

Structure of quantum and classical implementations of Popescu-Rohrlich box

Anna Jenčová

Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, Bratislava, Slovakia

Martin Plávala*

*Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, Bratislava, Slovakia and
Naturwissenschaftlich-Technische Fakultät, Universität Siegen, 57068 Siegen, Germany*

We construct implementations of the PR-box using quantum and classical channels as state spaces. In both cases our constructions are very similar and they share the same idea taken from general probabilistic theories and the square state space model. We construct all quantum qubit channels that maximally violate a given CHSH inequality, we show that they all are entanglement-breaking channels, that they have certain block-diagonal structure and we present some examples of such channels.

I. INTRODUCTION

The Bell non-locality is a well known topic in quantum theory, yet we still lack full understanding of its implications. The research of Bell non-locality was inspired by the famous paradox of Einstein, Podolsky and Rosen [1] that questioned the completeness of quantum theory.

As it was later shown by Bell [2] the EPR paradox does not question the completeness of quantum theory, but it rather separates it from any other classical theory. This was demonstrated by the well-known Bell inequalities, that constrain any classical theory but are violated by quantum theory. Probably the most well known and most studied of Bell inequalities is the CHSH inequality [3] that can be violated by quantum states and measurements, but this violation is constrained by the Tsirelson bound [4]. On one hand this shows that quantum theory is more non-local than any classical theory, on the other hand, Popescu and Rohrlich proved that the CHSH inequality may be violated even more by non-signaling correlations [5, 6]. Since then, it is a long-standing question whether the Tsirelson bound does have any operational meaning for quantum theory.

It was later shown by Barrett that the set of correlations identified in [5], now often called the PR-box, can be generated by a non-signaling theory [7], most commonly known as the Boxworld GPT. A realization of the PR-box in the real world would have several rather interesting implications, as shown in [8–10], see also [11] for a review.

It has been demonstrated several times that non-signaling classical and quantum channels provide realizations of the PR box [12–15]. As described in [14], such constructions can be put into the framework of general probabilistic theories (or GPT for short), in the following way: since the set of classical (or quantum) channels is compact and convex, it can be seen as a state space of some GPT. One can see non-signaling channels as elements of a joint state space, equivalent to entangled

states in quantum theory, and one can describe measurement procedures and CHSH experiments within this theory, yielding maximal violation of the CHSH inequality.

One has to be careful when using channels in this way. The formalism implies that we are able to use the channel only once, hence one can exploit the input state incompatibility of measurements on channels [16]. From a realistic view-point it is important to remember that using a non-local channel may take some time and resources, but the channel is considered to permit communication only if the channel is signaling.

In the present paper, we focus on the structure of the non-signaling channels such that the CHSH inequality is maximally violated for some choice of channel measurements. Such channels will be called the PR-channels. All the PR-channels obtained so far are in fact classical-to-classical and it is natural to ask whether there are some truly quantum non-signaling PR-channels. Another important question is the possibility of instantaneous implementation of such channels. Since PR-channels are non-signaling, their instantaneous implementation is not forbidden by special theory of relativity simply because no information is transferred, yet it is believed that such implementations do not exist.

We make a step toward addressing these questions, in that we present a characterization of the structure of all implementations of the PR-box in the framework of GPTs, especially for theories in which classical and quantum channels play the role of states. In any GPT, the pairs of measurements appearing in such implementations must be maximally incompatible and we show how the corresponding bipartite states are constructed from such pairs.

We apply the obtained results mainly to qubit bipartite non-signaling channels, where both parts of the input and output are qubit spaces. Here we give a full description of all possible pairs of maximally incompatible two-outcome channel measurements and of all qubit PR-channels. In particular, we prove that all these channels are necessarily entanglement-breaking. We believe that our results will bring more insight into the structure of PR-channels, in particular to the question of existence

* martin.plavala@mat.savba.sk

of their instantaneous implementation.

The article is organized as follows: in Sec. II we give a brief overview of general probabilistic theories as it will be used in later calculations. In Sec. III we present the method of finding all bipartite non-signaling states that maximally violate the CHSH inequality and show that maximally incompatible measurements are necessary, the main result is stated as Thm. 1. In Sec. IV, we present the (known) PR-box implementation by classical channels in the light of the results of Sec. III. The main purpose of Section V is to introduce the GPT of quantum channels and show how the known PR-box implementations (and their slight generalizations) are obtained by our construction. In Sec. VI we derive the structure of all qubit PR-channels. The main result is in Thm. 2 where we prove that all qubit PR-channels must be entanglement-breaking and we provide some examples. Some more technical proofs can be found in the appendices.

II. OVERVIEW OF GENERAL PROBABILISTIC THEORIES AND TENSOR PRODUCTS

General probabilistic theories (GPTs for short) provide a framework that uses operational axioms to describe various possible physical theories. GPTs include the classical and the quantum theory, hence this setting allows us to compare these two theories as well as to construct theories that are different from both. Among other things, GPTs provide a framework to describe measurements of a physical systems in a general and mathematically clear way. We briefly introduce the formalism below, which allows us to obtain general results applicable to various state spaces. Since it will be sufficient for all of our calculations, we will only consider theories with finite dimensional state spaces. This section is only intended to settle the notation that we use; for a full review of GPTs we refer the reader to [17]. A nice introduction to GPTs, including a historical account, can be found in [18]

The main idea behind GPTs is the following: both in (finite-dimensional) classical and quantum theory, the state space (i.e. the set of all preparation procedures) is a compact convex subset of a real finite-dimensional vector space. We will generalize both theories by assuming that the state space is some compact convex subset of a finite-dimensional vector space.

Any GPT can be described as a set of physical systems it contains. In operational terms, any system is determined by the set of allowed preparation procedures (states) and yes-no experiments (effects).

The set of states has a natural convex structure, because if x and y are preparable states of a system, then we can randomize the preparation to prepare the convex combination $\lambda x + (1 - \lambda)y$, which we again postulate to be a viable state. One can also argue that the state space must be closed (in a for now unspecified topology) as if we can prepare a series of states x_n converging to x , then

we should be also able to prepare x (albeit with infinite resources).

An effect is described by an affine function mapping states to probabilities of the "yes" outcome, represented by the interval $[0, 1]$. Recall that a function f is affine if it respects the convex structure: $f(\lambda x + (1 - \lambda)y) = \lambda f(x) + (1 - \lambda)f(y)$, for any states x and y and any $\lambda \in [0, 1]$. We will require that for any two distinct states there is an effect that distinguished these states strictly better than a random guess. This implies that the state space must be bounded. Indeed, if the state space would have a direction of recession in which it would go to infinity (i.e. it would be unbounded), any effect would have to be zero in this direction, so we would be unable to distinguish these states.

We also require the state space to be embedded in a real finite-dimensional vector space equipped with the Euclidean topology. This last assumption is only practical as it allows us to stick only to rather simple mathematics. Note that it is well known that any bounded and closed subset of a finite-dimensional real vector space is compact, hence the state space is compact.

A. Structure of GPTs

As we have seen above, the basic mathematical framework for GPT consists of compact convex sets in finite dimensional Euclidean spaces and affine functions on them. Below, we introduce the notations and further assumptions, applied throughout the paper.

Let V be a finite dimensional real vector space with the standard Euclidean topology and let $K \subset V$ be a compact convex set. Let $A(K)$ denote the linear space of real-valued affine functions on K . We will denote constant functions by the value they attain. Let $f, g \in A(K)$, then we introduce an ordering to $A(K)$ as follows: $f \geq g$ if and only if for every $x \in K$ we have $f(x) \geq g(x)$. Let $A(K)^+ = \{f \in A(K) : f \geq 0\}$ denote the convex, closed, generating, pointed cone of positive functions and let $E(K) = \{f \in A(K) : 0 \leq f \leq 1\}$ denote the set of effects on K , called the effect algebra. The effect algebra $E(K)$ is important in GPTs since (as will be explained below) the effects describe two-outcome measurements of the theory.

Denote $A(K)^*$ the dual of $A(K)$ and denote $A(K)^{**}$ the positive cone dual to $A(K)^+$, i.e.

$$A(K)^{**} = \{\psi' \in A(K)^* : \psi'(f) \geq 0, \forall f \in A(K)^+\}.$$

The cone $A(K)^{**}$ gives rise to an ordering on $A(K)^*$: let $\psi, \varphi \in A(K)^*$, then $\psi \geq \varphi$ if and only if $\psi - \varphi \in A(K)^{**}$, i.e. if $\psi - \varphi \geq 0$. The state space K is affinely isomorphic to the subset $\{\psi' \in A(K)^{**} : \psi'(1) = 1\}$, see [19, Chapter 1, Theorem 4.3]. For simplicity we will omit this isomorphism and we will treat K as a subset of $A(K)^*$. It follows that any $0 \neq \psi \in A(K)^{**}$ can be expressed uniquely as $\psi = \alpha x$ for $\alpha > 0$ and $x \in K$. The cone $A(K)^{**}$ is generating, so we can express every

$\psi \in A(K)^*$ as $\psi = \alpha x - \beta y$ for some $x, y \in K$ and $\alpha, \beta \in \mathbb{R}$, $\alpha, \beta \geq 0$.

Now we will present a simple definition of a two-outcome measurement in GPTs. Generally speaking, a measurement is a procedure that assigns probabilities to possible measurement outcomes. For simplicity we will restrict ourselves to outcomes labeled by the numbers $-1, 1$. Let A be such a measurement and let $P_x(\epsilon|A)$ denote the probability of obtaining the outcome labeled as $\epsilon \in \{-1, 1\}$ when we measure a system in the state $x \in K$. Since the outcome probabilities must respect probabilistic mixtures, the map $x \mapsto P_x(1|A)$ is an effect and $x \mapsto P_x(-1|A) = 1 - P_x(1|A)$ is an effect as well. Although in general the measurements in the theory may be restricted, in this work we will assume that any effect $f \in E(K)$ gives rise to such a measurement, determined for $x \in K$ as

$$\begin{aligned} P_x(1|A) &= f(x), \\ P_x(-1|A) &= (1 - f)(x) = 1 - f(x). \end{aligned}$$

Such an assumption is called the No-Restriction hypothesis [20], see also [21] for a recent treatment.

Measurements with a finite number of outcomes are similarly described by collections of effects $f_i \in E(K)$, such that $\sum_i f_i = 1$. For a more general treatment of measurements see e.g. [22, Section 2.2].

We now list some examples of state spaces, and relations between them, that are basic for the present work.

Example 1 (Classical bit). The simplest (nontrivial) example is the 1-dimensional simplex \mathfrak{S}_C . Let the extreme points (deterministic states) in \mathfrak{S}_C be denoted by s_0, s_1 , then $\mathfrak{S}_C = \text{conv}(s_0, s_1)$. Since all affine functionals are fully determined by their values at s_0, s_1 , we have $A(\mathfrak{S}_C) \simeq \mathbb{R}^2$ and the set of effects is identified with $E(\mathfrak{S}_C) \simeq [0, 1]^2$. Let $\pi \in E(\mathfrak{S}_C)$ be the effect determined by $\pi(s_0) = 0$, $\pi(s_1) = 1$, then $\pi : \mathfrak{S}_C \rightarrow [0, 1]$ is an affine isomorphism.

Example 2 (Quantum state space). Let \mathcal{H} be a finite dimensional complex Hilbert space. Let $B_h(\mathcal{H})$ denote the set of self-adjoint operators on \mathcal{H} and let $\mathfrak{S}_{\mathcal{H}} = \{\rho \in B_h(\mathcal{H}) : \rho \geq 0, \text{Tr}(\rho) = 1\}$ denote the set of density operators (or states) on \mathcal{H} , where $\rho \geq 0$ means that ρ is positive semi-definite and $\text{Tr}(\rho)$ denotes the trace of ρ . Let $\mathbb{1}$ denote the identity operator. Then $A(\mathfrak{S}_{\mathcal{H}}) \simeq B_h(\mathcal{H})$ and $E(\mathfrak{S}_{\mathcal{H}}) \simeq E(\mathcal{H})$, where $E(\mathcal{H}) = \{0 \leq E \leq \mathbb{1}, E \in B(\mathcal{H})\}$ is the set of quantum effects. The quantum measurements will be thus described by collections E_1, \dots, E_n of positive operators such that $\sum_i E_i = \mathbb{1}$, such a collection is called a POVM [23]. We will mostly treat the qubit case, that is $\dim(\mathcal{H}) = 2$.

Example 3 (Square state space). This state space is also called the gbit and appears as the state space of systems in a theory called GNST introduced in [7], now also known as the Boxworld GPT. This theory is one of the most commonly used examples of a GPT other than classical or quantum theories. The square S is a state space

with four extreme points $s_{00}, s_{10}, s_{01}, s_{11}$, such that

$$\frac{1}{2}(s_{00} + s_{11}) = \frac{1}{2}(s_{10} + s_{01}).$$

Let $\pi_0, \pi_1 \in E(S)$ be given, for $i, j \in \{0, 1\}$, as

$$\pi_0(s_{ij}) = i, \quad \pi_1(s_{ij}) = j,$$

then $A(S) = \text{span}(\{\pi_0, \pi_1, 1\})$ and $E(S) = \text{conv}(\{\pi_0, 1 - \pi_0, \pi_1, 1 - \pi_1, 0, 1\})$. Moreover, we can see that $\{s_{00}, s_{10} - s_{00}, s_{01} - s_{00}\}$ forms a linear basis of $A(S)^*$, which is dual to $\{\pi_0, \pi_1, 1\}$. Note also that the map

$$s \mapsto (\pi_0(s), \pi_1(s)), \quad s \in S$$

is an affine isomorphism of S onto $[0, 1]^2 \simeq \mathfrak{S}_C \times \mathfrak{S}_C$, its inverse is given by

$$(\lambda, \mu) \mapsto s_{00} + \lambda(s_{10} - s_{00}) + \mu(s_{01} - s_{00}), \quad \lambda, \mu \in [0, 1].$$

The notations π, π_0, π_1 introduced in the above examples will be kept throughout.

Example 4 (Classical bit channels). A channel $\Phi : \mathfrak{S}_C \rightarrow \mathfrak{S}_C$ is defined as an affine map of \mathfrak{S}_C into itself; a simple example of a channel is the identity channel $\text{id} : \mathfrak{S}_C \rightarrow \mathfrak{S}_C$. Let \mathcal{C}_C be the set of all channels (on the classical bit), then since $\pi \circ \Phi$ is an effect on \mathfrak{S}_C for every $\Phi \in \mathcal{C}_C$ and π is an isomorphism, we see that

$$\mathcal{C}_C \simeq E(\mathfrak{S}_C) \simeq [0, 1]^2 \simeq S.$$

Through this isomorphism, we establish a relation between classical bit channels and the Boxworld GPT.

The effect π_0 (π_1) on S corresponds to the effect in $E(\mathcal{C}_C)$, determined as $\Phi \mapsto \pi(\Phi(s_0))$ ($\Phi \mapsto \pi(\Phi(s_1))$). More generally, each effect in $E(\mathcal{C}_C)$ is a convex combination of effects of the form

$$F_{t_i, f_i}(\Phi) := f_i(\Phi(t_i)),$$

where $t_i \in \mathfrak{S}_C$ and $f_i \in E(\mathfrak{S}_C)$ (clearly, we may restrict to the extreme points $t_i \in \{s_0, s_1\}$). The corresponding measurement can be seen as a protocol where we choose the state t_i with some probability λ_i , input it into the measured channel Φ and apply the effect f_i to the output of the channel. For such a measurement A and a channel $\Phi \in \mathcal{C}_C$, we have

$$P_{\Phi}(1|A) = \sum_i \lambda_i P_{\Phi(t_i)}(1|M_i),$$

where M_i is the measurement given by f_i , see also Fig. 1.

Note that we can obtain similar relations for finite dimensional simplices, their products and channels between them. The (for us) most important example of quantum channels will be described later in Section V.

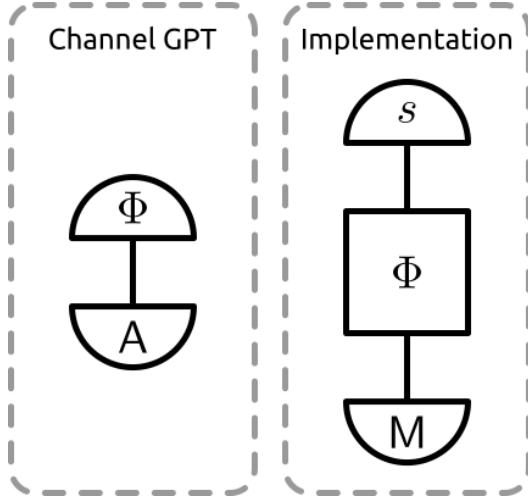


Figure 1. Two equivalent ways of describing the same measurement A on the channel $\Phi \in \mathcal{C}_C$. On the left we have the viewpoint of the GPT from example 4 where channels are considered as states and the channel is measured. On the right hand side we have the implementation of the measurement, where $s \in \mathfrak{S}_C$ is used as an input into the channel Φ and the outcome, which is $\Phi(s) \in \mathfrak{S}_C$ is measured by M .

B. Incompatible measurements in GPT

Let A, A' be two-outcome measurements on K given by the effects $f_A, f_{A'} \in E(K)$ respectively, then we say that they are compatible if there is a four-outcome measurement G with outcomes $(-1, -1), (-1, 1), (1, -1), (1, 1)$ such that for all $x \in K$ we have

$$\begin{aligned} P_x(1|A) &= P_x((1, 1)|G) + P_x((-1, 1)|G), \\ P_x(1|A') &= P_x((1, 1)|G) + P_x((-1, 1)|G). \end{aligned}$$

Equivalently, measurements A and A' are compatible if and only if there is $p \in E(K)$ such that

$$\begin{aligned} f_A &\geq p \\ f_{A'} &\geq p \\ 1 + p &\geq f_A + f_{A'} \end{aligned}$$

see [24] for a proof. This definition of compatibility of measurements generalizes to GPTs the well-known notion of compatibility (or joint measurability) of POVMs in quantum theory, which itself generalizes the notion of commutativity of projective measures, see [25] for a review.

The degree of compatibility of A and A' is defined as

$$\begin{aligned} \text{DegCom}(A, A') &= \sup_{T, T'} \{ \lambda \in [0, 1] : \lambda A + (1 - \lambda T), \\ &\quad \lambda A' + (1 - \lambda T') \text{ are compatible} \}, \end{aligned}$$

where T and T' are trivial or coin-toss measurements determined by constant effects $\mu 1, \mu' 1$ for $\mu, \mu' \in [0, 1]$. It is known that we always have $\text{DegCom}(A, A') \geq \frac{1}{2}$

[26] and we say that the measurements A and A' are maximally incompatible if $\text{DegCom}(A, A') = \frac{1}{2}$.

By [27], the effects $f_A, f_{A'} \in E(K)$ correspond to maximally incompatible measurements A, A' if and only if there are four points $x_{00}, x_{10}, x_{01}, x_{11} \in K$ satisfying

$$\frac{1}{2}(x_{00} + x_{11}) = \frac{1}{2}(x_{10} + x_{01}) \quad (1)$$

and such that

$$f_A(x_{00}) = f_A(x_{01}) = f_{A'}(x_{00}) = f_{A'}(x_{10}) = 0, \quad (2)$$

$$f_A(x_{10}) = f_A(x_{11}) = f_{A'}(x_{01}) = f_{A'}(x_{11}) = 1. \quad (3)$$

We will call such a set of points x_{ij} a witness square for $f_A, f_{A'}$, or equivalently for A, A' .

Example 5. Let A_0 denote the two-outcome measurement on the square determined by effect π_0 and let A_1 be determined by the effect π_1 , see Example 3. It is immediate that A_0 and A_1 are maximally incompatible, the witness square being formed by the extreme points $\{s_{ij}\}$. The existence of maximally incompatible measurements on S was first observed in [26].

We now present an equivalent characterization of maximally incompatible measurements which shows that all such pairs are, in some sense, isomorphic to the pair A_0, A_1 on S . This characterization was proved in a more general form in [28, Corollary 5], we include a short proof for the convenience of the reader.

Proposition 1. *The measurements A, A' on the state space K corresponding to the effects $f_A, f_{A'} \in E(K)$ respectively are maximally incompatible if and only if there are affine maps $\iota : S \rightarrow K$ and $\Pi : K \rightarrow S$ such that $\Pi \circ \iota = \text{id}$ and for $i, j \in \{0, 1\}$ we have*

$$f_A(\iota(s_{ij})) = \pi_0(s_{ij}), \quad (4)$$

$$f_{A'}(\iota(s_{ij})) = \pi_1(s_{ij}). \quad (5)$$

Proof. Assume that the maps with said properties exist, then the points $\iota(s_{ij}) \in K$ form the witness square which implies that the measurements are maximally incompatible.

If the measurements are maximally incompatible, then let $x_{ij} \in K$, $i, j \in \{0, 1\}$, be the corresponding witness square. Now define the maps $\iota : S \rightarrow K$ and $\Pi : K \rightarrow S$ as follows: ι is determined by the property that the images of the vertices of S form the witness square:

$$\iota(s_{ij}) = x_{ij} \quad (6)$$

and for $x \in K$ we put

$$\Pi(x) = f_A(x)(s_{10} - s_{00}) + f_{A'}(x)(s_{01} - s_{00}) + s_{00}. \quad (7)$$

We have

$$\Pi(\iota(s_{ij})) = \Pi(x_{ij}) = s_{ij}$$

and so $\Pi \circ \iota = \text{id}$ follows as well. Note that Eq. (4) and (5) are satisfied simply because $\iota(s_{ij}) = x_{ij}$. \square

Note that the maps in Prop. 1 can be uniquely extended to positive maps $\iota : A(S)^* \rightarrow A(K)^*$, $\Pi : A(K)^* \rightarrow A(S)^*$, that is, linear maps preserving the positive cones. Note that the map Π is uniquely determined by the given pair of measurements while ι is given by the choice of the witness square, which may be non-unique.

Corollary 1. *Let A, A' and B, B' be two pairs of maximally incompatible measurements on a state space K defined by the effects $f_A, f_{A'}, f_B, f_{B'}$. Let Π_A, ι_A and Π_B, ι_B be the corresponding maps. Then there is an affine map $T : K \rightarrow K$, such that $f_B = f_A \circ T$, $f_{B'} = f_{A'} \circ T$ and*

$$\iota_A = T \circ \iota_B, \quad \Pi_B = \Pi_A \circ T.$$

Proof. Put $T = \iota_A \circ \Pi_B$, all the properties are checked straightforwardly. \square

Remark 1. Note that the above results imply that $\{A_0, A_1\}$ is, up to affine isomorphisms, the unique maximally incompatible pair of two-outcome measurements on S .

C. Tensor products and bipartite systems in GPTs

To describe the state spaces of composite systems in GPTs, we need the notion of a tensor product of the state spaces. For simplicity, we will only consider the tensor product of a state space K with itself.

There are several ways to define the tensor product of compact convex sets, but there is a minimal and a maximal one. All of the possible tensor products are compact convex subsets in the tensor product $A(K)^* \otimes A(K)^*$.

The minimal tensor product, denoted by $K \dot{\otimes} K$, is the convex hull of the points of the form $x \otimes y$ for $x, y \in K$, i.e.

$$K \dot{\otimes} K = \text{conv}(\{x \otimes y : x, y \in K\}).$$

In other words, this is the smallest composite state space containing all locally prepared states. The maximal tensor product, denoted $K \hat{\otimes} K$, is the state space of all non-signaling states, that is

$$K \hat{\otimes} K = \{\psi \in A(K)^* \otimes A(K)^* : \psi(f \otimes g) \geq 0 \\ \forall f, g \in A(K)^+, (1 \otimes 1)(\psi) = 1\}.$$

To give more insight into these definitions, we will look at state spaces of quantum theory (Example 2). In this case, $\mathfrak{S}_{\mathcal{H}} \dot{\otimes} \mathfrak{S}_{\mathcal{H}}$ is the set of all separable states and $\mathfrak{S}_{\mathcal{H}} \hat{\otimes} \mathfrak{S}_{\mathcal{H}}$ is the set of all (normalized) entanglement witnesses [23, Definition 6.38].

Every GPT in which we want to describe bipartite systems must come equipped with a composition rule on how to form the joint state space of two (or more) systems. We will denote the joint state space as $K \tilde{\otimes} K$ and it will represent the set of all bipartite states for the given system. Note that $K \tilde{\otimes} K$ does not have any general definition as it is specified by the theory we are working with.

Some properties of the composition rule are sometimes imposed (such as it is given by a symmetric monoidal structure), but here we only require that it is a state space such that

$$K \dot{\otimes} K \subseteq K \tilde{\otimes} K \subseteq K \hat{\otimes} K.$$

In quantum theory, the joint state space is $\mathfrak{S}_{\mathcal{H}} \tilde{\otimes} \mathfrak{S}_{\mathcal{H}} = \mathfrak{S}_{\mathcal{H} \otimes \mathcal{H}}$ so that all of the above inclusions are strict. On the other hand, for the classical bit (or any simplex) we have $\mathfrak{S}_C \tilde{\otimes} \mathfrak{S}_C = \mathfrak{S}_C \hat{\otimes} \mathfrak{S}_C$, so that the joint state space is unique, denoted by $\mathfrak{S}_C \otimes \mathfrak{S}_C$. It can be easily seen that $\mathfrak{S}_C \otimes \mathfrak{S}_C$ can be identified with the 3-dimensional simplex, with vertices labeled by $\{(i, j), i, j = 0, 1\}$. In the Boxworld GPT, the composition rule is the maximal tensor product, so that $S \tilde{\otimes} S = S \hat{\otimes} S$. Here the minimal tensor product $S \dot{\otimes} S$ is quite different from $S \hat{\otimes} S$ (the GPT with the composition rule $S \tilde{\otimes} S = S \dot{\otimes} S$ is called GLT in [7]).

Consider a state space K and let $K \tilde{\otimes} K$ be the joint state space. Let A, B be two-outcome measurements on K given by effects $f_A, f_B \in E(K)$ and let $A \otimes B$ denote the four-outcome measurement obtained by measuring A on the first part and B on the other. This measurement is determined by the effects $\{f_A \otimes f_B, (1 - f_A) \otimes f_B, f_A \otimes (1 - f_B), (1 - f_A) \otimes (1 - f_B)\}$ and it is clear that $K \hat{\otimes} K$ is the largest state space such that all such locally prepared measurements are valid. Note also that for any $f \in E(K)$ and $x \in K \hat{\otimes} K$ one can define $(f \otimes \text{id})(x) \in A(K)^{*+}$ as the unique functional such that for any other $g \in E(K)$ we have

$$((f \otimes \text{id})(\phi))(g) = (f \otimes g)(\phi).$$

In particular, $(1 \otimes \text{id})(x)$ and $(\text{id} \otimes 1)(x)$ belong to K and define the two marginals of x which correspond to partial traces in quantum theory.

D. Review of the CHSH inequality

We provide a very short introduction to the CHSH inequality, in the setting of non-local boxes. These are defined as a black box, with two inputs with possible values A, A' and B, B' , respectively, and two outputs, each with values 1 or -1. It is assumed that such a box describes the situation when two experimenters (Alice and Bob), each on their part of a bipartite system, apply one of a given pair of two-outcome measurements $\{A, A'\}$ resp. $\{B, B'\}$.

Any non-local box x is fully described by the outcome probabilities $P_x(\epsilon, \eta | C, D)$, with $C = A$ or A' , $D = B$ or B' and $\epsilon, \eta \in \{-1, 1\}$. Assume that the measurements A and B are chosen. Then Alice will see the outcome 1 on her part with the probability $P_x(1, 1 | A, B) + P_x(1, -1 | A, B)$. For this to be a well-defined marginal outcome probability of A , we require that it stays the same if the other measurement is B' , that is,

$$P_x(1, 1 | A, B) + P_x(1, -1 | A, B) = \\ = P_x(1, 1 | A, B') + P_x(1, -1 | A, B'). \quad (8)$$

This condition, together with the analogical condition

$$\begin{aligned} P_x(1, 1|A, B) + P_x(-1, 1|A, B) = \\ = P_x(1, 1|A', B) + P_x(-1, 1|A', B) \end{aligned} \quad (9)$$

are called the non-signaling conditions [4, 5], because they mean that neither side can signal to the other by only using different local measurements and without announcing the outcome of the measurement. Non-local boxes satisfying these conditions, also called non-signaling boxes, are of particular interest in the theory of Bell inequalities, see also [11].

It is clear that if K is a state space in a GPT, then any $x \in K \hat{\otimes} K$ and any measurements $\{A, A'\}$ of the first part and $\{B, B'\}$ on the other implement a non-signaling box.

The central quantity for the formulation of the CHSH inequality is the correlation $E(A, B)$ between the measurements A and B , defined as

$$\begin{aligned} E(A, B) = P_x(1, 1|A, B) - P_x(1, -1|A, B) \\ - P_x(-1, 1|A, B) + P_x(-1, -1|A, B). \end{aligned}$$

It is straightforward that we have $-1 \leq E(A, B) \leq 1$. For the two pairs of measurements A, A' and B, B' , the CHSH quantity X_{CHSH} is given as

$$X_{\text{CHSH}} = E(A, B) + E(A, B') + E(A', B) - E(A', B').$$

It is known that in classical theories we have $|X_{\text{CHSH}}| \leq 2$, this is called the CHSH inequality. This inequality is violated in quantum theory, where the Tsirelson bound [4] gives $|X_{\text{CHSH}}| \leq 2\sqrt{2}$. The maximal value reachable by a non-signaling theory coincides with the algebraic maximum, which is $|X_{\text{CHSH}}| = 4$.

The non-signaling box attaining the value $X_{\text{CHSH}} = 4$ was defined by Popescu and Rohrlich [5], it is called the PR-box. The PR-box is determined by the outcome probabilities

$$P_x(\epsilon, \eta|C, D) = \begin{cases} \frac{1}{2} & \text{if } CD \neq A'B' \text{ and } \epsilon\eta = 1 \\ \frac{1}{2} & \text{or } CD = A'B' \text{ and } \epsilon\eta = -1 \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

It can be seen that the value $X_{\text{CHSH}} = -4$ is obtained from the PR-box by relabelling the outcomes on one of the sides and that this is the only possibility for maximal CHSH violation. We provide some more details in Appendix A for the convenience of the reader.

III. IMPLEMENTATIONS OF THE PR-BOX

The aim of this section is to characterize all implementations of the PR-box in the GPT framework, that is, for a state space K , we want to describe the states $\phi \in K \hat{\otimes} K$ and pairs of two-outcome measurements $\{A, A'\}$ and $\{B, B'\}$ such that the corresponding outcome probabilities maximally violate the CHSH inequality.

The best known implementation of the PR-box is provided in the Boxworld GPT. It was shown in [7] that all non-signaling boxes can be implemented with elements of the state space $S \hat{\otimes} S$ and the outcome probabilities are obtained by applying the measurements given by $A = B = A_0$ and $A' = B' = A_1$. The state in $S \hat{\otimes} S$ corresponding to the PR-box is given by

$$\phi_S = \frac{1}{2}((s_{00} - s_{10}) \otimes s_{00} + s_{11} \otimes s_{10} + s_{10} \otimes s_{01}), \quad (11)$$

up to local isomorphisms, this is the only implementation of the PR-box on S , see Appendix B for a proof.

It is known that for some systems the degree of compatibility of measurements is tied to violation of the CHSH inequality [29, 30] and, as we have seen in Prop. 1, all maximally incompatible pairs are obtained from the square state space by embedding it in other state spaces. It is therefore not surprising that all implementations of the PR-box are obtained from the above implementation on S .

Theorem 1. *Let K be a state space and let $\{A, A'\}$, $\{B, B'\}$ and $\phi \in K \hat{\otimes} K$ be an implementation of the PR-box. Then both pairs $\{A, A'\}$ and $\{B, B'\}$ are maximally incompatible. Moreover, let ι_A, Π_A and ι_B, Π_B be the maps for these pairs given by Prop. 1. Then*

$$\phi = (\iota_A \otimes \iota_B)(\phi_S) + \phi^\perp$$

where $\phi_S \in S \hat{\otimes} S$ is given by (11) and $\phi^\perp \in \ker(\Pi_A) \otimes \ker(\Pi_B)$.

Proof. The proof is in Appendix C. \square

We see from this result that maximally incompatible pairs of measurements are necessary for maximal CHSH violation, moreover, having such pairs of measurements on both sides, we can construct all possible states. While existence of such measurements is a property of K , we need also a joint state space $K \hat{\otimes} K$ that contains at least one of the candidate states. It follows from the next result that if we work with the maximal tensor product, maximal incompatibility is also sufficient for existence of an implementation of the PR-box.

Proposition 2. *Let K be a state space on which there exists a pair of maximally incompatible measurements and let $\iota : S \rightarrow K$ be the corresponding map. Then*

$$(\iota \otimes \iota)(\phi_S) \in K \hat{\otimes} K.$$

Proof. Since ι extends to a positive map $A(S)^* \rightarrow A(K)^*$ with the respective positive cones, the assertion follows by the fact that positive maps are completely positive for the maximal tensor product. In more details: note that the adjoint map $\iota^* : A(K) \rightarrow A(S)$, given by

$$\iota^*(f)(s) = f(\iota(s)), \quad s \in S, f \in A(K)$$

is again positive. Hence, for any $h_1, h_2 \in A(K)^+$ we have

$$(h_1 \otimes h_2)((\iota \otimes \iota)(\phi_S)) = ((\iota^* h_1) \otimes (\iota^* h_2))(\phi_S) \geq 0,$$

as $\phi_S \in S \hat{\otimes} S$ and $\iota^* h_i \in A(S)^+$. It follows that $(\iota \otimes \iota)(\phi_S) \in K \hat{\otimes} K$. \square

We finish this section by observing that the element $(\iota \otimes \iota)(\phi_S)$, corresponding to the special case $\phi^\perp = 0$ in Thm. 1, can be constructed from a witness square $\{x_{ij}\}$ of the maximally incompatible pair of measurements as

$$(\iota \otimes \iota)(\phi_S) = \frac{1}{2}((x_{00} - x_{10}) \otimes x_{00} + x_{11} \otimes x_{10} + x_{10} \otimes x_{01}).$$

This is immediate from (6) and (11). It is clear that we can construct a similar state for different measurements on each side, using some respective witness squares.

IV. CLASSICAL CHANNELS AS IMPLEMENTATIONS OF THE PR-BOX

As we have seen in Example 4, the set of classical bit channels is isomorphic to the square state space. By a similar reasoning as in the proof of Proposition 2, we can extend this to an isomorphism of the maximal tensor products $\mathcal{C}_C \hat{\otimes} \mathcal{C}_C$ and $S \hat{\otimes} S$. In this way, we may view the Boxworld as a GPT based on classical channels. In this section, we will describe the PR-box implementation (11) in this setting.

It is immediate from the definition that the non-local boxes can be identified with classical bipartite channels $\mathfrak{S}_C \otimes \mathfrak{S}_C \rightarrow \mathfrak{S}_C \otimes \mathfrak{S}_C$, where the four vertices of $\mathfrak{S}_C \otimes \mathfrak{S}_C$ are labeled by elements of $\{A, A'\} \times \{B, B'\}$ in the input space and by elements of $\{1, -1\}^2$ in the output. Since $S \hat{\otimes} S$ is identified with the non-signaling boxes, we see from the above remarks that $\mathcal{C}_C \hat{\otimes} \mathcal{C}_C$ can be described as the set of non-signaling classical bipartite channels.

As shown in Example 4, the measurements on \mathcal{C}_C corresponding to A_0, A_1 on S are given by the effects $F_{s_0, \pi}$ and $F_{s_1, \pi}$, respectively (where π is as in Example 1). To construct the bipartite channel Φ_C corresponding to the state ϕ_S of (11), we first find the elements of \mathcal{C}_C corresponding to the vertices of S . Looking at Example 4, the channels $\Phi_{ij} \simeq s_{ij}$ are determined by

$$\pi(\Phi_{ij}(s_0)) = i, \quad \pi(\Phi_{ij}(s_1)) = j,$$

hence $\Phi_{00} = 1_{\mathfrak{S}_C}(\cdot)s_0$ and $\Phi_{11} = 1_{\mathfrak{S}_C}(\cdot)s_1$ are constant channels, $\Phi_{01} = id$ and Φ_{10} is the negation channel, given by $s_0 \mapsto s_1$ and $s_1 \mapsto s_0$. It is now easily checked that Φ_C is given by

$$\begin{aligned} \Phi_C(s_0 \otimes s_0) &= \Phi_C(s_0 \otimes s_1) = \Phi_C(s_1 \otimes s_0) \\ &= \frac{1}{2}(s_0 \otimes s_0 + s_1 \otimes s_1) \end{aligned} \quad (12)$$

$$\Phi_C(s_1 \otimes s_1) = \frac{1}{2}(s_0 \otimes s_1 + s_1 \otimes s_0) \quad (13)$$

It can be also checked directly that this channel, together with the above pair of measurements applied on both

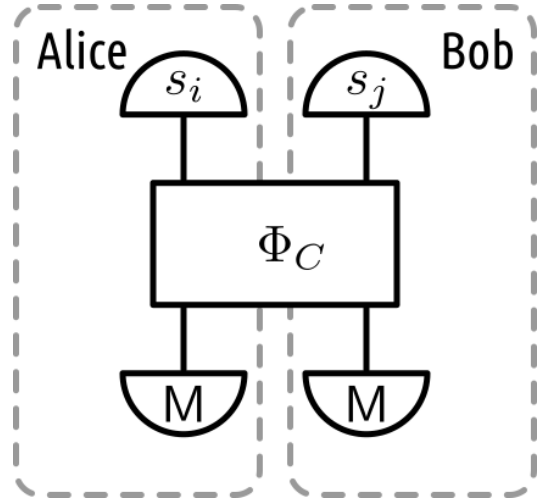


Figure 2. Alice and Bob can use the channel Φ_C to maximally violate the CHSH inequality as follows: Alice inputs the state s_i and Bob inputs the state s_j where $i, j \in \{0, 1\}$. This choice exactly corresponds to the choice of channel measurement $A = B$ or $A' = B'$ on either side. Finally both Alice and Bob apply the measurement M corresponding to the effect $\pi \in E(\mathfrak{S}_C)$. Note that for both Alice and Bob the procedure of inputting a testing state into the channel and measuring the outcome comprises a single measurement on the bipartite channel itself. Also note that both Alice and Bob can only access their respective input and output and since the channel Φ_C is non-signaling they can not obtain any information about each others choices of channel measurement.

sides, maximally violates the CHSH inequality. The protocol is depicted on Fig. 2. As in the case of S , this is the only implementation of the PR-box in this setting, up to local isomorphisms.

V. QUANTUM CHANNELS AS IMPLEMENTATIONS OF THE PR-BOX

We now get to the most important example of a state space for this work. Let \mathcal{H} be a finite dimensional Hilbert space and let $\mathcal{C}(\mathcal{H})$ denote the set of all quantum channels on $B(\mathcal{H})$, that is, all completely positive and trace preserving linear maps $B_h(\mathcal{H}) \rightarrow B_h(\mathcal{H})$. Let $\{|i\rangle\}_{i=1}^{\dim(\mathcal{H})}$ be an orthonormal basis of \mathcal{H} , then $|\psi_{\dim(\mathcal{H})}^+\rangle = \sum_{i=1}^{\dim(\mathcal{H})} |i\rangle \otimes |i\rangle \in \mathcal{H} \otimes \mathcal{H}$ is a multiple of the maximally entangled state. The Choi matrix of a channel Φ is given as $C(\Phi) = (\Phi \otimes id)(|\psi_{\dim(\mathcal{H})}^+\rangle \langle \psi_{\dim(\mathcal{H})}^+|) \in B(\mathcal{H} \otimes \mathcal{H})$. The set $\mathcal{C}(\mathcal{H})$ is isomorphic to the set of Choi matrices

$$\text{Choi}(\mathcal{H}) := \{A \in B_h(\mathcal{H} \otimes \mathcal{H}) : A \geq 0, \text{Tr}_1(A) = 1\}$$

where Tr_1 is the partial trace over the first Hilbert space. Clearly, $\mathcal{C}(\mathcal{H})$ is a compact convex subset of the finite dimensional real vector space of linear maps on $B_h(\mathcal{H})$ and can therefore be treated as a state space in some

GPT. The following description of the quantum channel GPT is based on the ideas of [31].

The measurements of quantum channels are in principle similar to measurements of classical channels described in Example 4. Clearly we can input any quantum state into the channel in question and measure the outcome, but this does not describe all of the possible measurements on quantum channels for a simple reason: the input state can also be entangled to other system. It turns out that all measurements on quantum channels are described as procedures where we input a bipartite and potentially entangled state $\rho \in B_h(\mathcal{H} \otimes \mathcal{H}')$ into the channel (possibly tensored with identity) and measure the outcome. The measurement on channels given by ρ and a quantum measurement described by the POVM E_1, \dots, E_n is determined by the outcome probabilities

$$P_\Phi(i|E) = \text{Tr}((\Phi \otimes \text{id})(\rho)E_i)$$

for $i \in \{1, \dots, n\}$ and any quantum channel $\Phi : B_h(\mathcal{H}) \rightarrow B_h(\mathcal{H})$. Any effect on channels has the form

$$F_{\rho,E}(\Phi) := \text{Tr}((\Phi \otimes \text{id})(\rho)E)$$

for some effect E and input state ρ . Note also that this expression is not unique. In particular, we can always assume the state ρ to be pure as we can always purify it by enlarging the Hilbert space \mathcal{H}' .

We can describe the measurements in an equivalent way, using the Choi matrices. Then there are operators $F_1, \dots, F_n \in B_h(\mathcal{H} \otimes \mathcal{H})$ such that

$$P_\Phi(i|E) = \text{Tr}(C(\Phi)F_i)$$

for $i \in \{1, \dots, n\}$. One can show that we may always choose $F_i \geq 0$ for all $i \in \{1, \dots, n\}$ and we must have $\sum_{i=1}^n F_i = \mathbb{1} \otimes \sigma$ where $\sigma \in \mathfrak{S}_{\mathcal{H}}$. Such a collection of operators F_1, \dots, F_n is called a process POVM, or PPOVM, or a quantum tester, which were first introduced in [31, 32]. This passage from quantum channels to Choi matrices is directly related to viewing the quantum channels in the GPT picture, see also Fig. 3.

We now need to specify the joint state space $\mathcal{C}(\mathcal{H}) \hat{\otimes} \mathcal{C}(\mathcal{H})$. This will be defined as the set of completely positive maps in $\mathcal{C}(\mathcal{H}) \hat{\otimes} \mathcal{C}(\mathcal{H})$, which is precisely the set $\mathcal{C}^{ns}(\mathcal{H} \otimes \mathcal{H})$ of bipartite non-signaling (or causal) channels, as defined in [12]. On the set of Choi matrices, it can be seen that the set of positive elements in $\text{Choi}(\mathcal{H}) \hat{\otimes} \text{Choi}(\mathcal{H})$ is isomorphic to the set of Choi matrices of elements in $\mathcal{C}^{ns}(\mathcal{H} \otimes \mathcal{H})$, via the map

$$\text{id}_1 \otimes \text{SWAP}_{23} \otimes \text{id}_4 : B(\mathcal{H}^{\otimes 4}) \rightarrow B(\mathcal{H}^{\otimes 4}),$$

where SWAP denotes the swap gate, that is for $X, Y \in B(\mathcal{H})$ we have

$$\text{SWAP}(X \otimes Y) = Y \otimes X.$$

We next describe a pair of maximally incompatible two-outcome measurements on $\mathcal{C}(\mathcal{H})$. Let $\rho_1, \rho_2, \sigma_1, \sigma_2 \in$

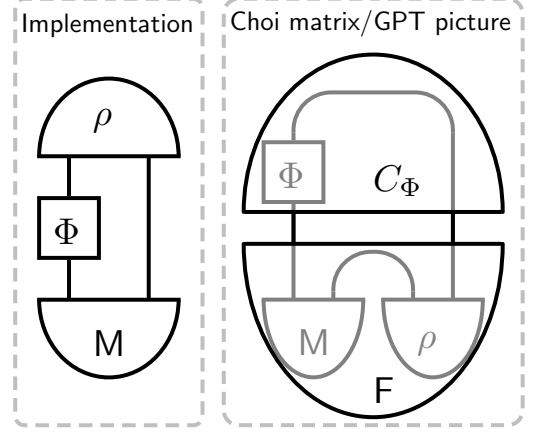


Figure 3. Description of a measurement on quantum channels. On the left, we have an implementation of the measurement with an input state ρ of the system composed with an ancilla and the final measurement M , which is a usual quantum measurement described by a POVM. Note that it is not necessary to consider convex combinations, since these can be included into the ancilla. The right side shows the same measurement represented by the PPOVM F applied on the Choi matrix C_Φ , obtained from Φ by applying the channel to (one part) of the state $|\psi_{\dim(\mathcal{H})}^+\rangle$. This is depicted as the "bending of the input wire" of Φ . Similarly, the PPOVM F is the *link product* of ρ and M [31] which is obtained by application of $|\psi_{\dim(\mathcal{H})}^+\rangle$ on parts of the inputs, as shown in the picture. The representation of Φ by its Choi matrix C_Φ can be straightforwardly interpreted as viewing the set of channels in the GPT picture.

$\mathfrak{S}_{\mathcal{H}}$, such that $\rho_1 \rho_2 = \sigma_1 \sigma_2 = 0$. Then there exist projections M, N , such that $M \rho_1 = \rho_1$, $M \rho_2 = 0$, $N \sigma_1 = \sigma_1$, $N \sigma_2 = 0$. Also let us denote $M^\perp = \mathbb{1} - M$, $N^\perp = \mathbb{1} - N$. Let the two-outcome measurements C_1, C_2 be given by the effects $F_{\sigma_1, M}$ and $F_{\sigma_2, M}$. Consider the channels $B_h(\mathcal{H}) \rightarrow B_h(\mathcal{H})$ given for $X \in B_h(\mathcal{H})$ as

$$\begin{aligned} \Phi_{00}(X) &= \text{Tr}(X)\rho_2, \\ \Phi_{10}(X) &= \text{Tr}(NX)\rho_1 + \text{Tr}(N^\perp X)\rho_2, \\ \Phi_{01}(X) &= \text{Tr}(NX)\rho_2 + \text{Tr}(N^\perp X)\rho_1, \\ \Phi_{11}(X) &= \text{Tr}(X)\rho_1 \end{aligned}$$

It is straightforward to verify that the measurements C_1, C_2 are maximally incompatible and the channels $\Phi_{00}, \Phi_{10}, \Phi_{01}, \Phi_{11}$ form a witness square for C_1, C_2 . Let $\iota : S \rightarrow \mathcal{C}(\mathcal{H})$ be the map (6) for this witness square.

Let $\Phi = (\iota \otimes \iota)(\Phi_S)$ be the tensor product element as described in Prop. 2. Let us denote $\rho_{cor} = \frac{1}{2}(\rho_1 \otimes \rho_1 + \rho_2 \otimes \rho_2)$ and $\rho_{ac} = \frac{1}{2}(\rho_1 \otimes \rho_2 + \rho_2 \otimes \rho_1)$. One can check that for $X, Y \in B_h(\mathcal{H})$, we have

$$\begin{aligned} \Phi(X \otimes Y) &= \frac{1}{2}((\Phi_{00}(X) - \Phi_{10}(X)) \otimes \Phi_{00}(Y) \\ &\quad + \Phi_{11}(X) \otimes \Phi_{10}(Y) + \Phi_{10}(X) \otimes \Phi_{01}(Y)) \\ &= \text{Tr}((N^\perp \otimes N^\perp)(X \otimes Y))\rho_{cor} \\ &\quad + \text{Tr}((\mathbb{1} \otimes \mathbb{1} - N^\perp \otimes N^\perp)(X \otimes Y))\rho_{ac}. \end{aligned}$$

By linearity, for every $\rho \in S_{\mathcal{H} \otimes \mathcal{H}}$ we have

$$\begin{aligned} \Phi(\rho) &= \text{Tr}((N^\perp \otimes N^\perp)(\rho))\rho_{ac} \\ &\quad + \text{Tr}((\mathbb{1} \otimes \mathbb{1} - N^\perp \otimes N^\perp)(\rho))\rho_{cor}. \end{aligned} \quad (14)$$

It is easy to see that this is indeed a quantum channel, so that we have $\Phi \in \mathcal{C}^{ns}(\mathcal{H} \otimes \mathcal{H}) = C(\mathcal{H}) \hat{\otimes} C(\mathcal{H})$. It follows by Thm. 1, but is also straightforward to verify, that the channel Φ and the measurements $A = B = C_1$ and $A' = B' = C_2$ are an implementation of the PR-box. The protocol is similar to the one used in the case of classical channels: Alice and Bob share the bipartite channel Φ and they both can choose the input state σ_1 or σ_2 each on their part of the channel, followed by the measurement $\{M, M^\perp\}$ on their part of the output.

Example 6. The above pair of maximally incompatible measurements is a generalization of an example studied in [16, 27], with $\dim(\mathcal{H}) = 2$ and $N = M = \rho_1 = \sigma_1 = |0\rangle\langle 0|$ and $N^\perp = M^\perp = \rho_2 = \sigma_2 = |1\rangle\langle 1|$. The corresponding channel $\Phi =: \Psi_S$ has the form

$$\begin{aligned} \Phi_S(\rho) &= \langle 11|\rho|11\rangle \frac{1}{2}(|01\rangle\langle 01| + |10\rangle\langle 10|) \\ &\quad + (1 - \langle 11|\rho|11\rangle) \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|). \end{aligned}$$

The resulting implementation of the PR-box was already observed in [12–14].

As the results of Thm. 1 suggest, even if the pair of maximally incompatible measurements is fixed, there can be more bipartite non-signaling channels that implement the PR-box. Any such channel will be called a PR-channel.

Observe that the PR-channel Φ given by (14) is of a special form, called an entanglement-breaking channel. An entanglement-breaking channel is such that for any state $\omega \in \mathfrak{S}_{\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}}$ we have that $(\Phi \otimes \text{id})(\omega)$ is a separable state. Such channels are also called measure-and-prepare, since we first perform a measurement, in this case the two-outcome measurement given by the effect $N^\perp \otimes N^\perp$, and according to the result, we prepare one of a given set of states, in this case ρ_{ac} or ρ_{cor} .

The structure of the channel Φ is even more simple. In fact, since N is a projection and the two states ρ_{ac} and ρ_{cor} commute, Φ is classical-to-classical. Moreover, note that both the measurement and the outcome states are separable. Even so, the resulting channel implements a PR-box.

Now we would like to see whether there are PR-channels of a more complicated structure. It would be quite hard to characterize all such channels in the general case. For this reason, in the next section we restrict to $\dim(\mathcal{H}) = 2$, i.e. to qubits.

VI. THE SPECIAL CASE OF QUBITS

In this section we restrict to qubits, i.e. $\dim(\mathcal{H}) = 2$ and we characterize all PR-channels that exist in this

scenario. So let \mathcal{H} be a complex Hilbert space, $\dim(\mathcal{H}) = 2$ and let $\{|0\rangle, |1\rangle\}$ be an orthonormal basis of \mathcal{H} . For vectors from $\mathcal{H} \otimes \mathcal{H}$, we will use the shorthand $|i\rangle \otimes |j\rangle = |ij\rangle$ for $i, j \in \{0, 1\}$. We will also use the notation \oplus for addition modulo 2.

We will begin by characterizing all maximally incompatible pairs of two-outcome measurements and their witness squares. Let the two measurements be given by effects $F_{\rho, M}$ and $F_{\sigma, N}$, for pure states $\rho, \sigma \in \mathfrak{S}_{\mathcal{H} \otimes \mathcal{H}'}$ and $M, N \in E(\mathcal{H} \otimes \mathcal{H}')$. Let $\ker(M)$ denote the projection onto the kernel of M . Then by (2) and (3), a witness square $\Phi_{00}, \Phi_{10}, \Phi_{01}, \Phi_{11} \in \mathcal{C}(\mathcal{H})$ must satisfy

$$\begin{aligned} \ker(M)(\Phi_{10} \otimes \text{id})(\rho) \ker(M) &= 0, \\ \ker(M)(\Phi_{11} \otimes \text{id})(\rho) \ker(M) &= 0, \end{aligned}$$

and

$$\begin{aligned} \ker(M)(\Phi_{00} \otimes \text{id})(\rho) \ker(M) &= (\Phi_{00} \otimes \text{id})(\rho), \\ \ker(M)(\Phi_{01} \otimes \text{id})(\rho) \ker(M) &= (\Phi_{01} \otimes \text{id})(\rho). \end{aligned}$$

From (1) it follows that

$$(\Phi_{00} \otimes \text{id})(\rho) + (\Phi_{11} \otimes \text{id})(\rho) = (\Phi_{10} \otimes \text{id})(\rho) + (\Phi_{01} \otimes \text{id})(\rho),$$

and after applying $\ker(M)$ we get

$$(\Phi_{00} \otimes \text{id})(\rho) = (\Phi_{01} \otimes \text{id})(\rho). \quad (15)$$

It follows that ρ cannot have maximal Schmidt rank as then (15) would imply $\Phi_{00} = \Phi_{01}$ which is impossible by (2) and (3). Since we have assumed $\dim(\mathcal{H}) = 2$, it follows that ρ must have Schmidt rank 1, i.e. ρ must be a pure product state, so we can assume $\rho = |x\rangle\langle x|$ for some $|x\rangle \in \mathcal{H}$, $\|x\| = 1$, and $M \in B_h(\mathcal{H})$, since ρ is not entangled and therefore we do not need the ancillary Hilbert space \mathcal{H}' . From

$$\text{Tr}(\Phi_{10}(|x\rangle\langle x|)M) = \text{Tr}(\Phi_{11}(|x\rangle\langle x|)M) = 1$$

and

$$\text{Tr}(\Phi_{00}(|x\rangle\langle x|)M) = \text{Tr}(\Phi_{01}(|x\rangle\langle x|)M) = 0$$

it follows that there must be an orthonormal basis $|\xi_0\rangle, |\xi_1\rangle$ of \mathcal{H} such that

$$M = |\xi_0\rangle\langle \xi_0|, \quad M^\perp = |\xi_1\rangle\langle \xi_1|,$$

and for $i, j \in \{0, 1\}$ we have

$$\Phi_{ij}(|x\rangle\langle x|) = |\xi_{i \oplus 1}\rangle\langle \xi_{i \oplus 1}|.$$

In a similar fashion, one can show that we must have $\sigma = |y\rangle\langle y|$ for some $|y\rangle \in \mathcal{H}$, $\|y\| = 1$, and that there is an orthonormal basis $|\eta_0\rangle, |\eta_1\rangle$ of \mathcal{H} such that

$$N = |\eta_0\rangle\langle \eta_0|, \quad N^\perp = |\eta_1\rangle\langle \eta_1|,$$

and for $i, j \in \{0, 1\}$ we have

$$\Phi_{ij}(|y\rangle\langle y|) = |\eta_{j \oplus 1}\rangle\langle \eta_{j \oplus 1}|.$$

Let $L \in B(\mathcal{H})$ be given by $L|0\rangle = |x\rangle$, $L|1\rangle = |y\rangle$ and let $\Phi_{ij}^L = \Phi_{ij}(L \cdot L^*)$. Then Φ_{ij}^L are completely positive maps satisfying (1), with Choi matrices

$$C(\Phi_{ij}^L) = \begin{pmatrix} |\xi_{i\oplus 1}\rangle\langle\xi_{i\oplus 1}| & X_{ij} \\ X_{ij}^* & |\eta_{j\oplus 1}\rangle\langle\eta_{j\oplus 1}| \end{pmatrix}$$

where $X_{ij} \in B(\mathcal{H})$. It follows by positivity of $C(\Phi_{ij}^L)$ that we must have

$$X_{(i\oplus 1)(j\oplus 1)} = z_{ij}|\xi_i\rangle\langle\eta_j|$$

where $z_{ij} \in \mathbb{C}$, $|z_{ij}| \leq 1$, see [33, 1.3.2 Proposition]. From Eq. (1) we get

$$z_{00}|\xi_0\rangle\langle\eta_0| + z_{11}|\xi_1\rangle\langle\eta_1| = z_{10}|\xi_1\rangle\langle\eta_0| + z_{01}|\xi_0\rangle\langle\eta_1|$$

from which it follows that $z_{ij} = 0$ for all $i, j \in \{0, 1\}$. This implies that $C(\Phi_{ij}^L)$ are block-diagonal matrices. In particular,

$$0 = \text{Tr}(\Phi_{ij}^L(|0\rangle\langle 1|)) = \text{Tr} \Phi_{ij}(|x\rangle\langle y|) = \langle x|y\rangle,$$

so that $\{|x\rangle, |y\rangle\}$ is an orthonormal basis of \mathcal{H} and we have proved that $C(\Phi_{ij}^L)$ are block-diagonal in this basis. Applying unitary transformations to the input resp. output space, transforming the basis $|x\rangle, |y\rangle$ resp. $|\xi_0\rangle, |\xi_1\rangle$ to $|0\rangle, |1\rangle$, we can summarize as follows.

Proposition 3. *Let A, A' be two-outcome measurements on qubit channels, given by PPOVMs $\{F_A, F_A^\perp\}$ and $\{F_{A'}, F_{A'}^\perp\}$. Then A, A' are maximally incompatible if and only if, up to unitary conjugation on the input and output spaces,*

$$\begin{aligned} F_A &= |00\rangle\langle 00|, & F_A^\perp &= |10\rangle\langle 10| \\ F_{A'} &= |\eta_0\rangle\langle\eta_0| \otimes |1\rangle\langle 1|, & F_{A'}^\perp &= |\eta_1\rangle\langle\eta_1| \otimes |1\rangle\langle 1| \end{aligned}$$

where $\{|\eta_0\rangle, |\eta_1\rangle\}$ is an orthonormal basis of \mathcal{H} . Moreover, there is a unique witness square for A, A' , with Choi matrices of the form

$$C_{ij} = \begin{pmatrix} |i\oplus 1\rangle\langle i\oplus 1| & 0 \\ 0 & |\eta_{j\oplus 1}\rangle\langle\eta_{j\oplus 1}| \end{pmatrix}.$$

By the above proposition, essentially any maximally incompatible pair is characterized by the choice of an orthonormal basis $|\eta_0\rangle, |\eta_1\rangle$. The corresponding channel measurements then consist of inputting $|0\rangle$ into the channel and measuring the basis $\{|i\rangle\}$ on the output, or inputting $|1\rangle$ and measuring the basis $\{|\eta_i\rangle\}$.

Let us pick some choice of the bases $|\eta_0^A\rangle, |\eta_1^A\rangle$ and $|\eta_0^B\rangle, |\eta_1^B\rangle$ on Alice's and Bobs part, respectively, and let A, A' and B, B' denote the corresponding maximally incompatible pairs. Note that the set of all PR-channels that give an implementation of the PR-box with these measurements is a face of $\mathcal{C}^{ns}(\mathcal{H} \otimes \mathcal{H})$. As we will see, all such faces consist entirely of entanglement-breaking channels.

Let Φ be a qubit PR-channel and let C be its Choi matrix. Then $C \in B(\mathcal{H}_{\text{out}} \otimes \mathcal{H}_{\text{in}})$, where both the input

and the output spaces are composed of Alice's and Bob's part: $\mathcal{H}_{\text{in}} = \mathcal{H}_{A,\text{in}} \otimes \mathcal{H}_{B,\text{in}}$ and $\mathcal{H}_{\text{out}} = \mathcal{H}_{A,\text{out}} \otimes \mathcal{H}_{B,\text{out}}$. We write C as a block matrix

$$C = \sum_{\alpha \in \{0,1\}^2} C_{\alpha,\beta} \otimes |\alpha\rangle\langle\beta|_{\text{in}},$$

where $C_{\alpha,\beta} \in B(\mathcal{H}_{\text{out}})$. To describe the structure of C , we need to introduce the following notations. For $x, y, z \in \mathbb{C}$, we denote

$$B_{\text{diag}}(z) = \begin{pmatrix} 1 & z \\ \bar{z} & 1 \end{pmatrix} = I_2 + z|0\rangle\langle 1| + \bar{z}|1\rangle\langle 0|$$

$$B_{\text{off}}(x, y) = \begin{pmatrix} 0 & x \\ y & 0 \end{pmatrix} = x|0\rangle\langle 1| + y|1\rangle\langle 0|, \quad i \neq j.$$

For $r \in \mathbb{N}$, we denote by B_r a block matrix in $B(\mathbb{C}^2 \otimes \mathbb{C}^r)$ of the form

$$\begin{aligned} B_r &= \frac{1}{2} \left(\sum_{p=1}^r B_{\text{diag}}(z_p) \otimes |p\rangle\langle p| \right. \\ &\quad \left. + \sum_{q \neq p=1}^r B_{\text{off}}(x_{p,q}, y_{p,q}) \otimes |p\rangle\langle q| \right). \end{aligned} \quad (16)$$

Proposition 4. *Let Φ be a qubit PR-channel. Then there are isometries $U_\alpha : \mathbb{C}^2 \rightarrow \mathcal{H}_{\text{out}}$, $\alpha \in \{0, 1\}^2$ and a decomposition $\{0, 1\}^2 = \Delta_0 \cup \Delta_1$, such that $U_\alpha = U_\beta =: V_{\text{out}}$ for $\alpha, \beta \in \Delta_1$ and C has the form*

$$\begin{aligned} C &= (V_{\text{out}} \otimes V_{\text{in}}) B_r (V_{\text{out}}^* \otimes V_{\text{in}}^*) \\ &\quad + \sum_{\alpha \in \Delta_0} U_\alpha B_{\text{diag}}(z_\alpha) U_\alpha^* \otimes |\alpha\rangle\langle\alpha|, \end{aligned} \quad (17)$$

here $|z_\alpha| \leq 1$, $r = |\Delta_1|$ and $V_{\text{in}} : \mathbb{C}^r \rightarrow \mathcal{H}_{\text{in}}$ is an isometry such that $V_{\text{in}}|p\rangle = |\alpha_p\rangle$, $\alpha_p \in \Delta_1$, $p = 1, \dots, r$.

The proof of this proposition is given in Appendix D.

Lemma 1. *Let $r \leq 3$ and let B_r be a matrix of the form (16). If B_r is positive, then it is separable.*

Proof. Let B_r be positive. Since $B_r \in B(\mathbb{C}^2 \otimes \mathbb{C}^r)$ and $r \leq 3$, we may apply the PPT criterion [34, 35], that is, B_r is separable if and only if it remains positive under partial transpose. We will apply the transpose to the first part, so we will show that the matrix

$$\begin{aligned} B_r^\Gamma &= \frac{1}{2} \left(\sum_{p=1}^r B_{\text{diag}}(z_p)^T \otimes |p\rangle\langle p| \right. \\ &\quad \left. + \sum_{q \neq p=1}^r B_{\text{off}}(x_{p,q}, y_{p,q})^T \otimes |p\rangle\langle q| \right) \end{aligned}$$

is positive. Let $V \in B(\mathbb{C}^2)$ be given as

$$V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

then V is unitary, $V = V^*$ and for any $t \in \mathbb{R}$ and $z_1, z_2 \in \mathbb{C}$ we have

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} t & z_1 \\ z_2 & t \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} t & z_2 \\ z_1 & t \end{pmatrix} = \begin{pmatrix} t & z_1 \\ z_2 & t \end{pmatrix}^T.$$

It follows that

$$B_r^\Gamma = (V \otimes \mathbb{1}_r) B_r (V \otimes \mathbb{1}_r)$$

so we have $B_r^\Gamma \geq 0$ and B_r is separable. \square

We now prove the main result of this section.

Theorem 2. *Let Φ be a qubit PR-channel. Then Φ is an entanglement-breaking channel.*

Proof. The channel Φ is entanglement-breaking if and only if its Choi matrix $C = C(\Phi)$ is separable. The assertion now follows by Prop. 4 and Lemma 1. \square

We will proceed by presenting a few examples of PR-channels. We concentrate on the choice $|\eta_i^A\rangle = |\eta_i^B\rangle = |i\rangle$. According to Appendix D, in this case we have

$$U_\alpha |i\rangle = \begin{cases} |ii\rangle & \text{if } \alpha \neq 11 \\ |i(i \oplus 1)\rangle & \text{otherwise} \end{cases}$$

$r = 3$ and $\Delta_0 = \{11\}$. So any such channel is specified by the choice of the parameters $z_p, x_{p,q}, y_{p,q}$, $p, q = 1, \dots, 3$ such that the matrix B_3 is positive, and any choice of $z_4 := z_{11}$ with $|z_4| \leq 1$.

An obvious choice is setting all these parameters to 0, in which case we obtain the channel of Example 6.

Example 7. Another possible choice of parameters is

$$z_p = \pm 1, \quad x_{p,q} = y_{p,q} = 0, \quad \forall p, q.$$

The resulting channels Φ_\pm are similar to Φ_S . We have

$$\Phi_\pm(\rho) = (1 - \langle 11|\rho|11\rangle) |\phi^\pm\rangle\langle\phi^\pm| + \langle 11|\rho|11\rangle |\psi^\pm\rangle\langle\psi^\pm|$$

where

$$\begin{aligned} |\phi^\pm\rangle &= \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle), \\ |\psi^\pm\rangle &= \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle). \end{aligned}$$

These channels are again classical-to-classical, but here the target states are pure and maximally entangled. A similar channel was also constructed by [15].

Example 8. Let $W : \mathbb{C}^2 \rightarrow V_{\text{out}}$ be the isometry given by $W|i\rangle \mapsto |ii\rangle$ and let $\tilde{W} := (V \otimes \mathbb{1})W$. Let $M_\ell \in B(\mathcal{H} \otimes \mathcal{H})$, $\ell = 1, \dots, k$ be effects such that $\sum_\ell M_\ell = I - |11\rangle\langle 11|$ and let $|w_\ell| \leq 1$, $\ell = 0, 1, \dots, k$. Then the entanglement-breaking channel

$$\begin{aligned} \Phi(\rho) &= \frac{1}{2} \left(\langle 11|\rho|11\rangle \tilde{W} B_{\text{diag}}(w_0) \tilde{W}^* \right. \\ &\quad \left. + \sum_{\ell=1}^k \text{Tr}(M_\ell \rho) W B_{\text{diag}}(w_\ell) W^* \right) \end{aligned}$$

is a PR-channel of the required form, with values of the parameters $z_4 = \omega_0$ and for $p, q \leq 3$

$$\begin{aligned} z_p &= \sum_{\ell=1}^k w_\ell \langle \alpha_p | M_\ell | \alpha_p \rangle \\ x_{p,q} = y_{p,q} &= \sum_{\ell=1}^k w_\ell \langle \alpha_p | M_\ell | \alpha_q \rangle. \end{aligned}$$

This example contains the above examples. Note that not all the PR-channels can be written in this form, since here $x_{p,q} = y_{p,q}$. Note also that we may choose M_ℓ and w_ℓ in such a way that the channel is not classical-to-classical and neither the measurement nor the target states are separable.

Example 9. We next look at an example where all the parameters have the same nonzero value, namely

$$\frac{1}{3} = z_p = x_{p,q} = y_{p,q}, \quad \forall p, q.$$

One can use numerical calculations to check that the corresponding matrix B_3 is positive. This example shows that we can have all of the parameters non-zero at the same time.

VII. CONCLUSIONS

We have shown that maximal violation of the CHSH inequality requires existence of maximally incompatible two-outcome measurements and described states in $K \hat{\otimes} K$ that lead to this violation. It follows that a GPT permits implementations of the PR-box if it contains a system with maximally incompatible measurements and such that the joint state space is large enough. We have applied the results to derive the implementations of PR-boxes by classical and quantum non-signaling channels. The derivation was carried out in the framework of GPTs, which opens the door for generalizations of our calculation. For the qubit case, we gave a full description of the PR-channels and proved that all such channels are necessarily entanglement-breaking.

The question of possibility of instantaneous implementation of these channels is out of the scope of this work. Obtaining some no-go theorems that would forbid such a possibility would provide further insight into Bell non-locality and our complete characterization of the qubit case might be useful for proving such results.

There is a plethora of further open questions and directions of research: one may ask about the structure of all implementations of PR-boxes for channels in higher dimensions and also for more general state spaces, one may also ask which states (and which measurements) violate the CHSH inequality more than a given number. One may also consider a resource theory of CHSH inequality violations.

Our results also raise the question of general applicability of the CHSH inequality as a test of quantumness

of a system, if having too big CHSH violation constrains us to entanglement-breaking channels, which can be seen as classical channels in a sense. This also suggests the existence of some kind of trade-off between CHSH violation and some notion of quantumness of non-signaling channels.

ACKNOWLEDGMENTS

We would like to thank the anonymous referee for giving us valuable hints and insights to improve the readability of the manuscript. This research was supported by grant VEGA 2/0142/20 and by the grant of the Slovak Research and Development Agency under contract APVV-16-0073. MP acknowledges support from the DFG and the ERC (Consolidator Grant 683107/TempoQ).

-
- [1] A. Einstein, B. Podolsky, and N. Rosen, Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?, *Phys. Rev.* **47**, 777 (1935).
 - [2] J. S. Bell, On the Einstein Podolski Rosen Paradox, *Physics* (College. Park. Md). **1**, 195 (1964).
 - [3] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed Experiment to Test Local Hidden-Variable Theories, *Phys. Rev. Lett.* **23**, 880 (1969).
 - [4] B. S. Cirel'son, Quantum generalizations of Bell's inequality, *Lett. Math. Phys.* **4**, 93 (1980).
 - [5] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, *Found. Phys.* **24**, 379 (1994).
 - [6] D. Rohrlich, PR-box correlations have no classical limit, *Quantum Theory A Two-Time Success Story* Yakir Aharonov Festschrift Glob. Financ. Cris. Indian Econ. , 205 (2014), arXiv:1407.8530.
 - [7] J. Barrett, Information processing in generalized probabilistic theories, *Phys. Rev. A* **75**, 032304 (2007), arXiv:0508211.
 - [8] Y. Qek and P. W. Shor, Quantum and superquantum enhancements to two-sender, two-receiver channels, *Phys. Rev. A* **95**, 052329 (2017), arXiv:1701.03813.
 - [9] W. Van Dam, Implausible consequences of superstrong nonlocality, *Nat. Comput.* **12**, 9 (2013), arXiv:0501159.
 - [10] H. Buhrman, M. Christandl, F. Unger, S. Wehner, and A. Winter, Implications of superstrong non-locality for cryptography, *Proc. R. Soc. A Math. Phys. Eng. Sci.* **462**, 1919 (2006), arXiv:0504133.
 - [11] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, *Rev. Mod. Phys.* **86**, 419 (2014), arXiv:1303.2849.
 - [12] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, Causal and localizable quantum operations, *Phys. Rev. A* **64**, 052309 (2001), arXiv:0102043 [quant-ph].
 - [13] M. J. Hoban and A. B. Sainz, A channel-based framework for steering, non-locality and beyond, *New J. Phys.* **20**, 053048 (2018), arXiv:1708.00750.
 - [14] M. Plávala and M. Ziman, Popescu-Rohrlich box implementation in general probabilistic theory of processes, *Physics Letters, Section A: General, Atomic and Solid State Physics* **384**, 126323 (2020), arXiv:1708.07425.
 - [15] C. Crépeau, J. Li, and N. Yang, private communication (2017).
 - [16] M. Sedlák, D. Reitzner, G. Chiribella, and M. Ziman, Incompatible measurements on quantum causal networks, *Phys. Rev. A* **93**, 052323 (2016).
 - [17] P. Janotta and H. Hinrichsen, Generalized probability theories: what determines the structure of quantum theory?, *J. Phys. A Math. Theor.* **47**, 323001 (2014), arXiv:1402.6562.
 - [18] L. Lami, *Non-classical correlations in quantum mechanics and beyond*, Ph.D. thesis, Universitat Autònoma de Barcelona (2017).
 - [19] L. Asimow and A. J. Ellis, *Convexity theory and its applications in functional analysis*, L.M.S. monographs (Academic Press, 1980).
 - [20] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Probabilistic theories with purification, *Phys. Rev. A - At. Mol. Opt. Phys.* **81**, 062348 (2010), arXiv:0908.1583.
 - [21] S. N. Filippov, S. Gudder, T. Heinosaari, and L. Lepäjärvä, Operational restrictions in general probabilistic theories, arXiv:1912.08538 (2019).
 - [22] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, Publications of the Scuola Normale Superiore (Scuola Normale Superiore, 2011).
 - [23] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory. From Uncertainty to Entanglement* (Cambridge University Press, 2012).
 - [24] M. Plávala, All measurements in a probabilistic theory are compatible if and only if the state space is a simplex, *Phys. Rev. A* **94**, 042108 (2016), arXiv:1608.05614.
 - [25] T. Heinosaari, T. Miyadera, and M. Ziman, An Invitation to Quantum Incompatibility, *J. Phys. A Math. Theor.* **49**, 123001 (2015), arXiv:1511.07548.
 - [26] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens, Comparing the degrees of incompatibility inherent in probabilistic physical theories, *Europhysics Lett.* **103**, 10002 (2013), arXiv:1210.4142.
 - [27] A. Jenčová and M. Plávala, Conditions on the existence of maximally incompatible two-outcome measurements in General Probabilistic Theory, *Phys. Rev. A* **96**, 022113 (2017), arXiv:1703.09447.
 - [28] A. Jenčová, Incompatible measurements in a class of general probabilistic theories, *Phys. Rev. A* **98**, 012133 (2018), arXiv:1705.08008.
 - [29] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory, *Phys. Rev. Lett.* **103**, 230402 (2009), arXiv:0905.2998.
 - [30] N. Stevens and P. Busch, Steering, incompatibility, and Bell-inequality violations in a class of probabilistic theories, *Phys. Rev. A* **89**, 022123 (2014), arXiv:1311.1462.
 - [31] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Theoretical framework for quantum networks, *Phys. Rev. A* **80**, 022339 (2009), arXiv:0904.4483.

- [32] M. Ziman, Process positive-operator-valued measure: A mathematical framework for the description of process tomography experiments, *Phys. Rev. A* **77**, 062112 (2008), arXiv:0802.3862.
- [33] R. Bhatia, *Positive Definite Matrices*, Princeton Series in Applied Mathematics (Princeton University Press, 2009).
- [34] S. Woronowicz, Positive maps of low dimensional matrix algebras, *Reports Math. Phys.* **10**, 165 (1976).
- [35] M. Horodecki, P. Horodecki, and R. Horodecki, Separability of mixed states: necessary and sufficient conditions, *Phys. Lett. A* **223**, 1 (1996).

Appendix A: The Popescu-Rohrlich box

Let x be a non-local box such that

$$X_{\text{CHSH}} = E(A, B) + E(A, B') + E(A', B) - E(A', B') = 4.$$

Since all the correlations are in $[-1, 1]$, we must have

$$E(A, B) = E(A, B') = E(A', B) = -E(A', B') = 1.$$

It is easily seen that this happens if and only if we have

$$\begin{aligned} P_x(1, 1|A, B) + P_x(-1, -1|A, B) &= 1, \\ P_x(1, 1|A, B') + P_x(-1, -1|A, B') &= 1, \\ P_x(1, 1|A', B) + P_x(-1, -1|A', B) &= 1, \\ P_x(1, -1|A', B') + P_x(-1, 1|A', B') &= 1, \end{aligned}$$

with all other probabilities equal to 0. From this and the non-signaling conditions, we obtain

$$\begin{aligned} P_x(1, 1|A, B) &= P_x(1, 1|A, B) + P_x(1, -1|A, B) \\ &= P_x(1, 1|A, B') + P_x(1, -1|A, B') \\ &= P_x(1, 1|A, B'). \end{aligned}$$

In a similar fashion one may show that all the nonzero probabilities must be equal, which implies the equality (10).

Appendix B: Implementation of the PR-box on the square state space

Let S be the square state space and let A_0, A_1 be the maximally incompatible two-outcome measurements corresponding to the effects $\pi_0, \pi_1 \in E(S)$ respectively. Since $\{1, \pi_0, \pi_1\}$ form a basis of $A(S)$, it is clear that any element $\phi \in S \hat{\otimes} S \subset A(S)^* \otimes A(S)^*$ is uniquely determined by the values $(f \otimes g)(\phi)$, where $f, g \in \{\pi_0, 1 - \pi_0, \pi_1, 1 - \pi_1\}$, which are exactly the outcome probabilities $P_\phi(\epsilon, \eta|C, D)$, $C, D \in \{A_0, A_1\}$, $\epsilon, \eta \in \{-1, 1\}$. So if there is an implementation of the PR-box with the measurements $\{A_0, A_1\}$ on both sides, it must be unique. Moreover, it follows by Corollary 1 that any other implementation, with other maximally incompatible pairs of measurements, is obtained by applying a local isomorphism on each copy of S .

Let us now find a state $\phi_S \in S \hat{\otimes} S$, satisfying the equalities (10). Every $\phi \in A(S)^* \otimes A(S)^*$ can be written as

$$\phi = \psi_{00} \otimes s_{00} + \psi_{10} \otimes s_{10} + \psi_{01} \otimes s_{01}$$

for some $\psi_{00}, \psi_{10}, \psi_{01} \in A(S)^*$. By the characterization of $E(S)$ in Example 3, we see by applying the maps $1 \otimes 1$, $\text{id} \otimes \pi_0$, $\text{id} \otimes (1 - \pi_0)$, $\text{id} \otimes \pi_1$, $\text{id} \otimes (1 - \pi_1)$ that $\phi \in S \hat{\otimes} S$ if and only if

$$1(\psi_{00} + \psi_{10} + \psi_{01}) = 1$$

and

$$\begin{aligned} \psi_{10} &\geq 0, & \psi_{00} + \psi_{10} &\geq 0, \\ \psi_{01} &\geq 0, & \psi_{00} + \psi_{01} &\geq 0, \end{aligned}$$

Writing $\psi_{00}, \psi_{01}, \psi_{10}$ in the basis $\{s_{00}, s_{10} - s_{00}, s_{01} - s_{00}\}$ and using the fact that $\{1, \pi_0, \pi_1\}$ is the dual basis (see Example 3), we see that the required equalities hold if we put

$$\psi_{00} = \frac{1}{2}(s_{00} - s_{10}), \quad \psi_{10} = \frac{1}{2}s_{11}, \quad \psi_{01} = \frac{1}{2}s_{10},$$

and it is easily checked that the conditions required for $\phi \in S \hat{\otimes} S$ are satisfied (the last inequality follows from $s_{00} + s_{11} = s_{10} + s_{01}$). This gives (11).

Appendix C: Proof of Thm. 1

Since $X_{\text{CHSH}} = 4$, we see that the outcome probabilities must satisfy (10). Let $f_A, f_{A'}, f_B, f_{B'} \in E(K)$ be the effects corresponding to the four measurements and put

$$\begin{aligned} x_{00} &= 2((1 - f_A) \otimes \text{id})(\phi), & x_{11} &= 2(f_A \otimes \text{id})(\phi) \\ x_{01} &= 2((1 - f_{A'}) \otimes \text{id})(\phi), & x_{10} &= 2(f_{A'} \otimes \text{id})(\phi). \end{aligned}$$

From the fixed values of the outcome probabilities, it can be checked that $x_{i,j} \in K$ and that for all $i, j \in \{0, 1\}$ we have

$$f_B(x_{ij}) = i, \quad f_{B'}(x_{ij}) = j$$

and

$$x_{00} + x_{11} = x_{10} + x_{01},$$

i.e. the points x_{ij} form a witness square for B, B' . It follows that B and B' are maximally incompatible and we can prove in the same way that A and A' are maximally incompatible as well. Let ι_A, Π_A and ι_B, Π_B be the affine maps from Prop. 1, resp. their positive linear extensions. Then it is easily seen from (7) that we have

$$\begin{aligned} \pi_0 \circ \Pi_A &= f_A, & \pi_1 \circ \Pi_A &= f_{A'}, \\ \pi_0 \circ \Pi_B &= f_B, & \pi_1 \circ \Pi_B &= f_{B'}. \end{aligned}$$

It follows that $(\Pi_A \otimes \Pi_B)(\phi) \in S \hat{\otimes} S$ with the pair of two-outcome measurements A_0, A_1 applied on both sides

is an implementation of the PR-box, hence we must have $(\Pi_A \otimes \Pi_B)(\phi) = \phi_S$.

Next, note that $P_A := \iota_A \circ \Pi_A$ ($P_B := \iota_B \circ \Pi_B$) is a positive projection on $A(K)^*$ onto the range of the map ι_A (ι_B). We then have

$$\begin{aligned}\phi &= (P_A \otimes P_B)(\phi) + (id - P_A \otimes P_B)(\phi) \\ &= (\iota_A \otimes \iota_B)(\phi_S) + \phi^\perp,\end{aligned}$$

where $(P_A \otimes P_B)(\phi^\perp) = 0$, which implies that

$$(\Pi_A \otimes \Pi_B)(\phi^\perp) = (\Pi_A \otimes \Pi_B) \circ (P_A \otimes P_B)(\phi^\perp) = 0.$$

Now note that using (7) and the definition of x_{ij} , we get

$$\begin{aligned}(\Pi_A \otimes id)(\phi) &= \frac{1}{2}((s_{00} - s_{10}) \otimes x_{00} \\ &\quad + s_{11} \otimes x_{10} + s_{10} \otimes x_{01}) \\ &= (id \otimes \iota_B)(\phi_S) = (\Pi_A \otimes id)((\iota_A \otimes \iota_B)(\phi_S))\end{aligned}$$

and hence $(\Pi_A \otimes id)(\phi^\perp) = 0$. In a similar manner, we obtain that $(id \otimes \Pi_B)(\phi^\perp) = 0$, which implies that $\phi^\perp \in \ker(\Pi_A) \otimes \ker(\Pi_B)$.

Appendix D: The proof of Prop. 4

Let A, A' be the pair of maximally incompatible two-outcome measurements corresponding to the choice of the ONB $|\eta_0^A\rangle, |\eta_1^A\rangle$ (Prop. 3) and similarly let B, B' be the measurements for $|\eta_0^B\rangle, |\eta_1^B\rangle$. Let $U_A, U_B : \mathcal{H} \rightarrow \mathcal{H}$ denote the unitaries given as

$$U_A|i\rangle = |\eta_i^A\rangle, \quad U_B|i\rangle = |\eta_i^B\rangle.$$

Let Φ be a PR-channel that maximally violates the CHSH inequality with this choice of measurements and let $C = C(\Phi)$ have the block-diagonal form

$$C = \sum_{\alpha, \beta \in \{0,1\}^2} C_{\alpha, \beta} \otimes |\alpha\rangle\langle\beta|_{\text{in}}.$$

Then Φ must satisfy (10), where for all $C \in \{A, A'\}$ and $D \in \{B, B'\}$, we have

$$\begin{aligned}P_\Phi(1, 1|C, D) &= \text{Tr}(\text{SWAP}_{23}(F_C \otimes F_D)C) \\ P_\Phi(-1, 1|C, D) &= \text{Tr}(\text{SWAP}_{23}(F_C^\perp \otimes F_D)C) \\ P_\Phi(1, -1|C, D) &= \text{Tr}(\text{SWAP}_{23}(F_C \otimes F_D^\perp)C) \\ P_\Phi(-1, -1|C, D) &= \text{Tr}(\text{SWAP}_{23}(F_C^\perp \otimes F_D^\perp)C),\end{aligned}$$

here F_C, F_C^\perp are the PPOVM operators corresponding to $C \in \{A, A'\}$ and F_D, F_D^\perp correspond to $D \in \{B, B'\}$. Let $Q_h, h = 1, \dots, 16$ be operators obtained as

$$Q_h = \text{SWAP}_{23}(H_A \otimes H_B),$$

where $H_A \in \{F_A, F_A^\perp, F_{A'}, F_{A'}^\perp\}$ and similarly for H_B . Then Q_h are mutually orthogonal rank 1 product projections and all the values of $\text{Tr}(Q_h C)$ are either 0 or 1/2. Let

$$P = \sum \{Q_h, \text{Tr}(Q_h C) = 1/2\}.$$

Since $C \geq 0$ and $\text{Tr}((\mathbb{1} - P)C) = 0$, we must have $C = PCP$. One can check by (10) that

$$P = \sum_{\alpha \in \{0,1\}^2} U_\alpha U_\alpha^* \otimes |\alpha\rangle\langle\alpha|_{\text{in}}$$

for the isometries $U_\alpha : \mathbb{C}^2 \rightarrow \mathcal{H}_{\text{out}}$, given as

$$U_{km} = (U_A^k \otimes U_B^m V^{m,k})W, \quad k, m = 0, 1,$$

where $V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $W : \mathbb{C}^2 \rightarrow \mathcal{H}_{\text{out}}$ is the isometry given as $W|i\rangle = |ii\rangle$. It follows that

$$C = PCP = \sum_{\alpha, \beta} U_\alpha U_\alpha^* C_{\alpha, \beta} U_\beta U_\beta^* \otimes |\alpha\rangle\langle\beta|_{\text{in}},$$

so that for $\alpha = km, \beta = ln$ we have

$$C_{\alpha, \beta} = \sum_{i, j} c_{ij}^{\alpha, \beta} U_A^k |i\rangle\langle j| U_A^{-l} \otimes U_B^m V^{m,k} |i\rangle\langle j| V^{-l,n} U_B^{-n}$$

for some coefficients $c_{ij}^{\alpha, \beta}$.

Since Φ is non-signaling, the Choi matrix must satisfy the conditions

$$\text{Tr}_{A, \text{out}}(C) = \mathbb{1}_{A, \text{in}} \otimes C_B, \quad \text{Tr}_{B, \text{out}}(C) = \mathbb{1}_{B, \text{in}} \otimes C_A,$$

where $C_B \in B^+(\mathcal{H}_{B, \text{out}} \otimes \mathcal{H}_{B, \text{in}})$ and $C_A \in B^+(\mathcal{H}_{A, \text{out}} \otimes \mathcal{H}_{A, \text{in}})$. This amounts to

$$\text{Tr}_A(C_{km, ln}) = 0, \quad k \neq l, \quad \forall m, n \quad (\text{D1})$$

$$\text{Tr}_A(C_{0m, 0n}) = \text{Tr}_A C_{1m, 1n}, \quad \forall m, n \quad (\text{D2})$$

$$\text{Tr}_B(C_{km, ln}) = 0, \quad m \neq n, \quad \forall k, l \quad (\text{D3})$$

$$\text{Tr}_B(C_{k0, l0}) = \text{Tr}_B(C_{k1, l1}), \quad \forall k, l. \quad (\text{D4})$$

If $k = l$, we have

$$\text{Tr}_A(C_{km, kn}) = \sum_i c_{ii}^{km, kn} U_B^m V^{m,k} |i\rangle\langle i| V^{-k,n} U_B^{-n}.$$

By (D2), we see that we must have $c_{ii}^{km, kn} = 0, \forall i$ whenever $m \neq n$. Similarly, by (D4) we obtain for $k \neq l$ that $c_{ii}^{k0, l0} = 0, \forall i$ and $c_{ij}^{k1, l1} = 0, \forall i \neq j$. Note also that $c_{ii}^{km, km} = 1/2$ by (10).

Looking at the conditions (D1) and (D3), we see that there are two different possibilities on each side: either $\langle i|U_A|j\rangle = 0$ for some i, j or $\langle i|U_A|j\rangle \neq 0$ for all i, j , the same for U_B . The first condition means that U_A is diagonal or off-diagonal (i.e. $\langle i|U_A|i\rangle = 0$). Since the diagonal (off-diagonal) elements in these cases only correspond to scalar factors of the basis vectors, they may safely be put to 1, so that $U_A = \mathbb{1}$ ($U_A = V$).

Assume that $U_A \neq \mathbb{1}, U_A \neq V$, then we obtain from (D1) that $C_{km, ln} = 0$ whenever $k \neq l$. There are essentially two off-diagonal blocks left:

$$C_{00, 01} = C_{01, 00}^* = \sum_{i \neq j} c_{ij}^{00, 01} |i\rangle\langle j| \otimes |i\rangle\langle j| U_B^*$$

Table I. All possible forms of C .

U_A/U_B	\neq	$\mathbb{1}$	V
\neq	$r = 1$	$r = 2$	$r = 2$
		$\Delta_1 = \{00, 01\}$	$\Delta_1 = \{10, 11\}$
		$V_{\text{out}} = W$	$V_{\text{out}} = (U_A \otimes \mathbb{1})W$
$\mathbb{1}$	$r = 2$	$r = 3$	$r = 3$
	$\Delta_1 = \{01, 11\}$	$\Delta_0 = \{11\}$	$\Delta_0 = \{01\}$
	$V_{\text{out}} = W$	$V_{\text{out}} = W$	$V_{\text{out}} = W$
V	$r = 2$	$r = 3$	$r = 3$
	$\Delta_1 = \{00, 10\}$	$\Delta_0 = \{10\}$	$\Delta_0 = \{00\}$
	$V_{\text{out}} = (\mathbb{1} \otimes U_B)W$	$V_{\text{out}} = W$	$V_{\text{out}} = (\mathbb{1} \otimes V)W$

and

$$C_{10,11}(= C_{11,10}^*) = \sum_{i \neq j} c_{ij}^{11,01} U_A |i\rangle \langle j| U_A^* \otimes |i\rangle \langle j| V^* U_B^*.$$

The condition (D3) implies that $C_{00,01} \neq 0$ only if U_B is diagonal, in which case $U_B = \mathbb{1}$ and

$$C_{00,01} = \sum_{i \neq j} c_{ij}^{00,01} |i\rangle \langle j| \otimes |i\rangle \langle j|.$$

Similarly, we can have $C_{11,01} \neq 0$ only if $U_B = V$, in which case

$$C_{11,01} = \sum_{i \neq j} c_{ij}^{11,01} U_A |i\rangle \langle j| U_A^* \otimes |i\rangle \langle j|.$$

It follows that if also $U_B \neq \mathbb{1}$, $U_B \neq V$, then $C = \sum_{\alpha} C_{\alpha,\alpha} \otimes |\alpha\rangle \langle \alpha|$ is block-diagonal, which means that C has the form (17) with $r = 1$. If $U_B = \mathbb{1}$ then we have $\Delta_1 = \{00, 01\}$ and $V_{\text{out}} = U_{00} = U_{01} = W$, for $U_B = V$ we obtain $\Delta_1 = \{10, 11\}$ and $V_{\text{out}} = U_{10} = U_{11} = (U_A \otimes \mathbb{1})W$. We may apply similar reasoning in all cases with $U_B \neq \mathbb{1}$, $U_B \neq V$.

Next we turn to the cases when $U_A, U_B \in \{\mathbb{1}, V\}$. We will provide a proof for $U_A = U_B = \mathbb{1}$, all other cases are similar. Here we obtain from (D1) - (D4) that all off-diagonal blocks $C_{\alpha,\beta}$ must have $c_{ii}^{\alpha,\beta} = 0$ and $C_{km,ln} = 0$ whenever $k.m \neq l.n$, which implies that $C_{\alpha,11} = 0$ for all $\alpha \neq 11$. Note that $U_{00} = U_{10} = U_{01} = W$, so that C has the form (17) with $\Delta_0 = \{11\}$ and $V_{\text{out}} = W$. All possible forms of C in the different cases are summarized in table I.