choose the standard representation in such a way that the following equalities hold:

$$S_{\varphi_m \otimes \varphi_n} = T_{\varphi_m \otimes \varphi_n}, \quad (51)$$

$$S_{\varphi_m \otimes \sigma_k} = T_{\varphi_m \otimes \sigma_k}, \quad k = x, y, \quad (52)$$

$$S_{\sigma_k \otimes \varphi_m} = T_{\sigma_k \otimes \varphi_m}, \quad k = x, y. \quad (53)$$

*Proof.* Let us choose single-qubit representations  $S^A$  and  $S^B$  that satisfy Eqs. (41), (42), and (43). On the other hand, choosing the states  $\{\varphi_n \otimes \varphi_n\}$  in lexicographic order as the four distinguishable states for the standard representation, we have

$$[S_{\varphi_1 \otimes \varphi_1}]_{rs} = \delta_{1r} \delta_{1s}, \quad [S_{\varphi_1 \otimes \varphi_2}]_{rs} = \delta_{2r} \delta_{2s},$$

$$[S_{\varphi_2 \otimes \varphi_1}]_{rs} = \delta_{3r} \delta_{3s}, \quad [S_{\varphi_2 \otimes \varphi_2}]_{rs} = \delta_{4r} \delta_{4s}.$$

With this choice we get  $S_{\varphi_m \otimes \varphi_n} = S_{\varphi_m}^A \otimes S_{\varphi_n}^B = T_{\varphi_m \otimes \varphi_n}$  for every  $m, n = 1, 2$ . This proves Eq. (51). Let us now prove Eqs. (52) and (53). Consider the two-dimensional face  $F_{11,12}$ , generated by the states  $\varphi_1 \otimes \varphi_1$  and  $\varphi_1 \otimes \varphi_2$ . This face is the face identified by the state  $\omega_{11,12} := \varphi_1 \otimes \chi_B$ , and we have  $F_{11,12} \simeq \{\varphi_1\} \otimes \text{St}_1(B)$ . Therefore we can choose the vectors  $\sigma_k^{11,12}$ ,  $k = x, y$  to satisfy the relation  $\sigma_k^{11,12} := \varphi_1 \otimes \sigma_k$ ,  $k = x, y$ . Now, in the standard representation we have

$$[S_{\sigma_x^{11,12}}]_{rs} = \delta_{r1} \delta_{s2} + \delta_{r2} \delta_{s1},$$

$$[S_{\sigma_y^{11,12}}]_{rs} = i\lambda(\delta_{r1} \delta_{s2} - \delta_{r2} \delta_{s1})$$

[cf. Eqs. (42) and (43)]. This implies  $S_{\sigma_k^{11,12}} = S_{\varphi_1}^A \otimes S_{\sigma_k}^B = T_{\varphi_1 \otimes \sigma_k}$  for  $k = x, y$ . Repeating the same argument for the face  $F_{22,21}$ ,  $F_{11,21}$ , and  $F_{21,22}$  we obtain the proof of Eqs. (52) and (53). ■

In order to prove that, with a suitable choice of axes, the standard representation coincides with the tensor representation—that is,  $S_\rho = T_\rho$  for every  $\rho \in \text{St}(AB)$ —it remains to find a choice of axes such that  $S_{\sigma_k \otimes \sigma_l} = T_{\sigma_k \otimes \sigma_l}$ ,  $k = x, y$ . This will be proved in the following.

*Lemma 71.* Let  $\Phi \in \text{St}_1(AB)$  be a pure state such that  $(a_1 \otimes a_1 | \Phi) = (a_2 \otimes a_2 | \Phi) = 1/2$  [such a state exists due to the superposition principle]. With a suitable choice of the matrix representation  $S^B$ , the state  $\Phi$  is represented by the matrix

$$T_\Phi = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \quad (54)$$

Moreover, one has

$$\Phi = \chi_A \otimes \chi_B + \frac{1}{4}(\sigma_x \otimes \sigma_x - \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z). \quad (55)$$

*Proof.* Let us start with the proof of Eq. (54). For every reversible transformation  $\mathcal{U} \in \mathbf{G}_A$ , let  $\mathcal{U}^* \in \mathbf{G}_B$  be the conjugate of  $\mathcal{U}$ , defined with respect to the state  $\Phi$ . Since all  $2 \times 2$  unitary (nontrivial) representations of  $\text{SU}(2)$  are unitarily equivalent, by a suitable choice of the standard representation  $S_\rho^B$  for system B, one has

$$S_{\mathcal{U}^*\rho}^B = U^* S_\rho^B U^T, \quad (56)$$

where  $U^*$  and  $U^T$  are the complex conjugate and the transpose of the matrix  $U \in \text{SU}(2)$  such that  $S_{\mathcal{U}\rho}^A = U S_\rho^A U^\dagger$ . Due to Eq. (56) and to lemma 69, the isotropic state  $\Phi$  must satisfy the condition  $(U \otimes U^*) T_\Phi (U^\dagger \otimes U^T) = T_\Phi, \forall U \in \text{SU}(2)$ . Now, the unitary representation  $\{U \otimes U^*\}$  has two irreducible subspaces and the projectors on them are given by the matrices

$$P_0 = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix},$$

$$P_1 = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} = I \otimes I - P_0,$$

where  $I$  is the  $2 \times 2$  identity matrix. The most general form for  $T_\Phi$  is then the following:

$$\begin{aligned} T_\Phi &= x_0 P_0 + x_1 P_1 \\ &= (x_0 - x_1) P_0 + x_1 I \otimes I \\ &= \begin{pmatrix} \alpha + \beta & 0 & 0 & \beta \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & \alpha & 0 \\ \beta & 0 & 0 & \alpha + \beta \end{pmatrix} \end{aligned}$$

having defined  $\alpha := x_1$  and  $\beta := (x_0 - x_1)/2$ . Now, by construction the state  $\Phi$  satisfies the condition

$$(a_m |_A | \Phi)_{AB} = \frac{1}{2} | \varphi_m \rangle_B, \quad m = 1, 2.$$

By definition of the tensor representation, the conditional states  $(a_m|_A|\Phi)_{AB}$  are described by the diagonal blocks of the matrix  $T_\Phi$ :

$$S_{(a_1|_A|\Phi)_{AB}}^B = \begin{pmatrix} \alpha + \beta & 0 \\ 0 & \alpha \end{pmatrix}, \quad S_{(a_2|_A|\Phi)_{AB}}^B = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha + \beta \end{pmatrix}. \quad (57)$$

Since the states  $\varphi_1$  and  $\varphi_2$  are pure, the above matrices must be rank-one. Moreover, their trace must be equal to  $(a_m \otimes e_B|\Phi) = 1/2(e_B|\varphi_m) = \frac{1}{2}$ ,  $m = 1, 2$ . Then we have two possibilities. Either (i)  $\alpha = 0$  and  $\beta = \frac{1}{2}$  or (ii)  $\alpha = -\beta = \frac{1}{2}$ . In case (i) Eq. (54) holds. In case (ii) to prove Eq. (54) we need to change our choice of matrix representation for the qubit B. Precisely, we make the following change:

$$\begin{aligned} S_{\sigma_x}^B &\mapsto \tilde{S}_{\sigma_x}^B = -S_{\sigma_x}^B, \\ S_{\sigma_y}^B &\mapsto \tilde{S}_{\sigma_y}^B = -S_{\sigma_y}^B, \\ S_{\sigma_z}^B &\mapsto \tilde{S}_{\sigma_z}^B = -S_{\sigma_z}^B, \end{aligned} \quad (58)$$

where  $\sigma_z := \varphi_1 - \varphi_2$ . Note that the inversion of the axes, sending  $\sigma_k$  to  $-\sigma_k$  for every  $k = x, y, z$  is not an allowed physical transformation, but this is not a problem here, because Eq. (58) is just a new choice of matrix representation, in which the set of states of system B is still represented by the Bloch sphere.

More concisely, the change of matrix representation  $S^B \mapsto \tilde{S}^B$  can be expressed as

$$S_\rho^B \mapsto \tilde{S}_\rho^B := Y [S_\rho^B]^T Y^\dagger, \quad Y := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Note that in the new representation  $\tilde{S}^B$  the physical transformation  $\mathcal{U}^*$  is still represented as  $\tilde{S}_{\mathcal{U}^*\rho}^B = U^* \tilde{S}_\rho^B U^T$ : indeed we have

$$\begin{aligned} \tilde{S}_{\mathcal{U}^*\rho}^B &= Y [S_{\mathcal{U}^*\rho}^B]^T Y^\dagger \\ &= Y (U^* S_\rho^B U^T)^T Y^\dagger \\ &= Y (U [S_\rho^B]^T U^\dagger) Y^\dagger \\ &= (Y U Y^\dagger) (Y [S_\rho^B]^T Y^\dagger) (Y U^\dagger Y^\dagger) \\ &= U^* (Y [S_\rho^B]^T Y^\dagger) U^T \\ &= U^* \tilde{S}_\rho^B U^T \end{aligned}$$

having used the relations  $Y^\dagger Y = I$  and  $Y U Y^\dagger = U^*$  for every  $U \in \text{SU}(2)$ . Clearly the change of standard representation  $S \rightarrow \tilde{S}$  for the qubit B induces a change of tensor representation  $T \rightarrow \tilde{T}$ , where  $\tilde{T}$  is the tensor representation defined by  $\tilde{T}_{\rho \otimes \sigma} := S_\rho^A \otimes \tilde{S}_\sigma^B$ . With this change of representation, we have

$$\tilde{T}_\Phi = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

This concludes the proof of Eq. (54).

Let us now prove Eq. (55). Using the fact that by definition  $T_{\rho \otimes \tau} = (S_\rho^A \otimes S_\tau^B)$  one can directly verify the relation

$$T_\Phi = S_\chi^A \otimes S_\chi^B + \frac{1}{4} (S_{\sigma_x}^A \otimes S_{\sigma_x}^B - S_{\sigma_y}^A \otimes S_{\sigma_y}^B + S_{\sigma_z}^A \otimes S_{\sigma_z}^B).$$

This is precisely the matrix version of Eq. (55). ■

Note that the choice of  $S^B$  needed in Eq. (54) is compatible with the choice of  $S^B$  needed in lemma 70: indeed, to prove compatibility we only have to show that the representation  $S_B$  used in Eq. (54) has the property  $[S_{\varphi_m}^B]_{rs} = \delta_{mr} \delta_{ms}$ ,  $m = 1, 2$ . This property is automatically guaranteed by the relation  $(a_m|_A|\Phi)_{AB} = 1/2|\varphi_m\rangle$ ,  $m = 1, 2$  and by Eq. (57) with  $\alpha = 0$  and  $\beta = 1/2$ .

*Corollary 46.* In the standard representation the state  $\Phi \in \text{St}_1(\text{AB})$  is represented by the matrix

$$S_\Phi = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & e^{i\theta} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ e^{-i\theta} & 0 & 0 & 1 \end{pmatrix}. \quad (59)$$

*Proof.* The thesis follows from theorem 17 and lemma 70. ■

We now define the reversible transformations  $\mathcal{U}_{x,\pi}$  and  $\mathcal{U}_{z,\frac{\pi}{2}}$  as follows:

$$\begin{aligned} S_{\mathcal{U}_{x,\pi}\rho} &= X S_\rho X, \quad X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ S_{\mathcal{U}_{z,\frac{\pi}{2}}\rho} &= e^{-i\frac{\pi}{4}Z} S_\rho e^{i\frac{\pi}{4}Z}, \quad Z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned} \quad (60)$$

Also, we define the states  $\Psi, \Phi_{z,\frac{\pi}{2}}$ , and  $\Psi_{z,\frac{\pi}{2}}$  as

$$\begin{aligned} |\Psi\rangle &:= (\mathcal{U}_{x,\pi} \otimes \mathcal{I}) |\Phi\rangle, \\ |\Phi_{z,\frac{\pi}{2}}\rangle &:= (\mathcal{U}_{z,\frac{\pi}{2}} \otimes \mathcal{I}) |\Phi\rangle, \\ |\Psi_{z,\frac{\pi}{2}}\rangle &:= (\mathcal{U}_{z,\frac{\pi}{2}} \otimes \mathcal{I}) |\Psi\rangle. \end{aligned}$$

*Lemma 72* The states  $\Psi, \Phi_{z,\frac{\pi}{2}}$ , and  $\Psi_{z,\frac{\pi}{2}}$  have the following tensor representation:

$$\begin{aligned} T_\Psi &= \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad T_{\Phi_{z,\frac{\pi}{2}}} = 12 \begin{pmatrix} 1 & 0 & 0 & -i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 1 \end{pmatrix}, \\ T_{\Psi_{z,\frac{\pi}{2}}} &= \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -i & 0 \\ 0 & i & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (61)$$

Moreover, one has

$$\begin{aligned} \Psi &= \chi_A \otimes \chi_B + \frac{1}{4} (\sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y - \sigma_z \otimes \sigma_z), \\ \Phi_{z,\frac{\pi}{2}} &= \chi_A \otimes \chi_B + \frac{1}{4} (\sigma_y \otimes \sigma_x + \sigma_x \otimes \sigma_y + \sigma_z \otimes \sigma_z), \\ \Psi_{z,\frac{\pi}{2}} &= \chi_A \otimes \chi_B + \frac{1}{4} (\sigma_y \otimes \sigma_x - \sigma_x \otimes \sigma_y - \sigma_z \otimes \sigma_z). \end{aligned} \quad (62)$$

*Proof.* Equation (61) is obtained from Eq. (54) by explicit calculation using lemma 69 and Eq. (60). Then, the validity of Eq. (62) is easily obtained from Eq. (55) using the relations

$$\begin{aligned} \mathcal{U}_{x,\pi} |\sigma_x\rangle &= |\sigma_x\rangle, \\ \mathcal{U}_{x,\pi} |\sigma_y\rangle &= -|\sigma_y\rangle, \\ \mathcal{U}_{x,\pi} |\sigma_z\rangle &= -|\sigma_z\rangle, \end{aligned}$$

and

$$\begin{aligned}\mathcal{U}_{z,\pi/2}|\sigma_x\rangle &= |\sigma_y\rangle, \\ \mathcal{U}_{z,\pi/2}|\sigma_y\rangle &= -|\sigma_x\rangle, \\ \mathcal{U}_{z,\pi/2}|\sigma_z\rangle &= |\sigma_z\rangle.\end{aligned}$$

**Lemma 73.** The states  $\Psi$ ,  $\Phi_{z,\frac{\pi}{2}}$ , and  $\Psi_{z,\frac{\pi}{2}}$  have a standard representation of the form

$$\begin{aligned}S_\Psi &= \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & e^{i\gamma} & 0 \\ 0 & e^{-i\gamma} & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\ S_{\Phi_{z,\frac{\pi}{2}}} &= \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & \lambda i e^{i\theta} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\lambda i e^{-i\theta} & 0 & 0 & 1 \end{pmatrix}, \\ S_{\Psi_{z,\frac{\pi}{2}}} &= \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & \mu i e^{i\gamma} & 0 \\ 0 & -\mu i e^{-i\gamma} & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.\end{aligned}\quad (63)$$

with  $\theta$  as in corollary 46,  $\gamma \in [0, 2\pi)$  and  $\lambda, \mu \in \{-1, 1\}$ .

*Proof.* Let us start from  $\Psi$ . First, from Eq. (62) it is immediate to obtain  $(a_1 \otimes a_1|\Psi) = (a_2 \otimes a_2|\Psi) = 0$  and  $(a_1 \otimes a_2|\Psi) = (a_2 \otimes a_1|\Psi) = 1/2$ . This gives the diagonal elements of  $S_\Psi$ . Then, using theorem 17 we obtain that  $S_\Psi$  must be as in Eq. (63), for some value of  $\gamma$ . Let us now consider  $\Phi_{z,\frac{\pi}{2}}$ . Again, the diagonal elements of the matrix  $S_{\Phi_{z,\frac{\pi}{2}}}$  are obtained from Eq. (62), which in this case yields  $(a_1 \otimes a_1|\Phi_{z,\frac{\pi}{2}}) = (a_2 \otimes a_2|\Phi_{z,\frac{\pi}{2}}) = 1/2$  and  $(a_1 \otimes a_2|\Phi_{z,\frac{\pi}{2}}) = (a_2 \otimes a_1|\Phi_{z,\frac{\pi}{2}}) = 0$ . Hence, by theorem 17 we must have

$$S_{\Phi_{z,\frac{\pi}{2}}} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & e^{i\theta'} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ e^{-i\theta'} & 0 & 0 & 1 \end{pmatrix}$$

for some value of  $\theta' \in [0, 2\pi)$ . Now, denote by  $A$  the effect such that  $(A|\Phi) = 1$ . We then have

$$\begin{aligned}(A|\Phi_{z,\frac{\pi}{2}}) &= \text{Tr}[E_A S_{\Phi_{z,\frac{\pi}{2}}}] = \text{Tr}[S_\Phi S_{\Phi_{z,\frac{\pi}{2}}}], \\ (A|\Psi_{z,\frac{\pi}{2}}) &= \text{Tr}[F_A T_{\Psi_{z,\frac{\pi}{2}}}] = \text{Tr}[T_\Phi T_{\Psi_{z,\frac{\pi}{2}}}] = \frac{1}{2}\end{aligned}$$

having used theorem 18, corollary 44, and Eq. (61). Hence we have  $\text{Tr}[S_\Phi S_{\Phi_{z,\frac{\pi}{2}}}] = 1/2$ , which implies  $\theta' = \theta \pm \frac{\pi}{2}$ , as in Eq. (63). Finally, the same arguments can be used for  $\Psi_{z,\frac{\pi}{2}}$ : The diagonal elements of  $S_{\Psi_{z,\frac{\pi}{2}}}$  are obtained from the relations  $(a_1 \otimes a_1|\Psi_{z,\frac{\pi}{2}}) = (a_2 \otimes a_2|\Psi_{z,\frac{\pi}{2}}) = 0$  and  $(a_1 \otimes a_2|\Psi_{z,\frac{\pi}{2}}) = (a_2 \otimes a_1|\Psi_{z,\frac{\pi}{2}}) = 1/2$ , which follow from Eq. (62). This implies that the matrix  $S_{\Psi_{z,\frac{\pi}{2}}}$  has the form

$$S_{\Psi_{z,\frac{\pi}{2}}} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & e^{i\gamma'} & 0 \\ 0 & e^{-i\gamma'} & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

for some  $\gamma' \in [0, 2\pi)$ . The relation  $\text{Tr}[S_\Psi S_{\Psi_{z,\frac{\pi}{2}}}] = \text{Tr}[T_\Psi T_{\Psi_{z,\frac{\pi}{2}}}] = 1/2$  then implies  $\gamma' = \gamma \pm \frac{\pi}{2}$ . ■

Let us now consider the four vectors  $\Sigma_x^{(11,22)}$ ,  $\Sigma_y^{(11,22)}$ ,  $\Sigma_x^{(12,21)}$ ,  $\Sigma_y^{(12,21)}$  defined as follows:

$$\begin{aligned}\Sigma_x^{(11,22)} &= 2 \left( \Phi - \chi_A \otimes \chi_B - \frac{1}{4} \sigma_z \otimes \sigma_z \right), \\ \Sigma_y^{(11,22)} &= 2 \left( \Phi_{z,\frac{\pi}{2}} - \chi_A \otimes \chi_B - \frac{1}{4} \sigma_z \otimes \sigma_z \right), \\ \Sigma_x^{(12,21)} &= 2 \left( \Psi - \chi_A \otimes \chi_B + \frac{1}{4} \sigma_z \otimes \sigma_z \right), \\ \Sigma_x^{(12,21)} &= 2 \left( \Psi_{z,\frac{\pi}{2}} - \chi_A \otimes \chi_B + \frac{1}{4} \sigma_z \otimes \sigma_z \right).\end{aligned}\quad (64)$$

By the previous results, it is immediate to obtain the matrix representations of these vectors. In the tensor representation, using Eqs. (54) and (61), we obtain

$$\begin{aligned}T_{\Sigma_x^{(11,22)}} &= \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, & T_{\Sigma_y^{(11,22)}} &= \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}, \\ T_{\Sigma_x^{(12,21)}} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, & T_{\Sigma_y^{(12,21)}} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},\end{aligned}$$

while in the standard representation, using Eqs. (46) and (63), we obtain

$$\begin{aligned}S_{\Sigma_x^{(11,22)}} &= \begin{pmatrix} 0 & 0 & 0 & e^{i\theta} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ e^{-i\theta} & 0 & 0 & 0 \end{pmatrix}, \\ S_{\Sigma_y^{(11,22)}} &= \begin{pmatrix} 0 & 0 & 0 & -\lambda i e^{i\theta} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \lambda i e^{-i\theta} & 0 & 0 & 0 \end{pmatrix}, \\ S_{\Sigma_x^{(12,21)}} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & e^{i\gamma} & 0 \\ 0 & e^{-i\gamma} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\ S_{\Sigma_x^{(12,21)}} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -\mu i e^{i\gamma} & 0 \\ 0 & \mu i e^{-i\gamma} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},\end{aligned}$$

Comparing the two matrix representations we are now in position to prove the desired result.

**Lemma 74.** With a suitable choice of axes, one has  $S_{\sigma_k \otimes \sigma_l} = T_{\sigma_k \otimes \sigma_l}$  for every  $k, l = x, y$ .

*Proof.* For the face (11, 22), using the freedom coming from Eqs. (43) and (44), we redefine the  $x$  and  $y$  axes so that  $\sigma_x^{(11,22)} := \Sigma_x^{(11,22)}$  and  $\lambda \sigma_y^{(11,22)} := \Sigma_y^{(11,22)}$ . In this way we have

$$S_{\Sigma_k^{(11,22)}} = T_{\Sigma_k^{(11,22)}} \quad \forall k = x, y.$$

Likewise, for the face (12,21) we redefine the  $x$  and  $y$  axes so that  $\sigma_x^{(12,21)} := \Sigma_x^{(12,21)}$  and  $\mu\sigma_y^{(12,21)} := \Sigma_y^{(12,21)}$ , so that we have

$$S_{\Sigma_k^{(12,21)}} = T_{\Sigma_k^{(12,21)}} \quad \forall k = x, y.$$

Finally, using Eqs. (55), (62), and (64) we have the relations

$$\begin{aligned} \sigma_x \otimes \sigma_x &= \Sigma_x^{(11,22)} + \Sigma_x^{(12,21)}, \\ \sigma_y \otimes \sigma_y &= \Sigma_x^{(11,22)} - \Sigma_x^{(12,21)}, \\ \sigma_x \otimes \sigma_y &= \Sigma_y^{(11,22)} - \Sigma_y^{(12,21)}, \\ \sigma_y \otimes \sigma_x &= \Sigma_y^{(11,22)} + \Sigma_y^{(12,21)}. \end{aligned}$$

Since  $S$  and  $T$  coincide on the right-hand side of each equality, they must also coincide on the left-hand side. ■

*Theorem 19.* With a suitable choice of axes, the standard representation coincides with the tensor representation, that is,  $S_\rho = T_\rho$  for every  $\rho \in \text{St}(\text{AB})$ .

*Proof.* Combining lemma 70 with lemma 74 we obtain that  $S$  and  $T$  coincide on the tensor products basis  $\mathcal{B} \times \mathcal{B}$ , where  $\mathcal{B} = \{\varphi_1, \varphi_2, \sigma_x, \sigma_y\}$ . By linearity,  $S$  and  $T$  coincide on every state. ■

From now on, whenever we will consider a composite system AB where A and B are two dimensional we will adopt the choice that guarantees that the standard representation coincides with the tensor representation.

#### D. Positivity of the matrices

In this paragraph we show that the states in our theory can be represented by positive matrices. This amounts to prove that for every system A, the set of states  $\text{St}_1(\text{A})$  can be represented as a subset of the set of density matrices in dimension  $d_A$ . This result will be completed in Sec. XIII E, where we will see that, in fact, every density matrix in dimension  $d_A$  corresponds to some state of  $\text{St}_1(\text{A})$ .

The starting point to prove positivity is the following.

*Lemma 75.* Let A and B be two-dimensional systems. Then, for every pure state  $\Psi \in \text{St}(\text{AB})$  one has  $S_\Psi \geq 0$ .

*Proof.* Take an arbitrary vector  $Z \in \mathbb{C}^2 \otimes \mathbb{C}^2$ , written in the Schmidt form as  $|Z\rangle = \sum_{n=1}^2 \sqrt{\lambda_n} |v_n\rangle |w_n\rangle$ . Introducing the unitaries  $U, V$  such that  $U|v_n\rangle = |n\rangle$  and  $V|w_n\rangle = |n\rangle$  for every  $n = 1, 2$  then we have  $|Z\rangle = (U^\dagger \otimes V^\dagger)|W\rangle$ , where  $|W\rangle = \sum_{n=1}^2 \sqrt{\lambda_n} |n\rangle |n\rangle$ . Therefore, we have

$$\langle Z|S_\Psi|Z\rangle = \langle W|S_{(\mathcal{U} \otimes \mathcal{V})\Psi}|W\rangle,$$

where  $\mathcal{U}$  and  $\mathcal{V}$  are the reversible transformations defined by  $S_{\mathcal{U}\rho} = U S_\rho U^\dagger$  and  $S_{\mathcal{V}\rho} = V S_\rho V^\dagger$ , respectively ( $\mathcal{U}$  and  $\mathcal{V}$  are physical transformations by virtue of corollary 30). Here we used the fact that the standard two-qubit representation coincides with the tensor representation and, therefore,  $S_{(\mathcal{U} \otimes \mathcal{V})\Psi} = (U \otimes V) S_\Psi (U \otimes V)^\dagger$ . Denoting the pure state  $(\mathcal{U} \otimes \mathcal{V})|\Psi\rangle$  by  $|\Psi'\rangle$  we then have

$$\begin{aligned} \langle Z|S_\Psi|Z\rangle &= \lambda_1 [S_{\Psi'}]_{11,11} + \lambda_2 [S_{\Psi'}]_{22,22} \\ &\quad + 2\sqrt{\lambda_1 \lambda_2} \text{Re}([S_{\Psi'}]_{11,22}). \end{aligned}$$

Since by theorem 17 we have  $[S_{\Psi'}]_{11,22} = \sqrt{[S_{\Psi'}]_{11,11} [S_{\Psi'}]_{22,22}} e^{i\theta}$ , we conclude

$$\begin{aligned} \langle Z|S_\Psi|Z\rangle &= \lambda_1 [S_{\Psi'}]_{11,11} + \lambda_2 [S_{\Psi'}]_{22,22} \\ &\quad + 2 \cos \theta \sqrt{\lambda_1 \lambda_2 [S_{\Psi'}]_{11,11} [S_{\Psi'}]_{22,22}} \\ &\geq (\sqrt{\lambda_1 [S_{\Psi'}]_{11,11}} - \sqrt{\lambda_2 [S_{\Psi'}]_{22,22}})^2 \geq 0. \end{aligned}$$

Finally, since the vector  $Z \in \mathbb{C}^2 \otimes \mathbb{C}^2$  is arbitrary, the matrix  $S_\Psi$  is positive. ■

*Corollary 47.* Let C be a system of dimension  $d_C = 4$ . Then, with a suitable choice of matrix representation the pure states of C are represented by positive matrices.

*Proof.* The system C is operationally equivalent to the composite system AB, where  $d_A = d_B = 2$ . Let  $\mathcal{U} \in \text{Transf}(\text{AB}, \text{C})$  be the reversible transformation implementing the equivalence. Now, we know that the states of AB are represented by positive matrices. If we define the basis vectors for C by applying  $\mathcal{U}$  to the basis for AB, then we obtain that the states of C are represented by the same matrices representing the states of AB. ■

*Corollary 48.* Let A be a system with  $d_A = 3$ . With a suitable choice of matrix representation, the matrix  $S_\varphi$  is positive for every pure state  $\varphi \in \text{St}(\text{A})$ .

*Proof.* Let C be a system with  $d_C = 4$ . By corollary 47 the states of C are represented by positive matrices. Define the state  $\omega := \frac{1}{3}(\varphi_1 + \varphi_2 + \varphi_3)$ , where  $\{\varphi_m\}_{m=1}^4$  are four perfectly distinguishable pure states. By the compression axiom, the face  $F_\omega$  can be encoded in a three-dimensional system D (corollary 40). In fact, since D is operationally equivalent to A, the face  $F_\omega$  can be encoded in A. Let  $\mathcal{E} \in \text{Transf}(\text{D}, \text{A})$  and  $\mathcal{D} \in \text{Transf}(\text{A}, \text{D})$  be the encoding and decoding operation, respectively. If we define the basis vectors for A by applying  $\mathcal{E}$  to the basis vectors for the face  $F_\omega$ , then we obtain that the states of A are represented by the same matrices representing the states in the face  $F_\omega$ . Since these matrices are positive, the thesis follows. ■

From now on, for every three-dimensional system A we will choose the  $x$  and  $y$  axes so that  $S_\rho$  is positive for every  $\rho \in \text{St}(\text{A})$ .

*Corollary 49.* Let  $\varphi \in \text{St}_1(\text{A})$  be a pure state with  $d_A = 3$ . Then, the corresponding matrix  $S_\varphi$ , given by

$$S_\varphi = \begin{pmatrix} p_1 & \sqrt{p_1 p_2} e^{i\theta_{12}} & \sqrt{p_1 p_3} e^{i\theta_{13}} \\ \sqrt{p_1 p_2} e^{-i\theta_{12}} & p_2 & \sqrt{p_2 p_3} e^{i\theta_{23}} \\ \sqrt{p_1 p_3} e^{-i\theta_{13}} & \sqrt{p_2 p_3} e^{-i\theta_{23}} & p_3 \end{pmatrix} \quad (65)$$

satisfies the property

$$e^{i\theta_{13}} = e^{i(\theta_{12} + \theta_{23})}.$$

Equivalently,  $S_\varphi = |v\rangle\langle v|$ , where  $v \in \mathbb{C}^3$  is the vector given by  $|v\rangle := (\sqrt{p_1}, \sqrt{p_2} e^{-i\theta_{12}}, \sqrt{p_3} e^{-i\theta_{13}})^T$ .

*Proof.* The relation can be trivially satisfied when  $p_i = 0$  for some  $i \in \{1, 2, 3\}$ . Hence let us assume  $p_1, p_2, p_3 > 0$ . Computing the determinant of  $S_\varphi$  one obtains  $\det(S_\varphi) = 2p_1 p_2 p_3 [\cos(\theta_{12} + \theta_{23} - \theta_{13}) - 1]$ . Since  $S_\varphi$  is positive, we must have  $\det(S_\varphi) \geq 0$ . If  $p_1, p_2, p_3 > 0$  the only possibility is  $\theta_{13} = \theta_{12} + \theta_{23} \pmod{2\pi}$ . ■

Corollary 49 can be easily extended to systems of arbitrary dimension. To this purpose, we choose the  $x$  and  $y$  axes in

such a way that the projection of every state  $\rho \in \text{St}_1(\mathbf{A})$  on a three-dimensional face is represented by a positive matrix.

*Lemma 76.* If  $\varphi \in \text{St}_1(\mathbf{A})$  is a pure state and  $d_{\mathbf{A}} = N$ , then  $S_{\varphi} = |v\rangle\langle v|$ , where  $v \in \mathbb{C}^N$  is the vector given by  $v := (\sqrt{p_1}, \sqrt{p_2}e^{-i\alpha_2}, \dots, \sqrt{p_N}e^{-i\alpha_N})^T$  with  $\alpha_i \in [0, 2\pi) \ \forall i = 2, \dots, N$ .

*Proof.* Consider a triple  $V = \{p, q, r\} \subseteq \{1, \dots, N\}$ . Then the state  $\Pi_V|\varphi\rangle$  is proportional to a pure state of a three-dimensional system, whose representation  $S_{\Pi_V\varphi}$  is the  $3 \times 3$  square submatrix of  $S_{\varphi}$  with elements  $[S_{\varphi}]_{kl} = \sqrt{p_k p_l} e^{i\theta_{kl}}$ ,  $(k, l) \in V \times V$ . Now, corollary 49 forces the relation  $e^{i\theta_{pr}} = e^{i(\theta_{pq} + \theta_{qr})}$ . Since this relation must hold for every choice of the triple  $V = \{p, q, r\}$ , if we define  $\alpha_p := \theta_{p1}$ , then we have  $e^{i\theta_{pq}} = e^{i(\theta_{p1} + \theta_{1q})} = e^{i(\theta_{p1} - \theta_{q1})} = e^{i(\alpha_p - \alpha_q)}$ . It is then immediate to verify that  $S_{\varphi} = |v\rangle\langle v|$ , where  $v = (\sqrt{p_1}, \sqrt{p_2}e^{-i\alpha_2}, \dots, \sqrt{p_N}e^{-i\alpha_N})^T$ . ■

In conclusion, we proved the following.

*Corollary 50.* For every system  $\mathbf{A}$ , the state space  $\text{St}_1(\mathbf{A})$  can be represented as a subset of the set of density matrices in dimension  $d_{\mathbf{A}}$ .

*Proof.* For every state  $\rho \in \text{St}_1(\mathbf{A})$  the matrix  $S_{\rho}$  is Hermitian by construction, with unit trace by corollary 42, and positive since it is a convex mixture of positive matrices. ■

### E. Quantum theory in finite dimensions

Here we conclude our derivation of quantum theory by showing that every density matrix in dimension  $d_{\mathbf{A}}$  corresponds to some state  $\rho \in \text{St}_1(\mathbf{A})$ .

We already know from the superposition principle (lemma 16) that for every choice probabilities  $\{p_i\}_{i=1}^{d_{\mathbf{A}}}$  there is a pure state  $\varphi \in \text{St}_1(\mathbf{A})$  such that  $\{p_i\}_{i=1}^{d_{\mathbf{A}}}$  are the diagonal elements of  $S_{\varphi}$ . Thus the set of density matrices corresponding to pure states contains at least one matrix of the form  $S_{\varphi} = |v\rangle\langle v|$ , with  $|v\rangle = (\sqrt{p_1}, \sqrt{p_2}e^{-i\beta_2}, \dots, \sqrt{p_{d_{\mathbf{A}}}}e^{-i\beta_{d_{\mathbf{A}}}})$ . It only remains to prove that every possible choice of phases  $\beta_i \in [0, 2\pi)$  corresponds to some pure state.

Recall that for a face  $F \subseteq \text{St}_1(\mathbf{A})$  we defined the group  $\mathbf{G}_{F, F^{\perp}}$  to be the group of reversible transformations  $\mathcal{U} \in \mathbf{G}_{\mathbf{A}}$  such that  $\mathcal{U} = \omega_F \mathcal{I}_{\mathbf{A}}$  and  $\mathcal{U} = \omega_{F^{\perp}} \mathcal{I}_{\mathbf{A}}$ . We then have the following.

*Lemma 77.* Consider a system  $\mathbf{A}$  with  $d_{\mathbf{A}} = N$ . Let  $\{\varphi_i\}_{i=1}^N \subseteq \text{St}_1(\mathbf{A})$  be a maximal set of perfectly distinguishable pure states,  $F$  be the face identified by  $\omega_F = 1/(N-1) \sum_{i=1}^{N-1} \varphi_i$  and  $F^{\perp}$  its orthogonal face, identified by the state  $\varphi_N$ . If  $\mathcal{U}$  is a reversible transformation in  $\mathbf{G}_{F, F^{\perp}}$ , then the action of  $\mathcal{U}$  is given by

$$S_{\mathcal{U}\rho} = U S_{\rho} U^{\dagger} \quad U = \begin{pmatrix} & & 0 \\ & I_{N-1} & \vdots \\ 0 & \dots & 0 & e^{-i\beta} \end{pmatrix}, \quad (66)$$

where  $I_{N-1}$  is the  $(N-1) \times (N-1)$  identity matrix and  $\beta \in [0, 2\pi)$ .

*Proof.* Consider an arbitrary state  $\rho \in \text{St}_1(\mathbf{A})$  and its matrix representation

$$S_{\rho} = \left( \begin{array}{c|c} S_{\Pi_F \rho} & \mathbf{f} \\ \hline \mathbf{f}^{\dagger} & S_{\Pi_{F^{\perp}} \rho} \end{array} \right),$$

where  $\mathbf{f} \in \mathbb{C}^{N-1}$  is a suitable vector. Since  $\mathcal{U} = \omega_F \mathcal{I}_{\mathbf{A}}$  and  $\mathcal{U} = \omega_{F^{\perp}} \mathcal{I}_{\mathbf{A}}$ , we have that

$$S_{\mathcal{U}\rho} = \left( \begin{array}{c|c} S_{\Pi_F \rho} & \mathbf{g} \\ \hline \mathbf{g}^{\dagger} & S_{\Pi_{F^{\perp}} \rho} \end{array} \right),$$

where  $\mathbf{g} \in \mathbb{C}^{N-1}$  is a suitable vector. To prove Eq. (66), we will now prove that  $\mathbf{g} = e^{i\beta} \mathbf{f}$  for some suitable  $\beta \in [0, 2\pi)$ .

Let us start from the case  $N = 3$ . Since  $\mathcal{U}|\varphi_i\rangle = |\varphi_i\rangle \ \forall i = 1, 2, 3$ , we have  $(a_i|\mathcal{U} = (a_i| \ \forall i = 1, 2, 3$  (lemma 51). This implies that  $\mathcal{U}$  sends states in the face  $F_{13}$  to states in the face  $F_{13}$ : indeed, for every  $\rho \in F_{13}$  one has  $(a_{13}|\mathcal{U}|\rho) = (a_{13}|\rho) = 1$ , which implies  $\mathcal{U}\rho \in F_{13}$  (lemma 46). In other words, the restriction of  $\mathcal{U}$  to the face  $F_{13}$  is a reversible qubit transformation. Therefore, the action of  $\mathcal{U}$  on a state  $\rho \in F_{13}$  must be given by

$$S_{\mathcal{U}\rho} = \begin{pmatrix} \rho_{11} & 0 & \rho_{13}e^{i\beta} \\ 0 & 0 & 0 \\ \rho_{31}e^{-i\beta} & 0 & \rho_{33} \end{pmatrix}$$

for some  $\beta \in [0, 2\pi)$ . Similarly, we can see that  $\mathcal{U}$  sends states in the face  $F_{23}$  to states in the face  $F_{23}$ . Hence, for every  $\sigma \in F_{23}$  we have

$$S_{\mathcal{U}\sigma} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_{22} & \sigma_{23}e^{i\beta'} \\ 0 & \sigma_{32}e^{-i\beta'} & \rho_{33} \end{pmatrix}$$

for some  $\beta' \in [0, 2\pi)$ . We now show that  $e^{i\beta'} = e^{i\beta}$ . To see that, consider a generic state  $\varphi \in \text{St}_1(\mathbf{A})$ , with the property that  $p_i = (a_i|\varphi) > 0$  for every  $i = 1, 2, 3$  (such state exists due to the superposition principle of theorem 16). Writing  $S_{\varphi}$  as in Eq. (65) we then have

$$S_{\mathcal{U}\varphi} = \begin{pmatrix} p_1 & \sqrt{p_1 p_2} e^{i\theta_{12}} & \sqrt{p_1 p_3} e^{i(\theta_{13} + \beta)} \\ \sqrt{p_1 p_2} e^{-i\theta_{12}} & p_2 & \sqrt{p_2 p_3} e^{i(\theta_{23} + \beta')} \\ \sqrt{p_1 p_3} e^{-i(\theta_{13} + \beta)} & \sqrt{p_2 p_3} e^{-i(\theta_{23} + \beta')} & p_3 \end{pmatrix}.$$

Now since  $\varphi$  and  $\mathcal{U}\varphi$  are pure states, by corollary 49 we must have

$$e^{i\theta_{13}} = e^{i(\theta_{12} + \theta_{23})}, \\ e^{i(\theta_{13} + \beta)} = e^{i(\theta_{12} + \theta_{23} + \beta')}.$$

By comparison we obtain  $e^{i\beta} = e^{i\beta'}$ . This proves Eq. (66) for  $N = 3$ . The proof for  $N > 3$  is then immediate: for every three-dimensional face  $F_{pqN}$  the action of  $\mathcal{U}$  is given Eq. (66) for some  $\beta_{pq}$ . However, since the two faces  $F_{pqN}$  and  $F_{p'q'N}$  overlap on  $\varphi_p$  we must have  $\beta_{pq} = \beta_{p'q'}$ . Similarly  $\beta_{pq} = \beta_{p'q}$ . We conclude that  $\beta_{pq} = \beta$  for every  $p, q$ . This proves Eq. (66) in the general case. ■

We now show that every possible phase shift in Eq. (66) corresponds to a physical transformation.

*Lemma 78.* A transformation  $\mathcal{U}$  of the form of Eq. (66) is a reversible transformation for every  $\beta \in [0, 2\pi)$ .

*Proof.* By lemma 77, the group  $\mathbf{G}_{F, F^{\perp}}$  is a subgroup of  $U(1)$ . Now, there are two possibilities: either  $\mathbf{G}_{F, F^{\perp}}$  is a (finite) cyclic group or  $\mathbf{G}_{F, F^{\perp}}$  coincides with  $U(1)$ . However, we know from theorem 15 that  $\mathbf{G}_{F, F^{\perp}}$  has a continuum of elements. Hence,  $\mathbf{G}_{F, F^{\perp}} \simeq U(1)$  and  $\beta$  can take every value in  $[0, 2\pi)$ . ■

An obvious corollary of the previous lemmas is the following.

*Corollary 51.* The transformation  $\mathcal{U}_\beta$  defined by

$$S_{\mathcal{U}_\beta \rho} = U S_\rho U^\dagger, \quad (67)$$

where  $U$  is the diagonal matrix with diagonal elements  $(1, e^{i\beta_1}, \dots, e^{i\beta_{N-1}})$  is a reversible transformation for every vector  $\beta := (\beta_2, \dots, \beta_N) \in [0, 2\pi) \times \dots \times [0, 2\pi)$ .

This leads directly to the conclusion of our derivation.

*Theorem 20.* For every system  $A$ , the state space  $\text{St}_1(A)$  is the set of all density matrices on the Hilbert space  $\mathbb{C}^{d_A}$ .

*Proof.* Let  $N = d_A$ . For every choice of probabilities  $\mathbf{p} = (p_1, \dots, p_N)$  there exists at least one pure state  $\varphi_{\mathbf{p}}$  such that  $p_k = (a_k | \varphi_{\mathbf{p}})$  for every  $k = 1, \dots, N$  (lemma 16). This state is represented by the matrix  $S_{\varphi_{\mathbf{p}}} = |v_{\mathbf{p}}\rangle\langle v_{\mathbf{p}}|$  with  $|v_{\mathbf{p}}\rangle = (\sqrt{p_1}, \sqrt{p_2}e^{-i\alpha_2}, \dots, \sqrt{p_N}e^{-i\alpha_N})^T$  (lemma 76). Finally, we can transform  $\varphi_{\mathbf{p}}$  with every reversible transformation  $\mathcal{U}_\beta$  defined in Eq. (67), thus obtaining  $S_{\mathcal{U}_\beta \varphi_{\mathbf{p}}} = U_\beta |v_{\mathbf{p}}\rangle\langle v_{\mathbf{p}}| U_\beta^\dagger$ , where  $U_\beta |v_{\mathbf{p}}\rangle = (\sqrt{p_1}, \sqrt{p_2}e^{-i(\alpha_2+\beta_2)}, \dots, \sqrt{p_N}e^{-i(\alpha_N+\beta_N)})^T$ . Since  $\mathbf{p}$  and  $\beta$  are arbitrary, this means that every rank-one density matrix corresponds to some pure state. Taking the possible convex mixtures we obtain that every  $N \times N$  density matrix corresponds to some state of system  $A$ . ■

Choosing a suitable representation  $\rho \mapsto S_\rho$ , we proved that for every system  $A$  the set of normalized states  $\text{St}_1(A)$  is the whole set of density matrices in dimension  $d_A$ . Thanks to the purification postulate, this is enough to prove that all the effects  $\text{Eff}(A)$  and all the transformations  $\text{Transf}(A, B)$  allowed in our theory are exactly the effects and the transformations allowed in quantum theory. Precisely we have the following.

*Corollary 52.* For every couple of systems  $A$  and  $B$  the set of physical transformations  $\text{Transf}(A, B)$  coincides with the set of all completely positive trace nonincreasing maps from  $M_{d_A}(\mathbb{C})$  to  $M_{d_B}(\mathbb{C})$ .

*Proof.* We proved that our theory has the same normalized states of quantum theory. On the other hand, quantum theory is a theory with purification and in quantum theory the possible physical transformations are quantum operations, that is, completely positive trace-preserving maps. The thesis then follows from the fact that two theories with purification that have the same set of normalized states are necessarily the same (theorem 3). ■

#### XIV. CONCLUSION

Quantum theory can be derived from purely informational principles. In particular, it belongs to a broad class of theories of information processing that includes classical and quantum information theory as special cases. Within this class, quantum theory is identified uniquely by the purification postulate, stating that the ignorance about a part is always compatible with the maximal knowledge of the whole in an essentially unique way. This postulate appears as the origin of the key features of quantum information processing, such as no-cloning, teleportation, and error correction (see also Ref. [22]). The general vision underlying the present work is that the main primitives of quantum information processing should be derived directly from the principles, without the abstract mathematics of Hilbert spaces, in order to make the

revolutionary aspects of quantum information immediately accessible and to place them in the broader context of the fundamental laws of physics.

Finally, we would like to comment on possible generalizations of our work. As in any axiomatic construction, one can ask how the results change when the principles are modified. For example, one may be interested in relaxing the local distinguishability axiom and in considering theories, like quantum theory on real Hilbert spaces, where global measurements are essential to characterize the state of a composite system. In this direction, the results of Ref. [22] suggest that also quantum theory on real Hilbert spaces can be derived from the purification principle, after that the local distinguishability requirement has been suitably relaxed. A possible way to weaken the local distinguishability requirement is to assume only the property of *local distinguishability from pure states* proposed in Ref. [22]: this property states that the probability of distinguishing two states by local measurements is larger than 1/2 whenever one of the two states is pure. A different way to relax local distinguishability would be to assume the property of *2-local tomography* proposed in Ref. [39], which requires that the state of a multipartite system can be completely characterized using only measurements on bipartite subsystems. This property is equivalent to *2-local distinguishability*, defined as the requirement that two different states of a multipartite system can be distinguished with probability of success larger than 1/2 using only local measurements or measurements on bipartite subsystems.

A more radical generalization of our work would be to relax the assumption of causality. This would be particularly important for the discussion of quantum gravity scenarios, where the causal structure is not given *a priori* but is part of the dynamical variables of the theory. In this respect, the contribution of our work is twofold. First, it makes evident how fundamental is the assumption of causality in the ordinary formulation of quantum theory: the whole formalism of quantum states as density matrices with unit trace, quantum measurements as resolutions of the identity, and quantum channels as trace-preserving maps is crucially based on it. Technically speaking, the fact that the normalization of a state is given by a single linear functional (the trace, in quantum theory) is the signature of causality. This partly explains the troubles and paradoxes encountered when trying to combine the formalism of density matrices with noncausal evolutions, as in Deutsch's model for close time-like curves [40,41]. Moreover, given that the usual notion of normalization has to be abandoned in the noncausal scenario, and that the ordinary quantum formalism becomes inadequate, one may ask in what sense a theory of quantum gravity would be "quantum". The suggestion coming from our work is that a quantum theory is a theory satisfying the purification principle, which can be suitably formulated even in the absence of causality [42]. The discussion of theories with purification in the noncausal scenario is an exciting avenue of future research.

#### ACKNOWLEDGMENTS

The authors are grateful to S Flammia for adding to the package qcircuits the commands needed for the

diagrammatic representation of states and effects. Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI. Research

at QUIT is supported by the University of Pavia and by INFN. Perimeter Institute and INFN supported the mutual exchange of scientific visits that made this work possible.

- 
- [1] D. Hilbert, J. von Neumann, and L. Nordheim, *Math. Ann.* **98**, 1 (1928).
- [2] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1932).
- [3] G. Birkhoff, in *Lattice Theory*, edited by R. P. Dilworth, Proceedings of the Second Symposium in Pure Mathematics of the American Mathematical Society (American Mathematical Society, Providence, 1961).
- [4] G. Birkhoff and J. von Neumann, *Ann. Math.* **37**, 743 (1936).
- [5] G. W. Mackey, *The Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, New York, 1963).
- [6] J. M. Jauch and C. Piron, *Helv. Phys. Acta* **36**, 837 (1963).
- [7] B. Coecke, D. Moore, and A. Wilce, *Current Research in Operational Quantum Logic: Algebras, Categories, Languages*, Fundamental theories of physics series (Kluwer Academic, Amsterdam, 2000).
- [8] G. Ludwig, *Foundations of Quantum Mechanics* (Springer, New York, 1983 and 1985), Vols 1 and 2.
- [9] W. K. Wootters and W. H. Zurek, *Nature (London)* **299**, 802 (1982).
- [10] D. Dieks, *Phys. Lett. A*, **92**, 271 (1982).
- [11] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Phys. Rev. Lett.* **70**, 1895 (1993).
- [12] S. Wiesner, *Sigact News* **15**, 78 (1983).
- [13] C. H. Bennett and G. Brassard, *Proceedings IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), pp. 175–179.
- [14] A. K. Ekert, *Phys. Rev. Lett.* **67**, 661 (1991).
- [15] P. W. Shor, *SIAM J. Comput.* **26**, 1484 (1997).
- [16] L. Hardy, e-print [arXiv:quant-ph/0101012](https://arxiv.org/abs/quant-ph/0101012).
- [17] B. Dakic and C. Bruckner, e-print [arXiv:0911.0695](https://arxiv.org/abs/0911.0695).
- [18] L. Masanes and M. Müller, e-print [arXiv:1004.1483v2](https://arxiv.org/abs/1004.1483v2).
- [19] G. M. D’Ariano, *AIP Conf. Proc.* **810**, 114 (2006); **844**, 101 (2006); **889**, 79 (2007); *Proceedings of the 8th International Conference on Quantum Communication, Measurement and Computing* (NICT, Japan, 2007), p. 345.
- [20] G. M. D’Ariano, in *Philosophy of Quantum Information and Entanglement*, edited by A. Bokulich and G. Jaeger (Cambridge University Press, Cambridge, 2010).
- [21] J. A. Wheeler, in *Complexity, Entropy, and the Physics of Information*, edited by W. Zurek (Addison-Wesley, Redwood City, CA, 1990).
- [22] G. Chiribella, G. M. D’Ariano, and P. Perinotti, *Phys. Rev. A* **81**, 062348 (2010).
- [23] E. Schrödinger, *Proc. Cambridge Philos. Soc.* **31**, 555 (1935).
- [24] In fact, the equivalence of all purifications under local reversible transformations is tightly related to what seems the main point of Schrödinger’s paper, namely that all possible ensembles of a mixed state can be “steered” by measurements on the purifying system. This property is an immediate consequence of the equivalence of all purifications.
- [25] W. Arveson, *An Invitation to C\*-Algebras* (Springer, New York, 1976).
- [26] W. F. Stinespring, *Proc. Am. Math. Soc.* **6**, 211 (1955).
- [27] M. A. Nielsen and I. L. Chuang, *Phys. Rev. Lett.* **79**, 321 (1997).
- [28] In quantum theory the trivial system I is represented by the one-dimensional Hilbert space  $\mathcal{H} \simeq \mathbb{C}$ .
- [29] B. Coecke, in *Quantum Theory: Reconsiderations of the Foundations III* (AIP, New York, 2005), p. 81.
- [30] B. Coecke, *Contemp. Phys.* **51**, 59 (2010).
- [31] G. Chiribella, G. M. D’Ariano, and P. Perinotti, *Europhys. Lett.* **83**, 30004 (2008).
- [32] G. Chiribella, G. M. D’Ariano, and P. Perinotti, *Phys. Rev. A* **80**, 022339 (2009).
- [33] L. Hardy, *J. Phys. A* **40**, 3081 (2007).
- [34] L. Hardy, e-print [arXiv:1005.5164](https://arxiv.org/abs/1005.5164).
- [35] B. Coecke and R. Lal, in *Proceedings of the 7th workshop on Quantum Physics and Logic*, edited by B. Coecke, P. Panangaden, and P. Selinger [[http://www.cs.ox.ac.uk/people/bob.coecke/QPL\\_proceedings.html](http://www.cs.ox.ac.uk/people/bob.coecke/QPL_proceedings.html)] (unpublished).
- [36] J. Barrett, *Phys. Rev. A* **75**, 032304 (2007).
- [37] A. Wilce, *Found. Phys.* **40**, 434 (2010).
- [38] In principle, the purifying system for the state  $R_i$  might depend on the value of  $i$ , let us call it  $C_i$ . However, since there is a finite number of outcomes one can always choose a single purifying system by defining the composite system  $C = C_1 \cdots C_N$  (here we labeled the outcomes with the natural numbers  $\{1, \dots, N\}$ ). If  $\Psi_i \in \text{St}(AC_i)$  is a purification of  $R_i$  and  $\psi_i$  is any pure state on the composite system including all system but the  $i$ th, then also  $\Psi'_i = \Psi_i \otimes \psi_i \in \text{St}(AC)$  is a purification of  $R_i$ .
- [39] L. Hardy and W. K. Wootters, e-print [arXiv:1005.4870](https://arxiv.org/abs/1005.4870).
- [40] D. Deutsch, *Phys. Rev. D* **44**, 3197 (1991).
- [41] C. H. Bennett, D. Leung, G. Smith, and J. A. Smolin, *Phys. Rev. Lett.* **103**, 170502 (2009).
- [42] G. Chiribella, G. M. D’Ariano, and P. Perinotti (in preparation).