Witnessing causal nonseparability

New Journal of Physics

The open access journal at the forefront of physics

CrossMark

FAST TRACK COMMUNICATION

Witnessing causal nonseparability

Mateus Araújo[1,2], **Cyril Branciard**[3], **Fabio Costa**[1,2,4], **Adrien Feix**[1,2], **Christina Giarmatzi**[4,5] and **Časlav Brukner**[1,2]

1 Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Vienna, Austria
2 Institute for Quantum Optics and Quantum Information (IQOQI), Boltzmanngasse 3, A-1090 Vienna, Austria
3 Institut Néel, CNRS and Université Grenoble Alpes, F-38042 Grenoble Cedex 9, France
4 Centre for Engineered Quantum Systems, School of Mathematics and Physics, The University of Queensland, St Lucia, QLD 4072, Australia
5 Centre for Quantum Computer and Communication Technology, School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia

E-mail: mateus.santos@univie.ac.at and cyril.branciard@neel.cnrs.fr
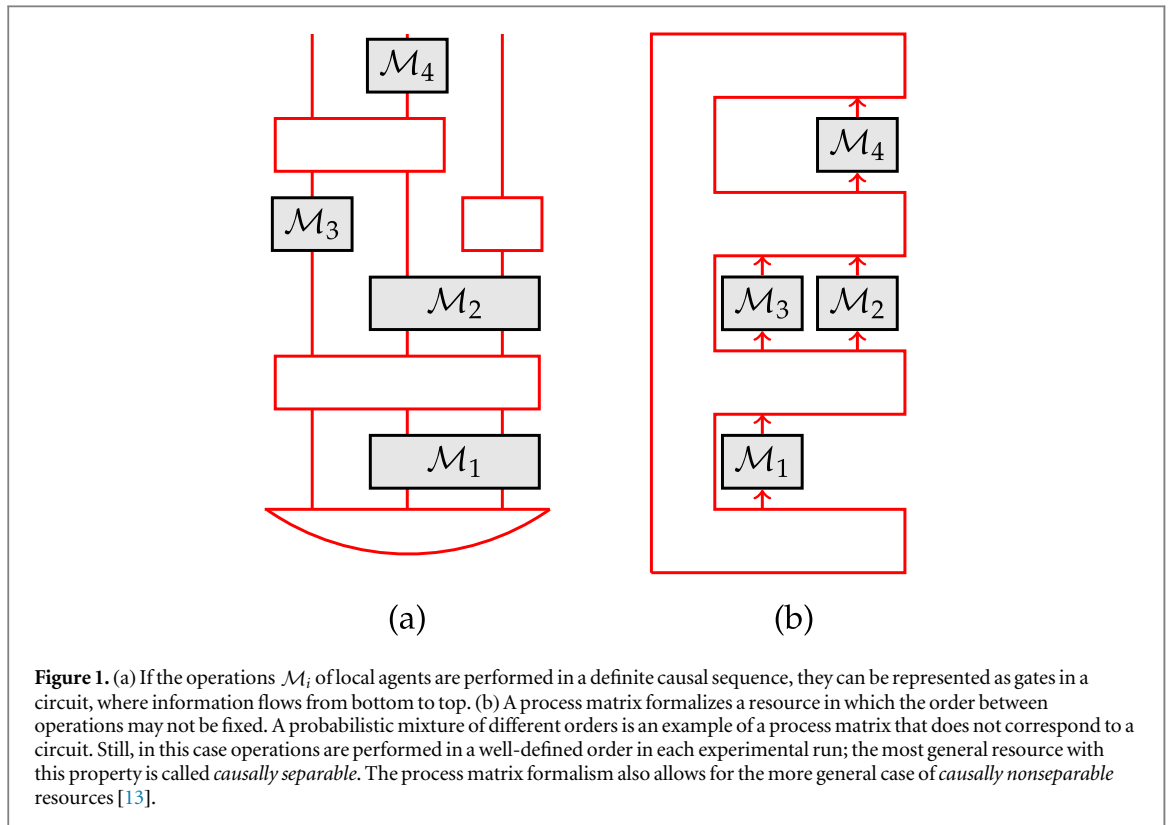
Abstract

Our common understanding of the physical world deeply relies on the notion that events are ordered with respect to some time parameter, with past events serving as causes for future ones. Nonetheless, it was recently found that it is possible to formulate quantum mechanics without any reference to a global time or causal structure. The resulting framework includes new kinds of quantum resources that allow performing tasks—in particular, the violation of *causal inequalities*—which are impossible for events ordered according to a global causal order. However, no physical implementation of such resources is known. Here we show that a recently demonstrated resource for quantum computation—the *quantum switch*—is a genuine example of 'indefinite causal order'. We do this by introducing a new tool—the *causal witness*—which can detect the *causal nonseparability* of any quantum resource that is incompatible with a definite causal order. We show however that the quantum switch does not violate any causal inequality.

1. Introduction

It is commonly assumed that information is processed through a series of operations which are performed according to a specific order. This is justified by the assumption of a global, underlying time parameter according to which all operations can be ordered. A convenient representation of this structure is that of a circuit [1], figure 1(a), in which systems are 'wires' that connect 'boxes', which represent operations performed on the systems. At a more abstract level, a circuit only imposes a given *causal structure* between operations, as the time order between operations that can be performed in parallel is irrelevant. The circuit framework is also ubiquitous in the study of quantum foundations to formalize generalized, possibly post-quantum, probabilistic theories [2–5].

It has been suggested that such a framework might be too restrictive to encompass the most general kinds of information processing allowed by quantum physics [6]. For example, one can consider protocols in which the order between different operations is controlled by a quantum degree of freedom. It has been shown that such protocols exploiting a so-called 'quantum switch' not only provide computational advantage over standard, time-ordered, ones [7, 8], but they are also physically realizable and a first experimental proof-of-principle has been recently demonstrated [9]. At a more fundamental level, an underlying time or causal order might not be well-defined in a theory that combines the dynamical causal structure of general relativity and the probabilistic nature of quantum mechanics [10–12].

It is therefore natural to ask what the most general resources allowed by quantum mechanics beyond the circuit model are. In [13] the *process matrix formalism* was proposed as a general framework to describe resources that can be accessed in 'local laboratories' and which are locally in agreement with quantum physics, see figure 1(b).

Figure 1. (a) If the operations $\mathcal{M}_i$ of local agents are performed in a definite causal sequence, they can be represented as gates in a circuit, where information flows from bottom to top. (b) A process matrix formalizes a resource in which the order between operations may not be fixed. A probabilistic mixture of different orders is an example of a process matrix that does not correspond to a circuit. Still, in this case operations are performed in a well-defined order in each experimental run; the most general resource with this property is called *causally separable*. The process matrix formalism also allows for the more general case of *causally nonseparable* resources [13].

Causal relations are defined operationally in this formalism. If, for example, through appropriate state preparations, an agent *A* can influence the outcomes of measurements performed by an agent *B*, whereas *B* is never able to influence *A*, then *A* causally precedes *B* by definition and, in this case, the physical resources available to them can in fact be represented as a circuit. A first example of a resource that cannot be represented as a circuit is a *probabilistic mixture* of circuits: a definite order still exists between *A* and *B* in each run of an experiment, but which order is realized in a given run is only specified according to some probability distribution. Resources compatible with a definite causal order, in this broader sense, are called *causally separable*. Surprisingly, the formalism also allows for *causally nonseparable* resources, which are incompatible with any definite order between operations. It was found that a set of agents with access to a specific causally nonseparable resource could perform a task, the violation of a *causal inequality*, which is impossible for arbitrary causally ordered strategies, even allowing probabilistic mixtures of orders [13]. However, there is no physical interpretation for such resources and no physically realizable protocol is known which can violate a causal inequality.

It is therefore not completely clear what is the precise relation between 'quantum correlations with no causal order', which violate causal inequalities, and physically implementable resources, such as the quantum switch, which outperform causally ordered ones. To understand this relation, a crucial observation is that the causal inequalities are *device-independent* constraints: they are formulated independently of the physics of the systems or the specific apparatuses employed. On the other hand, the tasks discussed in [7, 8] include additional assumptions, for example that in each laboratory quantum systems of a definite dimension have to be used. It is clear that, given additional restrictions, it is more difficult for causally ordered agents to perform certain tasks and, consequently, it can be easier to detect the lack of causal order in a physical resource.

The aim of the present work is to develop a general framework for the device-*dependent* detection of causal nonseparability. The central tool we introduce is what we call a *causal witness*, which represents a set of quantum operations, such as unitaries, channels, state preparations, and measurements, whose expectation value is non-negative as long as all the operations are performed in a definite causal order, i.e., as long as only causally separable resources are used. The observation of a negative expectation value is thus sufficient to conclude that the operations were not performed in a definite order. The concept is analogous to that of an entanglement witness: an observable that has a non-negative expectation value for separable states but can have a negative expectation value for specific entangled states.

We find that, for every causally nonseparable process, it is possible to construct a causal witness that detects it. Importantly, and differently from the case of entanglement witnesses, it is possible to use this method to write

necessary and sufficient conditions for causal separability in a form that can be checked efficiently using semidefinite programming (SDP).

The tools developed are applied to the study of the quantum switch as a resource within the process matrix formalism. We show that, indeed, the quantum switch corresponds to a causally nonseparable process. We show that the protocol of [7] can be reformulated as a causal witness which detects the causal nonseparability of the quantum switch. We also find new, more efficient witnesses, which could be useful for experimental implementations.

We finally address the question of whether the quantum switch can pass any device-independent test of causal nonseparability. As it turns out, this is not possible: we prove that a broad class of resources, including the quantum switch, cannot violate any causal inequality.

The paper is organized as follows: in section 2, we review the process matrix formalism, giving a convenient characterization of general and causally separable process matrices for the cases of interest. In section 3, we introduce and characterize the central concept of causal witness, and we present efficient algorithms for finding witnesses and for proving the causal (non)separability of a general process matrix. In section 4 we formalise the quantum switch as a process matrix. We proceed to prove its causal nonseparability in section 5, through the use of causal witnesses. One such witness is the task proposed in [7], that we optimize to increase its resistance to noise. Finally, we clarify in section 6 the link between causal witnesses and causal inequalities and show that the quantum switch cannot violate any causal inequality.

2. The process matrix formalism

In the general scenario we consider in this paper, N parties A^i establish correlations by exchanging physical systems between their laboratories. Each party opens their laboratory only once to let an incoming system enter and to send an outgoing system out; they can act on these systems by performing an arbitrary operation in their local laboratory, which can yield different measurement outcomes. The causal relations between the parties (i.e., the ordering of events) are not *a priori* specified. The most general situation compatible with the assumption that *the operations performed in each local laboratory can be described by the quantum formalism* can be conveniently represented in the 'process matrix' formalism introduced in [13]. This extends the 'comb' formalism of [14], which describes causally ordered quantum networks. The aim of the formalism is to characterize all possible probability distributions that can be obtained in our general scenario. The key concept is that of a *process*, which can be understood as the external resource determining the statistics of the local operations, and which generalizes both the notions of quantum state and of quantum channel. The *process matrix* is a useful mathematical representation of such a concept. We shall use these two terms interchangeably.
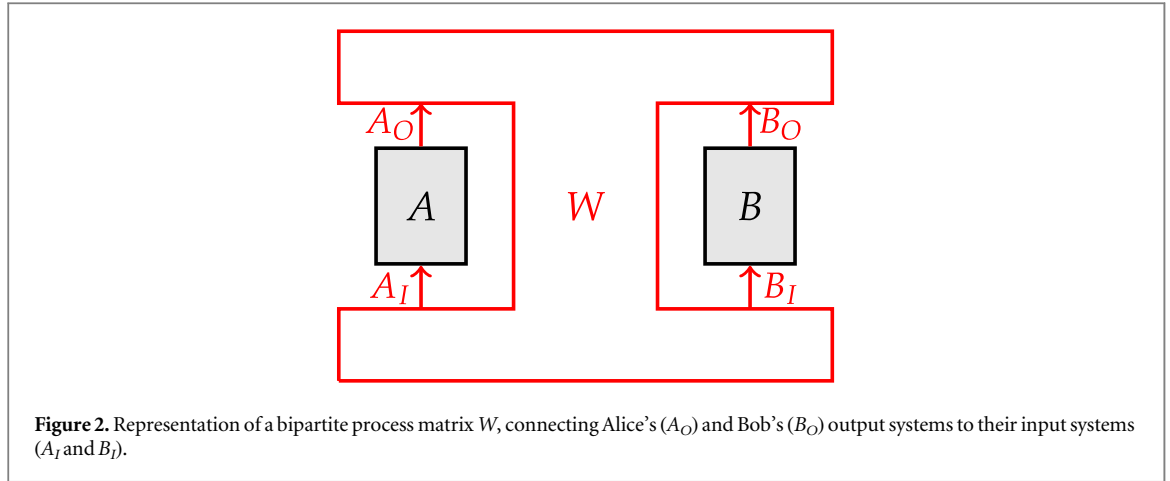
2.1. Local operations

Each party A acts in a *local quantum laboratory*, which can be identified by an input Hilbert space $\mathcal{H}^{A_I}$ and an output Hilbert space $\mathcal{H}^{A_O}$. The dimensions d_{A_I} and d_{A_O} of input and output spaces do not have to be equal, as ancillary systems can be added or discarded during an operation; we shall nevertheless assume throughout the paper that all Hilbert spaces are finite-dimensional. According to quantum theory, the most general local operation is described by a completely positive (CP), trace non-increasing map $\mathcal{M}^A : A_I \rightarrow A_O$ [15], where we write A_I, respectively A_O, for the space of Hermitian linear operators over the Hilbert space $\mathcal{H}^{A_I}$, resp. $\mathcal{H}^{A_O}$. Examples of CP maps are deterministic operations, such as unitaries or quantum channels, or (generalized) measurements. In general, a label a, denoting the measurement outcome, is associated with the CP map $\mathcal{M}_a^A$. The choice of operation (e.g. of measurement setting) is represented by an *instrument* [16], which is defined as the collection $\mathcal{J}^A = \{\mathcal{M}_a^A\}_{a=1}^m$ of CP maps associated to all measurement outcomes, characterized by the property that $\sum_{a=1}^m \mathcal{M}_a^A$ is CP and trace-preserving (CPTP). An instrument generalizes the notion of positive operator-valued measure (POVM) to include the transformations applied to the system; it reduces to a POVM for one-dimensional output spaces. When the choice of operation is described by a classical variable x, we will express such a dependence explicitly as $\mathcal{J}_x^A = \{\mathcal{M}_{a|x}^A\}_{a=1}^m$.

A convenient representation of CP maps is given by the Choi–Jamiołkowski (CJ) isomorphism [17]. For a CP map $\mathcal{M}_a^A : A_I \rightarrow A_O$, its corresponding CJ matrix is defined here as

$$M_a^{A_I A_O} := \left[\mathcal{I} \otimes \mathcal{M}_a^A (|\mathbb{1}\rangle\rangle\langle\langle\mathbb{1}|) \right]^T \in A_I \otimes A_O, \tag{1}$$

where $\mathcal{I}$ is the identity map, $|\mathbb{1}\rangle\rangle \equiv |\mathbb{1}\rangle\rangle^{A_I A_I} := \sum_j |j\rangle^{A_I} \otimes |j\rangle^{A_I} \in \mathcal{H}^{A_I} \otimes \mathcal{H}^{A_I}$ is a (non-normalized) maximally entangled state, and T denotes matrix transposition with respect to the chosen orthonormal basis $\{|j\rangle^{A_I}\}$ of $\mathcal{H}^{A_I}$. Some useful properties of the CJ isomorphism are given in appendix A.1. A map is CP if and only if its CJ representation is positive semidefinite, while the trace-preserving condition is equivalent to $\text{tr}_{A_O} M^{A_I A_O} = \mathbb{1}^{A_I}$ (where tr_{A_O} denotes the partial trace over A_O, and $\mathbb{1}^{A_I}$ is the identity matrix in A_I). An instrument is therefore

Figure 2. Representation of a bipartite process matrix W, connecting Alice's (A_O) and Bob's (B_O) output systems to their input systems (A_I and B_I).

equivalently represented as a set

$$\left\{ M_a^{A_I A_O} \right\}_{a=1}^m, \qquad M_a^{A_I A_O} \geqslant 0, \qquad \mathrm{tr}_{A_O} \sum_{a=1}^m M_a^{A_I A_O} = \mathbb{1}^{A_I}. \tag{2}$$

2.2. Process matrices

As discussed in [13], requiring that quantum mechanics holds locally implies that the probability that the N parties A^i observe the outcomes $a_1, \ldots, a_N$, for a choice of operations $x_1, \ldots, x_N$, is a multilinear function $P\left(\mathcal{M}_{a_1|x_1}^{A^1}, \ldots, \mathcal{M}_{a_N|x_N}^{A^N} \right)$ of the corresponding CP maps $\mathcal{M}_{a_1|x_1}^{A^1}, \ldots, \mathcal{M}_{a_N|x_N}^{A^N}$. Using the CJ representation, it was shown that these probabilities can then be expressed as

$$P\left(M_{a_1|x_1}^{A^1}, \ldots, M_{a_N|x_N}^{A^N} \right) = \mathrm{tr}\left[\left(M_{a_1|x_1}^{A_I^1 A_O^1} \otimes \ldots \otimes M_{a_N|x_N}^{A_I^N A_O^N} \right) W \right], \tag{3}$$

for some Hermitian operator $W \in A_I^1 \otimes A_O^1 \otimes \ldots \otimes A_I^N \otimes A_O^N$ called a *process matrix*, which describes the general quantum resource connecting the local laboratories.

The set of valid process matrices is defined by requiring that probabilities are well-defined—that is, they must be non-negative and must sum up to one—for all possible operations, including operations that involve, in each laboratory, local interactions with ancillary systems that may be entangled with the other laboratories. As we show in appendix B, these conditions are equivalent to

$$W \geqslant 0, \tag{4}$$

$$\mathrm{tr}\, W = d_O, \tag{5}$$

$$W = L_V(W), \tag{6}$$

where $d_O = d_{A_O^1} \ldots d_{A_O^N}$, and L_V is a projector onto the linear subspace $\mathcal{L}_V \subset A_I^1 \otimes A_O^1 \otimes \ldots \otimes A_I^N \otimes A_O^N$ defined in appendix B. We will denote the closed convex cone of non-normalized processes defined by (4) and (6) by $\mathcal{W}$.

In the case of two parties A (Alice) and B (Bob), see figure 2, these conditions on $W \in A_I \otimes A_O \otimes B_I \otimes B_O$ reduce to

$$W \geqslant 0, \tag{7}$$

$$\mathrm{tr}\, W = d_O, \tag{8}$$

$${}_{B_I B_O}W = {}_{A_O B_I B_O}W, \tag{9}$$

$${}_{A_I A_O}W = {}_{A_I A_O B_O}W, \tag{10}$$

$$W = {}_{B_O}W + {}_{A_O}W - {}_{A_O B_O}W, \tag{11}$$

where (here and throughout the paper) the operator ${}_X \cdot$ denotes the CPTP map consisting in tracing out the subsystem X and replacing it by the normalized identity operator, formally defined as

$${}_X W = \frac{\mathbb{1}^X}{d_X} \otimes \mathrm{tr}_X W. \tag{12}$$

2.2.1. Non-signalling and one-way-signalling process matrices

Two important particular cases of process matrices may shed light on the above definition. The first case is when the process matrix does not allow for any signalling, and the second one is when it allows for signalling only in one fixed direction between the parties. They are discussed in more details in appendix A.2.

The first case is described by process matrices W satisfying

$$W = {}_{A_O^1 \ldots A_O^N} W = \rho^{A_I^1 \ldots A_I^N} \otimes \mathbb{1}^{A_O^1 \ldots A_O^N},$$
(13)

where $\rho^{A_I^1 \ldots A_I^N}$ is a density matrix representing an ordinary quantum state. In this case, the probability rule (3) reduces to the standard Born rule

$$P\left(M_{a_1|x_1}^{A^1}, \ldots, M_{a_N|x_N}^{A^N}\right) = \text{tr}\left[\left(E_{a_1|x_1}^{A_I^1} \otimes \ldots \otimes E_{a_N|x_N}^{A_I^N}\right)\rho\right],$$
(14)

where $E_{a_i|x_i}^{A_I^i} := \text{tr}_{A_O^i} M_{a_i|x_i}^{A_I^i A_O^i}$ are POVM elements.

The second case, of which the first one is a particular case, is described by process matrices W satisfying

$$W = {}_{A_O^N} W,$$
$$ {}_{A_I^N A_O^N} W = {}_{A_O^{N-1} A_I^N A_O^N} W,$$
$$\vdots$$
$$ {}_{A_I^2 A_O^2 \ldots A_I^N A_O^N} W = {}_{A_O^1 A_I^2 A_O^2 \ldots A_I^N A_O^N} W.$$
(15)

These conditions, first found in [14, 18], mean that party A^i can only signal to party A^j if $i < j$. The process is therefore compatible with the causal order $A^1 \prec A^2 \prec \ldots \prec A^N$. When this is the case, we write as a mnemonic

$$W = W^{A^1 \prec A^2 \prec \ldots \prec A^N}.$$
(16)

Process matrices of this form (and the obvious permutations) are called *causally ordered*. As shown in [14, 18], they correspond to standard (causally ordered) quantum circuits and can be implemented as quantum channels with memory between the parties.

2.2.2. Bipartite causally separable processes

According to equation (15), a bipartite causally ordered process matrix $W^{A \prec B}$ compatible with the order $A \prec B$ satisfies $W^{A \prec B} = {}_{B_O} W^{A \prec B}$ and ${}_{B_I B_O} W^{A \prec B} = {}_{A_O B_I B_O} W^{A \prec B}$. Note that the latter relation corresponds to equation (9) above, and is therefore automatically satisfied if $W^{A \prec B} \in \mathcal{L}_V$. Thus, a given matrix $W^{A \prec B} \in A_I \otimes A_O \otimes B_I \otimes B_O$ is a valid causally ordered process matrix compatible with the order $A \prec B$ if and only if $W^{A \prec B} \geqslant 0, \text{tr } W^{A \prec B} = d_O$,

$$W^{A \prec B} \in \mathcal{L}_V \quad \text{and} \quad W^{A \prec B} = {}_{B_O} W^{A \prec B}.$$
(17)

The analogous condition holds for the order $B \prec A$.

Note that a non-signalling process matrix W must be compatible with both orders $A \prec B$ and $B \prec A$. It must therefore satisfy $W = {}_{B_O} W = {}_{A_O} W$, or equivalently $W = {}_{A_O B_O} W$; we indeed recover the form of equation (13).

Following [13], we say that a bipartite process matrix W is *causally separable* if it can be decomposed as a convex combination of causally ordered processes, i.e., if it is of the form

$$W^{\text{sep}} = q \, W^{A \prec B} + (1 - q) \, W^{B \prec A},$$
(18)

with $0 \leqslant q \leqslant 1$. Ignoring the normalization constraint, the set of causally separable process matrices is a convex cone, which we denote by $\mathcal{W}^{\text{sep}}$. A process matrix that cannot be decomposed as in (18) is called *causally nonseparable*.
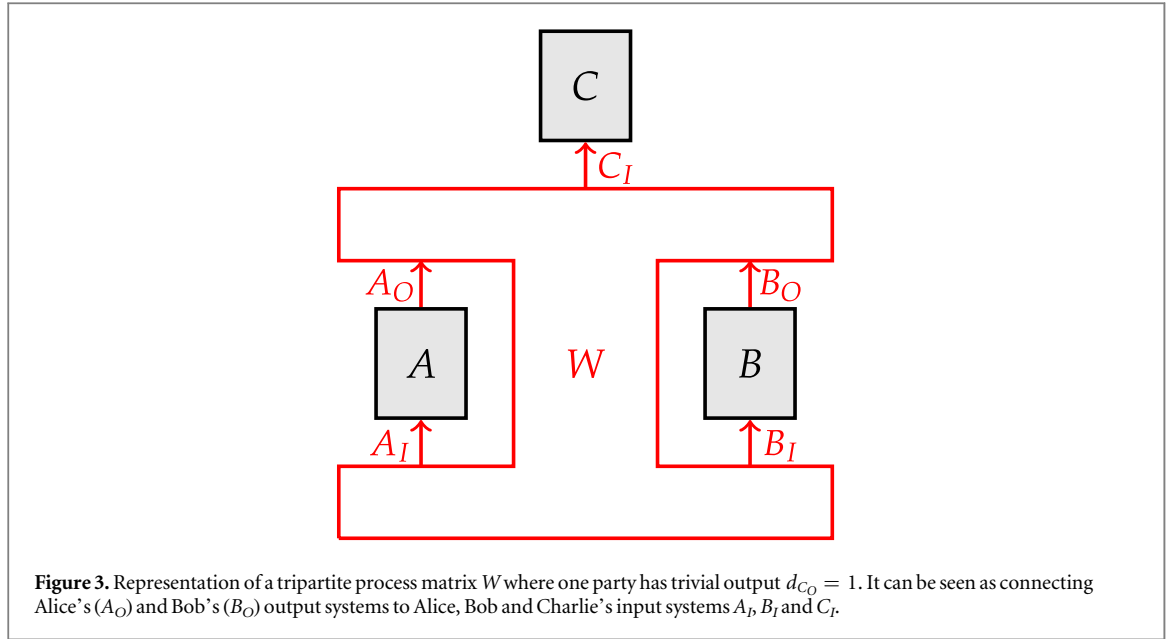
2.2.3. Tripartite causally separable processes

In this paper we will define tripartite causal separability only for processes where the output space of the third party C (Charlie) is trivial, i.e., $d_{C_O} = 1$ (see figure 3). As C cannot signal to the other parties, every process of this kind if compatible with C being last. Thus, only two causal orders are relevant in this case: $A \prec B \prec C$ and $B \prec A \prec C$. The conditions for process matrices being compatible with these orders are, according to equation (15),

$$W^{A \prec B \prec C} = {}_{C_O} W^{A \prec B \prec C},$$
(19)

$$ {}_{C_I C_O} W^{A \prec B \prec C} = {}_{B_O C_I C_O} W^{A \prec B \prec C},$$
(20)

$$ {}_{B_I B_O C_I C_O} W^{A \prec B \prec C} = {}_{A_O B_I B_O C_I C_O} W^{A \prec B \prec C},$$
(21)

Figure 3. Representation of a tripartite process matrix W where one party has trivial output $d_{C_O} = 1$. It can be seen as connecting Alice's (A_O) and Bob's (B_O) output systems to Alice, Bob and Charlie's input systems A_I, B_I and C_I.

and

$$W^{B \prec A \prec C} = {}_{C_O}W^{B \prec A \prec C}, \tag{22}$$

$$_{C_I C_O}W^{B \prec A \prec C} = {}_{A_O C_I C_O}W^{B \prec A \prec C}, \tag{23}$$

$$_{A_I A_O C_I C_O}W^{B \prec A \prec C} = {}_{B_O A_I A_O C_I C_O}W^{B \prec A \prec C}. \tag{24}$$

Since these three conditions together define a linear subspace, we can write them more succinctly as

$$W^{A \prec B \prec C} = L_{A \prec B \prec C}\left(W^{A \prec B \prec C}\right), \tag{25}$$

$$W^{B \prec A \prec C} = L_{B \prec A \prec C}\left(W^{B \prec A \prec C}\right), \tag{26}$$

where $L_{A \prec B \prec C}$ and $L_{B \prec A \prec C}$ are the projectors onto the aforementioned subspaces.

Therefore, when C's output space is trivial, we will call a tripartite process matrix W^{sep} causally separable if it is of the form

$$W^{\text{sep}} = q \, W^{A \prec B \prec C} + (1 - q) \, W^{B \prec A \prec C}, \tag{27}$$

with $0 \leqslant q \leqslant 1$. Ignoring the normalization constraint, this defines a convex cone $\mathcal{W}^{\text{sep}}_{3C}$. We will use this definition in section 5 to show that a recently introduced tripartite quantum resource, which yields information-processing advantages with respect to causally ordered processes [7, 8], is causally nonseparable.

The generalization of the notion of causal separability to a larger number of parties, with arbitrary dimensions of the output spaces, is not trivial. The reason is that one can consider situations in which an agent, through her local operations, could modify a classical variable that determines the causal order of agents in her future. In such a 'classical switch', operations would still be causally ordered in each run of an experiment, but it would not be possible to write the corresponding process matrix as a mixture of causally ordered ones. As this issue does not affect the cases treated here, we shall not consider it further. A more detailed analysis is presented in [19].

3. Causal witnesses

3.1. Definition and characterization

In this section we develop mathematical tools to identify, in the bipartite case, which process matrices are causally separable and which are not. In analogy with entanglement witnesses [20], we call a Hermitian operator S a *causal witness* (or *witness*, simply) if [6]

$$\text{tr}\left[S \, W^{\text{sep}}\right] \geqslant 0 \tag{28}$$

---

[6] Note that the bound 0 and the sign of the inequality are arbitrary; we choose them as in equation (28) for mathematical convenience.

for every causally separable process matrix W^{sep}. This definition is motivated by the separating hyperplane theorem [21]: since the set of causally separable processes is closed and convex, for every causally nonseparable process matrix W_{ns} there exists a causal witness $S_{W_{\text{ns}}}$ such that $\text{tr}[S_{W_{\text{ns}}} W_{\text{ns}}] < 0$.

To construct a witness for a given causally nonseparable process, we will start by characterizing the set of all causal witnesses in terms of linear constraints on a convex cone. This will allow us to cast the problem of finding a witness as an SDP problem. First, note that (28) is equivalent to

$$\text{tr}\Big[S\, W^{A \prec B} \Big] \geqslant 0 \quad \forall W^{A \prec B}, \tag{29a}$$

$$\text{tr}\Big[S\, W^{B \prec A} \Big] \geqslant 0 \quad \forall W^{B \prec A}. \tag{29b}$$

Let us focus on condition (29a). Using equation (17) and noting that for any valid process matrix W, $_{B_O}W$ is a valid causally ordered process matrix compatible with the order $A \prec B$, one finds that (29a) is equivalent to

$$\text{tr}\Big[S\big({}_{B_O}W \big) \Big] \geqslant 0 \quad \forall W \in \mathcal{L}_V,\ W \geqslant 0. \tag{30}$$

Thinking of the trace as the Hilbert–Schmidt inner product and noting that the map $_{B_O}\cdot$ is self-adjoint, we have that

$$\text{tr}\Big[S\big({}_{B_O}W \big) \Big] = \text{tr}\Big[\big({}_{B_O}S \big)\, W \Big], \tag{31}$$

and it is sufficient that $_{B_O}S \geqslant 0$ for the right-hand side to be non-negative for all valid W. An analogous argument shows that $_{A_O}S \geqslant 0$ is sufficient to satisfy condition (29b). We conclude that for S to be a causal witness, it is sufficient that

$$_{B_O}S \geqslant 0 \quad \text{and} \quad _{A_O}S \geqslant 0. \tag{32}$$

Note also that adding an operator $S^\perp$ belonging to the orthogonal complement $\mathcal{L}_V^\perp$ of $\mathcal{L}_V$ to any witness S gives another valid witness, since $\text{tr}[(S + S^\perp)W] = \text{tr}[SW]$ for any valid process matrix W. It turns out that this suffices to completely characterize the set of causal witnesses, as stated in the following theorem:

Theorem 1. *A Hermitian operator* $S \in A_I \otimes A_O \otimes B_I \otimes B_O$ *is a* causal witness *if and only if S can be written as*

$$S = S_P + S^\perp, \tag{33}$$

where S_P and $S^\perp$ are Hermitian operators such that

$$_{B_O}S_P \geqslant 0, \quad _{A_O}S_P \geqslant 0, \quad L_V\big(S^\perp\big) = 0. \tag{34}$$

The rather technical proof of this theorem is relegated to appendix C. This theorem provides a characterization of the closed convex cone of causal witnesses $\mathcal{S}$.

Since $S^\perp$ does not change the expectation value $\text{tr}[SW]$, it can freely be chosen to be for instance

$$S^\perp = L_V\big(S_P\big) - S_P, \tag{35}$$

so that $S = L_V(S_P)$. This has the effect of restricting witnesses to the subspace of valid processes $\mathcal{L}_V$, which have the following characterization:

Corollary 2. *A Hermitian operator* $S \in \mathcal{L}_V$ *is a* causal witness *if and only if there exists a Hermitian operator* $S_P \in A_I \otimes A_O \otimes B_I \otimes B_O$ *such that $S = L_V(S_P)$, $_{B_O}S_P \geqslant 0$, and $_{A_O}S_P \geqslant 0$.*

This restricted set of causal witnesses is also a closed convex cone, which we denote by $\mathcal{S}_V = \mathcal{S} \cap \mathcal{L}_V$.

One could define witnesses as belonging to $\mathcal{S}_V$ instead of $\mathcal{S}$, since both sets are as powerful in detecting causal nonseparability. However, some physically motivated witnesses, such as those presented in section 5.2 (for the tripartite case), do not belong to $\mathcal{S}_V$, which is why we use the more general definition that witnesses belong to $\mathcal{S}$.

3.2. Finding causal witnesses

The previous characterization of the convex cone of causal witnesses allows one to efficiently check the causal nonseparability of any process matrix W through algorithms for SDP [22]. They output a causal witness if W is causally nonseparable, and an explicit decomposition in terms of causally ordered process matrices otherwise.

The idea is simply to minimize $\text{tr}[SW]$ over the cone of causal witnesses[7] $\mathcal{S}_V$, and check whether we obtain a negative value or not. Note that in order to make $\text{tr}[S\,W]$ lower bounded (to avoid getting a value $-\infty$ for

---

[7] In principle minimizing over $\mathcal{S}$ instead of $\mathcal{S}_V$ would lead to the same value for $\text{tr}[SW]$, but this causes technical problems as explained in appendix E.

causally nonseparable process matrices) a normalization constraint on the witnesses has to be imposed. This normalization is arbitrary—any constraint that makes $\mathcal{S}_V$ compact suffices—and different normalization choices give rise to different interpretations for the value of $\mathrm{tr}[SW]$. We shall normalize the witnesses by imposing that $\mathrm{tr}\,[S\,\Omega] \leqslant 1$ for every (normalized) process matrix Ω, for $-\mathrm{tr}[SW]$ can then be interpreted as a measure of causal nonseparability, as we shall see later in this subsection. In order to be able to use it in the SDP problem we still need to write this normalization as a conic constraint. To do so, we extend the constraint $\mathrm{tr}\,[S\,\Omega] \leqslant 1$ to non-normalized process matrices by linearity:

$$\mathrm{tr}\,[S\,\Omega] \leqslant \mathrm{tr}[\Omega]/d_O, \tag{36}$$

which is equivalent to

$$\mathrm{tr}\left[\left(\mathbb{1}/d_O - S\right)\Omega\right] \geqslant 0 \tag{37}$$

for all $\Omega \in \mathcal{W}$. Recalling that S is assumed to be in $\mathcal{S}_V \subset \mathcal{L}_V$, this means that $\mathbb{1}/d_O - S \in \mathcal{W}_V^* := \mathcal{W}^* \cap \mathcal{L}_V$, where $\mathcal{W}^*$ is the dual cone of $\mathcal{W}$—that is, the cone of Hermitian operators that have non-negative trace with process matrices.

To test the causal nonseparability of a given process matrix W, we are thus led to define the following SDP problem:

$$\min \mathrm{tr}[SW] \ \mathrm{s.t.} \quad S \in \mathcal{S}_V, \quad \mathbb{1}/d_O - S \ \in \mathcal{W}_V^*, \tag{38}$$

which is written explicitly in terms of positive semidefinite constraints in appendix D.

If the solution of the SDP problem (38) leads to a negative expectation value of S, one can conclude that W is causally nonseparable, since SDP algorithms can be guaranteed[8] to find the optimal solution [22]. In such a case, the optimal solution S^* provides an explicit witness to verify the causal nonseparability of W. On the other hand, if $\mathrm{tr}[S^*W] = 0$, one concludes that W is causally separable, and an explicit decomposition of W into causally ordered processes is given by the SDP problem dual to (38) (this can be seen explicitly from the representation of the SDP problem (39) given in appendix D). As shown in appendix E, this dual is

$$\min \mathrm{tr}[\Omega]/d_O \ \mathrm{s.t.} \quad W + \Omega \in \mathcal{W}^{\mathrm{sep}}, \quad \Omega \in \mathcal{W}, \tag{39}$$

where $\mathcal{W}^{\mathrm{sep}}$ is the cone of non-normalized causally separable process matrices, as previously defined. Furthermore, the optimal value $\mathrm{tr}[\Omega^*]/d_O$ of problem (39) is related to the optimal value $\mathrm{tr}[S^*W]$ of problem (38) through

$$\mathrm{tr}\left[\Omega^*\right]/d_O = -\mathrm{tr}\left[S^*W\right]. \tag{40}$$

This gives an operational meaning to $-\mathrm{tr}[S^*W]$. As shown in appendix E, this quantity corresponds to the minimal $\lambda \geqslant 0$ such that

$$\frac{1}{1+\lambda}\left(W + \lambda\,\widetilde{\Omega}\right) \tag{41}$$

is causally separable, optimized over all valid, normalized processes $\widetilde{\Omega}$. In other words, it quantifies the resistance of W to the worst-case noise. This is an analogue of the measure of entanglement called generalized robustness, which quantifies the resistance of the entanglement of a quantum state to worst-case noise [23]. It turns out that for our case the interpretation of $-\mathrm{tr}[S^*W]$ as a *measure of causal nonseparability* is also tenable, as it respects some simple axioms that we propose in appendix F. For this reason, we define the *generalized robustness* of a process W as

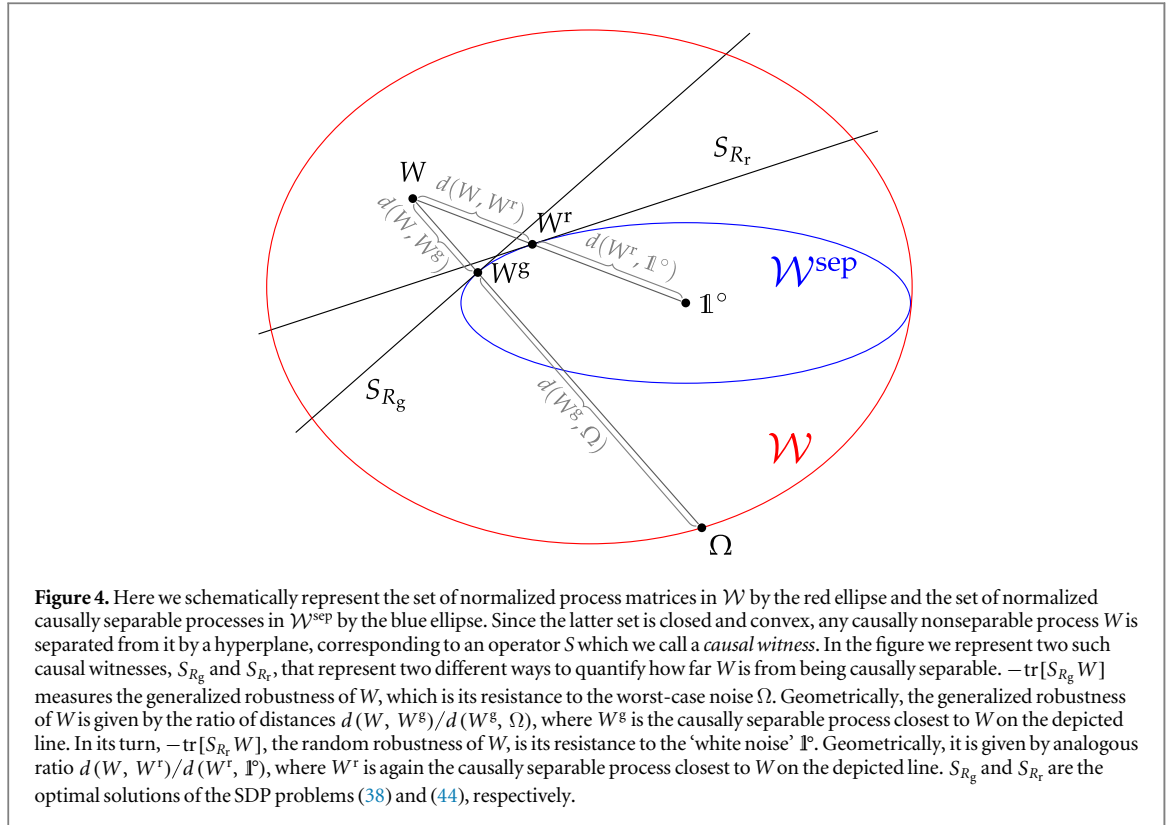$$R_{\mathrm{g}}(W) = -\mathrm{tr}\left[S^*W\right]. \tag{42}$$

Again in analogy with the case of entanglement measures, one can also define the *random robustness* [24] of W as is its resistance to 'white noise', which can be defined as the process that sends maximally mixed states to each laboratory, independently of the local operations:

$$\mathbb{1}^\circ := \frac{\mathbb{1}}{d_{A_I} d_{B_I}}. \tag{43}$$

The optimal witness with respect to random robustness can be found by solving an SDP problem analogous to (38):

$$\min \mathrm{tr}(SW) \ \mathrm{s.t.} \quad S \in \mathcal{S}_V, \quad \mathrm{tr}\left[S\mathbb{1}^\circ\right] \leqslant 1, \tag{44}$$

---

[8] When the assumptions of the duality theorem (8) are satisfied, which is the case for our SDP problems, as proven in appendix E.

Figure 4. Here we schematically represent the set of normalized process matrices in $\mathcal{W}$ by the red ellipse and the set of normalized causally separable processes in $\mathcal{W}^{\text{sep}}$ by the blue ellipse. Since the latter set is closed and convex, any causally nonseparable process W is separated from it by a hyperplane, corresponding to an operator S which we call a *causal witness*. In the figure we represent two such causal witnesses, S_{R_g} and S_{R_r}, that represent two different ways to quantify how far W is from being causally separable. $-\text{tr}[S_{R_g} W]$ measures the generalized robustness of W, which is its resistance to the worst-case noise Ω. Geometrically, the generalized robustness of W is given by the ratio of distances $d(W, W^g)/d(W^g, \Omega)$, where W^g is the causally separable process closest to W on the depicted line. In its turn, $-\text{tr}[S_{R_r} W]$, the random robustness of W, is its resistance to the 'white noise' $\mathbb{1}^\circ$. Geometrically, it is given by analogous ratio $d(W, W^r)/d(W^r, \mathbb{1}^\circ)$, where W^r is again the causally separable process closest to W on the depicted line. S_{R_g} and S_{R_r} are the optimal solutions of the SDP problems (38) and (44), respectively.

whose dual is

$$\min \lambda \quad \text{s.t.} \quad \lambda \geqslant 0, \quad W + \lambda \mathbb{1}^\circ \in \mathcal{W}^{\text{sep}}, \tag{45}$$

and random robustness itself is defined as

$$R_r(W) = -\text{tr}\Big[S^* W\Big], \tag{46}$$

where $\text{tr}(S^* W)$ is now the optimal value of the problem (44). This quantity can be used to compare witnesses in scenarios where white noise is an appropriate noise model. It cannot, however, be interpreted as a proper measure of causal nonseparability, as it does not respect all the axioms we propose in appendix F—more specifically, it is not monotonous under local operations.

A geometrical interpretation of the results of this section is shown in figure 4.

3.3. Implementing causal witnesses

Once a causal witness S has been obtained for a given causally nonseparable process matrix W, a natural question is how to 'measure' it, i.e., how to access the quantity $\text{tr}[S\,W]$—and, in particular, check its sign—experimentally.

To do so, note that as $S \in A_I \otimes A_O \otimes B_I \otimes B_O$ is a Hermitian operator, it can always be decomposed as a linear combination of the form[9]

$$S = \sum_{x,y,a,b} \gamma_{x,y,a,b}\, M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O}, \tag{47}$$

where $\gamma_{x,y,a,b}$ are real coefficients and $M_{a|x}^{A_I A_O}$ and $M_{b|y}^{B_I B_O}$ are positive semidefinite matrices that can be interpreted as the CJ representation of CP trace non-increasing maps (see section 2).

Expanding $\text{tr}[S\,W]$,

$$\text{tr}[S\,W] = \sum_{x,y,a,b} \gamma_{x,y,a,b}\, \text{tr}\Big[\Big(M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O}\Big) W\Big], \tag{48}$$

---

[9] In the decomposition (47), x, y, a and b should *a priori* simply be understood as labels for $M_{a|x}^{A_I A_O}$ and $M_{b|y}^{B_I B_O}$. We can however assume, without loss of generality, that $_{A_O}\big(\sum_a M_{a|x}^{A_I A_O}\big) \leqslant \mathbb{1}^{A_I A_O}/d_{A_O}$ and $_{B_O}\big(\sum_b M_{b|y}^{B_I B_O}\big) \leqslant \mathbb{1}^{B_I B_O}/d_{B_O}$ for all x, y (we can indeed always include scaling factors in the coefficients $\gamma_{x,y,a,b}$). Introducing, when required, some complementary positive semidefinite operators $M_{\varnothing|x}^{A_I A_O}$ and $M_{\varnothing|y}^{B_I B_O}$ (with null coefficients $\gamma_{x,y,a,b}$), so that now $_{A_O}\big(\sum_a M_{a|x}^{A_I A_O}\big) = \mathbb{1}^{A_I A_O}/d_{A_O}$ and $_{B_O}\big(\sum_b M_{b|y}^{B_I B_O}\big) = \mathbb{1}^{B_I B_O}/d_{B_O}$, the sets $\{M_{a|x}^{A_I A_O}\}_a$ and $\{M_{b|y}^{A_I A_O}\}_b$ can then be interpreted as the CJ representation of instruments, for which x, y are inputs and a, b are outputs.

where according to the generalized Born rule (3), the terms $\mathrm{tr}\left[\left(M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O}\right) W\right]$ represent the probabilities $P\left(M_{a|x}^{A_I A_O}, M_{b|y}^{B_I B_O}\right)$ that the maps $M_{a|x}^{A_I A_O}$ and $M_{b|y}^{B_I B_O}$ are realized. We assume that these CP maps can be implemented even if the causal order of the parties is not well-defined. The quantity $\mathrm{tr}[SW]$ can thus in principle be implemented experimentally by estimating the probabilities $P\left(M_{a|x}^{A_I A_O}, M_{b|y}^{B_I B_O}\right)$ and combining them as in equation (48).

The decomposition (47) is not unique. Furthermore, as noted before we can add to any witness S a term $S^\perp$ such that $L_V(S^\perp) = 0$ without changing its validity or its trace with any valid process. Hence, it actually suffices to find a decomposition for $S + S^\perp$ for some arbitrary $S^\perp$, implement the corresponding maps, and combine their statistics as above.

3.4. Example

Let us now illustrate the above considerations on an explicit example. Reference [13] introduced the following process matrix, for a case where all incoming and outgoing systems of A and B are two-dimensional (qubit) systems (i.e., $d_{A_I} = d_{A_O} = d_{B_I} = d_{B_O} = 2$):

$$W_{\mathrm{OCB}} = \frac{1}{4}\left[\mathbb{1} + \frac{\mathbb{1}^{A_I}Z^{A_O}Z^{B_I}\mathbb{1}^{B_O} + Z^{A_I}\mathbb{1}^{A_O}X^{B_I}Z^{B_O}}{\sqrt{2}}\right], \tag{49}$$

where Z and X are the Pauli matrices, and tensor products are implicit. One can easily check that $W_{\mathrm{OCB}} \geqslant 0$, that $\mathrm{tr}[W_{\mathrm{OCB}}] = 4 = d_O$, and that W_{OCB} satisfies equations (9)–(11), which ensures that it is indeed a valid process matrix. It was shown that W_{OCB} allows for a violation of a causal inequality (see section 6), which implies that it is causally nonseparable.

The concept of causal witnesses introduced here allows us to prove the causal nonseparability of W_{OCB} more directly. Solving the SDP problem (44) with YALMIP [25] and the solver MOSEK [26], we obtained, up to numerical precision, the optimal witness with respect to random robustness

$$S_{\mathrm{OCB}} = \frac{1}{4}\left[\mathbb{1} - \left(\mathbb{1}^{A_I}Z^{A_O}Z^{B_I}\mathbb{1}^{B_O} + Z^{A_I}\mathbb{1}^{A_O}X^{B_I}Z^{B_O}\right)\right]. \tag{50}$$

Applying it to W_{OCB}, we find that $-\mathrm{tr}[S_{\mathrm{OCB}}\, W_{\mathrm{OCB}}] = R_{\mathrm{r}}(W_{\mathrm{OCB}}) = \sqrt{2} - 1 > 0$ (where $R_{\mathrm{r}}(W_{\mathrm{OCB}})$ is the random robustness as defined in equation (46)). This proves that W_{OCB} is causally nonseparable.

This also implies that the process matrices of the form

$$W_{\mathrm{OCB}}(\lambda) = \frac{1}{1+\lambda}\left(W_{\mathrm{OCB}} + \lambda\, \mathbb{1}^\circ\right), \tag{51}$$

are causally nonseparable for

$$\lambda < R_{\mathrm{r}}\left(W_{\mathrm{OCB}}\right) = \sqrt{2} - 1 \tag{52}$$

(and their causal nonseparability is then witnessed by S_{OCB}). For $\lambda \geqslant \sqrt{2} - 1$, $W_{\mathrm{OCB}}(\lambda)$ is causally separable; the solution of the SDP problem (45) provides an explicit decomposition for $W_{\mathrm{OCB}}(R_{\mathrm{r}}(W_{\mathrm{OCB}}))$ (as can be seen when writing (45) in a form similar to equation (D.3)), from which we can derive an explicit decomposition for all $W_{\mathrm{OCB}}(\lambda)$ for $\lambda \geqslant \sqrt{2} - 1$, as

$$W_{\mathrm{OCB}}(\lambda) = \frac{1}{2}W_{\mathrm{OCB}}^{A \prec B}(\lambda) + \frac{1}{2}W_{\mathrm{OCB}}^{B \prec A}(\lambda), \tag{53}$$

where

$$W_{\mathrm{OCB}}^{A \prec B}(\lambda) := \frac{1}{4}\left[\mathbb{1} + \frac{\sqrt{2}}{1+\lambda}\,\mathbb{1}^{A_I}Z^{A_O}Z^{B_I}\mathbb{1}^{B_O}\right], \tag{54}$$

$$W_{\mathrm{OCB}}^{B \prec A}(\lambda) := \frac{1}{4}\left[\mathbb{1} + \frac{\sqrt{2}}{1+\lambda}\,Z^{A_I}\mathbb{1}^{A_O}X^{B_I}Z^{B_O}\right] \tag{55}$$

are causally ordered process matrices. (Note that for $\lambda < \sqrt{2} - 1$, $W_{\mathrm{OCB}}^{A \prec B}(\lambda)$ and $W_{\mathrm{OCB}}^{B \prec A}(\lambda)$ as defined above would not be positive semidefinite, which explains why equation (53) then fails to provide a valid causally separable decomposition of $W_{\mathrm{OCB}}(\lambda)$.)

To measure the witness S_{OCB} and obtain the quantity $\mathrm{tr}[S_{\mathrm{OCB}} \cdot W]$ experimentally, one can for instance decompose it in the following way: define, for x, y, y', a, $b = 0$, 1, the CJ matrices

$$M_{a|x}^{A_I A_O} := \left(\frac{\mathbb{1} + (-1)^a Z}{2}\right)^{A_I} \otimes \left(\frac{\mathbb{1} + (-1)^x Z}{2}\right)^{A_O}, \tag{56}$$

$$M_{b|y,y'=0}^{B_I B_O} := \left(\frac{\mathbb{1} + (-1)^b X}{2}\right)^{B_I} \otimes \left(\frac{\mathbb{1} + (-1)^{y+b} Z}{2}\right)^{B_O}, \tag{57}$$

$$M_{b|y,y'=1}^{B_I B_O} := \left(\frac{\mathbb{1} + (-1)^b Z}{2} \right)^{B_I} \otimes \frac{\mathbb{1}^{B_O}}{2}, \tag{58}$$

which represent measure-and-prepare maps (see appendix A.1). One can then check that

$$S_{\mathrm{OCB}} = 3 \cdot \mathbb{1}^\circ - 4\, G_{\mathrm{OCB}} \tag{59}$$

with

$$G_{\mathrm{OCB}} = \frac{1}{8} \sum_{x,y,a,b} \left[\delta_{a,y}\, M_{a|x}^{A_I A_O} \otimes M_{b|y,y'=0}^{B_I B_O} + \delta_{b,x}\, M_{a|x}^{A_I A_O} \otimes M_{b|y,y'=1}^{B_I B_O} \right], \tag{60}$$

where $\delta_{j,k}$ is the Kronecker delta. Thus, one can compute $\mathrm{tr}[S_{\mathrm{OCB}} \cdot W]$ by performing the maps above on W and combining the probabilities $P\left(M_{a|x}^{A_I A_O}, M_{b|y,y'}^{B_I B_O}\right) = \mathrm{tr}[M_{a|x}^{A_I A_O} \otimes M_{b|y,y'}^{B_I B_O} \cdot W]$ as follows:

$$\begin{aligned}
\mathrm{tr}\left[S_{\mathrm{OCB}} \cdot W \right] &= 3 - 4\, \mathrm{tr}\left[G_{\mathrm{OCB}} \cdot W \right] \\
&= 3 - 4 \cdot \frac{1}{8} \sum_{x,y,a,b} \left[\delta_{a,y}\, P\left(M_{a|x}^{A_I A_O}, M_{b|y,y'=0}^{B_I B_O} \right) + \delta_{b,x}\, P\left(M_{a|x}^{A_I A_O}, M_{b|y,y'=1}^{B_I B_O} \right) \right].
\end{aligned} \tag{61}$$

As one may recognize, the choice of CP maps in (56)–(58) is the same[10] as that considered in [13], so that the experimental procedure proposed here to measure the witness S_{OCB} would be the same as that suggested in [13] to violate a causal inequality. The labels x, y, y', a, b can be considered as inputs and outputs for the above maps (which indeed satisfy $_{A_O}\left(\sum_a M_{a|x}^{A_I A_O}\right) = \mathbb{1}^{A_I A_O}/d_{A_O}$ and $_{B_O}\left(\sum_b M_{b|y,y'}^{B_I B_O}\right) = \mathbb{1}^{B_I B_O}/d_{B_O}$ for all x, y, y'). As it turns out, in the causal inequality of [13] the probabilities $P(a, b|x, y, y') = P\left(M_{a|x}^{A_I A_O}, M_{b|y,y'}^{B_I B_O}\right)$ are actually combined in precisely the same way as above—namely, $\mathrm{tr}[G_{\mathrm{OCB}} \cdot W]$ above can be identified with the probability p_{succ} of winning the corresponding 'causal game'

$$p_{\mathrm{succ}} = \frac{1}{2}[P(a = y|y' = 0) + P(b = x|y' = 1)], \tag{62}$$

when the inputs x, y, $y' = 0,\, 1$ are given with equal probabilities.

Remarkably, in this particular case the bounds of the causal witness S_{OCB} and of the causal inequality (62) coincide, i.e., $\mathrm{tr}[S_{\mathrm{OCB}} \cdot W] \geqslant 0$ if and only if $p_{\mathrm{succ}} = \mathrm{tr}[G_{\mathrm{OCB}} \cdot W] \leqslant 3/4$, where $3/4$ is the upper bound on p_{succ} for any causal correlation (as defined in section 6 below). Furthermore, the noise threshold below which the noisy process matrix $W_{\mathrm{OCB}}(\lambda)$ (53) can violate the causal inequality is the same as the threshold $R_{\mathrm{r}}(W_{\mathrm{OCB}})$ below which $W_{\mathrm{OCB}}(\lambda)$ is causally nonseparable, as already noted in [28]. This is however not a general property of causal witnesses and causal inequalities: similarly to the case of entanglement vs. quantum nonlocality and of entanglement witnesses vs. Bell inequalities [27], there exist causally nonseparable process matrices that cannot yield any violation of any causal inequality—while there always exists a causal witness that detects their causal nonseparability. We will come back to this issue in section 6 below, with an explicit example in the tripartite case.

4. Quantum control of causal order

4.1. The quantum switch

It has recently been suggested that quantum computation can be extended beyond the framework of quantum circuits, which enforces a fixed order between the execution of quantum gates. The main idea is that the order in which gates are performed can be coherently controlled by a quantum system. The new resource that allows for such a control is the *quantum switch*, first proposed in [6]. It works as follows: consider a two-qubit system, composed of a control and of a target qubit. Two parties A and B act on the target qubit with the unitaries U_A, U_B respectively. If the control qubit is prepared in the state $|0\rangle$, U_A is applied to the target before U_B, while if the control is in state $|1\rangle$ the two unitaries are applied in the reversed order. The global unitary, acting on both the target and control qubits, is thus

$$V\left(U_A, U_B \right) = |0\rangle\langle 0| \otimes U_B U_A + |1\rangle\langle 1| \otimes U_A U_B, \tag{63}$$

where the first factor in each tensor product acts on the control system and the second factor acts on the target. For an initial state $\frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes |\psi\rangle$ of the control-target system, one gets, after applying V, the state

---

[10] Note that compared to [13], we exchanged in the present paper the notations x, y and a, b for inputs and outputs, so as to use here the same notations as most of the recent works on quantum and nonlocal correlations [27]. Furthermore, in [13] the state sent out by B when $y' = 1$ was arbitrary, while here we fixed it to be $\mathbb{1}^{B_O}/2$.

$\frac{1}{\sqrt{2}}(|0\rangle \otimes U_B U_A |\psi\rangle + |1\rangle \otimes U_A U_B |\psi\rangle)$, which can be interpreted as having applied the two unitaries on the target in a 'superposition of orders'[11].

Note that if the control system is discarded, one is left with the mixed state

$$\frac{1}{2}\left(U_B U_A |\psi\rangle \langle\psi| U_A^\dagger U_B^\dagger + U_A U_B |\psi\rangle \langle\psi| U_B^\dagger U_A^\dagger\right). \tag{64}$$

This can be produced by randomly exchanging the order in which U_A and U_B are applied and thus can be seen as an equal mixture of causally ordered processes. To make the situation more interesting, we shall be led to introduce a third party, C, who can perform measurements on the control qubit (and possibly also on the target qubit) in order to define a causally nonseparable process (using the definition (27)) using quantum control of causal order.

4.2. Process matrix representation of the quantum switch

For our purposes, we can formally represent the quantum switch (with fixed input state) as a tripartite process matrix: the two parties A and B perform an arbitrary CP map each on the target qubit, while C performs an arbitrary two-qubit POVM measurement on the resulting control-target state (with no outgoing system). The dimensions of input and output systems of the local laboratories are therefore

$$d_{A_I} = d_{A_O} = d_{B_I} = d_{B_O} = 2, \quad d_{C_I} = 4, \quad d_{C_O} = 1. \tag{65}$$

For clarity, we shall divide C's input space as $C_I = C_I^c \otimes C_I^t$, where C_I^c and C_I^t refer to the control and target qubits, respectively (with therefore $d_{C_I^c} = d_{C_I^t} = 2$).

In order to describe the process matrix of the quantum switch, we are first going to make use of the 'pure' version of the formalism, described in appendix A.2. An identity channel from a party's output space A_O to another party's input space B_I is described, as a process matrix, by the projector onto the 'process vector' $|\mathbb{1}\rangle\rangle^{A_O B_I} = \sum_{j=0,1} |j\rangle^{A_O} |j\rangle^{B_I}$. The situation where A receives a state $|\psi\rangle$, performs an arbitrary operation on it, and sends the output directly to B through an identity channel, who in turn sends the output of his operation to C_I^t, is represented by the process vector $|\psi\rangle^{A_I} |\mathbb{1}\rangle\rangle^{A_O B_I} |\mathbb{1}\rangle\rangle^{B_O C_I^t}$, see appendix A.2. Then the quantum switch, with the control qubit initially in the state $\frac{|0\rangle + |1\rangle}{\sqrt{2}}$ and the target qubit in the state $|\psi\rangle$, is represented by the process matrix $|w\rangle \langle w|$, where

$$|w\rangle = \frac{1}{\sqrt{2}}\left(|\psi\rangle^{A_I} |\mathbb{1}\rangle\rangle^{A_O B_I} |\mathbb{1}\rangle\rangle^{B_O C_I^t} |0\rangle^{C_I^c} + |\psi\rangle^{B_I} |\mathbb{1}\rangle\rangle^{B_O A_I} |\mathbb{1}\rangle\rangle^{A_O C_I^t} |1\rangle^{C_I^c}\right). \tag{66}$$

This can be checked by noting that

$$\langle\langle U_A^* |^{A_I A_O} \langle\langle U_B^* |^{B_I B_O} \cdot |w\rangle$$
$$= \frac{1}{\sqrt{2}}\left(|0\rangle^{C_I^c} \otimes \left(U_B U_A |\psi\rangle\right)^{C_I^t} + |1\rangle^{C_I^c} \otimes \left(U_A U_B |\psi\rangle\right)^{C_I^t}\right), \tag{67}$$

where $|U_A^*\rangle\rangle^{A_I A_O} := \mathbb{1} \otimes U_A^* |\mathbb{1}\rangle\rangle$ is the 'pure' CJ representation of U_A, and similarly for $|U_B^*\rangle\rangle^{B_I B_O}$ (and with $\langle\langle U^* | = |U^*\rangle\rangle^\dagger$); see appendix A.1. Note that the resulting state is the same as that obtained by applying (63) to an initial state $\frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes |\psi\rangle$.

Note that the process (66) itself is clearly causally nonseparable[12], since (i) it is a superposition of a pure process only compatible with the order $A \prec B \prec C$ and a pure process only compatible with the order $B \prec A \prec C$ and (ii) it is a projector onto a pure vector, thus it cannot be written as a nontrivial mixture of causally ordered processes.

From equation (66), one finds (using the facts that $\mathrm{tr}_{C_I^t}(|\mathbb{1}\rangle\rangle\langle\langle\mathbb{1}|^{B_O C_I^t}) = \mathbb{1}^{B_O}$ and $\mathrm{tr}_{C_I^t}(|\mathbb{1}\rangle\rangle\langle\langle\mathbb{1}|^{A_O C_I^t}) = \mathbb{1}^{A_O}$) that by tracing out C, one gets

$$\mathrm{tr}_{C_I} |w\rangle \langle w| = \mathrm{tr}_{C_I^c C_I^t} |w\rangle \langle w| = \frac{1}{2} W^{A \prec B} + \frac{1}{2} W^{B \prec A}, \tag{68}$$

where

$$W^{A \prec B} = |\psi\rangle \langle\psi|^{A_I} \otimes |\mathbb{1}\rangle\rangle\langle\langle\mathbb{1}|^{A_O B_I} \otimes \mathbb{1}^{B_O}, \tag{69}$$

---

[11] Since any CP map can be purified to a unitary evolution by introducing an ancillary system and a projective measurement on some subsystem of the original system and ancilla, the notion of superposition of orders can be easily extended from unitary operations to arbitrary CP maps by introducing an ancillary register for each party.

[12] Note that the results of [6] only show that the quantum switch cannot be realized by a circuit with a fixed order of gates, but the more general notion of causal (non)separability was not considered.

$$W^{B \prec A} = |\psi\rangle \langle\psi|^{B_I} \otimes |\mathbb{1}\rangle\rangle\langle\langle\mathbb{1}|^{B_O A_I} \otimes \mathbb{1}^{A_O}, \qquad (70)$$

are (bipartite) causally ordered process matrices; $\mathrm{tr}_{C_I} |w\rangle \langle w|$ indeed describes the situation of equation (64).

For some information-processing tasks, the quantum switch is known to provide an advantage over causally ordered processes [7, 8], even when C ignores the target system and only measures the control system. We will thus restrict our attention to witnesses of the form $S^{C_I^c} \otimes \mathbb{1}^{C_I^t}$, which can simplify the analysis and the experimental implementation. The reduced process we will be dealing with is the partial trace of the quantum switch (66) over the target system:

$$W_{\text{switch}} = \mathrm{tr}_{C_I^t} |w\rangle \langle w|. \qquad (71)$$

Note that the proof of causal nonseparability based on the purity of the switch does not extend to the reduced switch (71), since it is not an extremal process. We will therefore use the framework of causal witnesses to show that the reduced switch is also causally nonseparable.

5. Witnesses for the quantum switch

Since the quantum switch is a tripartite process where $d_{C_O} = 1$, we can use definition (27) to study its causal (non)separability. In this tripartite situation, we will define causal witnesses to be the Hermitian operators S such that

$$\mathrm{tr}\left[S\, W^{\text{sep}} \right] \geqslant 0 \qquad (72)$$

for every causally separable processes W^{sep} in the cone $\mathcal{W}_{3C}^{\text{sep}}$. The set of causal witnesses is thus the cone dual to $\mathcal{W}_{3C}^{\text{sep}}$, which we denote by $\mathcal{S}_{3C}$, or $\mathcal{S}_{3C,V}$ when restricted to $\mathcal{L}_V$. The characterization of $\mathcal{S}_{3C}$ is given by the following theorem:

Theorem 3. *A Hermitian operator* $S \in A_I \otimes A_O \otimes B_I \otimes B_O \otimes C_I \otimes C_O$ *with* $d_{C_O} = 1$ *is a causal witness if and only if S can be written as*

$$S = S_{ABC}^P + S_{ABC}^\perp = S_{BAC}^P + S_{BAC}^\perp, \qquad (73)$$

where

$$S_{ABC}^P \geqslant 0, \qquad L_{A \prec B \prec C}\left(S_{ABC}^\perp \right) = 0, \qquad (74)$$

$$S_{BAC}^P \geqslant 0, \qquad L_{B \prec A \prec C}\left(S_{BAC}^\perp \right) = 0, \qquad (75)$$

with $L_{A \prec B \prec C}$ *and* $L_{B \prec A \prec C}$ *as defined in section* 2.2.3.

The proof is given in appendix G. This characterization allows us to cast the problem of finding a witness for the quantum switch (or in fact for any process W with $d_{C_O} = 1$) as an SDP problem analogous to (38):

$$\min \mathrm{tr}(SW) \quad \text{s.t.} \quad S \in \mathcal{S}_{3C,V}, \quad \mathbb{1}/d_O - S \in \mathcal{W}_{3C,V}^*, \qquad (76)$$

where $\mathcal{W}_{3C,V}^* := \mathcal{W}_{3C}^* \cap \mathcal{L}_V$, with $\mathcal{W}_{3C}^*$ the dual of the cone $\mathcal{W}_{3C}$ of (non-normalized) tripartite process matrices with $d_{C_O} = 1$.

Analogously to problems (38), (39), the dual of (76) writes

$$\min \mathrm{tr}[\Omega]/d_O \quad \text{s.t.} \quad W + \Omega \in \mathcal{W}_{3C}^{\text{sep}}, \quad \Omega \in \mathcal{W}_{3C}, \qquad (77)$$

and the optimal values of (76) and (77) respect the duality relation (40), which allows us to interpret $-\mathrm{tr}(S^*W)$ as generalized robustness also in this case. Furthermore, (76) and (77) respect the assumptions of the duality theorem, and therefore SDP algorithms can find their optimal solutions efficiently. We shall, however, omit the proofs, as they are simply a slight modification of the ones already presented in appendix E.

5.1. Optimal witness

To find the optimal generalized robustness witness for the quantum switch we need to solve SDP problem (76) providing W_{switch} from equation (71) as an argument. Solving it using YALMIP and the solver MOSEK we obtain a witness S_{optimal} numerically; the generalized robustness of the quantum switch is found to be

$$R_{\text{g}}\left(W_{\text{switch}} \right) = -\mathrm{tr}\left[S_{\text{optimal}} W_{\text{switch}} \right] \approx 0.5454. \qquad (78)$$

Later in this section we will compare this number to that obtained from non-optimal witnesses. For this purpose, we shall use the amount of worst-case noise tolerated by a witness, i.e., the amount of worst-case noise that can be added to the quantum switch before the witness can no longer detect its causal nonseparability. It should be clear that, when the said witness is optimal, this number reduces to the generalized robustness of the quantum switch.

5.2. Chiribella's witness

In [7] Chiribella proposed an information-processing task for which the quantum switch had an advantage over causally ordered processes. We want to understand what this advantage means, and how it relates to causal nonseparability. For that we shall present a slightly modified version of his task and show how it can be understood as a causal witness.

Our version of the task is as follows: Alice (party A) receives a qubit in her lab, applies a unitary U_A to it, and sends it away. Bob (party B) receives a qubit in his lab, applies a unitary U_B to it, and sends it away. We assume that in each run of the experiment, U_A and U_B either commute or anticommute. Charlie (party C) receives a qubit in his lab, and makes a measurement on it to decide whether U_A and U_B commute or anticommute.

To construct a causal witness in relation to this task, we start with the CJ representation of the actions of the parties: Alice applying a unitary U_A, Bob applying a unitary U_B, and Charlie obtaining the result $\pm$ when measuring in the $|\pm\rangle = \dfrac{|0\rangle \pm |1\rangle}{\sqrt{2}}$ basis. Using the CJ representations $|U_A^*\rangle\rangle$ and $|U_B^*\rangle\rangle$ of U_A and U_B (see appendix A.1), the corresponding operator is

$$G_\pm^{U_A,U_B} = \left|U_A^*\right\rangle\!\!\left\rangle\!\left\langle\!\left\langle U_A^*\right| \otimes \left|U_B^*\right\rangle\!\!\right\rangle\!\left\langle\!\left\langle U_B^*\right| \otimes |\pm\rangle\langle\pm|. \tag{79}$$

The witness corresponding to the task is obtained by averaging over the cases where Charlie obtains $+$ when Alice and Bob apply commuting unitaries, and the cases where Charlie obtains $-$ when Alice and Bob apply anticommuting unitaries:

$$G_{\text{Chiribella}} = \frac{1}{2}\int d\mu_{[\,,\,]}\, G_+^{U_A,U_B} + \frac{1}{2}\int d\mu_{\{\,,\,\}}\, G_-^{U_A,U_B}, \tag{80}$$

where $d\mu_{[\,,\,]}$ is a measure over commuting unitaries, and $d\mu_{\{\,,\,\}}$ is a measure over anticommuting unitaries (we assume here that the cases where U_A and U_B commute and anticommute each appear with probability $\frac{1}{2}$). The probability of success in this task when the parties are using a strategy described by a process matrix W is then

$$p_{\text{succ}} = \text{tr}\big[G_{\text{Chiribella}}\, W\big]. \tag{81}$$

It is easy to check that for any choice of measures $d\mu_{[\,,\,]}$, $d\mu_{\{\,,\,\}}$ the probability of success is 1 when $W = W_{\text{switch}}$. The maximal probability of success for causally separable processes, however, depends crucially on the measures $d\mu_{[\,,\,]}$ and $d\mu_{\{\,,\,\}}$. If we were to choose, for example, measures that only produce pairs of Pauli matrices, then there is a causally separable circuit[13] that can decide the commutativity or anticommutativity with probability 1.

To avoid this problem we will first choose measures that can produce *any* pair of commuting or anticommuting unitaries (modulo global phases). Specifically, we choose the commuting measure $d\mu_{[\,,\,]}$ to pick up commuting unitaries of the form

$$U_A = U\begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta_1} \end{pmatrix}U^\dagger \quad \text{and} \quad U_B = U\begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta_2} \end{pmatrix}U^\dagger, \tag{82}$$

where U is uniformly distributed according to the Haar measure, and θ_i are uniformly distributed in the interval $[0, 2\pi]$. For the anticommuting measure $d\mu_{\{\,,\,\}}$, we will use $U_A = VXV^\dagger$ and $U_B = VZV^\dagger$, where V is also a Haar-random unitary (and X and Z are the Pauli matrices)[14].

With these measures $G_{\text{Chiribella}}$ turns out to be a valid causal witness, as the maximal probability of success for causally separable processes $p_{\text{succ}}^{\text{sep}}$ is bounded below one. To calculate it we need to solve the following SDP problem:

$$\max \text{tr}\big[G_{\text{Chiribella}}\, W\big] \text{ s.t. } \quad \text{tr}\, W = d_O, \quad W \in \mathcal{W}_{3C}^{\text{sep}}. \tag{83}$$

Solving it with YALMIP and MOSEK, we obtain

$$p_{\text{succ}}^{\text{sep}} \approx 0.9288. \tag{84}$$

The amount of worst-case noise that $G_{\text{Chiribella}}$ can tolerate is 0.0766, which is much worse than the 0.5454 tolerated by S_{optimal}.

An issue with $G_{\text{Chiribella}}$ is that it would take an infinite number of measurements to estimate each term of the sum in (80). Furthermore $d\mu_{[\,,\,]}$ and $d\mu_{\{\,,\,\}}$ were chosen arbitrarily, while it would be preferable to have a justification for the choice of a particular measure. Both problems are solved by restricting the unitaries U_A and U_B to come from a finite set. In this way we only need perform a finite number of measurements to estimate the

---

[13] One such circuit, described in [7], involves applying the Pauli unitaries to one half of a maximally entangled state and doing a measurement in the Bell basis.

[14] It turns out that with this choice of measures the witness $G_{\text{Chiribella}}$ is the same as we would obtain by translating the task from [7] directly into the language of causal witnesses; the only difference, then, is that in [7] the witness was decomposed in terms of measurements and repreparations, whereas we decomposed it using unitaries only.

witness, and it is possible to optimize the measures over commuting and anticommuting unitaries through SDP problems.

The best witness we found is obtained by choosing the following ten unitaries:

$$\mathcal{G} = \left\{ \mathbb{1}, X, Y, Z, \frac{X+Y}{\sqrt{2}}, \frac{X-Y}{\sqrt{2}}, \frac{X+Z}{\sqrt{2}}, \frac{X-Z}{\sqrt{2}}, \frac{Y+Z}{\sqrt{2}}, \frac{Y-Z}{\sqrt{2}} \right\} \tag{85}$$

(Y being the third Pauli matrix), and defining the witness to be

$$G_{\text{finite}} = \sum_{i,j=1}^{10} q_{ij}^{[\,,\,]} G_+^{U_i, U_j} + q_{ij}^{\{\,,\,\}} G_-^{U_i, U_j}, \tag{86}$$

where $U_k \in \mathcal{G}$, and $q_{ij}^{[\,,\,]}$, $q_{ij}^{\{\,,\,\}}$ are the input probability distributions over commuting and anticommuting unitaries, normalized such that $\sum_{i,j}(q_{ij}^{[\,,\,]} + q_{ij}^{\{\,,\,\}}) = 1$.

To obtain the weights $q_{ij}^{[\,,\,]}$, $q_{ij}^{\{\,,\,\}}$ and $p_{\text{succ}}^{\text{sep}}$ we solved an SDP problem presented in appendix H. We obtained

$$p_{\text{succ}}^{\text{sep}} \approx 0.8690, \tag{87}$$

and tolerance to worst-case noise 0.1507, which is higher than $G_{\text{Chiribella}}$'s 0.0766, but still lower than S_{optimal}'s 0.5454.

We want to emphasize that the witnesses obtained in this subsection are equivalent to the ones defined through (72) in the beginning of the present section—the only difference being the arbitary choice of the causal bound being $\geqslant 0$ vs $\leqslant p_{\text{succ}}^{\text{sep}}$. More precisely, let G be a witness such that

$$\text{tr}\left[G \, W^{\text{sep}} \right] \leqslant p_{\text{succ}}^{\text{sep}} \tag{88}$$

for every (normalized) causally separable W^{sep} and

$$T_0 \leqslant \text{tr}(G \, W) \leqslant T_1 \tag{89}$$

for every (normalized) process matrix W. Then

$$S = \frac{1}{p_{\text{succ}}^{\text{sep}} - T_0} \left(p_{\text{succ}}^{\text{sep}} \frac{\mathbb{1}}{d_O} - G \right) \tag{90}$$

is a valid generalized robustness witness. Furthermore, if S is the optimal witness for some process matrix W that saturates the upper bound $\text{tr}(G \, W) = T_1$, it follows that

$$R_g(W) = -\text{tr}[SW] = \frac{T_1 - p_{\text{succ}}^{\text{sep}}}{p_{\text{succ}}^{\text{sep}} - T_0}. \tag{91}$$

When G is either G_{finite} or $G_{\text{Chiribella}}$, we have that $T_0 = 0$ and $T_1 = 1$. And even though they are *not* optimal witnesses for W_{switch}, the relationship between $p_{\text{succ}}^{\text{sep}}$ and resistance to worst-case noise is valid for them, i.e., for both G_{finite} and $G_{\text{Chiribella}}$ the resistance to worst-case noise is equal to $1/p_{\text{succ}}^{\text{sep}} - 1$, as given by (91).

6. Causal inequalities

The notion of causal separability considered above relies on the quantum description of the local laboratories. One may ask what are the constraints imposed by a definite causal structure regardless of the specific description, or even the physics governing the devices performing the local operations. To study such restrictions, we will make use of so-called *causal inequalities* [13], which bound the possible correlations that can be established between events following a definite causal order. The violation of a causal inequality gives a stronger, device-independent signature of lack of causal order than the measurement of a witness. It is natural to ask whether it is possible to use the quantum switch to violate a causal inequality; we show below that this is not the case.

6.1. Device-independent causal relations

We still consider a multipartite scenario in which a set of N parties $\{A^i\}_{i=1}^N$ are located in different, separated laboratories. Each party can perform operations and obtain measurement outcomes. Contrary to the previous case however, we do not consider here any particular physical description of what happens in each lab; the 'settings' for the operations in the different laboratories and the measurement outcomes are labelled by some classical variables x_i and a_i (with $1 \leqslant i \leqslant N$), respectively; for simplicity we assume that the x_i's and a_i's take a finite number of values. Defining the vector of settings $\vec{x} = (x_1, \dots x_N)$ and the vector of outcomes $\vec{a} = (a_1, \dots, a_N)$, the device-independent description of the correlations established in such an experiment is encoded in the conditional probability $P(\vec{a} \mid \vec{x})$.

Causal inequalities [13] are constraints on $P(\vec{a}|\vec{x})$ derived from the assumption that there exists an underlying causal structure defining the order between parties. To be more precise, let us represent the causal order in which the parties act by a permutation σ, defined such that party i acts before party j if and only if $\sigma(i) < \sigma(j)$. This leads to a total ordering of the parties, namely $A^{\sigma(1)} \prec A^{\sigma(2)} \prec \ldots \prec A^{\sigma(N)}$. We then say that a probability distribution $P(\vec{a}|\vec{x})$ is *compatible with the causal order* σ if no party signals to those before her[15], namely if for every i the marginal distribution

$$P\left(a_{\sigma(1)}, \ldots, a_{\sigma(i)} \middle| \vec{x}\right) := \sum_{\substack{a_{\sigma(j)} \\ j > i}} P\left(\vec{a} \middle| \vec{x}\right) \tag{92}$$

does not depend on the inputs $x_{\sigma(j)}$ with $j > i$; i.e.

$$\begin{aligned}
&P\left(a_{\sigma(1)}, \ldots, a_{\sigma(i)} \middle| x_{\sigma(1)}, \ldots, x_{\sigma(i)}, x_{\sigma(i+1)}, \ldots, x_{\sigma(N)}\right) \\
&= P\left(a_{\sigma(1)}, \ldots, a_{\sigma(i)} \middle| x_{\sigma(1)}, \ldots, x_{\sigma(i)}, x'_{\sigma(i+1)}, \ldots, x'_{\sigma(N)}\right) \\
&\forall\, x_{\sigma(j)}, x'_{\sigma(j)}.
\end{aligned} \tag{93}$$

A probability distribution that is compatible with at least one causal order σ is said to be *causally ordered*.

More generally, we allow the parties to share randomness to agree on a specific order of sending signals between them before the inputs of the game are given to them. This allows for convex combinations of causally ordered probability distributions:

$$P\left(\vec{a} \middle| \vec{x}\right) = \sum_{\sigma} q_\sigma\, P_\sigma\left(\vec{a} \middle| \vec{x}\right), \quad q_\sigma \geqslant 0, \quad \sum_{\sigma} q_\sigma = 1, \tag{94}$$

where each P_σ is compatible with a fixed order σ. These are still not the most general correlations compatible with the assumption of a definite causal structure, as one party could control the causal order of a set of parties in its future [19, 29, 30]. Correlations compatible with this most general scenario of definite causal order are called simply *causal*. In the bipartite case, the set of causal correlations forms a convex polytope, delimited by a finite number of facets that define causal inequalities [31]. The explicit definition of causal correlations in the general N-partite case is, however, rather cumbersome, and for the purposes of this article it will be enough to consider probability distributions of the form (94), which is a sufficient (although not necessary) condition for causal separability.

As causally separable processes can only generate causal correlations, the violation of a causal inequality can also be used to detect the causal nonseparability of a process. While causal witnesses are *device-dependent* and can only detect causal nonseparability if each party trusts her operation's implementation, causal inequalities are completely *device-independent*: even if each party distrusts her laboratory, they can still detect causal nonseparability from the statistics of their experimental outcomes, if those violate a causal inequality. While for every causally nonseparable process there is causal witness that will detect its nonseparability, there are causally nonseparable processes cannot be used to violate any causal inequalities: in the next subsection we will prove that the quantum switch provides such an example. There is an analogy here with entanglement witnesses, which allow for a device-dependent way of detecting entanglement, and Bell inequalities, which provide a device-independent entanglement certification—'nonlocality' [27]. The important difference is that states violating Bell inequalities are physically implementable, while no example of a physically implementable process violating causal inequalities is known.

6.2. Quantum control of orders and causal inequalities

One might first wonder if the quantum switch allows for a causal inequality violation between A and B (such as the bipartite causal inequalities of [13, 31]); this is however clearly not the case since, as pointed out before, ignoring (i.e., tracing out) the third party C makes the process matrix of the quantum switch causally separable.

One might still hope that the quantum switch can be used to violate a tripartite inequality (see e.g. [30]), explicitly involving party C; as it turns out, this is also impossible, as a consequence of the following theorem[16].

Theorem 4. *Consider $N + 1$ parties $\{A^1, \ldots, A^N, C\}$ with settings $\{x_1, \ldots, x_N, z\}$ and outcomes $\{a_1, \ldots a_N, c\}$. If the marginal distribution*

---

[15] Note that this condition is strictly stronger than no-signalling to each individual party, since it is possible to signal to a group of parties without signalling to any individual party.

[16] A similar conclusion based on the same example has been obtained by Oreshkov and Giarmatzi independently of the other authors of this paper and is presented in [19]

$$P\left(\vec{a}\,\middle|\,\vec{x},\,z\right) := \sum_c P\left(\vec{a},\,c\,\middle|\,\vec{x},\,z\right) \tag{95}$$

is such that

(1) $P\left(\vec{a}\,|\,\vec{x},\,z\right) = P\left(\vec{a}\,|\,\vec{x}\right)$—*i.e., it does not depend on z: C does not signal to any other (group of) parties;*

(2) $P\left(\vec{a}\,|\,\vec{x}\right) = \sum_\sigma q_\sigma\, P_\sigma\left(\vec{a}\,|\,\vec{x}\right)$, *where* $q_\sigma \geqslant 0, \sum_\sigma q_\sigma = 1$, *and the probability distributions* P_σ *are causally ordered,*

then the full $(N+1)$-*partite probability distribution* $P\left(\vec{a},\,c\,|\,\vec{x},\,z\right)$ *is causal.*

Proof. Using Bayes' rule and the assumptions of the theorem, we can write

$$P\left(\vec{a},\,c\,\middle|\,\vec{x},\,z\right) = P\left(\vec{a}\,\middle|\,\vec{x},\,z\right) P\left(c\,\middle|\,\vec{a},\,\vec{x},\,z\right), \tag{96}$$

$$= \sum_\sigma q_\sigma\, P_\sigma\left(\vec{a}\,\middle|\,\vec{x}\right) P\left(c\,\middle|\,\vec{a},\,\vec{x},\,z\right), \tag{97}$$

$$= \sum_\sigma q_\sigma\, \widetilde{P}_\sigma\left(\vec{a},\,c\,\middle|\,\vec{x},\,z\right), \tag{98}$$

where $\widetilde{P}_\sigma(\vec{a},\,c\,|\,\vec{x},\,z) := P_\sigma(\vec{a}\,|\,\vec{x})\, P(c\,|\,\vec{a},\,\vec{x},\,z)$ is compatible with the order $A^{\sigma(1)} \prec \ldots \prec A^{\sigma(N)} \prec C$; this shows that $P(\vec{a},\,c\,|\,\vec{x},\,z)$ is causal. $\qquad\square$

To see that the correlations generated by the quantum switch (equation (66)) respect assumptions 1 and 2 of the previous theorem, let us calculate the marginal probability distribution defined in equation (95) through the generalized Born rule (3), when the three parties perform operations $M_{a|x}^{A_I A_O}$, $M_{b|y}^{B_I B_O}$ and $M_{c|z}^{C_I}$:

$$P(a,\,b|x,\,y,\,z) = \sum_c \mathrm{tr}\left[M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O} \otimes M_{c|z}^{C_I} \cdot |w\rangle\langle w|\right]$$

$$= \mathrm{tr}\left[M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O} \otimes \left(\sum_c M_{c|z}^{C_I}\right) \cdot |w\rangle\langle w|\right]. \tag{99}$$

Since the third party C has no output space ($d_{C_O} = 1$), then for any instrument $\{M_{c|z}^{C_I}\}$ we have $\sum_c M_{c|z}^{C_I} = \mathbb{1}^{C_I}$, so that

$$P(a,\,b|x,\,y,\,z) = \mathrm{tr}\left[M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O} \cdot W^{AB}\right] \tag{100}$$

with

$$W^{AB} := \mathrm{tr}_{C_I} |w\rangle\langle w|. \tag{101}$$

This implies that $P(a,\,b|x,\,y,\,z)$ does not depend on z, as required. As argued before, tracing out C from the process matrix representing the quantum switch leads to a causally separable process matrix of the form $W^{AB} = \frac{1}{2}W^{A\prec B} + \frac{1}{2}W^{B\prec A}$ with causally ordered process matrices $W^{A\prec B}$ and $W^{B\prec A}$, which can only generate causally ordered probability distributions $P_{A\prec B}$ and $P_{B\prec A}$. Hence, $P(a,\,b|x,\,y,\,z)$ can be decomposed as $\frac{1}{2}P_{A\prec B}(a,\,b|x,\,y,\,z) + \frac{1}{2}P_{B\prec A}(a,\,b|x,\,y,\,z)$, so that the second assumption of theorem 4 is also satisfied.

Therefore, the quantum switch represents an example of a causally nonseparable process that can only generate causal correlations, and hence cannot be used to violate any causal inequality[17]. Note that the quantum switch (as defined in equation (66)) is a pure process, so this result shows a stark dissimilarity between entangled states and causally nonseparable processes: while every pure entangled state violates a Bell inequality [32], the analogous result does not hold for process matrices.

It is noteworthy that all the examples of causally nonseparable processes for which a physical interpretation is known, including those generated by space–time superpositions [33], fall into this category. This raises the question of whether causally nonseparable processes that do violate causal inequalities can be physically implemented at all.

7. Conclusion

The process matrix formalism was originally conceived as a rather speculative extension of quantum mechanics to possibly include the indefinite causal structures expected in a quantized theory of gravity [10]. The results of this work show that, in fact, it is a natural framework to study a class of quantum resources which cannot be

---

[17] Note that theorem 4 implies that this is also true for the N-partite generalization of the quantum switch defined in [8].

captured by the circuit model, but nonetheless are physically realizable and can provide powerful computational advantages. We have shown that the quantum switch, a recently demonstrated resource for quantum computation, can be conveniently represented as a causally non-separable process matrix. We have also presented causal witnesses that can verify the causal nonseparability of the switch. As they only require performing unitaries in a 'superposition of order' and a final measurement of a control qubit, such witnesses can be easily implemented in quantum-optics setups, as the one employed in [9].

The theory of causal witnesses developed here has close resemblances with the theory of entanglement witnesses. In both cases, one is interested in finding ways to certify that a resource is outside some convex set, the set of separable states in the latter case, that of causally nonseparable process matrices in the former case. Following this analogy, causal inequalities can be seen as the counterpart to the Bell inequalities, as they both provide device-independent tests regarding the existence of some classical variable: local hidden variables for measurement outcomes in one case, classical variables determining the causal order in the other. A significant difference between the two frameworks is that the problem of determining causal separability can be solved numerically with efficient algorithms, whereas characterizing entanglement has been proven to be an NP-hard problem [34].

As one could expect from the analogy with entanglement, there exist causally nonseparable processes that cannot violate causal inequalities. What is striking, in the case of process matrices, is that a physical interpretation is known only for resources in this category. As one of the main open problems in this field is the characterization of physical process matrices, it is tempting to speculate whether the (im)possibility to violate causal inequalities could provide a useful guidance in this respect.

Acknowledgments

Appendix A. Details of the formalism

Here we explore in more details the properties of the CJ isomorphism and of the process matrix formalism. Note that other existing definitions of the CJ isomorphism differ by a transposition or a partial transposition from the one given here, which follows the convention in [13] and allows a direct identification of non-signaling processes with quantum states.

A.1. CJ isomorphism

A.1.1. Pure CJ isomorphism.   It is convenient to distinguish two versions of the CJ isomorphism: one for maps over density matrices and one for linear operators on pure state. The latter—the 'pure CJ isomorphism'—can be represented via the 'double-ket' notation [35, 36]. For a linear operator $A : \mathcal{H}^{A_I} \rightarrow \mathcal{H}^{A_O}$, we define[18]

$$|A^*\rangle\!\rangle^{A_I A_O} := \mathbb{1} \otimes A^* \, |\mathbb{1}\rangle\!\rangle, \tag{A.1}$$

where $|\mathbb{1}\rangle\!\rangle \equiv |\mathbb{1}\rangle\!\rangle^{A_I A_I} := \sum_j |j\rangle^{A_I} \otimes |j\rangle^{A_I} \in \mathcal{H}^{A_I} \otimes \mathcal{H}^{A_I}$ (with also, of course, the usual notation $\langle\!\langle \mathbb{1}| = |\mathbb{1}\rangle\!\rangle^\dagger$), and the complex conjugation * is defined with respect to the chosen orthonormal basis $\{ |j\rangle^{A_I} \}$ of $\mathcal{H}^{A_I}$. The inverse map is given by

$$A \, |\psi\rangle = \left[\langle\psi|^{A_I} \otimes \mathbb{1}^{A_O} \cdot |A^*\rangle\!\rangle^{A_I A_O} \right]^*. \tag{A.2}$$

We say that $|A^*\rangle\!\rangle$ is the CJ representation (or CJ vector) of A. The cumbersome complex conjugation in the definition allows us to have a simpler representation for the process matrix.

A.1.2. Maximally entangled states and unitaries.   Consider here the case where the input and output spaces have equal dimensions, $d_{A_I} = d_{A_O}$. The state obtained by applying a local unitary to one subsystem of a maximally entangled state is also maximally entangled. in reverse, it is possible to generate any (bipartite) maximally

---

[18] Superscripts on CJ vectors and CJ matrices indicate the systems they refer to (they may be omitted when the context makes it clear enough).

entangled state by applying a local unitary to one subsystem of a reference maximally entangled state. Therefore, the CJ vector $|U^*\rangle\!\rangle^{A_IA_O} = \mathbb{1} \otimes U^* |\mathbb{1}\rangle\!\rangle$ *is maximally entangled if and only if U is a unitary.* More explicitly, an operator

$$U = \sum_{jk} u_{jk} |j\rangle\langle k| \tag{A.3}$$

is unitary if and only if $\sum_l u_{jl} u_{kl}^* = \sum_l u_{lk}^* u_{lj} = \delta_{j,k}$ for all j, k. One can check that this is also a necessary and sufficient condition for which

$$\left| U^* \right\rangle\!\rangle^{A_IA_O} = \sum_{jk} u_{jk}^* |k\rangle^{A_I} |j\rangle^{A_O} \tag{A.4}$$

is maximally entangled.

A.1.3. Measurement-preparation. Another useful linear operator is $|\psi\rangle\langle\phi|$, which describes the observation of an outcome $|\phi\rangle$ in a projective measurement, followed by the repreparation of a state $|\psi\rangle$. Plugging this into the definition (A.1), we find the CJ representation

$$\left| (|\psi\rangle\langle\phi|)^* \right\rangle\!\rangle^{A_IA_O} = |\phi\rangle^{A_I} \otimes \left(|\psi\rangle^* \right)^{A_O}. \tag{A.5}$$

Reciprocally, every pure product CJ vector represents a measurement-preparation operation.

An important particular case is when $|\psi\rangle = |\phi\rangle$, which corresponds to the ideal non-demolition von Neumann measurement:

$$\left| (|\phi\rangle\langle\phi|)^* \right\rangle\!\rangle^{A_IA_O} = |\phi\rangle^{A_I} \otimes \left(|\phi\rangle^* \right)^{A_O}. \tag{A.6}$$

A.1.4. Mixed CJ operators. For the general case of a linear map $\mathcal{M}^A : A_I \to A_O$, we define the CJ isomorphism as

$$M^{A_IA_O} := \left[\mathcal{I} \otimes \mathcal{M}^A (|\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|) \right]^T. \tag{A.7}$$

It is easy to verify that the definition (A.7) reduces to (A.1) for operators of the form $\mathcal{M}^A(\rho) = A\rho A^\dagger$, i.e. that, in such a case

$$M = \left| A^* \right\rangle\!\rangle\langle\!\langle A^* \right| \tag{A.8}$$

(with $|A^*\rangle\!\rangle \equiv |A^*\rangle\!\rangle^{A_IA_O}$ and $\langle\!\langle A^*| = |A^*\rangle\!\rangle^\dagger$).

According to Choi's theorem [37], a linear map $\mathcal{M}^A : A_I \to A_O$ is CP if and only if its CJ matrix is positive semidefinite, $M^{A_IA_O} \geqslant 0$. A characterization of the trace-preserving condition can be found using the inverse CJ isomorphism

$$\mathcal{M}^A(\rho) = \left[\mathrm{tr}_{A_I}\!\left[\rho^{A_I} \otimes \mathbb{1}^{A_O} \cdot M^{A_IA_O} \right] \right]^T. \tag{A.9}$$

By taking the trace of both sides of the equation, it can be readily verified that the map $\mathcal{M}^A$ is trace-preserving if and only if

$$\mathrm{tr}_{A_O} M^{A_IA_O} = \mathbb{1}^{A_I}. \tag{A.10}$$

Note that a CP map can be part of an instrument only if it is *trace-non-increasing*, a condition that translates to

$$\mathbb{1}^{A_I} - \mathrm{tr}_{A_O} M^{A_IA_O} \geqslant 0. \tag{A.11}$$

A useful example is the CPTP map $\mathcal{M}^A(\sigma) = \rho \,\mathrm{tr}\, \sigma$, which corresponds to the preparation of a (normalized) state ρ independently of the input state σ. Its CJ representation is found to be

$$M^{A_IA_O} = \mathbb{1}^{A_I} \otimes \left(\rho^T \right)^{A_O}. \tag{A.12}$$

A second relevant case is the CP (not trace-preserving) map that gives the probability of observing a POVM element E in a measurement: $\mathcal{M}^A(\rho) = \mathrm{tr}[E\rho]$ (here $d_{A_O} = 1$). Its CJ representation is simply

$$M^{A_I} = E^{A_I}. \tag{A.13}$$

Finally, the situation where a POVM element E is measured on the state σ in A_I and a state ρ is prepared in A_O corresponds to the CP map $\mathcal{M}^A(\sigma) = \rho \,\mathrm{tr}\,[E\sigma]$, which has CJ representation

$$M^{A_IA_O} = E^{A_I} \otimes \left(\rho^T \right)^{A_O}. \tag{A.14}$$

A.2. Process matrices

Here we discuss in more detail some examples and properties of process matrices.

A.2.1. Quantum states. Consider a bipartite process matrix of the form

$$W^{A_I A_O B_I B_O} = \rho^{A_I B_I} \otimes \mathbb{1}^{A_O B_O}. \tag{A.15}$$

According to the generalized Born rule, equation (3), the probability for the two parties A and B to perform trace non-increasing CP maps with CJ matrices $M^{A_I A_O}$ and $M^{B_I B_O}$, respectively, is given by

$$
\begin{aligned}
P\left(M^{A_I A_O}, M^{B_I B_O}\right) &= \mathrm{tr}\left[\left(M^{A_I A_O} \otimes M^{B_I B_O}\right)W\right], \\
&= \mathrm{tr}\left[\left(E^{A_I} \otimes E^{B_I}\right)\rho^{A_I B_I}\right],
\end{aligned}
\tag{A.16}
$$

where $E^{A_I} := \mathrm{tr}_{A_O} M^{A_I A_O}$ and $E^{B_I} := \mathrm{tr}_{B_O} M^{B_I B_O}$. These operators are positive semidefinite and, because of equation (A.11), they can be completed to form a POVM. Thus equation (A.16) corresponds to the probability of observing the POVM element $E^{A_I} \otimes E^{B_I}$ given the state $\rho^{A_I B_I}$; in other words, the process matrix (A.15) describes a bipartite state. Notice that a process matrix of this form does not allow signalling in either direction and therefore, being compatible with both $A \prec B$ and $B \prec A$, it is causally separable. This is irrespective of the state ρ, which can be entangled or separable. Note also the difference between the process matrix (A.15) and the CJ representation of state preparation, equation (A.12).

A.2.2. Channels. Consider a bipartite situation where a party A only performs state preparations, while the second party B only performs measurements. In this case, the local laboratory of A is characterized by a trivial input space, $d_{A_I} = 1$, while B has a trivial output space, $d_{B_O} = 1$. The process matrix shared by A and B, which represents here a quantum channel, is then defined on the space $A_O \otimes B_I \ni W$. The probability that B observes a POVM element E when A prepares a state ρ is given by

$$P(E|\rho) = \mathrm{tr}\left[\left(\rho^T\right)^{A_O} \otimes E^{B_I} \cdot W^{A_O B_I}\right], \tag{A.17}$$

where we used (A.12) and (A.13) for the local operations. This is equivalent to saying that B measures E in the state $\mathrm{tr}_{A_O}[(\rho^T)^{A_O} \otimes \mathbb{1}^{B_I} \cdot W^{A_O B_I}] = [\mathrm{tr}_{A_O}[\rho^{A_O} \otimes \mathbb{1}^{B_I} \cdot (W^T)^{A_O B_I}]]^T$. Comparing this with the inverse CJ transformation (A.9), we find that the process matrix W corresponds to a channel with CJ representation W^T. In other words, a channel $\mathcal{C}$ from A_O to B_I is represented by the process matrix

$$W^{A_O B_I} = \mathcal{I} \otimes \mathcal{C}(|\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|) \tag{A.18}$$

(with here $|\mathbb{1}\rangle\!\rangle \equiv |\mathbb{1}\rangle\!\rangle^{A_O A_O}$). Note that the CJ representation of a channel, equation (A.7), differs by a transposition from the corresponding process matrix (A.18).

A.2.3. Reduced process matrices. Given a multipartite process $W = W^{A_I^1 A_O^1 \ldots A_I^N A_O^N}$ and a CPTP map for the jth party with CJ matrix $M^{A_I^j A_O^j}$, we define the *reduced process matrix* for the remaining $N - 1$ parties, given $M^{A_I^j A_O^j}$, as

$$\overline{W}\left(M^{A_I^j A_O^j}\right) := \mathrm{tr}_{A_I^j A_O^j}\left[\left(\mathbb{1}^{A_I^1 A_O^1} \otimes \ldots M^{A_I^j A_O^j} \otimes \ldots \mathbb{1}^{A_I^N A_O^N}\right) \cdot W\right]. \tag{A.19}$$

With the usual generalized Born rule (3), the reduced process matrix gives the probability for the remaining $N - 1$ parties to measure arbitrary CP maps, given that the jth party performs $M^{A_I^j A_O^j}$. The explicit dependence of $\overline{W}$ on $M^{A_I^j A_O^j}$ accounts for the possibility of signalling: the remaining parties observe different probability distributions depending on the choice of CPTP map performed by party j. As an example, consider a process matrix of the form (A.18). If A prepares a state ρ, the reduced process matrix for B is

$$
\begin{aligned}
\overline{W}^{B_I}(\rho) &= \mathrm{tr}_{A_O}\left[\left(\rho^T\right)^{A_O} \otimes \mathbb{1}^{B_I} \cdot W^{A_O B_I}\right] \\
&= \sum_{jk} \langle k| \rho^T |j\rangle \, \mathcal{C}(|j\rangle\langle k|) = \mathcal{C}(\rho).
\end{aligned}
\tag{A.20}
$$

Thus, for a process that represents a channel from A to B, the reduced process for B, given that A prepares ρ, is simply the channel applied to ρ, as should be expected.

A.2.4. Pure process matrices. In some cases, the process matrix turns out to be a rank-one projector: $W = |w\rangle\langle w|$ for some 'process vector' $|w\rangle$. If the CJ operators representing the local operations are also rank-one projectors, as is the case for unitaries and projective measurements followed by pure repreparations, it is convenient to work at the level of vectors and of probability amplitudes: given the local operations $A_1, \ldots, A_N$ represented by the CJ vectors $|A_1^*\rangle\!\rangle^{A_I^1 A_O^1}, \ldots, |A_N^*\rangle\!\rangle^{A_I^N A_O^N}$, the overall probability amplitude is given (up to global

phase, which we choose to be 0) by

$$\langle\!\langle A_1^* |^{A_I^1 A_O^1} \otimes ... \otimes \langle\!\langle A_N^* |^{A_I^N A_O^N} \cdot |w\rangle^{A_I^1 A_O^1 ... A_I^N A_O^N}. \tag{A.21}$$

The probability is then obtained as the modulus square of the amplitude and conforms to the general expression (3). Given that party j performs the unitary U_j, the reduced process is clearly given by the partial scalar product

$$\mathbb{1}^{A_I^1 A_O^1} \otimes ... \langle\!\langle U_j^* |^{A_I^j A_O^j} \otimes ... \mathbb{1}^{A_I^N A_O^N} \cdot |w\rangle^{A_I^1 A_O^1 ... A_I^N A_O^N}. \tag{A.22}$$

The process matrix describing a unitary channel U from A_O to B_I is of particular interest. Using (A.18), we find that it is given by

$$|w\rangle^{A_O B_I} = \mathbb{1} \otimes U |\mathbb{1}\rangle\!\rangle = |U\rangle\!\rangle^{A_O B_I}. \tag{A.23}$$

Note again the difference between this expression and the CJ representation (A.1). Generalizing this to a sequence of parties $A_1, ..., A_N$, with the output of party j connected to the input of party $j + 1$ via the unitary U_j, we find

$$|w\rangle^{A_O^1 ... A_I^N} = |U_1\rangle\!\rangle^{A_O^1 A_I^2} \otimes ... \otimes |U_N\rangle\!\rangle^{A_O^{N-1} A_I^N}. \tag{A.24}$$

Appendix B. Valid process matrices

The conditions for an operator $W \in A_I \otimes A_O \otimes B_I \otimes B_O$ to be a valid process matrix were first found in [13], where they were formulated in a basis-dependent way. Here we derive the equivalent characterization of valid process matrices given in equations (4)–(6); we formulate it in a basis-independent way, which we find to be more convenient for our purposes.

We present below the derivation in the bipartite case, and also write explicitly, for ease of reference, the characterization in the tripartite case. The N-partite case follows from a straightforward generalization.

B.1. Bipartite process matrices

Recall that a given operator $W \in A_I \otimes A_O \otimes B_I \otimes B_O$ is a valid process matrix if and only if it yields, through the generalized Born rule (3), only well-defined probabilities—that is, the probabilities must be non-negative and must sum up to 1.

Non-negativity. As recalled previously, a map is CP if and only if its CJ representation is positive semidefinite. Including the possibility that A and B's operations involve interactions with a (possibly entangled) ancillary system in a state $\rho^{A_I' B_I'}$, the non-negativity of probabilities is thus equivalent to[19]

$$\text{tr}\left[\left(M^{A_I' A_I A_O} \otimes M^{B_I' B_I B_O} \right) \cdot \left(\rho^{A_I' B_I'} \otimes W^{A_I A_O B_I B_O} \right) \right] \geqslant 0$$

$$\forall \ M^{A_I' A_I A_O} \geqslant 0, M^{B_I' B_I B_O} \geqslant 0, \rho^{A_I' B_I'} \geqslant 0. \tag{B.1}$$

For the case where the ancillary spaces A_I' and B_I' are isomorphic to $A_I \otimes A_O$, and $M^{A_I' A_I A_O}$ and $\rho^{A_I' B_I'}$ are both projectors onto the maximally entangled state $|\mathbb{1}\rangle\!\rangle^{A_I A_O / A_I A_O} := \sum_{j,k} |j, k\rangle^{A_I A_O} \otimes |j, k\rangle^{A_I A_O}$ (where $\{|j, k\rangle^{A_I A_O}\}$ is an orthonormal basis of $\mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O}$), we find that the trace in (B.1) is equal to $\text{tr}[M^{B_I' B_I B_O} \cdot W^{A_I A_O B_I B_O}]$. Requiring that its value is non-negative for all $M^{B_I' B_I B_O} \geqslant 0$ implies that W must be positive semidefinite.

Reciprocally, $W \geqslant 0$ clearly implies that (B.1) is satisfied. Hence, the non-negativity of probabilities is equivalent to W being positive semidefinite, equation (4).

Normalization. The fact that probabilities must sum up to 1 for all instruments is equivalent to the constraint that the probability of realization of any CPTP map is 1. Now, recall that a CP map $\mathcal{M}^A : A_I \to A_O$ is trace-preserving if and only if its CJ matrix $M^{A_I A_O}$ satisfies $\text{tr}_{A_O} M^{A_I A_O} = \mathbb{1}^{A_I}$—or equivalently, using the notation of equation (12), ${}_{A_O}M^{A_I A_O} = \mathbb{1}^{A_I A_O}/d_{A_O}$. Ignoring here for simplicity the possible use of an ancillary system (which leads to the same conclusion[20]), the normalization of probabilities is thus equivalent to

---

[19] Note that ignoring the possibility of an ancillary system, one would only find that W must be 'positive on pure tensors' (with respect to the partition $A_I A_O / B_I B_O$)—a class strictly larger than positive semidefinite matrices [38].

[20] Taking into account a possible ancillary state $\rho^{A_I' B_I'}$, the same reasoning as below leads to equations (B.4) and (B.15), with W replaced by $\rho^{A_I' B_I'} \otimes W^{A_I A_O B_I B_O}$ (and where the definitions of the maps L_A and L_B should include the ancillary systems B_I' and A_I', resp.), which must hold for all $\rho^{A_I' B_I'}$ such that $\text{tr} \ \rho^{A_I' B_I'} = 1$. One can easily check that these are indeed equivalent to (B.4) and (B.15) in their original form.

$$\mathrm{tr}\left[\left(M^{A_I A_O} \otimes M^{B_I B_O}\right) \cdot W^{A_I A_O B_I B_O}\right] = 1$$

$$\forall \ M^{A_I A_O} \geqslant 0, M^{B_I B_O} \geqslant 0,$$

$$\text{s.t. } {}_{A_O}M^{A_I A_O} = \mathbb{1}^{A_I A_O}/d_{A_O}, \ {}_{B_O}M^{B_I B_O} = \mathbb{1}^{B_I B_O}/d_{B_O}. \tag{B.2}$$

First of all, note that the positivity of $M^{A_I A_O}$ and $M^{B_I B_O}$ is irrelevant here, since the set of positive semidefinite operators is a full dimensional subset of the space of Hermitian operators[21]. The only relevant conditions are the normalization constraints. Defining the maps ${}_{[1-A_O]}M = M - {}_{A_O}M$ and ${}_{[1-B_O]}M = M - {}_{B_O}M$, and noting that for any Hermitian operators x and y, the operators ${}_{[1-A_O]}x + \mathbb{1}/d_{A_O}$ and ${}_{[1-B_O]}y + \mathbb{1}/d_{B_O}$ satisfy the above normalization constraints (where from now on we are omitting the superscripts to reduce cluttering), we find that equation (B.2) is equivalent to

$$\mathrm{tr}\left[\left({}_{[1-A_O]}x + \mathbb{1}/d_{A_O}\right) \otimes \left({}_{[1-B_O]}y + \mathbb{1}/d_{B_O}\right)W\right] = 1 \quad \forall \ x, y. \tag{B.3}$$

For $x = y = 0$, this yields the normalization condition of equation (5),

$$\mathrm{tr}[W] = d_{A_O}d_{B_O}. \tag{B.4}$$

For $y = 0$ and $x = 0$, respectively, this in turn implies

$$\mathrm{tr}\left[\left({}_{[1-A_O]}x \otimes \mathbb{1}\right)W\right] = 0 \quad \forall \ x, \tag{B.5}$$

$$\mathrm{tr}\left[\left(\mathbb{1} \otimes {}_{[1-B_O]}y\right)W\right] = 0 \quad \forall \ y, \tag{B.6}$$

which then imply

$$\mathrm{tr}\left[\left({}_{[1-A_O]}x \otimes {}_{[1-B_O]}y\right)W\right] = 0 \quad \forall \ x, y. \tag{B.7}$$

Reciprocally, equations (B.4)–(B.7) clearly imply (B.3), so that these are equivalent to equation (B.2).

Thinking of the trace as the Hilbert–Schmidt inner product

$$\langle M, W \rangle = \mathrm{tr}[M \cdot W]$$

(for Hermitian operators M, W) and noting that the maps ${}_{[1-A_O]}\cdot$ and ${}_{[1-B_O]}\cdot$ are self-adjoint, the conditions (B.5)–(B.7) are equivalent to

$${}_{[1-A_O]}\left(\mathrm{tr}_{B_I B_O} W\right) = 0, \tag{B.8}$$

$${}_{[1-B_O]}\left(\mathrm{tr}_{A_I A_O} W\right) = 0, \tag{B.9}$$

$${}_{[1-A_O]}{}_{[1-B_O]}W = 0, \tag{B.10}$$

which we can rewrite as

$${}_{B_I B_O}W = {}_{A_O B_I B_O}W, \tag{B.11}$$

$${}_{A_I A_O}W = {}_{A_I A_O B_O}W, \tag{B.12}$$

$$W = {}_{B_O}W + {}_{A_O}W - {}_{A_O B_O}W, \tag{B.13}$$

which are conditions (9)–(11).

Note that each condition (B.8)–(B.10) defines a linear subspace, and the intersection of these three linear subspaces is the smallest subspace that contains all valid bipartite process matrices, which we denote by[22]

$$\mathcal{L}_V = \left\{ W \in A_I \otimes A_O \otimes B_I \otimes B_O \ \middle|\ W = L_V(W) \right\}, \tag{B.14}$$

The projector onto this subspace, L_V, shall be used quite often in the paper, so it is useful to find an explicit expression for it. To do that, first we rewrite conditions (B.11)–(B.13) explicitly as projections onto subspaces, i.e., as

---

[21] More precisely: for any Hermitian operator $M^{A_I A_O}$, there always exists $\alpha > 0$ such that $M^{A_I A_O} + \alpha \, \mathbb{1}^{A_I A_O}/d_{A_O} \geqslant 0$. Assuming that ${}_{A_O}M^{A_I A_O} = \mathbb{1}^{A_I A_O}/d_{A_O}$, one can thus decompose $M^{A_I A_O}$ as $M^{A_I A_O} = (\alpha + 1) M_+^{A_I A_O} - \alpha \, M_-^{A_I A_O}$, with $M_+^{A_I A_O} = \frac{1}{\alpha+1}(M^{A_I A_O} + \alpha \, \mathbb{1}^{A_I A_O}/d_{A_O}) \geqslant 0$ and $M_-^{A_I A_O} = \mathbb{1}^{A_I A_O}/d_{A_O} \geqslant 0$ satisfying ${}_{A_O}M_+^{A_I A_O} = {}_{A_O}M_-^{A_I A_O} = \mathbb{1}^{A_I A_O}/d_{A_O}$. Similarly, any Hermitian operator $M^{B_I B_O}$ such that ${}_{B_O}M^{B_I B_O} = \mathbb{1}^{B_I B_O}/d_{B_O}$ can be decomposed as $M^{B_I B_O} = (\beta + 1) M_+^{B_I B_O} - \beta \, M_-^{B_I B_O}$, with $M_+^{B_I B_O} \geqslant 0, M_-^{B_I B_O} \geqslant 0$ and ${}_{B_O}M_+^{B_I B_O} = {}_{B_O}M^{B_I B_O} = \mathbb{1}^{B_I B_O}/d_{B_O}$. Note that the four pairs $(M_\pm^{A_I A_O}, M_\pm^{B_I B_O})$ satisfy the assumptions of equation (B.2), and therefore $\mathrm{tr}\,[(M_\pm^{A_I A_O} \otimes M_\pm^{B_I B_O}) \cdot W^{A_I A_O B_I B_O}] = 1$. Expanding $\mathrm{tr}\,[(M^{A_I A_O} \otimes M^{B_I B_O}) \cdot W^{A_I A_O B_I B_O}]$ using the decomposition just constructed, we find that its value is also 1. Hence, if equation (B.2) holds, then it also holds without the positivity constraints on $M^{A_I A_O}$ and $M^{B_I B_O}$; the converse is of course trivially true.

[22] Note that although we do not write that explicitly, the projectors we define below (e.g. L_V), and of course the subspaces they define (e.g. $\mathcal{L}_V$), depend on the number of parties N.

$$W = L_A(W), \; W = L_B(W), \; W = L_{AB}(W), \tag{B.15}$$

where the projectors L_A, L_B, and L_{AB} are given by

$$L_A(W) = W - {}_{B_I B_O}W + {}_{A_O B_I B_O}W, \tag{B.16}$$

$$L_B(W) = W - {}_{A_I A_O}W + {}_{A_I A_O B_O}W, \tag{B.17}$$

$$L_{AB}(W) = {}_{B_O}W + {}_{A_O}W - {}_{A_O B_O}W, \tag{B.18}$$

Since the three projectors above commute, the projector onto the intersection of their subspaces L_V is given simply by the composition of L_A, L_B, and L_{AB}, i.e.,

$$L_V(W) = L_A \circ L_B \circ L_{AB}(W), \tag{B.19}$$

which, after simplification, can be written as

$$L_V(W) = {}_{A_O}W + {}_{B_O}W - {}_{A_O B_O}W - {}_{B_I B_O}W + {}_{A_O B_I B_O}W - {}_{A_I A_O}W + {}_{A_I A_O B_O}W. \tag{B.20}$$

Summing up, we conclude that an operator $W \in A_I \otimes A_O \otimes B_I \otimes B_O$ is a valid bipartite process matrix if and only if $W \geqslant 0$, $\mathrm{tr}\, W = d_{A_O} d_{B_O}$, and $W = L_V(W)$, as in equations (4)–(6).

B.2. Tripartite process matrices

A similar reasoning leads to the conclusion that an operator $W \in A_I \otimes A_O \otimes B_I \otimes B_O \otimes C_I \otimes C_O$ is a valid tripartite process matrix if and only if $W \geqslant 0$, $\mathrm{tr}\, W = d_{A_O} d_{B_O} d_{C_O}$, and

$$
\begin{aligned}
W = L_A(W), &\quad W = L_B(W), &\quad W = L_C(W), \\
W = L_{AB}(W), &\quad W = L_{AC}(W), &\quad W = L_{BC}(W), \\
W = L_{ABC}(W),
\end{aligned}
\tag{B.21}
$$

where the maps $L_A, L_B, L_C, L_{AB}, L_{AC}, L_{BC}$, and L_{ABC} are now commuting projectors onto linear subspaces of $A_I \otimes A_O \otimes B_I \otimes B_O \otimes C_I \otimes C_O$, defined by

$$L_A(W) = \left[{}_{1-(1-A_O)B_I B_O C_I C_O} \right] W,$$

$$L_B(W) = \left[{}_{1-(1-B_O)A_I A_O C_I C_O} \right] W,$$

$$L_C(W) = \left[{}_{1-(1-C_O)A_I A_O B_I B_O} \right] W,$$

$$L_{AB}(W) = \left[{}_{1-(1-A_O)(1-B_O)C_I C_O} \right] W,$$

$$L_{AC}(W) = \left[{}_{1-(1-A_O)(1-C_O)B_I B_O} \right] W,$$

$$L_{BC}(W) = \left[{}_{1-(1-B_O)(1-C_O)A_I A_O} \right] W,$$

$$L_{ABC}(W) = \left[{}_{1-(1-A_O)(1-B_O)(1-C_O)} \right] W,$$

where we used the shorthand notation

$$\left[\sum_X \alpha_X X \right] W = \sum_X \alpha_X \cdot {}_X W \tag{B.22}$$

for a sum over products of subsystems X with coefficients α_X (and with ${}_1 W := W$).

The constraints in (B.21) are equivalent to

$$W = L_V(W), \tag{B.23}$$

where the map L_V is obtained here by composing the 7 maps $L_A, L_B, L_C, L_{AB}, L_{AC}, L_{BC}$, and L_{ABC}. One finds in this tripartite case, after simplification

$$L_V(W) = \left[{}_{1-(1-A_O+A_I A_O)(1-B_O+B_I B_O)(1-C_O+C_I C_O)+A_I A_O B_I B_O C_I C_O} \right] W, \tag{B.24}$$

which defines a projector onto the linear subspace

$$\mathcal{L}_V = \left\{ W \in A_I \otimes A_O \otimes B_I \otimes B_O \otimes C_I \otimes C_O \;\middle|\; W = L_V(W) \right\}. \tag{B.25}$$

B.3. N-partite process matrices

The generalization to the N-partite case is rather straightforward. We find that an operator $W \in A_I^1 \otimes A_O^1 \otimes \ldots \otimes A_I^N \otimes A_O^N$ is a valid N-partite process matrix if and only if $W \geqslant 0$, $\mathrm{tr}\, W = d_{A_O^1} \ldots d_{A_O^N}$, and for all $2^N - 1$ non-empty subsets $\mathcal{X}$ of $\{1, \ldots, N\}$,

$$W = L_{\mathcal{X}}(W) := \left[1 - \prod_{i \in \mathcal{X}} \left(1 - A_O^i \right) \prod_{i \notin \mathcal{X}} A_I^i A_O^i \right] W. \tag{B.26}$$

Note that the $2^N - 1$ maps $L_{\mathcal{X}}$ are commuting projectors onto linear subspaces of $A_I^1 \otimes A_O^1 \otimes \dots \otimes A_I^N \otimes A_O^N$. The constraints (B.26) are equivalent to

$$W = L_V(W), \tag{B.27}$$

where the map L_V is obtained this time by composing the $2^N - 1$ maps $L_{\mathcal{X}}$. More explicitly, one finds in the N-partite case the general expression

$$L_V(W) = \left[1 - \prod_i \left(1 - A_O^i + A_I^i A_O^i \right) + \prod_i A_I^i A_O^i \right] W, \tag{B.28}$$

which again defines a projector onto the linear subspace

$$\mathcal{L}_V = \left\{ W \in A_I^1 \otimes A_O^1 \otimes \dots \otimes A_I^N \otimes A_O^N \ \middle| \ W = L_V(W) \right\}. \tag{B.29}$$

Proof. For any subset $\mathcal{X}$ of $\{1, \dots, N\}$, define $P_{\mathcal{X}} = \prod_{i \in \mathcal{X}} \left(1 - A_O^i \right) \prod_{i \notin \mathcal{X}} A_I^i A_O^i$. For $\mathcal{X} \neq \mathcal{X}'$, note that there exists (at least one) i_0 such that the product $P_{\mathcal{X}} P_{\mathcal{X}'}$ contains the factor $(1 - A_O^{i_0}) A_I^{i_0} A_O^{i_0}$. Now, $\left[\left(1 - A_O^{i_0} \right) A_O^{i_0} \right] W = 0$, so that $_{[P_{\mathcal{X}} P_{\mathcal{X}'}]} W = 0$.

Developing L_V, we thus find $L_V(W) = \left[\prod_{\mathcal{X} \neq \varnothing} (1 - P_{\mathcal{X}}) \right] W = \left[1 - \sum_{\mathcal{X} \neq \varnothing} P_{\mathcal{X}} \right] W = \left[1 - \sum_{\mathcal{X}} P_{\mathcal{X}} + \prod_i A_I^i A_O^i \right] W$. Now, one can write $\sum_{\mathcal{X}} P_{\mathcal{X}} = \sum_{k_1=0}^1 \dots \sum_{k_N=0}^1 (1 - A_O^1)^{k_1} (A_I^1 A_O^1)^{1-k_1} \dots (1 - A_O^N)^{k_N} (A_I^N A_O^N)^{1-k_N} = \prod_i \sum_{k_i=0}^1 (1 - A_O^i)^{k_i} (A_I^i A_O^i)^{1-k_i} = \prod_i (1 - A_O^i + A_I^i A_O^i)$, from which equation (B.28) follows. $\qquad \square$

Appendix C. Characterization of bipartite causal witnesses

To characterize below the set of causal witnesses in the bipartite case, we shall make use of some basic definitions and facts from convex analysis, which we state here without any proof; the interested reader can find them for instance in sections 2, 14 and 16 of [21] [23].

Let E be a vector space equipped with an inner product $\langle \cdot, \cdot \rangle$, and let E' be the space of all linear functionals on E. In the finite-dimensional case that interests us, E' is isomorphic to E.

(1) A subset $\mathcal{K}$ of E is a convex cone if and only if for every $x, y \in \mathcal{K}$ we also have that $\lambda x + \mu y \in \mathcal{K}$, for any $\lambda, \mu > 0$.

(2) Let $\mathcal{K} \subseteq E$ be a convex cone. Then its dual cone $\mathcal{K}^* \subseteq E'$ is defined as

$$\mathcal{K}^* = \left\{ x^* \in E' \ \middle| \ \langle x^*, x \rangle \geqslant 0 \quad \forall x \in \mathcal{K} \right\}. \tag{C.1}$$

(3) The dual of a linear subspace $\mathcal{L} \subseteq E$ is its orthogonal complement:

$$\mathcal{L}^* = \mathcal{L}^\perp = \left\{ x^* \in E' \ \middle| \ \langle x^*, x \rangle = 0 \quad \forall x \in \mathcal{L} \right\}. \tag{C.2}$$

(4) Let $\mathcal{K}_1, \mathcal{K}_2 \subseteq E$ be closed convex cones that contain the origin. Then

$$\left[\mathrm{conv} \left(\mathcal{K}_1 \bigcup \mathcal{K}_2 \right) \right]^* = \mathcal{K}_1^* \bigcap \mathcal{K}_2^*, \tag{C.3a}$$

$$\left(\mathcal{K}_1 \bigcap \mathcal{K}_2 \right)^* = \mathrm{conv} \left(\mathcal{K}_1^* \bigcup \mathcal{K}_2^* \right), \tag{C.3b}$$

where conv denotes the convex hull.

We shall furthermore use below the following characterization of bipartite causally separable process matrices:

---

[23] Note a difference in language. This reference uses the polar cone K° instead of the dual K^*. They are simply related by $K^\circ = -K^*$.

Lemma 5. *A given matrix $W \in A_I \otimes A_O \otimes B_I \otimes B_O$ is a valid causally separable process matrix if and only if* tr $W = d_O, W \in \mathcal{L}_V$ *(i.e., it satisfies equations ([9])–([11])), and it can be written as*

$$W = W^{A \prec B} + W^{B \prec A} \tag{C.4}$$

$$\text{with} \quad W^{A \prec B} \geqslant 0, \quad W^{A \prec B} = {}_{B_O}W^{A \prec B}, \tag{C.5}$$

$$W^{B \prec A} \geqslant 0, \quad W^{B \prec A} = {}_{A_O}W^{B \prec A}. \tag{C.6}$$

Proof. The 'only if' direction is straightforward (simply replace $q\ W^{A \prec B} \rightarrow W^{A \prec B}$ and $(1 - q)\ W^{B \prec A} \rightarrow W^{B \prec A}$ to go from ([18]) to ([C.4]), so that $W^{A \prec B}$ and $W^{B \prec A}$ in ([C.4]) are not normalized).

To see that the converse also holds, first note that $W^{A \prec B} \geqslant 0$ and $W^{B \prec A} \geqslant 0$ imply that $W \geqslant 0$, so that W is indeed a valid process matrix. Note furthermore that $W^{B \prec A} = {}_{A_O}W^{B \prec A}$ implies that ${}_{B_I B_O}W^{B \prec A} = {}_{A_O B_I B_O}W^{B \prec A}$, i.e., that $W^{B \prec A}$ satisfies ([9]). Since $W \in \mathcal{L}_V$ also satisfies ([9]), so does $W^{A \prec B} = W - W^{B \prec A}$. Similarly, $W^{A \prec B} = {}_{B_O}W^{A \prec B}$, together with the assumption that $W \in \mathcal{L}_V$, implies that both $W^{A \prec B}$ and $W^{B \prec A}$ satisfy ([10]). Lastly, $W^{A \prec B} = {}_{B_O}W^{A \prec B}$ and $W^{B \prec A} = {}_{A_O}W^{B \prec A}$ directly imply that both $W^{A \prec B}$ and $W^{B \prec A}$ satisfy ([11]). All in all, this shows that $W^{A \prec B}$ and $W^{B \prec A}$ are, up to normalization (which can easily be dealt with as above so as to recover the form ([18])), valid causally ordered process matrices. □

We are now in a position to prove theorem [1]:

Proof of theorem 1. We want to characterize the set of all Hermitian operators $S \in A_I \otimes A_O \otimes B_I \otimes B_O$ such that

$$\text{tr}\left[SW^{\text{sep}}\right] \geqslant 0 \tag{C.7}$$

for all causally separable process matrices W^{sep}. Note that the condition $\text{tr}[W^{\text{sep}}] = d_O$ is not relevant for the characterization of the witnesses, so we shall lift it. Without this restriction the set of (non-normalized) causally separable process matrices becomes a convex cone, which we denote by $\mathcal{W}^{\text{sep}}$. If we consider the duality relations with respect to the Hilbert–Schmidt inner product in the space $A_I \otimes A_O \otimes B_I \otimes B_O$, then the convex cone of causal witnesses $\mathcal{S}$ is the dual cone of $\mathcal{W}^{\text{sep}}$. To characterize it we are going to use the representation of $\mathcal{W}^{\text{sep}}$ that follows from lemma [5]:

$$\mathcal{W}^{\text{sep}} = \text{conv}\left[\left(\mathcal{P} \bigcap \mathcal{L}_{B_O}\right) \bigcup \left(\mathcal{P} \bigcap \mathcal{L}_{A_O}\right)\right] \bigcap \mathcal{L}_V, \tag{C.8}$$

where $\mathcal{P}$ is the self-dual cone of positive semidefinite matrices and $\mathcal{L}_{B_O}$ and $\mathcal{L}_{A_O}$ are the linear subspaces

$$\mathcal{L}_{B_O} = \left\{W \mid W = {}_{B_O}W\right\}, \tag{C.9}$$

$$\mathcal{L}_{A_O} = \left\{W \mid W = {}_{A_O}W\right\}, \tag{C.10}$$

and with $\mathcal{L}_V$ defined in equation ([B.14]). Their orthogonal complements within the subspace $A_I \otimes A_O \otimes B_I \otimes B_O$ are simply given by

$$\mathcal{L}_{B_O}^{\perp} = \left\{S \mid {}_{B_O}S = 0\right\}, \tag{C.11}$$

$$\mathcal{L}_{A_O}^{\perp} = \left\{S \mid {}_{A_O}S = 0\right\}, \tag{C.12}$$

$$\mathcal{L}_V^{\perp} = \left\{S \mid L_V(S) = 0\right\}. \tag{C.13}$$

Taking then the dual of $\mathcal{W}^{\text{sep}}$ using the duality relations ([C.2]), ([C.3]), we get that the cone of causal witnesses is

$$S = \text{conv}\left[\left[\left(\mathcal{P} \bigcap \mathcal{L}_{B_O}\right)^* \bigcap \left(\mathcal{P} \bigcap \mathcal{L}_{A_O}\right)^*\right] \bigcup \mathcal{L}_V^{\perp}\right]. \tag{C.14}$$

Focusing on $(\mathcal{P} \bigcap \mathcal{L}_{B_O})^*$, using again the duality relations ([C.2]), ([C.3]), we see that

$$\left(\mathcal{P} \bigcap \mathcal{L}_{B_O}\right)^* = \text{conv}\left(\mathcal{P} \bigcup \mathcal{L}_{B_O}^{\perp}\right) = \left\{S_+ + S_0 \mid S_+ \geqslant 0, {}_{B_O}S_0 = 0\right\} = \left\{S \mid {}_{B_O}S \geqslant 0\right\}, \tag{C.15}$$

where the last equality is stating the fact that $S = S_+ + S_0$ with $S_+ \geqslant 0$ and ${}_{B_O}S_0 = 0$ if and only if ${}_{B_O}S \geqslant 0$. To see that this is true, let S be such that ${}_{B_O}S \geqslant 0$. Define then $S_+ = {}_{B_O}S$ and $S_0 = S - {}_{B_O}S$. Then $S = S_+ + S_0$, $S_+ \geqslant 0$ by assumption, and ${}_{B_O}S_0 = 0$ since the map ${}_{B_O}\cdot$ is a projector. To prove the other direction, let $S = S_+ + S_0$ with $S_+ \geqslant 0$ and ${}_{B_O}S_0 = 0$. The map ${}_{B_O}\cdot$ being positive, it follows that ${}_{B_O}S = {}_{B_O}S_+ \geqslant 0$.

Similarly,

$$\left(\mathcal{P} \cap \mathcal{L}_{A_O}\right)^* = \left\{ S \mid {}_{A_O}S \geqslant 0 \right\}. \tag{C.16}$$

Putting equations (C.14)–(C.16) together, we see that a causal witness can be written as $S = S_P + S^\perp$ with ${}_{B_O}S_P \geqslant 0$, ${}_{A_O}S_P \geqslant 0$, and $L_V(S^\perp) = 0$. □

Appendix D. Explicit positive semidefinite constraints

For the convenience of the reader, we present the SDP problems (38) and (39) with all conic constraints rewritten in terms of the positive semidefinite cone, to facilitate implementation. To rewrite (38), we need a characterization of the cones $\mathcal{S}_V$ and $\mathcal{W}_V^*$. The first one is given by corollary 2. The second one is obtained as follows: since $\mathcal{W} = \mathcal{P} \cap \mathcal{L}_V$, we have that

$$\begin{aligned}
\mathcal{W}_V^* &= \mathcal{W}^* \cap \mathcal{L}_V \\
&= \left(\mathcal{P} \cap \mathcal{L}_V\right)^* \cap \mathcal{L}_V \\
&= \mathrm{conv}\left(\mathcal{P} \cup \mathcal{L}_V^\perp\right) \cap \mathcal{L}_V \\
&= \left\{ L_V\left(\Sigma_P\right) \mid \Sigma_P \geqslant 0 \right\},
\end{aligned} \tag{D.1}$$

where equation (D.1) follows from an argument analogous to the one used to derive corollary 2.

With this characterization, the SDP problem (38) then becomes

$$\begin{aligned}
& \min \ \mathrm{tr}(SW) \\
& \text{s.t.} \quad S = L_V\left(S_P\right), \ {}_{A_O}S_P \geqslant 0, \ {}_{B_O}S_P \geqslant 0, \\
& \mathbb{1}/d_O - S = L_V\left(\Sigma_P\right), \ \Sigma_P \geqslant 0.
\end{aligned} \tag{D.2}$$

To rewrite the SDP problem (39), we use the characterization of $\mathcal{W}^{\mathrm{sep}}$ given in lemma 5:

$$\begin{aligned}
& \min \ \mathrm{tr}(\Omega)/d_O \\
& \text{s.t.} \quad W + \Omega = W^{A \prec B} + W^{B \prec A}, \\
& W^{A \prec B} \geqslant 0, \quad W^{A \prec B} = {}_{B_O}W^{A \prec B}, \\
& W^{B \prec A} \geqslant 0, \quad W^{B \prec A} = {}_{A_O}W^{B \prec A}, \\
& \Omega \geqslant 0, \quad \Omega = L_V(\Omega).
\end{aligned} \tag{D.3}$$

Note that we could use directly the definition of $\mathcal{W}^{\mathrm{sep}}$ from section 2.2.2, which would give us a slightly more complicated SDP problem.

Appendix E. Duality for conic problems

In this appendix we show that the two problems defined in section 3.2 are SDP problems and they are dual to each other. We show, furthermore, that the duality theorem applies to them, which implies that the optimal solutions can be found efficiently and that equation (40) holds.

Let us first recall the definitions of primal and dual conic problems (definition 4.2.1 in [22]), of which SDP problems are a particular case:

Definition 6. Let E be a finite-dimensional vector space, $\mathcal{K}$ a closed convex pointed cone in E with a nonempty interior, and $\mathcal{L}$ a linear subspace of E. Let also $b \in E$ and $c \in E'$. The data E, $\mathcal{K}$, $\mathcal{L}$, b, and c define a pair of conic problems

$$\begin{aligned}
(P): \quad & \min \ \langle c, x \rangle \quad \text{s.t.} \quad x \in \mathcal{K} \cap (\mathcal{L} + b), \\
(D): \quad & \min \ \langle y, b \rangle \quad \text{s.t.} \quad y \in \mathcal{K}^* \cap \left(\mathcal{L}^\perp + c\right),
\end{aligned}$$

where $\mathcal{K}^* \subset E'$ is the cone dual to $\mathcal{K}$, $\mathcal{L}^\perp \subset E$ is the orthogonal complement to $\mathcal{L}$, $\mathcal{L} + b \subset E$ and $\mathcal{L}^\perp + c \subset E'$ are affine subspaces. (P) and (D) are called, respectively, the primal and dual problems associated with the above data.

We want our SDP problems to measure how much worst-case noise needs to be added to a given process matrix W to make it causally separable, i.e., the minimal $\lambda \geqslant 0$ for which

$$\frac{1}{1+\lambda}\left(W + \lambda\,\widetilde{\Omega}\right) \tag{E.1}$$

is a causally separable process, optimized over all valid (normalized) processes $\widetilde{\Omega}$. First note that we can get rid of the quadratic variable $\lambda\,\widetilde{\Omega}$ by defining $\Omega = \lambda\,\widetilde{\Omega}$, which makes the objective λ equal to $\mathrm{tr}\,\Omega/d_O$. Remembering also that the normalization $1/(1+\lambda)$ is irrelevant for conic constraints, the problem reduces to minimizing $\mathrm{tr}\,\Omega/d_O$ such that

$$W + \Omega \in \mathcal{W}^{\mathrm{sep}}, \quad \Omega \in \mathcal{W}. \tag{E.2}$$

To translate this SDP problem into the language of definition 6, let us define

$$E = \mathcal{L}_V \times \mathcal{L}_V, \tag{E.3}$$

$$\mathcal{K} = \mathcal{W}^{\mathrm{sep}} \times \mathcal{W}, \tag{E.4}$$

$$\mathcal{L} = \left\{ (\Omega, \Omega) \,\middle|\, \Omega \in \mathcal{L}_V \right\}, \tag{E.5}$$

$$b = (W, 0), \tag{E.6}$$

$$c = \left(0, \mathbb{1}/d_O\right), \tag{E.7}$$

and the inner product

$$\langle (S, \Sigma), (W, \Omega) \rangle = \mathrm{tr}(SW) + \mathrm{tr}(\Sigma\Omega). \tag{E.8}$$

With these definitions, and denoting by $x = (\omega, \Omega)$ its variable, the primal SDP problem becomes

$$\begin{aligned} \min \quad & \left\langle \left(0, \mathbb{1}/d_O\right), (\omega, \Omega) \right\rangle \\ \mathrm{s.t.} \quad & (\omega, \Omega) \in \left(\mathcal{W}^{\mathrm{sep}} \times \mathcal{W}\right) \cap \{\omega = W + \Omega\}, \end{aligned} \tag{E.9}$$

which indeed corresponds to the SDP problem (39).

To construct the dual SDP problem, first note that

$$E' = \mathcal{L}_V \times \mathcal{L}_V, \tag{E.10}$$

$$\mathcal{K}^* = \mathcal{S}_V \times \mathcal{W}_V^*, \tag{E.11}$$

$$\mathcal{L}^\perp = \left\{ (S, -S) \,\middle|\, S \in \mathcal{L}_V \right\} \tag{E.12}$$

where we used the property that $(\mathcal{K}_1 \times \mathcal{K}_2)^* = \mathcal{K}_1^* \times \mathcal{K}_2^*$ in equation (E.11). Denoting by $y = (S, \Sigma)$ its variable, the dual SDP problem is then

$$\begin{aligned} \min \quad & \langle (S, \Sigma), (W, 0) \rangle \\ \mathrm{s.t.} \quad & (S, \Sigma) \in \left(\mathcal{S}_V \times \mathcal{W}_V^*\right) \cap \left\{ \Sigma = \mathbb{1}/d_O - S \right\}, \end{aligned} \tag{E.13}$$

which corresponds to the SDP problem (38).

Let us emphasize that here the duals of $\mathcal{W}^{\mathrm{sep}}$ and $\mathcal{W}$ are, respectively, $\mathcal{S}_V$ and $\mathcal{W}_V^*$, instead of $\mathcal{S}$ and $\mathcal{W}^*$, which is a consequence of choosing the vector space E to be $E = F \times F$ with $F = \mathcal{L}_V$ instead of $F = A_I \otimes A_O \otimes B_I \otimes B_O$. We did this because as subsets of $A_I \otimes A_O \otimes B_I \otimes B_O$, the cones $\mathcal{W}^{\mathrm{sep}}$ and $\mathcal{W}$ (and therefore $\mathcal{K} = \mathcal{W}^{\mathrm{sep}} \times \mathcal{W}$) have empty interiors, and therefore these cones would not satisfy the requirements of definition 6. This is problematic because the duals of cones with empty interior are not pointed (in our case, $\mathcal{S}$ and $\mathcal{W}^*$ are not pointed), and algorithms that solve SDP problems are numerically unstable when optimizing over non-pointed cones.

This definition is indeed satisfied by the cones we chose, i.e., $\mathcal{W}^{\mathrm{sep}} \times \mathcal{W} \subseteq \mathcal{L}_V \times \mathcal{L}_V$ is indeed pointed and has nonempty interior, as we shall check now. A pointed cone $\mathcal{K}$ is a cone such that $\mathcal{K} \cap (-\mathcal{K}) = \{0\}$. This indeed satisfied for $\mathcal{W}^{\mathrm{sep}} \times \mathcal{W}$, as both cones require their elements to be positive semidefinite, and $W \geqslant 0$ and $-W \geqslant 0$ imply that $W = 0$. To show that $\mathcal{W}^{\mathrm{sep}} \times \mathcal{W}$ has nonempty interior, it is enough[24] to find an operator that belongs to $\mathrm{int}\mathcal{W}^{\mathrm{sep}}$. This is done through the following lemma:

Lemma 7. $\mathbb{1} + \Omega \in \mathrm{int}\mathcal{W}^{\mathrm{sep}}$ *for any* $\Omega \in \mathcal{L}_V$ *such that* $\|\Omega\|_2 < \dfrac{1}{2d_I}$, *where* $d_I = d_{A_I} d_{B_I}$ *and* $\| \cdot \|_2$ *is the Hilbert–Schmidt norm.*

---

[24] Since $\mathrm{int}(\mathcal{W}^{\mathrm{sep}} \times \mathcal{W}) = \mathrm{int}\mathcal{W}^{\mathrm{sep}} \times \mathrm{int}\mathcal{W}$ and $\mathcal{W}^{\mathrm{sep}} \subseteq \mathcal{W}$.

Proof. Since $\Omega \in \mathcal{L}_V$, the discussion in section 2.2.2 implies that the operators

$$\omega^{A \prec B} := {}_{B_O}\Omega, \qquad \omega^{B \prec A} := \Omega - {}_{B_O}\Omega \tag{E.14}$$

are causally ordered (in the sense that they satisfy equation (17) and the analogous relation for the order $B \prec A$, respectively), and so are the operators

$$W^{A \prec B} := \frac{1}{2}\mathbb{1}^\circ + \omega^{A \prec B}, \tag{E.15}$$

$$W^{B \prec A} := \frac{1}{2}\mathbb{1}^\circ + \omega^{B \prec A}. \tag{E.16}$$

Since, furthermore

$$W^{A \prec B} + W^{B \prec A} = \mathbb{1}^\circ + \Omega, \tag{E.17}$$

we have that $\mathbb{1}^\circ + \Omega \in \mathcal{W}^{\text{sep}}$ if $W^{A \prec B}$ and $W^{B \prec A}$ are positive semidefinite. This is the case if

$$\|\omega^{A \prec B}\| \leqslant \frac{1}{2d_I} \quad \text{and} \quad \|\omega^{B \prec A}\| \leqslant \frac{1}{2d_I}, \tag{E.18}$$

where $\|\cdot\|$ is the standard operator norm (i.e., the maximum singular value). To be able to enforce that, first note that $\omega^{A \prec B}$ and $\omega^{B \prec A}$ are orthogonal, and therefore Pythagoras' theorem implies that

$$\|\Omega\|_2^2 = \|\omega^{A \prec B}\|_2^2 + \|\omega^{B \prec A}\|_2^2, \tag{E.19}$$

which implies that

$$\max \left\{ \|\omega^{A \prec B}\|, \|\omega^{B \prec A}\| \right\} \leqslant \max \left\{ \|\omega^{A \prec B}\|_2, \|\omega^{B \prec A}\|_2 \right\} \leqslant \|\Omega\|_2, \tag{E.20}$$

and therefore

$$\|\Omega\|_2 \leqslant \frac{1}{2d_I} \tag{E.21}$$

implies that $\mathbb{1}^\circ + \Omega \in \mathcal{W}^{\text{sep}}$. This in turn implies that the interior of the ball composed of operators $\mathbb{1}^\circ + \Omega$ with Ω satisfying (E.21) belongs to the interior of $\mathcal{W}^{\text{sep}}$, i.e., $\|\Omega\|_2 < \frac{1}{2d_I}$ implies that $\mathbb{1}^\circ + \Omega \in \text{int}\mathcal{W}^{\text{sep}}$.   □

This concludes the proof that problems (38) and (39) are SDP problems dual to each other. We shall now proceed to show that the duality theorem (theorem 4.2.1 in [22]) applies to them:

Theorem 8. *Let* (P), (D) *be a primal-dual pair of conic problems as defined above, and let the pair be such that*

(1) The set of primal solutions $\mathcal{K} \cap (\mathcal{L} + b)$ *intersects* $\text{int}\mathcal{K}$;

(2) The set of dual solutions $\mathcal{K}^* \cap (\mathcal{L}^\perp + c)$ *intersects* $\text{int}\mathcal{K}^*$;

(3) $\langle c, x \rangle$ *is lower bounded for all* $x \in \mathcal{K} \cap (\mathcal{L} + b)$.
Then both the primal and the dual problems are solvable, and the optimal solutions x^* *and* y^* *satisfy the relation*

$$\langle c, b \rangle = \langle c, x^* \rangle + \langle y^*, b \rangle. \tag{E.22}$$

Let us check that for the SDP problems (E.9) and (E.13), the three assumptions of the duality theorem are indeed satisfied.

To see that (1) is satisfied, we need to find $\Omega \in \text{int}\mathcal{W}$ such that $W + \Omega \in \text{int}\mathcal{W}^{\text{sep}}$. Take $\Omega = \lambda\mathbb{1}^\circ$; then $W + \lambda\mathbb{1}^\circ \in \text{int}\mathcal{W}^{\text{sep}}$ iff $\frac{1}{\lambda}W + \mathbb{1}^\circ \in \text{int}\mathcal{W}^{\text{sep}}$. Using lemma 7, we conclude that this is true if

$$\left\| \frac{1}{\lambda}W \right\|_2 < \frac{1}{2d_I}. \tag{E.23}$$

Since

$$\left\| \frac{1}{\lambda}W \right\|_2 \leqslant \frac{1}{\lambda} \|W\|_1 = \frac{d_O}{\lambda} \tag{E.24}$$

it is enough to choose

$$\lambda > 2d_I d_O \tag{E.25}$$

to satisfy inequality (E.23), and we are done.

To see that (2) is satisfied, we need to exhibit a witness S such that $S \in \text{int}\mathcal{S}_V$ and $\mathbb{1}/d_O - S \in \text{int}\mathcal{W}_V^*$. Since the cone $\mathcal{P} \cap \mathcal{L}_V$ of positive semidefinite matrices in $\mathcal{L}_V$ is a full-dimensional subset of both $\mathcal{S}_V$ and $\mathcal{W}_V^*$, it is enough to find an operator S such that $S > 0$ and $\mathbb{1}/d_O - S > 0$. One can take $S = \mathbb{1}/(2d_O)$.

To see that (3) is satisfied, note that $\Omega \geqslant 0$ implies that $\text{tr}\,\Omega/d_O \geqslant 0$.

All in all, the three assumptions of the duality theorem above are thus satisfied. Applying the identity (E.22) to our pair of conic problems, we have, for the optimal solutions Ω^* and S^*:

$$0 = \text{tr}\big[\Omega^*\big]\big/d_O + \text{tr}\big[S^*W\big], \tag{E.26}$$

as claimed in equation (40). As discussed in section 3.2, a value $\text{tr}[\Omega^*]/d_O = -\text{tr}[S^*W] > 0$ guarantees that the process matrix W is causally nonseparable, and the solution S^* of the dual problem provides an explicit causal witness; a value $\text{tr}[\Omega^*]/d_O = 0$ proves that the process matrix W is causally separable, and the primal problem provides a decomposition of W in terms of causally ordered process matrices $W^{A \prec B}$ and $W^{B \prec A}$ (again, this is easier to see in the representation of the primal problem shown in (D.3)).

Appendix F. Measuring causal nonseparability

A causal witness can be used not only to detect the causal nonseparability of a given process, but also to *measure* it. This is analogous to the situation with entanglement witnesses and entanglement measures [39]. First of all, we need to define what we mean by a measure of causal nonseparability. In analogy with the case of entanglement, we suggest that a proper measure of causal nonseparability $\mathcal{N}$ should satisfy the following properties:

Discrimination: $\mathcal{N}(W) \geqslant 0$ for every process matrix W, with $\mathcal{N}(W) = 0$ if and only if W is causally separable.

Convexity: $\mathcal{N}\big(\sum_i p_i W_i\big) \leqslant \sum_i p_i \mathcal{N}(W_i)$ for any process matrices W_i and any $p_i \geqslant 0$, with $\sum_i p_i = 1$.

Monotony: $\mathcal{N}(\$(W)) \leqslant \mathcal{N}(W)$, where $\$(W)$ is any process obtainable from W by composing it with local CPTP maps.

Now we shall prove that both $R_g(W)$ and $R_r(W)$ as defined in equations (42) and (46) respect the properties of *Discrimination* and *Convexity*, whereas $R_g(W)$ respects *Monotony* but $R_r(W)$ does not.

Discrimination follows from the definition of the SDP problems (38), (39) and (44), (45). Note that since they satisfy the assumptions of the duality theorem (8), there are algorithms that actually find the optimal solutions efficiently.

To demonstrate *Convexity*, let us denote by S_W the optimal witness for a given process matrix W; because of its optimality, one has, for any process matrices W_i and any $p_i \geqslant 0$,

$$\text{tr}\big[S_{W_j} W_j\big] \leqslant \text{tr}\bigg[\bigg(S_{\sum_i p_i W_i}\bigg) W_j\bigg] \tag{F.1}$$

and therefore

$$-\text{tr}\bigg[\bigg(S_{\sum_i p_i W_i}\bigg)\sum_i p_i W_i\bigg] \leqslant -\sum_i p_i\,\text{tr}\big[S_{W_i} W_i\big], \tag{F.2}$$

that is, $\mathcal{N}\big(\sum_i p_i W_i\big) \leqslant \sum_i p_i \mathcal{N}(W_i)$.

Now we show that *Monotony* does hold for $R_g(W)$. For that, first we need to define the map $\$(\cdot)$ that composes a process W with local operations. More specifically, the map $\$(\cdot)$ composes a process with the CPTP map M_1^A applied to Alice's input, the CPTP map M_3^A applied to Alice's output, the CPTP map M_1^B applied to Bob's input, and the CPTP map M_3^B applied to Bob's output. We can then define $\$(\cdot)$ as the map such that for all processes W and all CP maps C_2^A and C_2^B we have that

$$\text{tr}\bigg[\big(C_2^A \otimes C_2^B\big) \cdot \$(W)\bigg] = \text{tr}\bigg[\big(C_{123}^A \otimes C_{123}^B\big)W\bigg], \tag{F.3}$$

Figure F1. (a) The situation where the parties share a bipartite process W (in red) and apply the CPTP maps $\mathcal{M}_1^X$ and $\mathcal{M}_3^X$ (in blue) to their inputs and outputs can be equivalently described by (b) a single bipartite process $\$(W)$ (in red).
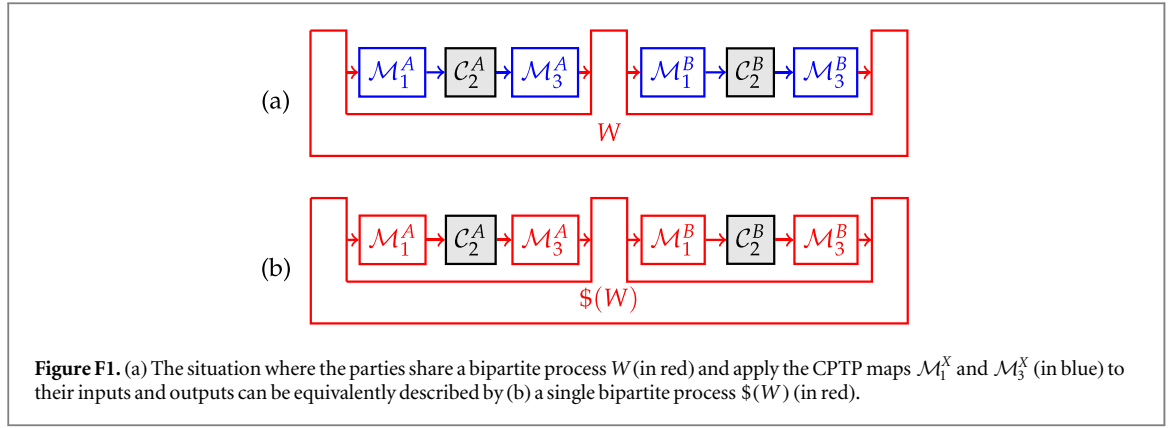
where

$$C_{123}^X := \left[\mathcal{I} \otimes \left(\mathcal{M}_3^X \circ \mathcal{C}_2^X \circ \mathcal{M}_1^X \right) \left(|\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}| \right) \right]^T \tag{F.4}$$

is the CJ operator of the composition of the each party's operations. The processes W and $\$(W)$ are illustrated in figure F1.

It follows from this definition that $\$(W)$ is a valid process. To see this, note that the validity of W implies that the probabilities

$$P\left(\mathcal{C}_{123}^A, \mathcal{C}_{123}^B \right) = \operatorname{tr}\left[\left(C_{123}^A \otimes C_{123}^B \right) W \right] \tag{F.5}$$

are positive and normalized. By definition, these are equal to the probabilities

$$P\left(\mathcal{C}_2^A, \mathcal{C}_2^B \right) = \operatorname{tr}\left[\left(C_2^A \otimes C_2^B \right) \cdot \$(W) \right], \tag{F.6}$$

and the arguments in appendix B show that requiring the probabilities $P(\mathcal{C}_2^A, \mathcal{C}_2^B)$ to be positive and normalized is enough to imply the validity of the process $\$(W)$.

Furthermore, if W is causally separable so is $\$(W)$. This follows from the linearity of $\$(\,\cdot\,)$ and from the fact that $\$(\,\cdot\,)$ preserves the causal order when applied to a causally ordered process, which follows directly from the analogous property for quantum combs [14].

We want to show that for all $\$(\,\cdot\,)$ (i.e., for all CPTP maps M_1^A, M_3^A, M_1^B and M_3^B) and W,

$$R_g(\$(W)) \leqslant R_g(W), \tag{F.7}$$

or equivalently that

$$-\operatorname{tr}\left[\left(S_{\$(W)} \right)\$(W) \right] \leqslant -\operatorname{tr}\left[\left(S_W \right) W \right]. \tag{F.8}$$

By duality, this is equivalent to

$$-\operatorname{tr}\left[\$^*\left(S_{\$(W)} \right) W \right] \leqslant -\operatorname{tr}\left[\left(S_W \right) W \right] \tag{F.9}$$

(where $\$^*$ is the adjoint of $\$$), which follows from the optimality of S_W if $\$^*\left(S_{\$(W)} \right)$ is a valid causal witness that respects the normalization condition for generalized robustness (as defined in SDP problem (38)). Therefore, we need to show it has the two following properties:

$$\operatorname{tr}\left[\$^*\left(S_{\$(W)} \right) W^{\mathrm{sep}} \right] \geqslant 0 \quad \forall\, W^{\mathrm{sep}}, \quad \mathbb{1}\big/d_O - \$^*\left(S_{\$(W)} \right) \in \mathcal{W}^*. \tag{F.10}$$

The first one follows from duality

$$\operatorname{tr}\left[\$^*\left(S_{\$(W)} \right) W^{\mathrm{sep}} \right] = \operatorname{tr}\left[S_{\$(W)} \$\left(W^{\mathrm{sep}} \right) \right] \tag{F.11}$$

and the fact that $\$(W^{\mathrm{sep}})$ is causally separable and $S_{\$(W)}$ is a causal witness.

The second one is equivalent to

$$\operatorname{tr}\left[\left(\mathbb{1}\big/d_O - \$^*\left(S_{\$(W)} \right) \right) \Omega \right] \geqslant 0 \tag{F.12}$$

for every (not necessarily normalized) process matrix Ω. From duality and linearity this is equivalent to

$$\mathrm{tr}\left[S_{\$(W)}\,\$(\Omega)\right] \leqslant \mathrm{tr}[\Omega]/d_O, \tag{F.13}$$

and this follows from the fact that $\$(\,\cdot\,)$ is trace-preserving and that $\mathbb{1}/d_O - S_{\$(W)} \in \mathcal{W}^*$ (which is the normalization condition from the SDP problem (38)).

An analogous proof fails for random robustness, as the adjoint map $\$^*(\,\cdot\,)$ can increase the trace of a witness, and therefore make it fail to satisfy the normalization condition for SDP problem (44). To show that $R_\mathrm{r}(W)$ does not in fact satisfy *Monotony*, it is enough to find a process and local operations such that $R_\mathrm{r}(\$(W)) > R_\mathrm{r}(W)$.

A concrete counterexample can be obtained by considering W_{OCB} and S_{OCB} from section 3.4. Let

$$W_1 = W_{\mathrm{OCB}} \otimes \frac{\mathbb{1}^{A_I{}'}}{2} \tag{F.14}$$

be the process obtained from W_{OCB} by adding a maximally mixed qubit to Alice's input space. Then its random robustness is (up to numerical precision)

$$R_\mathrm{r}\left(W_1\right) = -\mathrm{tr}\, S_{W_1}\, W_1 = \sqrt{2} - 1, \tag{F.15}$$

where

$$S_{W_1} = 2\, S_{\mathrm{OCB}} \otimes |0\rangle\langle 0|^{A_I{}'} \tag{F.16}$$

is its optimal random robustness witness. Now, we can obtain the process

$$\$\left(W_1\right) = W_{\mathrm{OCB}} \otimes |0\rangle\langle 0|^{A_I{}'} \tag{F.17}$$

from W_1 simply by discarding the system in Alice's input space A_I' and replacing it with $|0\rangle\langle 0|$, which is clearly a local operation. Then its random robustness is (up to numerical precision)

$$R_\mathrm{r}\left(\$\left(W_1\right)\right) = -\mathrm{tr}\, S_{\$\left(W_1\right)}\, \$\left(W_1\right) = 2\left(\sqrt{2} - 1\right), \tag{F.18}$$

where $S_{\$(W_1)} = S_{W_1}$. Thus we have shown that

$$R_\mathrm{r}\left(\$\left(W_1\right)\right) > R_\mathrm{r}\left(W_1\right), \tag{F.19}$$

so random robustness is not monotonous under local operations.

Appendix G. Characterization of tripartite causal witnesses

Proof of theorem 3. As defined in section 2.2.3, the cone of tripartite causally separable processes with $d_{C_O} = 1$ is

$$\mathcal{W}_{3C}^{\mathrm{sep}} = \mathrm{conv}\left[\left(\mathcal{P} \cap \mathcal{L}_{A \prec B \prec C}\right) \cup \left(\mathcal{P} \cap \mathcal{L}_{B \prec A \prec C}\right)\right], \tag{G.1}$$

where $\mathcal{L}_{A \prec B \prec C}$ and $\mathcal{L}_{B \prec A \prec C}$ are the linear subspaces defined by the projectors $L_{A \prec B \prec C}$ and $L_{B \prec A \prec C}$. The cone of causal witnesses $\mathcal{S}_{3C}$ is its dual

$$\mathcal{S}_{3C} = \mathcal{W}_{3C}^{\mathrm{sep}*}. \tag{G.2}$$

Using duality relations (C.3) and (C.2) and the fact that the cone of positive semidefinite matrices is self-dual, we get

$$\begin{aligned}
\mathcal{S}_{3C} &= \left(\mathcal{P} \cap \mathcal{L}_{A \prec B \prec C}\right)^* \cap \left(\mathcal{P} \cap \mathcal{L}_{B \prec A \prec C}\right)^* \\
&= \left[\mathrm{conv}\left(\mathcal{P} \cup \mathcal{L}_{A \prec B \prec C}^\perp\right)\right] \cap \left[\mathrm{conv}\left(P \cup \mathcal{L}_{B \prec A \prec C}^\perp\right)\right],
\end{aligned} \tag{G.3}$$

with

$$\mathrm{conv}\left(\mathcal{P} \cup \mathcal{L}_{A \prec B \prec C}^\perp\right) = \left\{S_{ABC}^P + S_{ABC}^\perp \;\middle|\; S_{ABC}^P \geqslant 0,\, L_{A \prec B \prec C}\left(S_{ABC}^\perp\right) = 0\right\} \tag{G.4}$$

and

$$\mathrm{conv}\left(\mathcal{P} \cup \mathcal{L}_{B \prec A \prec C}^\perp\right) = \left\{S_{BAC}^P + S_{BAC}^\perp \;\middle|\; S_{BAC}^P \geqslant 0,\, L_{B \prec A \prec C}\left(S_{BAC}^\perp\right) = 0\right\}. \tag{G.5}$$

$\square$

Appendix H. Optimizing Chiribella's task

We want to optimize the weights $q_{ij}^{[,]}$, $q_{ij}^{\{,\}}$ so as to minimize the maximal probability of success[25] for causally separable processes p_{succ}^{sep}, i.e., we want to minimize the upper bound

$$\mathrm{tr}\left(G_{finite} W^{sep} \right) \leqslant p_{succ}^{sep}. \tag{H.1}$$

This is relevant because, according to equation (91), a lower p_{succ}^{sep} corresponds to a larger resistance to worst-case noise.

To do this, note that $\mathrm{tr}(G_{finite} W) \leqslant p_{succ}^{sep}$ if and only if $\mathrm{tr}[(p_{succ}^{sep} \mathbb{1}/d_O - G_{finite})W] \geqslant 0$. Imposing that this holds for all causally separable processes $W \in \mathcal{W}_{3C}^{sep}$ amounts to imposing that $p_{succ}^{sep} \frac{\mathbb{1}}{d_O} - G_{finite} \in \mathcal{S}_{3C}$, where $\mathcal{S}_{3C}$ is the cone of causal witnesses (characterized through theorem 3).

We are thus led to define the following SDP problem:

$$\begin{aligned}
&\min\ p_{succ}^{sep}\\
&\text{s.t.}\quad p_{succ}^{sep} \mathbb{1}/d_O - G_{finite} \in \mathcal{S}_{3C},\\
&q_{ij}^{[,]} \geqslant 0, \quad q_{ij}^{\{,\}} \geqslant 0, \quad \sum_{i,j=1}^{10} q_{ij}^{[,]} + q_{ij}^{\{,\}} = 1,\\
&q_{ij}^{[,]} = 0 \qquad \forall\ i,j \quad \text{s.t.}\quad \left[U_i, U_j \right] \neq 0,\\
&q_{ij}^{\{,\}} = 0 \qquad \forall\ i,j \quad \text{s.t.}\quad \left\{ U_i, U_j \right\} \neq 0,
\end{aligned} \tag{H.2}$$

where in order to keep the interpretation of the task as guessing whether the unitaries commute or anticommute, we imposed that $q_{ij}^{[,]} = 0$ for non-commuting U_i, U_j and $q_{ij}^{\{,\}} = 0$ for non-anticommuting U_i, U_j.

Solving this problem numerically, we found

$$p_{succ}^{sep} \approx 0.8690 \tag{H.3}$$

We omit the optimal $q_{ij}^{[,]}$, $q_{ij}^{\{,\}}$ for brevity.

References

[1] Deutsch D 1989 Quantum computational networks *Proc. R. Soc. London* A **425** 73–90
[2] Hardy L 2009 Foliable operational structures for general probabilistic theories *Deep Beauty: Understanding the Quantum World through Mathematical Innovation* ed H Halvorson (Cambridge: Cambridge University Press) p 409
[3] Coecke B 2010 Quantum picturalism *Contemp. Phys.* **51** 59–83
[4] Chiribella G, D'Ariano G M and Perinotti P 2010 Probabilistic theories with purification *Phys. Rev.* A **81** 062348
[5] Chiribella G, D'Ariano G M and Perinotti P 2011 Informational derivation of quantum theory *Phys. Rev.* A **84** 012311
[6] Chiribella G, D'Ariano G M, Perinotti P and Valiron B 2013 Quantum computations without definite causal structure *Phys. Rev.* A **88** 022318
[7] Chiribella G 2012 Perfect discrimination of no-signalling channels via quantum superposition of causal structures *Phys. Rev.* A **86** 040301
[8] Araújo M, Costa F and Brukner Č 2014 Computational advantage from quantum-controlled ordering of gates *Phys. Rev. Lett.* **113** 250402
[9] Procopio L M, Moqanaki A, Araújo M, Costa F, Calafell I A, Dowd E G, Hamel D R, Rozema L A, Brukner Č and Walther P 2015 Experimental superposition of orders of quantum gates *Nat. Commun.* **6** 7913
[10] Hardy L 2007 Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure *J. Phys. A: Math. Theor.* **40** 3081–99
[11] Rovelli C 1990 Quantum mechanics without time: a model *Phys. Rev.* D **42** 2638–46
[12] Ashtekar A 1996 Large quantum gravity effects: unforeseen limitations of the classical theory *Phys. Rev. Lett.* **77** 4864–7
[13] Oreshkov O, Costa F and Brukner Č 2012 Quantum correlations with no causal order *Nat. Commun.* **3** 1092
[14] Chiribella G, D'Ariano G M and Perinotti P 2009 Theoretical framework for quantum networks *Phys. Rev.* A **80** 022339
[15] Nielsen M and Chuang I 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[16] Davies E and Lewis J 1970 An operational approach to quantum probability *Commun. Math. Phys.* **17** 239–60
[17] Życzkowski K and Bengtsson I 2006 *Geometry of Quantum States* (Cambridge: Cambridge University Press)
[18] Gutoski G and Watrous J 2006 Toward a general theory of quantum games *Proc. 39th ACM STOC* pp 565–74 (arXiv:quant-ph/0611234)
[19] Oreshkov O and Giarmatzi C 2015 Causal and causally separable processes (arXiv:1506.05449)
[20] Gühne O and Tóth G 2009 Entanglement detection *Phys. Rep.* **474** 1–75
[21] Rockafellar R T 1970 *Convex Analysis* (Princeton, NJ: Princeton University Press)
[22] Nesterov Y and Nemirovskii A 1987 *Interior Point Polynomial algorithms in Convex Programming* (*Stud. Appl. Math.* vol 32) (Philadelphia: SIAM)
[23] Steiner M 2003 Generalized robustness of entanglement *Phys. Rev.* A **67** 054305
[24] Vidal G and Tarrach R 1999 Robustness of entanglement *Phys. Rev.* A **59** 141

---

[25] Remember that the probability of success for the quantum switch is always equal to one.

[25] Lofberg J 2004 YALMIP : a toolbox for modeling and optimization in MATLAB *Proc. CACSD Conf.* (*Taipei, Taiwan*)(http://users.isy.liu.se/johanl/yalmip)

[26] The MOSEK optimization toolbox for MATLAB http://mosek.com/

[27] Brunner N, Cavalcanti D, Pironio S, Scarani V and Wehner S 2014 Bell nonlocality *Rev. Mod. Phys.* **86** 419–78

[28] Brukner C 2015 Bounding quantum correlations with indefinite causal order *New J. Phys.* **17** 083034

[29] Baumeler Ä and Wolf S 2014 Perfect signaling among three parties violating predefined causal order *2014 IEEE Int. Symp. on Information Theory* (*ISIT*) pp 526–30 (arXiv:quant-ph/1312.5916)

[30] Baumeler Ä, Feix A and Wolf S 2014 Maximal incompatibility of locally classical behavior and global causal order in multi-party scenarios *Phys. Rev.* A **90** 042106

[31] Branciard C, Araújo M, Feix A, Costa F and Brunker Č 2015 The simplest causal inequalities and their violation (arXiv:1508.01704)

[32] Gisin N 1991 Bell's inequality holds for all non-product states *Phys. Lett.* A **154** 201–2

[33] Zych M, Costa F, Pikovski I and Brukner Č 2015 Bell inequalities for temporal order (in preparation)

[34] Gurvits L 2003 Classical deterministic complexity of Edmonds' Problem and quantum entanglement *Proc. 35th Annual ACM Symp. on Theory of Computing* (New York: ACM) pp 10–19

[35] Royer A 1991 Wigner function in Liouville space: a canonical formalism *Phys. Rev.* A **43** 44–56

[36] Braunstein S L, D'Ariano G M, Milburn G J and Sacchi M F 2000 Universal teleportation with a twist *Phys. Rev. Lett.* **84** 3486–9

[37] Choi M-D 1975 Completely positive linear maps on complex matrices *Linear Algebr. Appl.* **10** 285–90

[38] Barnum H, Fuchs C A, Renes J M and Wilce A 2005 Influence-free states on compound quantum systems (arXiv:quant-ph/0507108)

[39] Brandão F G S L 2005 Quantifying entanglement with witness operators *Phys. Rev.* A **72** 022310