universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Quantum computation with indefinite causal structures“

verfasst von / submitted by

Mateus de Araújo Santos

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2016 / Vienna 2016

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 796 605 411
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Physik
Betreut von / Supervisor:	Univ.-Prof. Dr. Časlav Brukner

Contents

Listen to your crazy laugh
Before you hang a right
And disappear from sight

Belle & Sebastian

Agradecimentos

Despite appearances to the contrary, physicists don't actually like to work in a frictionless vacuum, and I'm no exception. The work done in my PhD was only possible and pleasant due to the collaboration and support of several people. First of all, I'd like to thank Časlav, for making the University of Vienna the most chilled and stimulating research environment I know. His suggestions and objections were fundamental in obtaining the results inside and outside this thesis. Eu também gostaria de agradecer os meus antigos orientadores, Marcelo Terra Cunha e Reinaldo Oliveira Vianna, por me preparar para estar aqui desde o tempo em que eu nem sabia o que era um Hamiltoniano (não que algum apareça nessa tese =). E também ao Dutty, por me contar que tinha um grupo de pesquisa massa em Viena que procurava um aluno de doutorado.

Special thanks to my friends and collaborators Fabio Costa, Igor Pikovski, and Adrien Feix, for the countless nights we spent in the office working and playing video games.

Ao meu grande amigo Marco Túlio, com quem eu tenho o prazer de colaborar em projetos difíceis. Infelizmente projetos que são realmente difíceis frequentemente falham. Ainda vamos arrancar um resultado foda segurando B.

To Cyril Branciard, for his steely determination in getting to the bottom of every technical and not-so-technical issue. I'm sure our (sometimes overheated) discussions made our work better in the end.

To Harvey Brown, for showing me how valuable philosophy is in elucidating the most confusing physical questions.

Für Daniela, dafür dass du mich glücklich machst.

Last but not least, I'd like to thank the funding bodies that made this PhD possible by paying me a salary: the John Templeton Foundation, the CoQuS doctoral program, the Austrian Science Fund, and the EU project RAQUEL.

Abstract

This thesis explores which causal relations are allowed by the laws of quantum information processing. It is motivated by the conundrum that although we expect any theory that unifies gravity and quantum mechanics to feature indefinite causal structures due to quantum uncertainty in the metric, we currently have poor understanding of the notion of an indefinite causal structure. The tool we use to approach this problem is the process matrix, which encodes the most general way that some parties that respect quantum mechanics can signal to each other without creating logical contradictions such as the grandfather paradox. This generality allows process matrices to encode both definite and indefinite causal orders, without passing judgement on their physical plausibility.

One aspect of the thesis is to develop technical tools to study types of causal relations within the process matrix formalism. Using semidefinite programming we define witnesses – analogous to entanglement witnesses – that can distinguish between process matrices that encode definite and indefinite causal orders. We also use semidefinite programming and polyhedral computation to systematize the study of causal inequalities which, analogously to Bell inequalities, allow one to certify in a device-independent way that a given causal structure is indefinite. These tools allows us to find the simplest possible causal inequalities and to develop algorithms to find process matrices and quantum operations that violate them.

Another aspect of the thesis is to investigate which process matrices encode physical causal structures and which might turn out to be mathematical artefacts. We do this in two ways: the first is to identify which causal structures are compatible with the reversibility of physical laws and which are not. Taking the consensus view that all physical laws are ultimately reversible allows us to identify as unphysical the causal structures that are fundamentally irreversible. The other way we investigate their physicality is from the point of view of quantum computation: if they turn out to be too powerful, it is a reason to be suspicious of their plausibility. We find that even though process matrices do provide an asymptotic advantage in query complexity for a specific problem, we can prove that they are weaker than other models of quantum computation based on closed timelike curves, and thus more physically plausible.

Zusammenfassung

In dieser Dissertation wird untersucht, welche Kausalbeziehungen mit den Gesetzen der Quanteninformationsverarbeitung kompatibel sind. Sie ist durch die Erwartung motiviert, dass jede Theorie, die Gravitation und Quantenmechanik vereint, unbestimmte kausale Strukturen beinhaltet, die durch die quantenmechanische Unschärfe der Metrik hervorgerufen wird. Das gegenwärtige Verständnis solcher unbestimmter kausaler Strukturen ist jedoch noch sehr unvollständig. Als Werkzeug zu ihrer näheren Untersuchung verwenden wir die sogenannte "process matrix", eine Matrix die die allgemeinste Möglichkeit der Signalübertragung zwischen Parteien enthält, die nicht zu logischen Paradoxien (wie etwa dem Großvaterparadoxon) führt. Diese Allgemeinheit erlaubt es, sowohl bestimmte als auch unbestimmte kausale Strukturen zu beschreiben, ohne a priori ihre physikalische Plausibilität zu beurteilen.

Der erste Aspekt der Doktorarbeit ist die Entwicklung eines theoretischen Instrumentariums zur Untersuchung von kausalen Beziehungen im "process matrix"-Formalismus. Mithilfe von Semidefiniter Programmierung können wir Zeugenoperatoren definieren – analog zu Zeugenoperatoren für Verschränkung – die zwischen "process matrices" bestimmter und unbestimmter kausaler Ordnung unterscheiden können. Wir verwenden auch Semidefiniter Programmierung und polyhedrale Programmierung um systematisch kausale Ungleichungen zu untersuchen, die, wie Bell-Ungleichungen für Verschränkung, die unbestimmte kausale Struktur nur aufgrund der statistischen Korrelationen nachzuweisen erlaubt. Diese Herangehensweise erlaubt es uns, die einfachsten kausalen Ungleichungen herzuleiten und Algorithmen zu entwickeln, um "process matrices" und quantenmechanische Operationen zu finden, welche diese verletzen.

Der zweite Aspekt der Dissertation betrifft die Unterscheidung derjenigen "process matrices", die physikalisch relevant sind von denen, die vielleicht nur mathematische Artefakte sind. Dafür haben wir zwei Ansätze entwickelt. Der erste Ansatz identifiziert kausale Strukturen, die mit der Reversibilität von physikalischen Gesetzen vereinbar sind. Unter der Annahme, dass alle physikalischen Gesetze prinzipiell reversibel sind, können grundlegend irreversible kausale Strukturen als unphysikalisch betrachtet werden. Der zweite Ansatz betrachtet kausale Strukturen aus der Sicht der Theorie der Quantencomputer: Wenn jene zu mächtig wären, hätte man Grund, ihre physikalische Plausibilität in Zweifel zu ziehen. Obwohl einige "process matrices", für ein spezifisches Problem, eine asymptotische Reduktion der Anzahl der Queries erlauben, sind sie doch weniger mächtig als andere Modelle von Quantencomputern, die auf geschlossenen zeitartigen Weltlinien beruhen und deshalb physikalisch plausibler.

Preamble

We want to understand which causal relations are allowed by the laws of quantum mechanics. This is motivated by the realization that any future theory that unifies quantum mechanics with gravity will display quantum uncertainty in the metric tensor, which should lead to some sort of coherent superposition of different causal structures. Since we currently lack a concrete theory that displays such indefinite causal structures, we cannot study them from a more traditional, bottom-up approach, that would analyse their properties as they appear in the theory. Instead we shall use a top-down approach, that uses general principles of quantum mechanics to constrain which causal structures are possible, and examines their information-processing capabilities while abstracting away physical details such as particles and Hamiltonians.

This is, broadly speaking, the perspective of quantum information. Accordingly, we shall focus on the capabilities of indefinite causal structures to do communication and, most crucially, quantum computation. The focus on quantum computation has two motivations: one is that a new quantum algorithm that uses indefinite causal structures to solve a practical problem would be of even technological interest. Another is that if some indefinite causal structures are found to be implausibly powerful, this can be used as evidence that they are actually mathematical pathologies and not causal structures that we should expect to find in Nature.

We start in Chapter 1 by analysing the computational power of the most basic example of a indefinite causal structure – the quantum switch – first introduced in Ref. [1]. It can be understood informally as a coherent superposition of different causal orders, or, more precisely, a physical setup where a quantum system controls the order of signalling between some parties. It is an ideal starting point, as the quantum switch not only is sure to exist in Nature (in fact, it has already been implemented experimentally [2]), but also is simple enough that we can define it without needing first to develop a general formalism to describe indefinite causal structures. In this Chapter we find that the quantum switch allows a specific computational problem to be solved with $O(n)$ queries, whereas any quantum circuit that has no access to indefinite causal orders is likely to require $\Omega(n^2)$ queries to solve the same problem. We furthermore prove that this is the maximal advantage that the quantum switch can provide, as $\Theta(n^2)$ queries are sufficient to simulate it exactly.

To gain a deeper understanding of indefinite causality, we want to connect the promising results shown in Chapter 1 with a general formalism capable of describing it. The one we use is the process matrix framework, introduced in Ref. [3]. Its central object – the process matrix – is defined as the most general way of transmitting information between some isolated laboratories that obey quantum mechanics, that still does not generate any logical paradoxes that might arise from

that, such as the grandfather paradox. There are then two kinds of process matrices: the first, vanilla kind, describes situations where the communication between the laboratories happens in a well-defined causal order (or an incoherent mixture of well-defined causal orders), and is called causally separable. The second kind describes situations where this is not the case, and is called causally nonseparable. That these causally nonseparable process matrices do describe, in a strong sense, indefinite causality, is proven via the violation of causal inequalities. These are device-independent inequalities, that take into account only the correlations produced by laboratories, and are derived from the assumption that there exists a well-defined past, and that you cannot signal to it.

In Chapter 2 we show then how to write the quantum switch in the process matrix formalism, and prove that despite being causally nonseparable, it cannot be used to violate any causal inequality. We show, however, that it can violate a weaker, device-dependent notion of definite causality: a causal witness. In the case of the quantum switch we find a particularly nice witness, that can be understood as a game of guessing whether some unitaries commute or anti-commute. In the general case we do not try to formulate witnesses as games, but we can prove, using semidefinite programming, that for any causally nonseparable process matrix there exists a causal witness that can detect it. Furthermore, these witness are always implementable experimentally, independently of whether the process matrix itself is, and we have in fact already implemented one of them [4].

Chapter 3 is dedicated to the systematic study of causal inequalities. We show that, analogously to Bell inequalities, causal inequalities can be defined as the facets of the polytope of causal correlations. Characterizing these polytopes allows us therefore to find out which are the simplest possible causal inequalities, which turn out to be variants of a game where one party has to guess the input of the other. We show also that these inequalities can be violated with process matrices, through a variant of the see-saw algorithm augmented with semidefinite programming. This algorithm also allows us to make a tentative characterization of the set of correlations that can be generated by process matrices.

At this point the technical aspects of the process matrix formalism were pretty well-understood, but its connection with actual physics was not well explored. In fact, the only causally nonseparable process matrix that was sure to be physical was the quantum switch; the other process matrices remained as mathematical constructs with unknown physical status. They were, after all, defined only from the requirement of not generating logical contradictions, and it would be unsurprising if it turned out that Nature demand more than mere logical consistency.

Exploring the consequences of a possible such demand is the goal of Chapter 4: inspired by the successful use of the principle of reversibility in the reconstructions of quantum mechanics, we ask which process matrices are compatible with the idea that transformations between quantum states must be reversible. To do that, we need to rework the definition of a process matrix: instead of connecting only the isolated laboratories between themselves, it also connects them with a global past and a global future. We can then define a “pure” process matrix as one that induces a reversible transformation from the global past to the global future whenever all laboratories implement reversible transformations. Then the precise question about whether a given process matrix is compatible with reversibility is about whether it can be extended into a pure process matrix, that is, whether it can be purified. We find that several of the previously introduced process matrices are not purifiable, including the ones from Ref. [3] and Chapter 3, and thus are likely to be unphysical.

The reversibility principle is, however, not enough to decide the question of physicality for all process matrices, as there are several which are purifiable but we don't know how to implement experimentally, including some that can violate causal inequalities. To gather additional evidence about their plausibility, we return to the main theme of this thesis, and investigate their computational power.

This is the subject of Chapter 5. To be able to study the computational power of process matrices as a whole, as opposed to particular cases as done in Chapter 1, we need to define a proper computational model that uses them. It turns out that redefinition of process matrices done in Chapter 4 is well-suited to this task, as giving process matrices a global past and a global future allows us to use them as subroutines in a normal quantum circuit, and we can as such define their computational model as an extension of the circuit model.

We can then show that this computational model is a simple particular case of quantum circuits extended with closed timelike curves: it consists precisely of the circuits made with *linear* closed timelike curves. Since their nonlinearity is the source of major pathologies, such as the ability to solve NP-complete problems and to distinguish non-orthogonal states, one can expect process matrices to be free of such problems. We still do not have, however, fully characterized their computational power.

Chapter 1

Computational advantage from quantum-controlled ordering of gates

This chapter was published as

- M. Araújo, F. Costa and Č. Brukner.
  'Computational Advantage from Quantum-Controlled Ordering of Gates'.
  Phys. Rev. Lett. 113 250402 (2014). arXiv:1401.8127 [quant-ph].

Abstract

It is usually assumed that a quantum computation is performed by applying gates in a specific order. One can relax this assumption by allowing a control quantum system to switch the order in which the gates are applied. This provides a more general kind of quantum computing, that allows transformations on blackbox quantum gates that are impossible in a circuit with fixed order. Here we show that this model of quantum computing is physically realizable, by proposing an interferometric setup that can implement such a quantum control of the order between the gates. We show that this new resource provides a reduction in computational complexity: we propose a problem that can be solved using $O(n)$ blackbox queries, whereas the best known quantum algorithm with fixed order between the gates requires $O(n^2)$ queries. Furthermore, we conjecture that solving this problem in a classical computer takes exponential time, which may be of independent interest.

1.1 Introduction

A useful tool to calculate the complexity of a quantum algorithm is the blackbox model of quantum computation. In this model, the input to the computation is encoded in a unitary gate – treated as a blackbox – and the complexity of the algorithm is the number of times this gate has to be queried to solve the problem.

Typically, black-box computation is studied within the quantum circuit formalism [5]. A quantum circuit consists of a collection of wires, representing quantum systems, that connect boxes, representing unitary transformations. In this framework, wires are assumed to connect the various gates in a fixed structure, thus the

order in which the gates are applied is determined in advance and independently of the input states. It was first proposed in [1] that such a constraint can be relaxed: one can consider situations where the wires, and thus the order between gates, can be controlled by some extra variable. This is natural if one thinks of the circuit's wires as quantum systems that can be in superposition.

Such "superpositions of orders" allow performing information-theoretical tasks that are impossible in the quantum circuit model: it was shown in [6] that it is possible to decide whether a pair of blackbox unitaries commute or anticommute with a single use of each unitary, whereas in a circuit with a fixed order at least one of the unitaries must be used twice. (The same task was considered in a quantum optics context in [7], where a less efficient protocol was found.)

It was not known, however, whether this advantage can be translated into more efficient algorithms for quantum computing, i.e., if a quantum computer that can control the order between gates can solve a computational problem with asymptotically less resources than a quantum computer with fixed circuit structure.

Here we present such a problem: given a set of n unitary matrices and the promise that they satisfy one out of $n!$ specific properties, find which property is satisfied. The essential resource to solve this problem is the quantum control over the order of n blackboxes, first introduced in Ref. [8]. We show that, by using this resource, the problem can be solved with $O(n)$ queries to the blackboxes, while the best known algorithm with fixed order requires $O(n^2)$ queries. Furthermore, while both quantum methods of solving the problem run in polynomial time, the best known classical algorithm to solve it runs in exponential time, which may be of independent interest.

We further discuss a possible interferometric implementation of the protocol. For the superposition of the order of just two gates, a realization with current quantum optics techniques is possible. For a higher number of gates practical implementations become more challenging.

1.2 Algorithm

The quantum control of the order between n unitary gates can be formalized by introducing the n-SWITCH gate. As in Ref. [8], we consider a d-dimensional target system, initialized in some state $|\psi\rangle$, and an $n!$-dimensional control system. Let $\{U_i\}_0^{n-1}$ be a set of unitaries and

$$\Pi_x = U_{\sigma_x(n-1)} \dots U_{\sigma_x(1)} U_{\sigma_x(0)} \tag{1.1}$$

for some permutation σ_x, where $x = 0, \dots, n! - 1$ is a chosen labelling of permutations[1]. Then the n-SWITCH S_n is a controlled quantum gate: its effect is to apply the product of unitaries Π_x to the state $|\psi\rangle$ conditioned on the value of the control register $|x\rangle$. In symbols,

$$S_n|x\rangle|\psi\rangle = |x\rangle\Pi_x|\psi\rangle. \tag{1.2}$$

Using this gate we can introduce an algorithm that exploits the quantum control of orders to achieve a reduction in query complexity for the solution of a specific problem. The algorithm is based on the standard Hadamard test. The idea is to initiate the control system in a state corresponding to a uniform superposition of all permutations, apply S_n, and then measure the control system in the Fourier basis.

[1] More precisely, $\sigma_x(j)$ is the image of the jth element under the permutation x

With a suitable choice of the unitaries, we can make the result of this measurement deterministic and, since there are $n!$ different results, this means that we can differentiate between $n!$ different properties of n unitaries.

To be more precise, let $\omega = e^{i\frac{2\pi}{n!}}$. We say that the set of unitaries $\{U_i\}_0^{n-1}$ has property $\mathbf{P}_y$ if it is true that

$$\forall x \; \Pi_x = \omega^{xy}\Pi_0, \tag{1.3}$$

for the given y. For example, property $\mathbf{P}_0$ is the property that $\Pi_x = \Pi_0$ for all x, i.e., that all the matrices commute with each other.

Note that it is not possible to satisfy property $\mathbf{P}_1$ if the dimension of the unitaries d is less than $n!$. To see that, consider $x = y = 1$, and take the determinant on both sides of equation (1.3):

$$\det \Pi_1 = \omega^d \det \Pi_0. \tag{1.4}$$

Since $\det \Pi_x = \det \Pi_0$, it follows that $\omega^d = 1$, and therefore d must be at least $n!$.

The computational problem is defined as follows: given a set $\{U_i\}_0^{n-1}$ of unitary matrices of dimension $d \geq n!$, decide which of the properties $\mathbf{P}_y$ is satisfied by this set, given the promise that one of these $n!$ properties is satisfied.

The protocol for solving this problem is the following: we initialize the target system in *any* state $|\psi\rangle$, and the control system in the state $|C\rangle$ which corresponds to an equal superposition of all permutations:

$$|C\rangle|\psi\rangle = \frac{1}{\sqrt{n!}} \sum_{x=0}^{n!-1} |x\rangle|\psi\rangle. \tag{1.5}$$

Then, we apply the n-SWITCH:

$$S_n|C\rangle|\psi\rangle = \frac{1}{\sqrt{n!}} \sum_{x=0}^{n!-1} |x\rangle\Pi_x|\psi\rangle. \tag{1.6}$$

Now we apply the Fourier transform over $\mathbf{Z}_{n!}$ to our control qudit

$$\mathcal{F}_{n!}S_n|C\rangle|\psi\rangle = \frac{1}{n!} \sum_{x,s=0}^{n!-1} |s\rangle\omega^{-xs}\Pi_x|\psi\rangle. \tag{1.7}$$

and measure the control qudit in the computational basis, with outcome probabilities

$$p_s = \frac{1}{n!^2}\left\| \sum_{x=0}^{n!-1} \omega^{-xs}\Pi_x|\psi\rangle \right\|^2. \tag{1.8}$$

Using the promise $\Pi_x = \omega^{xy}\Pi_0$ we get that

$$p_s = \frac{1}{n!^2}\left\| \sum_{x=0}^{n!-1} \omega^{x(y-s)}\Pi_0|\psi\rangle \right\|^2 = \delta_{sy}, \tag{1.9}$$

that is, if property $\mathbf{P}_y$ is true, the result of the measurement is going to be y with probability one, so we can find out which property the unitaries have in a single run of the protocol.

We should notice that the problem is not trivial, i.e. there exist, for every n, infinitely many sets of unitary matrices that satisfy each of the $n!$ properties $\mathbf{P}_y$ (see Appendix A) The problem, and the corresponding protocol, can be also modified to tolerate possible experimental error. This modification is shown in Appendix B.

1.3 Query Complexity

We are interested in determining the number of times that the unitaries U_i must be used to run the algorithm. Clearly this depends only on the implementation of the n-SWITCH gate, since the unitaries are not used anywhere else. As proposed in [1], the SWITCH can in principle be implemented by adding quantum control to the connections between the unitaries. In such an implementation it is sufficient to use a single copy of each unitary, while the control system determines the order in which the target system passes through the unitaries.

Since the implementation with quantum control of the connections between gates is explicitly outside the quantum circuit formalism, we cannot simply calculate the number of uses of the unitaries by counting the number of times they appear in a circuit. Nevertheless, we can formulate the notion of "gate uses" in a precise, operational, way. Imagine we append, to each gate, an additional "flag" quantum system that counts the number of times that gate is used. This can be done in a reversible way: the j-th flag is initialized in the state $|0\rangle_j$ and, whenever the unitary U_j is used, it is updated through the unitary transformation $|f\rangle_j \to |f+1\rangle_j$. It is easy to see that, after applying the n-SWITCH, the state of the flags factorizes, with each flag in the state $|1\rangle_j$. According to this definition, the total number of queries necessary to run the algorithm is n.

In comparison, the optimal simulation of the n-SWITCH gate with a fixed circuit has query complexity $\Omega(n^2)$ [2]. To see this, first note that one can assume without loss of generality that all blackbox unitaries are applied each in a different time step, since if two blackboxes are applied in parallel, we can always introduce a time delay between them, without changing the action of the circuit. More technically, a circuit defines a partial order for its gates, which can always be completed into a total order. Then, let $\{A_i\}_0^{m-1}$ be the m blackbox unitaries appearing in the circuit, with $A_j \in \{U_i\}_0^{n-1}$, queried in the order $A_0 \preceq \cdots \preceq A_{m-1}$. From the Appendix B of Ref. [1], it follows that it is only possible to apply the Π_x to $|\psi\rangle$ if the unitaries $U_{\sigma_x(0)}, \ldots, U_{\sigma_x(n-1)}$ are present in the circuit in the order defined by σ_x. The lower bound on the query complexity is then the minimal m for which all $n!$ permutations of $\{U_0, \ldots, U_{n-1}\}$ are present as subsequences of the sequence $\{A_0, \ldots, A_{m-1}\}$.

It turns out that this is an open problem in combinatorics [9, 10]. However, it is known that the optimal m respects the bounds

$$n^2 - C_\epsilon n^{7/4+\epsilon} \leq m \leq \left\lceil n^2 - \frac{7}{3}n + \frac{19}{3} \right\rceil$$

for any $\epsilon > 0$, where C_ϵ is a constant that depends on ϵ. This concludes the proof.

It is also possible to construct a quantum circuit that simulates the n-SWITCH gate from such a sequence. We shall, however, refrain from doing so. Instead, for completeness, we present a simple circuit that simulates the n-SWITCH gate using $m = n^2$ queries in the Appendix C.

Of course, it might not be necessary to use the n-SWITCH gate in order to determine which property $\mathbf{P}_y$ the unitaries satisfy. For example, it is possible to solve the problem by directly measuring the phase obtained when applying the permutation σ_1. Since $\Pi_1 = \omega^y \Pi_0$, this is sufficient to determine y. However, this

[2] We say that $f(n)$ is $\Omega(g(n))$ if there exists a constant M such that $f(n) \geq Mg(n)$ for sufficiently large n.

protocol can work only if the relative phase is measured with an error smaller than $\frac{2\pi}{n!}$, and for blackbox unitaries this can only be done with an exponential amount of queries.

This is the case, for example, for Kitaev's phase estimation algorithm [11]. This algorithm is not usually applied to blackbox unitaries, but this can be done using the techniques in [12–14]. In this case, to calculate the phase with the required $O(n \log n)$ bits of precision, one would need to implement the matrices controlled-$U_i^{2^k}$, with $k = 1, \ldots, n \log n$, which would require an exponential amount of queries to the blackboxes U_i. Even if one assumes that it is possible to apply controlled-$U_i^{2^k}$ efficiently – a necessary assumption to make Kitaev's algorithm efficient – one would need $O(n \log n)$ queries to each $U_i^{2^k}$ oracle. Since there are n unitaries U_i, the query complexity would be $O(n^2 \log n)$, which is still less efficient than simulating the n-SWITCH with a fixed circuit.

1.4 Running time

Instead of query complexity we may want to consider the running time of the algorithm. If we assume that this is dominated by applying the unitaries U_i, then there is no difference between the implementation with superposition of orders or the fixed quantum circuit: both run in time $O(n)$ (see Appendix C).

It is interesting, nevertheless, to compare the time required to solve the problem between quantum and classical computers. If we assume that the unitaries U_i are decomposed in a polynomial amount of elementary gates, they can be given as an input of polynomial size to a classical algorithm, and it makes sense to compare the classical and quantum running times.

As argued above, the problem of determining y reduces to the problem of calculating the relative phase between Π_1 and Π_0, which may differ by the permutation of a single pair of unitaries. However, as discussed before, the dimension of the unitaries must be at least $n!$ for this problem to be nontrivial, and it seems unlikely that one could extract the phase from these exponentially large unitary matrices on a classical computer in polynomial time. On the other hand, the running time of the quantum algorithm is clearly polynomial for unitaries decomposed in a polynomial amount of elementary gates. Therefore we conjecture that for the problem presented there is an exponential separation between classical and quantum complexity, which may be of independent interest.

Note that our algorithm is based on the quantum Fourier transform, as are several algorithms that show an exponential separation between classical and quantum complexity, but there does not appear to be a more direct connection with specific classes of quantum algorithms, such as those that solve the hidden subgroup problem (see Appendix D).

1.5 Physical implementation

In Ref. [1] it was proposed to apply the superposition principle to the physical components of a quantum computer that determine the order between gates. Since this requires a quantum control over macroscopic systems, it seems outside of the reach of current technology and could be practically unfeasible. Here we propose an implementation of the n-SWITCH that, although experimentally challenging, might be feasible.

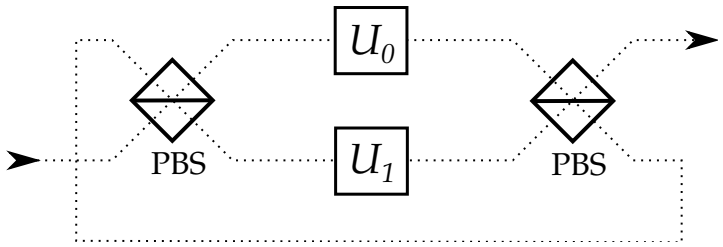


Figure 1.1: Linear optical implementation of the 2-SWITCH. The unitaries U_0 and U_1 act on internal degrees of freedom of a single photon, such as space bins, time bins, or angular momentum modes. The polarization state of the photon determines the order in which the unitaries are applied. A photon with polarization $|H\rangle$ is transmitted by the polarizing beam splitters (PBSs), so that U_0 is applied before U_1. For a photon with polarization $|V\rangle$, reflected by the PBSs, U_1 is applied first and U_0 second.

We first consider an optical implementation of a 2-SWITCH for 2×2 unitaries, illustrated in Fig. 1.1 (this implementation was independently developed in [15]). The control system is the polarization of a photon and the target system some internal degree of freedom of the same photon, such as space bins, time bins, or angular momentum modes. If the photon is prepared in a horizontally polarized state $|H\rangle$, it is transmitted by both polarizing beam splitters (PBSs), resulting in the application of the unitary U_0 first and of U_1 second. A photon in a vertically polarized state $|V\rangle$ is reflected by both PBSs, thus the two unitaries are applied in the reversed order. For an arbitrary polarization state $\alpha|H\rangle+\beta|V\rangle$, the photon exits the interferometer in the state $\alpha|H\rangle U_1 U_0|\psi\rangle + \beta|V\rangle U_0 U_1|\psi\rangle$, which corresponds to the output of the 2-SWITCH.

The extension of this scheme to the general case of an n-SWITCH can be obtained with a generalization of the PBS to an element, which we call n-ROUTER, with n input modes and n output modes (see Fig. 1.2). If the control system is in a state $|x\rangle$, the n-ROUTER sends the input mode j to the output mode $\sigma_x(j)$. The unitary $U_{\sigma_x(j)}$ is applied to a system in the mode $\sigma_x(j)$, which then enters a second router that performs the inverse permutation. The output mode j of the second router is then directed to the input mode $j+1$ of the first one. It is straightforward to check that a system entering mode 0 of the first router in the state $|x\rangle|\psi\rangle$ exits mode $n-1$ of the second router in the state $|x\rangle\Pi_x|\psi\rangle$. (In Appendix E we show how to construct an n-ROUTER with $O(n^2)$ binary routers.)

This higher-dimensional routing can be achieved, for example, with orbital angular momentum of light [16, 17]. However, the main limitation of an optical implementation of the n-ROUTER is that it is not scalable in an obvious way, since it requires encoding an exponential number of degrees of freedom in a single photon ($n!$ for the control system and, as argued before, at least $n!$ for the target system). A scalable implementation could be obtained by encoding the degrees of freedom in $O(n \log n)$ particles, each carrying a constant number of degrees of freedom (e.g., one qubit each). The main challenge is then to implement a router that, conditioned on the multiparticle state, coherently directs all the particles in a specific mode. This is in principle possible if the particles are bound together, e.g. as atoms in a molecule. Recent progress in matter-wave interferometry suggests that such a

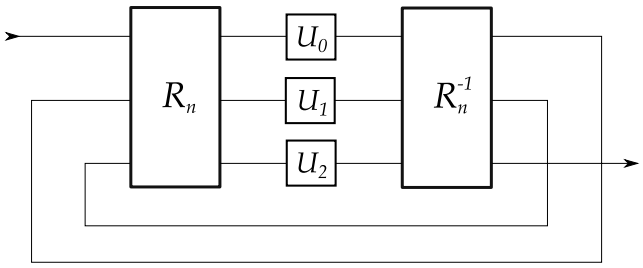


Figure 1.2: Implementation of the n-SWITCH (in the figure, $n = 3$). Both control and target, in state $|x\rangle$ and $|\psi\rangle$ respectively, are encoded in a single (possibly multi-particle) system. When the system enters the n-ROUTERs (R_n) in mode j, it is redirected to mode $\sigma_x(j)$ and the unitary $U_{\sigma_x(j)}$ is applied to $|\psi\rangle$. R_n^{-1} performs the inverse permutation and sends the system to mode $j+1$ of the first router. In this way, a system entering mode 0 of the first router, eventually exits mode $n-1$ of the second router with target in the state $\Pi_x|\psi\rangle$.

quantum control of composite systems could be achievable in the future [18, 19].

Other realizations of superposition of orders, based on different models of computation, could also be possible. For example, an implementation of the 2-SWITCH within adiabatic quantum computing was proposed recently [20].

1.6 Conclusion

We have shown that extending the quantum circuit model by allowing quantum control of the order between gates provides a reduction in the number of queries needed to solve a computational problem. Furthermore, we have proposed a physically realizable experimental scheme to implement such a control.

While the reduction is only polynomial, and thus does not create a new complexity class, the result shows that extending the quantum circuit model is possible and can provide a computational advantage. Besides, the computational problem introduced has no known efficient solution by a classical algorithm, which may be of independent interest.

Other extensions of the quantum circuit model of blackbox computation have been proposed [14]: it was shown recently [14, 21, 22] that a quantum circuit cannot apply blackbox gates conditioned on the state of a control qubit. However, such a control is physically realizable [12, 13] and therefore should be allowed by the formalism. It is also intriguing to ask what computational advantages might be achieved once the restrictions imposed by the fixed causal structure of quantum mechanics are relaxed [3].

Chapter 2

Witnessing causal nonseparability

This chapter was published as

- M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi and Č. Brukner. 'Witnessing causal nonseparability'. New J. Phys. 17 102001 (2015). arXiv:1506.03776 [quant-ph].

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.

2.1 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 [23], Fig. 2.1(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 [24–27].

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 [1]. 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 [6, 28], but they are also physically realizable and a first experimental proof-of-principle has been recently demonstrated [2]. 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 [29–31].

It is therefore natural to ask what the most general resources allowed by quantum mechanics beyond the circuit model are. In Ref. [3] 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, Fig. 2.1(b).

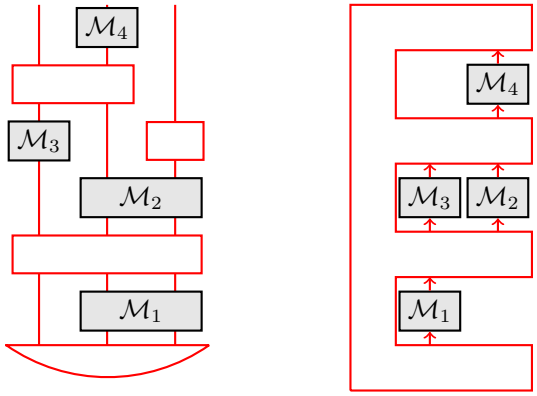


Figure 2.1: (a) If the operations $\mathcal{M}_i$ of local agents are performed in a definite causal sequence, they can be represented as gates in a circuit, where information flows from bottom to top. (b) A process matrix formalizes a resource in which the order between operations may not be fixed. 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 [3].

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 [3]. 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 Refs. [6, 28] include additional assumptions, as 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 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 Ref. [6] 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 2.2, we review the process matrix formalism, giving a convenient characterization of general and causally separable process matrices for the cases of interest. In Section 2.3, 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 2.4 we formalise the quantum switch as a

process matrix. We proceed to prove its causal nonseparability in Section 2.5, through the use of causal witnesses. One such witness is the task proposed in Ref. [6], that we optimize to increase its resistance to noise. Finally, we clarify in Section 2.6 the link between causal witnesses and causal inequalities and show that the quantum switch cannot violate any causal inequality.

2.2 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 Ref. [3]. This extends the "comb" formalism of Ref. [32], 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.

2.2.1 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 $\mathcal{M}^A : A_I \to A_O$ [33], 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* [34], which is defined as the collection $\mathcal{J}^A = \left\{\mathcal{M}_a^A\right\}_{a=1}^m$ 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 POVM (positive operator-valued measure) to include the transformations applied to the system; it reduces to a POVM for 1-dimensional output spaces. When the choice of operation is described by a classical variable x, we will express such a dependence explicitly as $\mathcal{J}_x^A = \left\{\mathcal{M}_{a|x}^A\right\}_{a=1}^m$.

A convenient representation of CP maps is given by the Choi-Jamiołkowski (CJ) isomorphism [35]. For a CP map $\mathcal{M}_a^A : A_I \to A_O$, its corresponding CJ matrix is

defined here as

$$M_a^{A_I A_O} := \left[\mathcal{I} \otimes \mathcal{M}_a^A(|\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|)\right]^T \in A_I \otimes A_O, \tag{2.1}$$

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 F.1. A map is completely positive if and only if its CJ representation is positive semidefinite, while the trace-preserving condition is equivalent to $\mathrm{tr}_{A_O} M^{A_I A_O} = \mathbb{1}^{A_I}$ (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 equivalently represented as a set

$$\left\{M_a^{A_I A_O}\right\}_{a=1}^m, \quad M_a^{A_I A_O} \geq 0, \quad \mathrm{tr}_{A_O} \sum_{a=1}^m M_a^{A_I A_O} = \mathbb{1}^{A_I}. \tag{2.2}$$

2.2.2 Process matrices

As discussed in Ref. [3], 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 $P\big(\mathcal{M}^{A^1}_{a_1|x_1}, \ldots, \mathcal{M}^{A^N}_{a_N|x_N}\big)$ of the corresponding CP maps $\mathcal{M}^{A^1}_{a_1|x_1}, \ldots, \mathcal{M}^{A^N}_{a_N|x_N}$. Using the CJ representation, it was shown that these probabilities can then be expressed as

$$\begin{aligned} &P\big(M^{A^1}_{a_1|x_1}, \ldots, M^{A^N}_{a_N|x_N}\big) \\ &\quad = \mathrm{tr}\Big[\Big(M^{A_I^1 A_O^1}_{a_1|x_1} \otimes \ldots \otimes M^{A_I^N A_O^N}_{a_N|x_N}\Big) W\Big], \end{aligned} \tag{2.3}$$

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 1 – 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 G, these conditions are equivalent to

$$W \geq 0, \tag{2.4}$$
$$\mathrm{tr}\, W = d_O, \tag{2.5}$$
$$W = L_V(W), \tag{2.6}$$

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 G. We will denote the closed convex cone of non-normalized processes defined by (2.4) and (2.6) by $\mathcal{W}$.

In the case of two parties A (Alice) and B (Bob), see Figure 2.2, these conditions on $W \in A_I \otimes A_O \otimes B_I \otimes B_O$ reduce to

$$W \geq 0\,, \tag{2.7}$$
$$\mathrm{tr}\, W = d_O\,, \tag{2.8}$$
$${}_{B_I B_O} W = {}_{A_O B_I B_O} W\,, \tag{2.9}$$
$${}_{A_I A_O} W = {}_{A_I A_O B_O} W\,, \tag{2.10}$$
$$W = {}_{B_O} W + {}_{A_O} W - {}_{A_O B_O} W\,, \tag{2.11}$$

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

$$ {}_X W = \frac{\mathbb{1}^X}{d_X} \otimes \mathrm{tr}_X W \,. \tag{2.12}$$

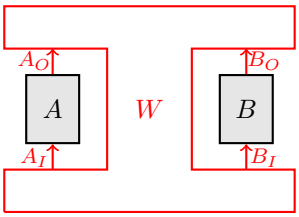


Figure 2.2: Representation of a bipartite process matrix W, connecting Alice's (A_O) and Bob's (B_O) output systems to their input systems (A_I and B_I).

Non-signalling and 1-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 F.2.

The first case is described by process matrices W satisfying

$$ W = {}_{A_O^1 \ldots A_O^N} W = \rho^{A_I^1 \ldots A_I^N} \otimes \mathbb{1}^{A_O^1 \ldots A_O^N}, \tag{2.13}$$

where $\rho^{A_I^1 \ldots A_I^N}$ is a density matrix representing an ordinary quantum state. In this case, the probability rule (2.3) reduces to the standard Born rule

$$ P\big(M^{A^1}_{a_1|x_1}, \ldots, M^{A^N}_{a_N|x_N}\big) = \mathrm{tr}\Big[\Big(E^{A_I^1}_{a_1|x_1} \otimes \ldots \otimes E^{A_I^N}_{a_N|x_N}\Big)\rho\Big], \tag{2.14}$$

where $E^{A_I^i}_{a_i|x_i} := \mathrm{tr}_{A_O^i}\, M^{A_I^i A_O^i}_{a_i|x_i}$ are POVM elements.

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

$$\begin{aligned} W &= {}_{A_O^N} W, \\ {}_{A_I^N A_O^N} W &= {}_{A_O^{N-1} A_I^N A_O^N} W, \\ &\vdots \\ {}_{A_I^2 A_O^2 \ldots\, A_I^N A_O^N} W &= {}_{A_O^1 A_I^2 A_O^2 \ldots\, A_I^N A_O^N} W \,. \end{aligned} \tag{2.15}$$

These conditions, first found in [32, 36], 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

$$ W = W^{A^1 \prec A^2 \prec \ldots \prec A^N}. \tag{2.16}$$

Process matrices of this form (and the obvious permutations) are called *causally ordered*. As shown in Refs. [32, 36], they correspond to standard (causally ordered) quantum circuits and can be implemented as quantum channels with memory between the parties.

Bipartite causally separable processes

According to Eq. (2.15), 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 Eq. (2.9) 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 $W^{A\prec B} \geq 0$, $\mathrm{tr}\, W^{A\prec B} = d_O$,

$$W^{A\prec B} \in \mathcal{L}_V \quad \text{and} \quad W^{A\prec B} = {}_{B_O}W^{A\prec B}. \tag{2.17}$$

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 Eq. (2.13).

Following Ref. [3], 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

$$W^{\text{sep}} = q\, W^{A\prec B} + (1-q)\, W^{B\prec A}, \tag{2.18}$$

with $0 \leq q \leq 1$. 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 (2.18) is called *causally nonseparable*.

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 2.3). 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 (2.15),

$$W^{A\prec B\prec C} = {}_{C_O}W^{A\prec B\prec C}, \tag{2.19}$$

$${}_{C_I C_O}W^{A\prec B\prec C} = {}_{B_O C_I C_O}W^{A\prec B\prec C}, \tag{2.20}$$

$${}_{B_I B_O C_I C_O}W^{A\prec B\prec C} = {}_{A_O B_I B_O C_I C_O}W^{A\prec B\prec C}, \tag{2.21}$$

and

$$W^{B\prec A\prec C} = {}_{C_O}W^{B\prec A\prec C}, \tag{2.22}$$

$${}_{C_I C_O}W^{B\prec A\prec C} = {}_{A_O C_I C_O}W^{B\prec A\prec C}, \tag{2.23}$$

$${}_{A_I A_O C_I C_O}W^{B\prec A\prec C} = {}_{B_O A_I A_O C_I C_O}W^{B\prec A\prec C}. \tag{2.24}$$

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

$$W^{A\prec B\prec C} = L_{A\prec B\prec C}(W^{A\prec B\prec C}), \tag{2.25}$$

$$W^{B\prec A\prec C} = L_{B\prec A\prec C}(W^{B\prec A\prec C}), \tag{2.26}$$

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

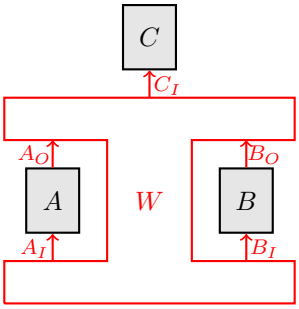


Figure 2.3: Representation of a tripartite process matrix W where one party has trivial output $d_{C_O} = 1$. It can be seen as connecting Alice's (A_O) and Bob's (B_O) output systems to Alice, Bob and Charlie's input systems A_I, B_I and C_I.

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

$$W^{\text{sep}} = q\, W^{A\prec B\prec C} + (1-q)\, W^{B\prec A\prec C}, \tag{2.27}$$

with $0 \leq q \leq 1$. Ignoring the normalization constraint, this defines a convex cone $\mathcal{W}_{3C}^{\text{sep}}$. We will use this definition in Section 2.5 to show that a recently introduced tripartite quantum resource, which yields information-processing advantages with respect to causally ordered processes [6, 28], 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 wouldn't 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 will be presented in an upcoming work [37].

2.3 Causal witnesses

2.3.1 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 [38], we call a hermitian operator S a *causal witness* (or *witness*, simply) if[1]

$$\text{tr}[S\, W^{\text{sep}}] \geq 0 \tag{2.28}$$

for every causally separable process matrix W^{sep}. This definition is motivated by the separating hyperplane theorem [39]: 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 $\text{tr}[S_{W_{\text{ns}}} W_{\text{ns}}] < 0$.

[1] Note that the bound 0 and the sign of the inequality are arbitrary; we choose them as in Eq. (2.28) for mathematical convenience.

To construct a witness for a given 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 (2.28) is equivalent to

$$\mathrm{tr}[S\,W^{A\prec B}] \geq 0 \quad \forall\, W^{A\prec B}\,, \tag{2.29a}$$

$$\mathrm{tr}[S\,W^{B\prec A}] \geq 0 \quad \forall\, W^{B\prec A}\,. \tag{2.29b}$$

Let us focus on condition (2.29a). Using Eq. (2.17) 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 (2.29a) is equivalent to

$$\mathrm{tr}[S({}_{B_O}W)] \geq 0 \quad \forall\, W \in \mathcal{L}_V,\ W \geq 0\,. \tag{2.30}$$

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

$$\mathrm{tr}[S({}_{B_O}W)] = \mathrm{tr}[({}_{B_O}S)\,W], \tag{2.31}$$

and it is sufficient that ${}_{BO}S \geq 0$ for the right-hand-side to be non-negative for all valid W. An analogous argument shows that ${}_{A_O}S \geq 0$ is sufficient to satisfy condition (2.29b). We conclude that for S to be a causal witness, it is sufficient that

$${}_{B_O}S \geq 0 \quad \text{and} \quad {}_{A_O}S \geq 0. \tag{2.32}$$

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 $\mathrm{tr}[(S+S^\perp)W] = \mathrm{tr}[SW]$ 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*

$$S = S_P + S^\perp, \tag{2.33}$$

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

$${}_{B_O}S_P \geq 0, \quad {}_{A_O}S_P \geq 0, \quad L_V(S^\perp) = 0\,. \tag{2.34}$$

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

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

$$S^\perp = L_V(S_P) - S_P, \tag{2.35}$$

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 $S = L_V(S_P)$, ${}_{B_O}S_P \geq 0$, and ${}_{A_O}S_P \geq 0$.*

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 2.5.2 (for the tripartite case), do not belong to $\mathcal{S}_V$, which is why we use the more general definition that witnesses belong to $\mathcal{S}$.

2.3.2 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 semidefinite programming (SDP) [40]. 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 $\mathrm{tr}[S\,W]$ over the cone of causal witnesses[2] $\mathcal{S}_V$, and check whether we obtain a negative value or not. Note that in order to make $\mathrm{tr}[S\,W]$ lower bounded (to avoid getting a value $-\infty$ for causally nonseparable process matrices) a normalisation constraint on the witnesses has to be imposed. This normalisation is arbitrary – any constraint that makes $\mathcal{S}_V$ compact suffices – and different normalisation choices give rise to different interpretations for the value of $\mathrm{tr}[S\,W]$. We shall normalise the witnesses by imposing that $\mathrm{tr}[S\,\Omega] \leq 1$ for every (normalised) process matrix Ω, for $-\mathrm{tr}[S\,W]$ 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 normalisation as a conic constraint. To do so, we extend the constraint $\mathrm{tr}[S\,\Omega] \leq 1$ to non-normalised process matrices by linearity:

$$\mathrm{tr}[S\,\Omega] \leq \mathrm{tr}[\Omega]/d_O, \tag{2.36}$$

which is equivalent to

$$\mathrm{tr}[(\mathbb{1}/d_O - S)\,\Omega] \geq 0 \tag{2.37}$$

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:

$$\begin{aligned} &\min \mathrm{tr}[SW] \\ \text{s.t.} \quad &S \in \mathcal{S}_V, \quad \mathbb{1}/d_O - S \ \in \mathcal{W}_V^*, \end{aligned} \tag{2.38}$$

which is written explicitly in terms of positive semidefinite constraints in Appendix I.

If the solution of the SDP problem (2.38) leads to a negative expectation value of S, one can conclude that W is causally nonseparable, since SDP algorithms can be guaranteed[3] to find the optimal solution [40]. In such a case, the optimal solution S^* provides an explicit witness to verify the causal nonseparability of W.

[2] In principle minimizing over $\mathcal{S}$ instead of $\mathcal{S}_V$ would lead to the same value for $\mathrm{tr}[S\,W]$, but this causes technical problems as explained in Appendix J.

[3] When the assumptions of the Duality Theorem (13) are satisfied, which is the case for our SDP problems, as proven in Appendix J.

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 (2.38) (this can be seen explicitly from the representation of the SDP problem (2.39) given in Appendix I). As shown in Appendix J, this dual is

$$\begin{aligned} &\min \mathrm{tr}[\Omega]/d_O \\ \text{s.t.} \quad & W + \Omega \in \mathcal{W}^{\text{sep}}, \quad \Omega \in \mathcal{W}, \end{aligned} \tag{2.39}$$

where $\mathcal{W}^{\text{sep}}$ is the cone of non-normalized causally separable process matrices, as previously defined. Furthermore, the optimal value $\mathrm{tr}[\Omega^*]/d_O$ of problem (2.39) is related to the optimal value $\mathrm{tr}[S^*W]$ of problem (2.38) through

$$\mathrm{tr}[\Omega^*]/d_O = -\mathrm{tr}[S^*W]. \tag{2.40}$$

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

$$\frac{1}{1+\lambda}\left(W + \lambda \widetilde{\Omega}\right) \tag{2.41}$$

is causally separable, optimized over all valid, normalised 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 generalised robustness, which quantifies the resistance of the entanglement of a quantum state to worst-case noise [41]. 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 K. For this reason, we define the *generalised robustness* of a process W as

$$R_{\mathrm{g}}(W) = -\mathrm{tr}(S^*W). \tag{2.42}$$

Again in analogy with the case of entanglement measures, one can also define the *random robustness* [42] 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:

$$\mathbb{1}^\circ := \frac{\mathbb{1}}{d_{A_I} d_{B_I}}. \tag{2.43}$$

The optimal witness with respect to random robustness can be found by solving an SDP problem analogous to (2.38):

$$\begin{aligned} &\min \mathrm{tr}(SW) \\ \text{s.t.} \quad & S \in \mathcal{S}_V\,, \quad \mathrm{tr}(S\mathbb{1}^\circ) \leq 1\,, \end{aligned} \tag{2.44}$$

whose dual is

$$\begin{aligned} &\min \lambda \\ \text{s.t.} \quad & \lambda \geq 0\,, \quad W + \lambda \mathbb{1}^\circ \in \mathcal{W}^{\text{sep}}\,, \end{aligned} \tag{2.45}$$

and random robustness itself is defined as

$$R_{\mathrm{r}}(W) = -\mathrm{tr}(S^*W)\,, \tag{2.46}$$

where $\mathrm{tr}(S^*W)$ is now the optimal value of the problem (2.44). This quantity can be used to compare witnesses in scenarios where white noise is an appropriate noise model, however, it cannot be interpreted as a proper measure of causal nonseparability, as it does not respect all the axioms we propose in appendix K – more specifically, it is not monotonous under local operations.

A geometrical interpretation of the results of this section is shown in Figure 2.4.

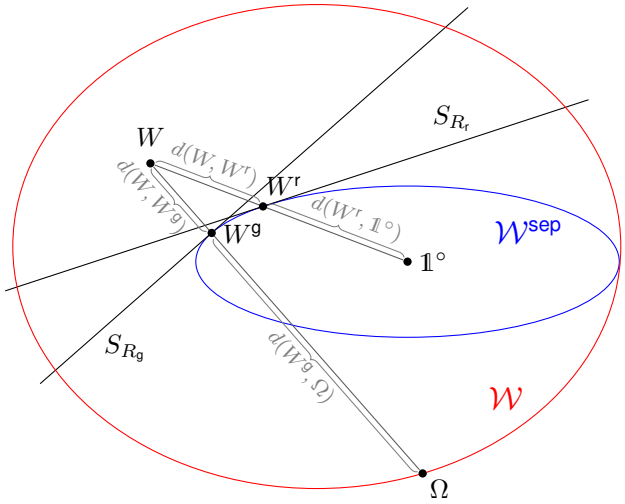


Figure 2.4: Here we schematically represent the set of normalised process matrices in $\mathcal{W}$ by the red ellipse and the set of normalised causally separable processes in $\mathcal{W}^{\mathsf{sep}}$ by the blue ellipse. Since the latter set is closed and convex, any causally nonseparable process 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, $S_{R_{\mathsf{g}}}$ and $S_{R_{\mathsf{r}}}$, that represent two different ways to quantify how far W is from being causally separable. $-\operatorname{tr}(S_{R_{\mathsf{g}}}W)$ measures the generalised robustness of W, which is its resistance to the worst-case noise Ω. Geometrically, the generalised robustness of W is given by the ratio of distances $d(W, W^{\mathsf{g}})/d(W^{\mathsf{g}}, \Omega)$, where W^{g} is the causally separable process closest to W on the depicted line. In its turn, $-\operatorname{tr}(S_{R_{\mathsf{r}}}W)$, the random robustness of W, is its resistance to the "white noise" $\mathbb{1}^\circ$. Geometrically, it is given by analogous ratio $d(W, W^{\mathsf{r}})/d(W^{\mathsf{r}}, \mathbb{1}^\circ)$, where W^{r} is again the causally separable process closest to W on the depicted line. $S_{R_{\mathsf{g}}}$ and $S_{R_{\mathsf{r}}}$ are the optimal solutions of the SDP problems (2.38) and (2.44), respectively.

2.3.3 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 $\operatorname{tr}[S\,W]$ – 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[4]

$$S = \sum_{x,y,a,b} \gamma_{x,y,a,b}\, M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O}\,, \tag{2.47}$$

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 Choi-Jamiołkowski representation of CP trace non-increasing maps (see Section 2.2).

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

$$\mathrm{tr}[S\,W] \;\; = \sum_{x,y,a,b} \gamma_{x,y,a,b}\; \mathrm{tr}\left[\left(M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O}\right) W\right], \tag{2.48}$$

where according to the generalized Born rule (2.3), 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 $\mathrm{tr}[S\,W]$ 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 Eq. (2.48).

The decomposition (2.47) 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.

2.3.4 Example

Let us now illustrate the above considerations on an explicit example. Ref. [3] introduced the following process matrix, for a case where all incoming and outgoing systems of A and B are 2-dimensional (qubit) systems (i.e., $d_{A_I} = d_{A_O} = d_{B_I} = d_{B_O} = 2$):

$$W_{\text{OCB}} = \frac{1}{4}\left[\mathbb{1} + \frac{\mathbb{1}^{A_I} Z^{A_O} Z^{B_I} \mathbb{1}^{B_O} + Z^{A_I} \mathbb{1}^{A_O} X^{B_I} Z^{B_O}}{\sqrt{2}}\right], \tag{2.49}$$

where Z and X are the Pauli matrices, and tensor products are implicit. One can easily check that $W_{\text{OCB}} \geq 0$, that $\mathrm{tr}[W_{\text{OCB}}] = 4 = d_O$, and that W_{OCB} satisfies Eqs. (2.9)–(2.11), 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 2.6), which implies that it is causally nonseparable.

[4] In the decomposition (2.47), 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 ${}_{A_O}\left(\sum_a M_{a|x}^{A_I A_O}\right) \leq \mathbb{1}^{A_I A_O}/d_{A_O}$ and ${}_{B_O}\left(\sum_b M_{b|y}^{B_I B_O}\right) \leq \mathbb{1}^{B_I B_O}/d_{B_O}$ 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 ${}_{A_O}\left(\sum_a M_{a|x}^{A_I A_O}\right) = \mathbb{1}^{A_I A_O}/d_{A_O}$ and ${}_{B_O}\left(\sum_b M_{b|y}^{B_I B_O}\right) = \mathbb{1}^{B_I B_O}/d_{B_O}$, 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.

The concept of causal witnesses introduced here allows us to prove the causal nonseparability of W_{OCB} more directly. Solving the SDP problem (2.44) with YALMIP [43] and the solver MOSEK [44], we obtained, up to numerical precision, the optimal witness with respect to random robustness

$$S_{\text{OCB}} = \frac{1}{4}\left[\mathbb{1} - \left(\mathbb{1}^{A_I} Z^{A_O} Z^{B_I} \mathbb{1}^{B_O} + Z^{A_I} \mathbb{1}^{A_O} X^{B_I} Z^{B_O}\right)\right]. \tag{2.50}$$

Applying it to W_{OCB}, we find that $-\operatorname{tr}[S_{\text{OCB}}\, W_{\text{OCB}}] = R_{\text{r}}(W_{\text{OCB}}) = \sqrt{2} - 1 > 0$ (where $R_{\text{r}}(W_{\text{OCB}})$ is the random robustness as defined in Equation (2.46)). This proves that W_{OCB} is causally nonseparable.

This also implies that the process matrices of the form

$$W_{\text{OCB}}(\lambda) = \frac{1}{1+\lambda}\left(W_{\text{OCB}} + \lambda\, \mathbb{1}^{\circ}\right), \tag{2.51}$$

are causally nonseparable for

$$\lambda < R_{\text{r}}(W_{\text{OCB}}) = \sqrt{2} - 1 \tag{2.52}$$

(and their causal nonseparability is then witnessed by S_{OCB}). For $\lambda \geq \sqrt{2} - 1$, $W_{\text{OCB}}(\lambda)$ is causally separable; the solution of the SDP problem (2.45) provides an explicit decomposition for $W_{\text{OCB}}(R_{\text{r}}(W_{\text{OCB}}))$ (as can be seen when writing (2.45) in a form similar to Eq. (I.6)), from which we can derive an explicit decomposition for all $W_{\text{OCB}}(\lambda)$ for $\lambda \geq \sqrt{2} - 1$, as

$$W_{\text{OCB}}(\lambda) = \frac{1}{2} W_{\text{OCB}}^{A \prec B}(\lambda) + \frac{1}{2} W_{\text{OCB}}^{B \prec A}(\lambda), \tag{2.53}$$

where

$$W_{\text{OCB}}^{A \prec B}(\lambda) := \frac{1}{4}\left[\mathbb{1} + \frac{\sqrt{2}}{1+\lambda}\, \mathbb{1}^{A_I} Z^{A_O} Z^{B_I} \mathbb{1}^{B_O}\right], \tag{2.54}$$

$$W_{\text{OCB}}^{B \prec A}(\lambda) := \frac{1}{4}\left[\mathbb{1} + \frac{\sqrt{2}}{1+\lambda}\, Z^{A_I} \mathbb{1}^{A_O} X^{B_I} Z^{B_O}\right] \tag{2.55}$$

are causally ordered process matrices. (Note that for $\lambda < \sqrt{2} - 1$, $W_{\text{OCB}}^{A \prec B}(\lambda)$ and $W_{\text{OCB}}^{B \prec A}(\lambda)$ as defined above would not be positive semidefinite, which explains why Eq. (2.53) then fails to provide a valid causally separable decomposition of $W_{\text{OCB}}(\lambda)$.)

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

$$M_{a|x}^{A_I A_O} := \left(\frac{\mathbb{1} + (-1)^a Z}{2}\right)^{A_I} \otimes \left(\frac{\mathbb{1} + (-1)^x Z}{2}\right)^{A_O}, \tag{2.56}$$

$$M_{b|y,y'=0}^{B_I B_O} := \left(\frac{\mathbb{1} + (-1)^b X}{2}\right)^{B_I} \otimes \left(\frac{\mathbb{1} + (-1)^{y+b} Z}{2}\right)^{B_O}, \tag{2.57}$$

$$M_{b|y,y'=1}^{B_I B_O} := \left(\frac{\mathbb{1} + (-1)^b Z}{2}\right)^{B_I} \otimes \frac{\mathbb{1}^{B_O}}{2}, \tag{2.58}$$

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

$$S_{\text{OCB}} = 3 \cdot \mathbb{1}^{\circ} - 4\, G_{\text{OCB}} \tag{2.59}$$

with

$$G_{\text{OCB}} = \frac{1}{8} \sum_{x,y,a,b} \left[\delta_{a,y}\, M_{a|x}^{A_I A_O} \otimes M_{b|y,y'=0}^{B_I B_O} + \delta_{b,x}\, M_{a|x}^{A_I A_O} \otimes M_{b|y,y'=1}^{B_I B_O} \right], \quad (2.60)$$

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

$$\begin{aligned} \mathrm{tr}[S_{\text{OCB}} \cdot W] &= 3 - 4\ \mathrm{tr}[G_{\text{OCB}} \cdot W] \\ &= 3 - 4 \cdot \frac{1}{8} \sum_{x,y,a,b} \left[\delta_{a,y}\, P\big(M_{a|x}^{A_I A_O}, M_{b|y,y'=0}^{B_I B_O}\big) + \delta_{b,x}\, P\big(M_{a|x}^{A_I A_O}, M_{b|y,y'=1}^{B_I B_O}\big) \right]. \end{aligned} \quad (2.61)$$

As one may recognize, the choice of CP maps in (2.56)–(2.58) is the same[5] as that considered in Ref. [3], so that the experimental procedure proposed here to measure the witness S_{OCB} would be the same as that suggested in [3] 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 ${}_{A_O}\big(\sum_a M_{a|x}^{A_I A_O}\big) = \mathbb{1}^{A_I A_O}/d_{A_O}$ and ${}_{B_O}\big(\sum_b M_{b|y,y'}^{B_I B_O}\big) = \mathbb{1}^{B_I B_O}/d_{B_O}$ for all x, y, y'). As it turns out, in the causal inequality of Ref. [3] the probabilities $P(a, b|x, y, y') = P\big(M_{a|x}^{A_I A_O}, M_{b|y,y'}^{B_I B_O}\big)$ are actually combined in precisely the same way as above – namely, $\mathrm{tr}[G_{\text{OCB}} \cdot W]$ above can be identified with the probability p_{succ} of winning the corresponding "causal game",

$$p_{\text{succ}} = \frac{1}{2}[P(a = y|y' = 0) + P(b = x|y' = 1)], \quad (2.62)$$

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 (2.62) coincide, i.e., $\mathrm{tr}[S_{\text{OCB}} \cdot W] \geq 0$ if and only if $p_{\text{succ}} = \mathrm{tr}[G_{\text{OCB}} \cdot W] \leq 3/4$, where $3/4$ is the upper bound on p_{succ} for any causal correlation (as defined in Section 2.6 below). Furthermore, the noise threshold below which the noisy process matrix $W_{\text{OCB}}(\lambda)$ (2.53) can violate the causal inequality is the same as the threshold $R_{\text{r}}(W_{\text{OCB}})$ below which $W_{\text{OCB}}(\lambda)$ is causally nonseparable, as already noted in Ref. [46]. 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 [45], 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 2.6 below, with an explicit example in the tripartite case.

[5] Note that compared to Ref. [3], 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 [45]. Furthermore, in [3] the state sent out by B when $y' = 1$ was arbitrary, while here we fixed it to be $\mathbb{1}^{B_O}/2$.

2.4 Quantum control of causal order

2.4.1 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 Ref. [1]. 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

$$V(U_A, U_B) = |0\rangle\langle 0| \otimes U_B U_A + |1\rangle\langle 1| \otimes U_A U_B, \tag{2.63}$$

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 $\frac{1}{\sqrt{2}}\big(|0\rangle \otimes U_B U_A|\psi\rangle + |1\rangle \otimes U_A U_B|\psi\rangle\big)$, which can be interpreted as having applied the two unitaries on the target in a "superposition of orders"[6].

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

$$\frac{1}{2}\Big(U_B U_A|\psi\rangle\langle\psi|U_A^\dagger U_B^\dagger + U_A U_B|\psi\rangle\langle\psi|U_B^\dagger U_A^\dagger\Big). \tag{2.64}$$

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 (2.27)) using quantum control of causal order.

2.4.2 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

$$d_{A_I} = d_{A_O} = d_{B_I} = d_{B_O} = 2, \quad d_{C_I} = 4, \quad d_{C_O} = 1. \tag{2.65}$$

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 F.2. An

[6] 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.

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 F.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

$$|w\rangle = \frac{1}{\sqrt{2}} \Big(|\psi\rangle^{A_I} |\mathbb{1}\rangle\!\rangle^{A_O B_I} |\mathbb{1}\rangle\!\rangle^{B_O C_I^t} |0\rangle^{C_I^c} + |\psi\rangle^{B_I} |\mathbb{1}\rangle\!\rangle^{B_O A_I} |\mathbb{1}\rangle\!\rangle^{A_O C_I^t} |1\rangle^{C_I^c} \Big). \tag{2.66}$$

This can be checked by noting that

$$\langle\!\langle U_A^*|^{A_I A_O} \langle\!\langle U_B^*|^{B_I B_O} \cdot |w\rangle = \frac{1}{\sqrt{2}} \Big(|0\rangle^{C_I^c} \otimes \big(U_B U_A |\psi\rangle\big)^{C_I^t} + |1\rangle^{C_I^c} \otimes \big(U_A U_B |\psi\rangle\big)^{C_I^t} \Big), \tag{2.67}$$

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 F.1. Note that the resulting state is the same as that obtained by applying (2.63) to an initial state $\frac{|0\rangle+|1\rangle}{\sqrt{2}} \otimes |\psi\rangle$.

Note that the process (2.66) itself is clearly causally nonseparable[7], 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 Eq. (2.66), 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

$$\mathrm{tr}_{C_I} |w\rangle\langle w| = \mathrm{tr}_{C_I^c C_I^t} |w\rangle\langle w| = \frac{1}{2} W^{A \prec B} + \frac{1}{2} W^{B \prec A}, \tag{2.68}$$

where

$$W^{A \prec B} = |\psi\rangle\langle\psi|^{A_I} \otimes |\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|^{A_O B_I} \otimes \mathbb{1}^{B_O}, \tag{2.69}$$

$$W^{B \prec A} = |\psi\rangle\langle\psi|^{B_I} \otimes |\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|^{B_O A_I} \otimes \mathbb{1}^{A_O}, \tag{2.70}$$

are (bipartite) causally ordered process matrices; $\mathrm{tr}_{C_I} |w\rangle\langle w|$ indeed describes the situation of Eq. (2.64).

For some information-processing tasks, the quantum switch is known to provide an advantage over causally ordered processes [6, 28], 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 (2.66) over the target system:

$$W_{\text{switch}} = \mathrm{tr}_{C_I^t} |w\rangle\langle w|. \tag{2.71}$$

Note that the proof of causal nonseparability based on the purity of the switch does not extend to the reduced switch (2.71), since it is not an extremal process. We will

[7] Note that the results of Ref. [1] 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.

therefore use the framework of causal witnesses to show that the reduced switch is also causally nonseparable.

2.5 Witnesses for the quantum switch

Since the quantum switch is a tripartite process where $d_{C_O} = 1$, we can use definition (2.27) to study its causal (non)separability. In this tripartite situation, we will define causal witnesses to be the hermitian operators S such that

$$\mathrm{tr}[S\, W^{\mathsf{sep}}] \geq 0 \tag{2.72}$$

for every causally separable processes W^{sep} in the cone $\mathcal{W}_{3C}^{\mathsf{sep}}$. The set of causal witnesses is thus the cone dual to $\mathcal{W}_{3C}^{\mathsf{sep}}$, 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*

$$S = S^P_{ABC} + S^{\perp}_{ABC} = S^P_{BAC} + S^{\perp}_{BAC}, \tag{2.73}$$

where

$$S^P_{ABC} \geq 0, \quad L_{A\prec B\prec C}(S^{\perp}_{ABC}) = 0, \tag{2.74}$$
$$S^P_{BAC} \geq 0, \quad L_{B\prec A\prec C}(S^{\perp}_{BAC}) = 0, \tag{2.75}$$

with $L_{A\prec B\prec C}$ and $L_{B\prec A\prec C}$ as defined in Subsection 2.2.2.

The proof is given in Appendix L. 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 (2.38):

$$\begin{aligned} &\min \mathrm{tr}(SW) \\ \text{s.t.} \quad &S \in \mathcal{S}_{3C,V}, \quad \mathbb{1}/d_O - S \in \mathcal{W}^*_{3C,V}, \end{aligned} \tag{2.76}$$

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 (2.38)–(2.39), the dual of (2.76) writes

$$\begin{aligned} &\min \mathrm{tr}[\Omega]/d_O \\ \text{s.t.} \quad &W + \Omega \in \mathcal{W}_{3C}^{\mathsf{sep}}, \quad \Omega \in \mathcal{W}_{3C}, \end{aligned} \tag{2.77}$$

and the optimal values of (2.76) and (2.77) respect the duality relation (2.40), which allows us to interpret $-\mathrm{tr}(S^*W)$ as generalised robustness also in this case. Furthermore, (2.76) and (2.77) 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 J.

2.5.1 Optimal witness

To find the optimal generalised robustness witness for the quantum switch we need to solve SDP problem (2.76) providing W_{switch} from Eq. (2.71) as an argument. Solving it using YALMIP and the solver MOSEK we obtain a witness S_{optimal} numerically; the generalised robustness of the quantum switch is found to be

$$R_{\text{g}}(W_{\text{switch}}) = -\operatorname{tr} S_{\text{optimal}} W_{\text{switch}} \approx 0.5454\,. \tag{2.78}$$

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 generalised robustness of the quantum switch.

2.5.2 Chiribella's witness

In Ref. [6] 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 Choi-Jamiołkowski 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 $|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}$ basis. Using the CJ representations $|U_A^*\rangle\!\rangle$ and $|U_B^*\rangle\!\rangle$ of U_A and U_B (see Appendix F.1), the corresponding operator is

$$G_{\pm}^{U_A,U_B} = |U_A^*\rangle\!\rangle\langle\!\langle U_A^*| \otimes |U_B^*\rangle\!\rangle\langle\!\langle U_B^*| \otimes |\pm\rangle\langle\pm|_{\pm}\,. \tag{2.79}$$

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:

$$G_{\text{Chiribella}} = \frac{1}{2}\int \mathrm{d}\mu_{[\,,\,]}\, G_{+}^{U_A,U_B} + \frac{1}{2}\int \mathrm{d}\mu_{\{\,,\,\}}\, G_{-}^{U_A,U_B}\,, \tag{2.80}$$

where $\mathrm{d}\mu_{[\,,\,]}$ is a measure over commuting unitaries, and $\mathrm{d}\mu_{\{\,,\,\}}$ 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

$$p_{\text{succ}} = \operatorname{tr}[G_{\text{Chiribella}} W]\,. \tag{2.81}$$

It is easy to check that for any choice of measures $\mathrm{d}\mu_{[\,,\,]}, \mathrm{d}\mu_{\{\,,\,\}}$ 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 $\mathrm{d}\mu_{[\,,]}$ and $\mathrm{d}\mu_{\{\,,\}}$. If we were to choose, for example, measures that only produce pairs of Pauli matrices, then there is a causally separable circuit[8] 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 $\mathrm{d}\mu_{[\,,]}$ to pick up commuting unitaries of the form

$$U_A = U\begin{pmatrix}1 & 0\\ 0 & e^{i\theta_1}\end{pmatrix}U^\dagger \quad \text{and} \quad U_B = U\begin{pmatrix}1 & 0\\ 0 & e^{i\theta_2}\end{pmatrix}U^\dagger, \tag{2.82}$$

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 $\mathrm{d}\mu_{\{\,,\}}$, 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)[9].

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

$$\begin{aligned} &\max \operatorname{tr}[G_{\text{Chiribella}} W] \\ \text{s.t.}\quad &\operatorname{tr} W = d_O, \quad W \in \mathcal{W}^{\text{sep}}_{3C}. \end{aligned} \tag{2.83}$$

Solving it with YALMIP and MOSEK, we obtain

$$p^{\text{sep}}_{\text{succ}} \approx 0.9288\,. \tag{2.84}$$

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 (2.80). Furthermore $\mathrm{d}\mu_{[\,,]}$ and $\mathrm{d}\mu_{\{\,,\}}$ 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 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:

$$\mathcal{G} = \{\mathbb{1}, X, Y, Z, \frac{X+Y}{\sqrt{2}}, \frac{X-Y}{\sqrt{2}}, \frac{X+Z}{\sqrt{2}}, \frac{X-Z}{\sqrt{2}}, \frac{Y+Z}{\sqrt{2}}, \frac{Y-Z}{\sqrt{2}}\} \tag{2.85}$$

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

$$G_{\text{finite}} = \sum_{i,j=1}^{10} q_{ij}^{[\,,]}\, G_+^{U_i,U_j} + q_{ij}^{\{\,,\}}\, G_-^{U_i,U_j}, \tag{2.86}$$

[8] One such circuit, described in Ref. [6], involves applying the Pauli unitaries to one half of a maximally entangled state and doing a measurement in the Bell basis.

[9] 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 Ref. [6] directly into the language of causal witnesses; the only difference, then, is that in [6] the witness was decomposed in terms of measurements and repreparations, whereas we decomposed it using unitaries only.

where $U_k \in \mathcal{G}$, and $q_{ij}^{[,]}, q_{ij}^{\{,\}}$ are the input probability distributions over commuting and anticommuting unitaries, normalised 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 M. We obtained

$$p_{\text{succ}}^{\text{sep}} \approx 0.8690\,, \tag{2.87}$$

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 (2.72) in the beginning of the present section – the only difference being the arbitary choice of the causal bound being ≥ 0 vs $\leq p_{\text{succ}}^{\text{sep}}$. More precisely, let G be a witness such that

$$\operatorname{tr}(G\,W^{\text{sep}}) \leq p_{\text{succ}}^{\text{sep}} \tag{2.88}$$

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

$$T_0 \leq \operatorname{tr}(G\,W) \leq T_1 \tag{2.89}$$

for every (normalised) process matrix W. Then

$$S = \frac{1}{p_{\text{succ}}^{\text{sep}} - T_0}\left(p_{\text{succ}}^{\text{sep}}\frac{\mathbb{1}}{d_O} - G\right) \tag{2.90}$$

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

$$R_{\text{g}}(W) = -\operatorname{tr}[SW] = \frac{T_1 - p_{\text{succ}}^{\text{sep}}}{p_{\text{succ}}^{\text{sep}} - T_0}\,. \tag{2.91}$$

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 (2.91).

2.6 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* [3], 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.

2.6.1 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 $1 \leq i \leq N$), 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 $P(\vec{a}|\vec{x})$.

Causal inequalities [3] 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 $A^{\sigma(1)} \prec A^{\sigma(2)} \prec \dots \prec A^{\sigma(N)}$. 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[10], namely if for every i the marginal distribution

$$P(a_{\sigma(1)}, \dots, a_{\sigma(i)}|\vec{x}) := \sum_{\substack{a_{\sigma(j)} \\ j>i}} P(\vec{a}|\vec{x}) \tag{2.92}$$

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

$$\begin{aligned} P(a_{\sigma(1)}, \dots, a_{\sigma(i)}|x_{\sigma(1)}, \dots, x_{\sigma(i)}, x_{\sigma(i+1)}, \dots, x_{\sigma(N)}) \\ = P(a_{\sigma(1)}, \dots, a_{\sigma(i)}|x_{\sigma(1)}, \dots, x_{\sigma(i)}, x'_{\sigma(i+1)}, \dots, x'_{\sigma(N)}) \\ \forall\, x_{\sigma(j)}, x'_{\sigma(j)}\,. \end{aligned} \tag{2.93}$$

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:

$$P(\vec{a}|\vec{x}) = \sum_{\sigma} q_\sigma\, P_\sigma(\vec{a}|\vec{x}), \quad q_\sigma \geq 0, \quad \sum_{\sigma} q_\sigma = 1\,, \tag{2.94}$$

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 [37, 47, 48]. 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 [49]. 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 (2.94), 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

[10] 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.

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" [45]. 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.

2.6.2 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 Refs. [3, 49]); 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. [48]), explicitly involving party C; as it turns out, this is also impossible, as a consequence of the following theorem[11]:

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*

$$P(\vec{a}|\vec{x}, z) := \sum_c P(\vec{a}, c|\vec{x}, z) \tag{2.95}$$

is such that

1. *$P(\vec{a}|\vec{x}, z) = P(\vec{a}|\vec{x})$ – i.e., it does not depend on z: C does not signal to any other (group of) parties;*
2. *$P(\vec{a}|\vec{x}) = \sum_\sigma q_\sigma P_\sigma(\vec{a}|\vec{x})$, where $q_\sigma \geq 0$, $\sum_\sigma q_\sigma = 1$, and the probability distributions P_σ are causally ordered,*

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

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

$$\begin{aligned} P(\vec{a}, c|\vec{x}, z) &= P(\vec{a}|\vec{x}, z)\, P(c|\vec{a}, \vec{x}, z) && (2.96)\\ &= \sum_\sigma q_\sigma\, P_\sigma(\vec{a}|\vec{x})\, P(c|\vec{a}, \vec{x}, z) && (2.97)\\ &= \sum_\sigma q_\sigma\, \widetilde{P}_\sigma(\vec{a}, c|\vec{x}, z), && (2.98) \end{aligned}$$

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. □

[11] 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 Ref. [37]

To see that the correlations generated by the quantum switch (Eq. (2.66)) respect assumptions *1.* and *2.* of the previous theorem, let us calculate the marginal probability distribution defined in Eq. (2.95) through the generalized Born rule (2.3), when the three parties perform operations $M_{a|x}^{A_I A_O}$, $M_{b|y}^{B_I B_O}$ and $M_{c|z}^{C_I}$:

$$\begin{aligned} P(a,b|x,y,z) &= \sum_c \mathrm{tr}\Big[M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O} \otimes M_{c|z}^{C_I} \cdot |w\rangle\langle w|\Big] \\ &= \mathrm{tr}\left[M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O} \otimes \Big(\sum_c M_{c|z}^{C_I}\Big) \cdot |w\rangle\langle w|\right]. \end{aligned} \tag{2.99}$$

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

$$P(a,b|x,y,z) = \mathrm{tr}\Big[M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O} \cdot W^{AB}\Big] \tag{2.100}$$

with

$$W^{AB} := \mathrm{tr}_{C_I} |w\rangle\langle w|. \tag{2.101}$$

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[12]. 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 [50], fall into this category. This raises the question of whether causally nonseparable processes that do violate causal inequalities can be physically implemented at all.

2.7 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 [29]. The results of this work show that, in fact, it is a natural framework to study a class of quantum resources which cannot be 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 Ref. [2].

[12] Note that Theorem 4 implies that this is also true for the N-partite generalization of the quantum switch defined in [28].

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 [51].

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.

Acknowledgements

We thank Michal Sedlák for useful discussions. We acknowledge support from the European Commission project RAQUEL (No. 323970); the Austrian Science Fund (FWF) through the Special Research Program Foundations and Applications of Quantum Science (FoQuS), the doctoral programme CoQuS, and Individual Project (No. 2462); FQXi; the John Templeton Foundation; the Templeton World Charity Foundation (grant TWCF 0064/AB38); the French National Research Agency through the 'Retour Post-Doctorants' program (ANR-13-PDOC-0026); and the European Commission through a Marie Curie International Incoming Fellowship (PIIF-GA-2013-623456).

Chapter 3

The simplest causal inequalities

This chapter was published as

- C. Branciard, M. Araújo, A. Feix, F. Costa and Č. Brukner.
  'The simplest causal inequalities and their violation'.
  New J. Phys. 18 013008 (2015). arXiv:1508.01704 [quant-ph].

My contribution to this paper was defining the causal polytope, finding the causal inequalities, and developing the algorithm to find violations of them.

Abstract

In a scenario where two parties share, act on and exchange some physical resource, the assumption that the parties' actions are ordered according to a definite causal structure yields constraints on the possible correlations that can be established. We show that the set of correlations that are compatible with a definite causal order forms a polytope, whose facets define *causal inequalities*. We fully characterize this *causal polytope* in the simplest case of bipartite correlations with binary inputs and outputs. We find two families of nonequivalent causal inequalities; both can be violated in the recently introduced framework of *process matrices*, which extends the standard quantum formalism by relaxing the implicit assumption of a fixed causal structure. Our work paves the way to a more systematic investigation of causal inequalities in a theory-independent way, and of their violation within the framework of process matrices.

3.1 Introduction

In our common understanding of the world, we typically perceive events as being embedded in some causal structure, where events happening earlier can influence events happening later but not vice versa. Correlations can be established in such a picture by physical systems that may be shared or exchanged by different parties, and which may be used to communicate or convey causal influences.

It is well known, however, that this view is challenged by quantum correlations: Bell's theorem [52] shows for instance that these conflict with Reichenbach's common cause principle [53, 54], so that quantum mechanics forces us to generalize the notion of causal influence [55–60]. Another implication of this picture is that if one assumes that the parties interact only once with the physical medium, then

only one-way influences (i.e., one-way signaling) are possible, which restricts—independently of any assumptions on the physics of the involved systems—the possible correlations that can be observed.

But is this view that events should comply with a definite causal structure, and causal influences can only be unidirectional, necessary in any physical theory? Or could one envisage theories where the causal relations between events are not necessarily well defined [29, 61]? To answer these questions, Oreshkov, Costa and Brukner developed the framework of *process matrices* as an extension of quantum theory, where the assumption of a fixed causal structure is relaxed [3]. Process matrices describe the physical resource that allows different parties to establish correlations. Oreshkov *et al.* showed that certain so-called *causally nonseparable process matrices* indeed do not comply with a definite causal structure.

The incompatibility of a certain causally nonseparable process matrix with a definite causal structure was proven in Ref. [3] by its ability to generate correlations that are incompatible with a definite causal order, as demonstrated by the violation of a so-called *causal inequality*. This can be tested in a device-independent manner, by just looking at the statistics observed in an experiment. It was recently shown that causal nonseparability could also be detected in a device-dependent manner by using so-called *causal witnesses* [62]. This approach is more powerful as it can detect all causally nonseparable process matrices, while not all causally nonseparable process matrices can violate a causal inequality [37, 62]. Furthermore, physical implementations of certain (multipartite) causally nonseparable process matrices, and of corresponding causal witnesses that detect their causal nonseparability, have been proposed [1, 37, 62] and even realized experimentally [2], while it is still not known whether there actually exist any physically realizable process that violates a causal inequality. Nevertheless, the device-independent approach is still of interest as it relaxes the requirement to trust the functioning and the operations implemented by one's devices in an experiment. It is furthermore also theory-independent: causal inequalities can in principle be tested, and correlations with no definite causal order can be identified whatever the description of the physical resource is—whether we use the process matrix framework or any other theory to be discovered in the future. A related open question is whether the ability to violate causal inequalities can—in analogy with Bell nonlocality [45]—be exploited as a resource, just like causally non-separable process matrices provide advantages for information-theoretical [6] and computational [28] tasks.

Our paper aims at providing a better understanding of the device-independent characterization of correlations that are compatible with a definite causal order or not. We show that bipartite correlations with a definite causal order form a convex polytope, whose facets correspond to causal inequalities (Section 3.2). We characterize this *causal polytope* in the simplest scenario where the two parties observe correlations with binary inputs and outputs, which gives us two families of new causal inequalities. We then investigate their possible violation in the framework of process matrices, and find that these can indeed be violated (Section 3.3). This provides an example of *"noncausal"* process matrix correlations in a simpler scenario than that considered in Ref. [3], where one party had two input bits, or in Refs. [47, 48], where more parties were involved.

3.2 Correlations with definite causal order

3.2.1 "Causal correlations"

We consider an experiment with two parties, Alice and Bob, each of them having control over some closed laboratory. They both open their lab, let some physical system in, interact with it and send a physical system out, only once during each run of the experiment. Alice and Bob are given some classical inputs labeled by x and y, and return some classical outputs a and b, respectively. We assume that all inputs and outputs have a finite number of possible values. The correlation that Alice and Bob establish in such an experiment is described by the joint conditional probability distribution $p(a,b|x,y)$.

In a situation where at each run of the experiment Alice's events precede Bob's events (denoted $A \prec B$), Alice could send her input and output to Bob, but not vice versa; hence, there cannot be any signaling from Bob to Alice, and their correlation, which we shall denote in this case $p^{A\prec B}$, must therefore satisfy

$$\forall\, x,y,y',a, \quad p^{A\prec B}(a|x,y) = p^{A\prec B}(a|x,y')\,, \tag{3.1}$$

with $p^{A\prec B}(a|x,y^{(\prime)}) = \sum_b p^{A\prec B}(a,b|x,y^{(\prime)})$. Similarly, in a situation where Bob's events precede Alice's ($B \prec A$), their correlation $p^{B\prec A}$ must satisfy the no-signaling-to-Bob constraint

$$\forall\, x,x',y,b, \quad p^{B\prec A}(b|x,y) = p^{B\prec A}(b|x',y)\,, \tag{3.2}$$

with $p^{B\prec A}(b|x^{(\prime)},y) = \sum_a p^{B\prec A}(a,b|x^{(\prime)},y)$. Note that non-signaling correlations satisfy both Eqs. (3.1) and (3.2), and are compatible with both causal orders $A \prec B$ and $B \prec A$. More generally, if the correlation is compatible with the causal order $A \prec B$ with probability q, and with $B \prec A$ with probability $1-q$, then the correlation will be of the form

$$p(a,b|x,y) = q\, p^{A\prec B}(a,b|x,y) + (1-q)\, p^{B\prec A}(a,b|x,y)\,. \tag{3.3}$$

Following Refs. [37, 62, 63], we call the bipartite probability distribution $p(a,b|x,y)$ (or the correlation it describes, equivalently) *"causal"* if it can be written as in Eq. (3.3), with $q \in [0,1]$ and $p^{A\prec B}$ and $p^{B\prec A}$ valid (i.e., nonnegative and normalized) probability distributions satisfying Eqs. (3.1) and (3.2), respectively. Causal correlations are those that can be obtained in a situation where every run of the experiment is compatible with a definite causal order ($A \prec B$ or $B \prec A$), which may however vary for each run, and is only determined probabilistically. Note that the decomposition (3.3) is in general not unique, as non-signaling contributions can be included in either $p^{A\prec B}$ or $p^{B\prec A}$.

3.2.2 Causal polytopes and causal inequalities

Correlations that are compatible with the causal order $A \prec B$ satisfy nonnegativity ($p^{A\prec B}(a,b|x,y) \geq 0\ \forall x,y,a,b$) and normalization ($\sum_{a,b} p^{A\prec B}(a,b|x,y) = 1\ \forall x,y$) constraints, together with the no-signaling-to-Alice constraint (3.1). As these constitute a finite number of linear constraints on a bounded probability space[1], it

[1] The probability space can for instance be understood geometrically as the set of points in a high enough dimensional space, whose coordinates are the values $p(a,b|x,y)$. Clearly, the nonnegativity and normalization constraints make it bounded.

follows that the set of correlations $p^{A\prec B}$ is a (convex) polytope [64]. Similarly, the set of correlations $p^{B\prec A}$ that are compatible with the causal order $B \prec A$ is also a polytope. Now, according to Eq. (3.3), the set of causal correlations is simply the convex hull of the sets of correlations $p^{A\prec B}$ and $p^{B\prec A}$, and is therefore itself a polytope, which we call the *causal polytope*.

By construction, the extremal points of the causal polytope are extremal points of either the polytope of $p^{A\prec B}$ correlations, or of the polytope of $p^{B\prec A}$ correlations (or of both polytopes); in Appendix N we show that these correspond to deterministic correlations compatible with either causal order (or both, in the case of nonsignaling correlations). From this "$\mathcal{V}$-representation" of the causal polytope in terms of its vertices, for a given number of inputs and outputs, one can in principle determine its equivalent "$\mathcal{H}$-representation" in terms of its facets [64] (although in practice, this is a hard problem to solve when the number of inputs and outputs increase). Some of its facets are trivial, in the sense that they only correspond to the nonnegativity constraints $p(a, b|x, y) \geq 0$; its other, nontrivial facets define so-called *causal inequalities* [3]—inequalities that are satisfied by any causal correlation.

The above characterization hints of course at a strong analogy with Bell inequalities, which may be obtained as facets of the "local polytope" [45, 52, 65] (or may not; in the same way that not all Bell inequalities are facets of the local polytope, not all causal inequalities are facets of the causal polytope, as they can also correspond to some external hyperplanes[2]). Causal inequalities are written as linear combinations of the conditional probabilities $p(a, b|x, y)$, constrained by some *"causal bounds"*. They can also be translated in the language of *"causal games"* by considering for instance the linear combination to define the score, or possibly the probability of success (for some specific distribution of inputs), of some game. They can be tested experimentally in a device-independent way—i.e., by just considering the observed statistics, without making any assumptions on the functioning of the physical devices used in the experiment: a violation of a causal inequality guarantees that the observed correlation is incompatible with a definite causal order—or, in short, is *noncausal*.

3.2.3 The simplest causal polytope

To illustrate the previous discussion, we now turn to the characterization of the simplest nontrivial causal polytope. Note that causal inequalities can only be nontrivial if each party has nontrivial inputs and outputs—i.e., if they can take at least two different values. Indeed, if one party only has trivial inputs or outputs, then clearly either (3.1) or (3.2) holds, so that any correlation is compatible with a definite causal order.

Hence, the simplest candidate for a nontrivial causal polytope is the case with a single bit of input and output for each of the two parties[3] (which we shall denote by

[2] E.g., one can check that the original causal inequality of Ref. [3] is not a facet of the causal polytope for 1 input bit for Alice, 2 input bits for Bob, and 1 output bit for each of them (the 320 vertices of the causal polytope that saturate that inequality, out of 5056 vertices, only span a 21-dimensional affine subspace, while a facet of this 24-dimensional polytope should have dimension 23). Likewise, the causal inequalities (3.21–3.24) below are not facets of the causal polytope for binary inputs and outputs (they are only facets of its projection onto the plane considered in Subsection 3.3.3).

[3] Actually, one also has a nontrivial causal polytope in a scenario where one of Alice and/or Bob's input yields a binary output, while the other always gives the same output (or has no output, equivalently). In such a case the only nontrivial causal inequalities are of the LGYNI type, Eq. (3.5) or (3.7), and its analysis below—e.g. its violation by process matrix correlations—remains valid (note on the other hand

0 or 1), reminiscent of the scenario considered by Clauser–Horne–Shimony–Holt (CHSH) in the case of nonlocality [67]. We generated the list of its 112 deterministic vertices (see Appendix N), and enumerated its 48 facets using the software `lrs` [68].

16 of these facets are trivial, corresponding to the nonnegativity constraints $p(ab|xy) \geq 0$. By relabeling the inputs and outputs, the 32 remaining, non-trivial facets can be grouped in two non-equivalent families of causal inequalities: 16 facets are relabelings of the inequality

$$\frac{1}{4} \sum_{x,y,a,b} \delta_{a,y}\, \delta_{b,x}\, p(a,b|x,y) \;\leq\; \frac{1}{2}\,, \tag{3.4}$$

where $\delta_{i,j}$ is the Kronecker delta, while the last 16 facets are relabelings of the inequality

$$\frac{1}{4} \sum_{x,y,a,b} \delta_{x(a\oplus y),0}\, \delta_{y(b\oplus x),0}\, p(a,b|x,y) \;\leq\; \frac{3}{4}\,, \tag{3.5}$$

where $\oplus$ denotes addition modulo 2.

The causal inequality (3.4) can be interpreted as a bound on the maximal probability of success for a bipartite *"guess your neighbor's input"* (GYNI) game [69] with uniform input bits x, y (such that $p(x,y) = \frac{1}{4}$), where Alice and Bob's task is to guess each other's input, i.e., to output $a = y$ and $b = x$. Implicitly assuming uniform input bits[4], inequality (3.4) can indeed be written in a more compact form as

$$p_{\text{GYNI}} \;:=\; p(a = y, b = x) \;\leq\; \frac{1}{2}\,. \tag{3.6}$$

This causal bound on the probability of success p_{GYNI} can easily be understood: assuming that the correlation is compatible with the causal order $A \prec B$, Alice cannot know anything about Bob's input bit and can therefore only make a random guess, so that $p(a = y) = \frac{1}{2}$ and therefore $p(a = y, b = x) \leq \frac{1}{2}$; a similar reasoning holds for the causal order $B \prec A$, and a convex mixture cannot increase the bound on p_{GYNI}.

Similarly, the causal inequality (3.5) can be interpreted as a bound on the maximal probability of success for what we shall call a *"lazy GYNI"* (LGYNI) game, still with uniformly random input bits, where Alice and Bob's task is now to guess each other's input only when their respective input is 1 (for an input 0, their output can be arbitrary). Implicitly assuming uniform input bits, inequality (3.5) can then also be written in a more compact form as

$$p_{\text{LGYNI}} \;:=\; p\big(x(a \oplus y) = 0, y(b \oplus x) = 0\big) \;\leq\; \frac{3}{4}\,. \tag{3.7}$$

This causal bound on the probability of success p_{LGYNI} can also easily be understood with a similar reasoning as above (taking into account that Alice for instance is only asked to guess Bob's input half of the time, when her input is 1).

---

that the local polytope for such scenarios is trivial [66]). For simplicity however, we choose to impose throughout the paper that all inputs should have the same number of outputs.

[4] Note that the assumption of uniform inputs is only necessary to justify the shorthand notation $p(a = y, b = x)$ for the left hand side of Eq. (3.4), and to interpret it as the success probability for the GYNI game. Whether an inequality written as a combination of conditional probabilities (like (3.4) or (3.5) for instance) defines a causal inequality or not depends of course in no way on the distribution of inputs.

3.3 Process matrix correlations with no definite causal order

In this section we study the violation of our simplest causal inequalities in the framework of *process matrices*, introduced recently by Oreshkov, Costa and Brukner [3]. Let us first start with a brief overview of this framework.

3.3.1 The process matrix framework

The basic assumption of the framework is that quantum theory correctly describes what happens *locally* in Alice and Bob's laboratories; however, no assumption is being made about the *global* causal structure in which the parties operate.

More specifically, it is assumed that Alice and Bob can perform any operation allowed by the standard formulation of quantum theory, as described by quantum *instruments* [34] from some input Hilbert spaces $\mathcal{H}^{A_I}$ and $\mathcal{H}^{B_I}$ (for Alice and Bob, respectively) to some output Hilbert spaces $\mathcal{H}^{A_O}$ and $\mathcal{H}^{B_O}$. An instrument is a set of completely positive (CP), trace non-increasing maps from $\mathcal{L}(\mathcal{H}^{X_I})$ to $\mathcal{L}(\mathcal{H}^{X_O})$ (for $X = A, B$), where $\mathcal{L}(\mathcal{H}^{X_I})$ and $\mathcal{L}(\mathcal{H}^{X_O})$ are the spaces of linear operators over the Hilbert spaces $\mathcal{H}^{X_I}$ and $\mathcal{H}^{X_O}$. Each CP map of a given instrument is associated with a given classical output, which we shall again denote by a and b for Alice and Bob, and all CP maps of an instrument must sum up to a trace-preserving map. The various instruments that the parties can choose to apply shall be labeled by some classical "inputs" x and y.

Using the Choi-Jamiołkowski (CJ) isomorphism [70, 71], one can represent Alice's maps as some operators[5] $M_{a|x}^{A_I A_O}$ on the tensor product space $\mathcal{L}(\mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O})$. The conditions for the collection of operators $\{M_{a|x}^{A_I A_O}\}_a$ (for some fixed input x) to be a valid instrument translate to

$$M_{a|x}^{A_I A_O} \geq 0 \quad \forall\, a \quad \text{and} \quad \mathrm{tr}_{A_O} \sum_a M_{a|x}^{A_I A_O} = \mathbb{1}^{A_I}, \tag{3.8}$$

where tr_{A_O} denotes the partial trace over Alice's output system, and $\mathbb{1}^{A_I}$ is the identity operator in Alice's input Hilbert space. Similarly, Bob's maps can be represented as operators $M_{b|y}^{B_I B_O}$ on $\mathcal{L}(\mathcal{H}^{B_I} \otimes \mathcal{H}^{B_O})$, and a collection of operators $\{M_{b|y}^{B_I B_O}\}_b$ (for some fixed input y) must satisfy analogous constraints to Eq. (3.8) to be a valid instrument.

As shown in Ref. [3], the assumption of local consistency with quantum theory implies that the probability $p(a, b|x, y)$ of observing the classical outputs a, b for a choice of instruments labeled by x, y is a bilinear function of Alice and Bob's maps, which can be written as

$$p(a, b|x, y) = \mathrm{tr}\left[(M_{a|x}^{A_I A_O} \otimes M_{b|y}^{B_I B_O}) \cdot W\right], \tag{3.9}$$

for some hermitian matrix $W \in \mathcal{L}(\mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O} \otimes \mathcal{H}^{B_I} \otimes \mathcal{H}^{B_O})$. Requiring that the probabilities given by (3.9) are nonnegative and normalized for all possible choice of quantum operations (including operations involving possibly entangled ancillary systems) imposes some restrictions on the possible W matrices [3]. As shown in

[5] Throughout the paper, superscripts on operators refer to the Hilbert space they act on.

Ref. [62], these constraints can be expressed as follows:

$$W \geq 0\,, \tag{3.10a}$$

$$\mathrm{tr}\, W = d_{A_O}\, d_{B_O}\,, \tag{3.10b}$$

$${}_{B_I B_O} W = {}_{A_O B_I B_O} W\,, \tag{3.10c}$$

$${}_{A_I A_O} W = {}_{A_I A_O B_O} W\,, \tag{3.10d}$$

$$W = {}_{B_O} W + {}_{A_O} W - {}_{A_O B_O} W\,, \tag{3.10e}$$

where the last three conditions are written using the operation ${}_X\cdot$ defined by

$${}_X W = \frac{\mathbb{1}^X}{d_X} \otimes \mathrm{tr}_X\, W \tag{3.11}$$

for $X = A_I, A_O, B_I, B_O$, with $\mathbb{1}^X$ and tr_X denoting the identity operator and the partial trace over the Hilbert space $\mathcal{H}^X$, respectively, and d_X denoting its dimension.

Operators W that satisfy these conditions are called *process matrices*. They generalize the notions of "quantum strategies" [36] and of "quantum combs" [32] (which do assume a causal order), and represent the most general way to "connect" the output spaces $\mathcal{H}^{A_O} \otimes \mathcal{H}^{B_O}$ to the input spaces $\mathcal{H}^{A_I} \otimes \mathcal{H}^{B_I}$ (see Fig. 3.1) in a way that is locally consistent with quantum theory. While the above conditions do not impose a global causal order *a priori* and therefore allow in general for two-way signaling, the nonnegativity and normalization constraints on the probabilities guarantee that no logical paradoxes, like the grandfather paradox for instance [72, 73], appear. In the following we will refer to correlations of the form (3.9), with Alice and Bob's instruments satisfying Eq. (3.8) (together with its analogous form for Bob) and W satisfying the conditions (3.10), as *process matrix correlations*.

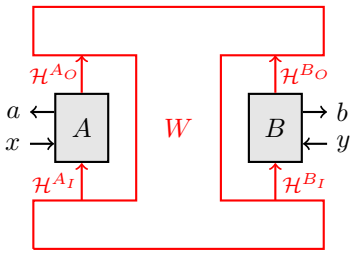


Figure 3.1: A process matrix W represents the physical resource which connects Alice's ($\mathcal{H}^{A_O}$) and Bob's ($\mathcal{H}^{B_O}$) output Hilbert spaces to their input Hilbert spaces ($\mathcal{H}^{A_I}$, $\mathcal{H}^{B_I}$) in such a way that what happens in Alice and Bob's labs is locally consistent with quantum theory [3]. Process matrices generalize in particular quantum states and quantum channels.

3.3.2 Violation of the simplest causal inequalities by process matrix correlations

It was shown in Ref. [3] that certain process matrices[6] could generate correlations with no definite causal order. A specific process matrix and specific instruments were indeed found, which violate a particular causal inequality with one input bit for Alice and two for Bob, and one output bit for each. Remarkably, one of Bob's input bits could be used to distinguish some runs of the experiment where signaling happened in one direction, and some runs where it happened in the other direction. It remained an open question, whether this special input bit for Bob was necessary to obtain noncausal correlations in the process matrix framework, or whether any simpler causal inequality (with fewer inputs) could be violated. Here we answer this question positively, by exhibiting violations of both our GYNI and LGYNI inequalities (3.6, 3.7) by process matrix correlations.

Let us start with a simple example with two-dimensional input and output systems—"qubits"—for Alice and Bob (i.e., $d_{A_I} = d_{A_O} = d_{B_I} = d_{B_O} = 2$). One can check that the matrix

$$W = \frac{1}{4}\left[\mathbb{1}^{\otimes 4} + \frac{Z^{A_I} Z^{A_O} Z^{B_I} \mathbb{1}^{B_O} + Z^{A_I} \mathbb{1}^{A_O} X^{B_I} X^{B_O}}{\sqrt{2}}\right], \tag{3.12}$$

where Z and X are the Pauli matrices and where tensor products are implicit, satisfies the constraints (3.10), so that it defines a valid process matrix. We choose Alice and Bob's operations to be the same, defined by

$$M_{0|0}^{A_I A_O} = M_{0|0}^{B_I B_O} = 0\,, \tag{3.13}$$

$$M_{1|0}^{A_I A_O} = M_{1|0}^{B_I B_O} = 2\left|\Phi^+\right\rangle\!\left\langle\Phi^+\right|, \tag{3.14}$$

$$M_{0|1}^{A_I A_O} = M_{0|1}^{B_I B_O} = |0\rangle\langle 0| \otimes |0\rangle\langle 0|\,, \tag{3.15}$$

$$M_{1|1}^{A_I A_O} = M_{1|1}^{B_I B_O} = |1\rangle\langle 1| \otimes |0\rangle\langle 0|\,, \tag{3.16}$$

with $\{|0\rangle, |1\rangle\}$ denoting the computational basis (i.e., the eigenbasis of Z), and $|\Phi^+\rangle := (|00\rangle + |11\rangle)/\sqrt{2}$. These indeed satisfy (3.8), and thus constitute valid instruments. These operations can be interpreted as follows: when their input is 0, Alice and Bob simply transmit their incoming physical system, untouched ($2\left|\Phi^+\right\rangle\!\left\langle\Phi^+\right|$ being indeed the CJ representation of an identity channel), and output the value 1; when their input is 1, Alice and Bob perform a measurement in the Z basis, whose result defines their classical output, and send out the fixed state $|0\rangle\langle 0|$. With these definitions, one can calculate the success probabilities of the GYNI and LGYNI games using Eqs. (3.6, 3.7)—or more explicitly (3.4, 3.5)—and Eq. (3.9). One finds

$$p_{\text{GYNI}} = \frac{5}{16}\left(1 + \frac{1}{\sqrt{2}}\right) \approx 0.5335 > \frac{1}{2}\,, \tag{3.17}$$

$$p_{\text{LGYNI}} = \frac{5}{16}\left(1 + \frac{1}{\sqrt{2}}\right) + \frac{1}{4} \approx 0.7835 > \frac{3}{4}\,, \tag{3.18}$$

[6] A necessary condition for a process matrix to allow for a causal inequality violation is that it is *causally nonseparable* [3]—i.e., that it is itself incompatible with a definite causal order. In the multipartite case this is known however not to be a sufficient condition [37, 62]. It remains an open question whether there can be bipartite causally nonseparable process matrices that only generate causal correlations.

which indeed violate the causal inequalities (3.6, 3.7).

One may now wonder, what the largest possible violation of these two causal inequalities by process matrix correlations is. To optimize the violations for some input and output Hilbert spaces of a given dimension, we used a See-Saw algorithm inspired by that of Werner and Wolf [74], as described in Appendix O. Note that because the optimization problem is nonconvex, the algorithm is not guaranteed to converge to a global maximum. Nevertheless, for small enough dimensions (at least, for qubits), the repeatability of our results for different random starting points of the algorithm makes us confident that we indeed found the global maxima. From our numerical results, we thus conjecture that the maximal violations of our causal inequalities achievable with qubit systems are

$$p_{\text{GYNI}}^{\max,d=2} \approx 0.5694 > \frac{1}{2}\,, \tag{3.19}$$

$$p_{\text{LGYNI}}^{\max,d=2} \approx 0.8194 = p_{\text{GYNI}}^{\max,d=2} + \frac{1}{4} > \frac{3}{4}\,. \tag{3.20}$$

In Appendix P we give an analytical description of the process matrices that reach these values.

Going to larger dimensions, we found that the maximal value of p_{GYNI} could increase, as summarized in Table 3.1 for dimensions up to 5; however, we did not find any larger value for p_{LGYNI} than $p_{\text{LGYNI}}^{\max,d=2}$ above, which—provided our See-Saw algorithm did find the global maxima—reveals some fundamental difference between the two inequalities, despite their similarities (and in addition to the fact that contrary to GYNI, the outputs corresponding to certain inputs are irrelevant in the LGYNI game; see also footnote 3). It remains an open question, which values are the true "Tsirelson bounds" [75] for these two causal inequalities, in the sense of the largest possible values for p_{GYNI} and p_{LGYNI} reachable with quantum process matrices of any dimension.

d	2	3	4	5
$p_{\text{GYNI}}^{\max,d}$	0.5694	0.6104	0.6201	0.6218

Table 3.1: Maximal values of p_{GYNI} found through numerical optimization, as a function of the dimension of Alice and Bob's input and output Hilbert spaces $d = d_{A_I} = d_{A_O} = d_{B_I} = d_{B_O}$.

3.3.3 Boundary of the set of process matrix correlations

To finish with, let us picture the set of process matrix correlations vs that of causal correlations. In order to visualize the two, we shall project them onto the plane with coordinates $\big(p(a{=}y), p(b{=}x)\big)$, where we again implicitly assume uniformly random inputs for ease of notations[7]. In this plane the complementarity between the two directions of signaling, from Alice to Bob and from Bob to Alice, is conspicuous; the projected causal polytope is bounded here by the four causal inequalities[8] (see Figure 3.2)

$$\frac{1}{2}\,p(a=y)+\frac{1}{2}\,p(b=x)\;\leq\;\frac{3}{4}\,, \tag{3.21}$$

$$\frac{1}{2}\,p(a=y)-\frac{1}{2}\,p(b=x)\;\leq\;\frac{1}{4}\,, \tag{3.22}$$

$$-\frac{1}{2}\,p(a=y)+\frac{1}{2}\,p(b=x)\;\leq\;\frac{1}{4}\,, \tag{3.23}$$

$$-\frac{1}{2}\,p(a=y)-\frac{1}{2}\,p(b=x)\;\leq\;-\frac{1}{4}\,. \tag{3.24}$$

To obtain a lower bound for the boundary of the set of process matrix correlations, we again used the See-Saw algorithm described in Appendix O to maximize quantities of the form $\alpha\,p(a{=}y)+\beta\,p(b{=}x)$, with various weights α,β. Different runs of the algorithm gave us different lower bounds (recall that the See-Saw algorithm is not guaranteed to always find the global optimum), which we combined to obtain the bounds represented on Figure 3.2 for dimensions $d=2,3$ and 4 of the input and output Hilbert spaces—and which we believe are close to the actual boundaries of the process matrix correlations for these dimensions. The largest violations of the causal inequality (3.21) we found are $\frac{1}{2}\,p(a{=}y)+\frac{1}{2}\,p(b{=}x)=0.7715$ for $d=2$; 0.7892 for $d=3$; and 0.8001 for $d=4$.

A surprising feature of the set of process matrix correlations for dimension 2 is that it does not seem to be convex (see Figure 3.2, red region). We believe this is a true characteristic of it, not only a numerical artifact due to some failure to find global optima. Note, nevertheless, that the boundary of the set of process matrix correlations for arbitrary dimensions *is* convex, as proven in Appendix Q.

3.4 Conclusion

We have shown that the set of correlations compatible with a definite causal order ("causal correlations") forms a convex polytope, which we fully characterized in the simplest nontrivial bipartite scenario with binary inputs and outputs. Two nonequivalent families of causal inequalities were obtained, Eqs. (3.6–3.7), for which we

[7] Without assuming uniformly random inputs, $p(a=y)$ and $p(b=x)$ in Eqs. (3.21–3.24) and in Figure 3.2 should be replaced by $\frac{1}{4}\sum_{x,y,a,b}\delta_{a,y}\,p(a,b|x,y)$ and $\frac{1}{4}\sum_{x,y,a,b}\delta_{b,x}\,p(a,b|x,y)$, respectively.

[8] Note that inequality (3.21) looks quite similar to Oreshkov *et al.*'s original causal inequality [3] (which motivated us in particular to write it with the $\frac{1}{2}$ factors). Oreshkov *et al.*'s inequality includes some additional conditioning on a second input bit for Bob. Averaging that inequality with its equivalent version where Bob's second input bit is flipped yields inequality (3.21). Similarly, Brukner's "Tsirelson-like bound" for Oreshkov *et al.*'s inequality [46] (for a limited set of possible instruments) also yields the same Tsirelson-like bound $\frac{1+1/\sqrt{2}}{2}\simeq 0.8536$ for inequality (3.21) (with the same restriction); there remains a large gap between that bound and the lower bounds we obtained numerically for dimensions $2,3$ and 4.

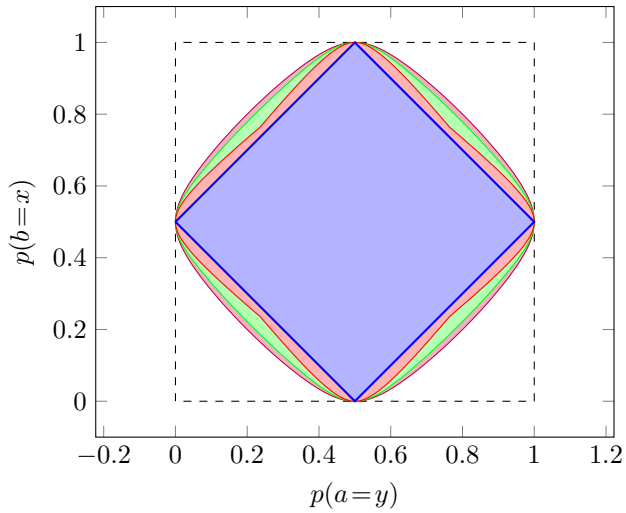


Figure 3.2: Projection of the probability space for binary inputs and outputs onto the plane $\big(p(a=y), p(b=x)\big)$ (see main text). The causal polytope is projected onto the blue diamond, delimited by the causal inequalities (3.21–3.24). The red, green and purple regions correspond to process matrix correlations that are reachable with input and output Hilbert spaces of dimensions $d = 2, 3$ and 4, respectively. The outer dashed square delimits all valid probabilities $p(a=y), p(b=x) \in [0, 1]$. Its upper right corner corresponds for instance to a correlation such that Alice's output is always equal to Bob's input ($a=y$) and Bob's output is always equal to Alice's input ($b=x$), which requires perfect 2-way signaling and violates Eqs. (3.6), (3.7) and (3.21) up to their algebraic maximum. This correlation may somehow be thought of as being analogous to the Popescu-Rohrlich (PR) box considered in the context of nonlocal correlations [76] (one difference, however, is that this correlation is deterministic, while the PR box correlations are not).

gave intuitive interpretations in terms of "causal games". These allow for a device-independent characterization of correlations with or without definite causal order, and can be tested independently of the physical theory under consideration. We exhibited in particular violations of these inequalities by process matrix correlations, which generalize standard quantum correlations. Because of their simplicity (and despite the violations we found being somewhat less intuitive), we expect these new inequalities—in particular the first one, interpreted as a "guess your neighbor's input" game—to become archetypical examples of causal inequalities, just like the CHSH inequality is the archetype of Bell inequalities [45, 67].

Our approach can be used to characterize (non)causal correlations in more complex scenarios as well. It should be noted that because of the large dimension of the probability space and the large number of vertices of the causal polytope (see Appendix N), the full facet enumeration problem rapidly becomes intractable in practice as the number of inputs and outputs increases beyond the simplest binary case. Nevertheless, one could adapt some of the tricks developed for the derivation of Bell inequalities (see Ref. [45] for a review) to construct new causal inequalities for various scenarios of interest. Violations of these inequalities in the process matrix framework can then be investigated using our See-Saw algorithm.

An interesting question is whether it would also be possible to derive nontrivial bounds on such violations from certain information-theoretic principles [77], along analogous lines to the research program that aims at restricting quantum nonlocal correlations from various principles [45].

In the present paper we focused for simplicity on the bipartite case. Our work can naturally also be extended to the case of more parties. With a proper generalization of the concept of noncausal correlations (see for instance Ref. [37]), it can also be shown that multipartite noncausal correlations form a convex polytope. Similar techniques can be used to characterize this polytope, construct causal inequalities and investigate their possible violation. Note that a remarkable new feature in the multipartite case is that violations are also possible with "classical process matrices" [48].

One of the main open questions along the line of research presented here is whether it would actually be possible, in practice, to observe correlations with no definite causal order and a violation of a causal inequality. As an extension of standard quantum theory, the framework of process matrices—which does indeed predict such violations theoretically—appears as a good candidate to provide such a possibility. However, to the best of our knowledge, no practical implementation has been identified for any of the process matrices that are known to violate a causal inequality [3, 47, 48] (including the ones presented here)—while in contrast, a *causally nonseparable* quantum process has been recently demonstrated experimentally [2]. It is likely that more complex scenarios need to be considered, and a systematic investigation of causal inequalities and their violation by process matrix correlations might prove useful to find practical violations—or, should it be the case, to clarify why such violations cannot be observed in practice. Our work makes the first crucial step in this direction.

Note added.— While finishing writing up this manuscript, we became aware that the concept of causal polytopes introduced here was also referred to (with proper reference to our work) in Ref. [37], where the emphasis was put on multipartite scenarios, and in Ref. [78], where the authors also introduced, for the multipartite case as well, larger polytopes of logically consistent but possibly noncausal classical processes.

Acknowledgements

We thank Ognyan Oreshkov and Michal Sedlák for useful discussions. We acknowledge support from the French National Research Agency through the 'Retour Post-Doctorants' program (ANR-13-PDOC-0026); the European Commission through a Marie Curie International Incoming Fellowship (PIIF-GA-2013-623456); the European Commission project RAQUEL (No. 323970); the Austrian Science Fund (FWF) through the Special Research Program Foundations and Applications of Quantum Science (FoQuS), the doctoral programme CoQuS, and Individual Project (No. 2462); FQXi; the John Templeton Foundation; and the Templeton World Charity Foundation (grant TWCF 0064/AB38).

Chapter 4

A purification postulate for quantum mechanics with indefinite causal order

This chapter will be submitted for publication in the near future as

- M. Araújo, A. Feix, M. Navascués and Č. Brukner.
  'A purification postulate for quantum mechanics with indefinite causal order'.

Abstract

To study which are the most general causal structures which are compatible with local quantum mechanics, Oreshkov *et al.* [3] introduced the notion of a process: a resource shared between some parties that allows for quantum communication between them without a predetermined causal order. These processes can be used to perform several tasks that are impossible in standard quantum mechanics: they allow for the violation of causal inequalities, and provide an advantage for computational and communication complexity. Nonetheless, no process that can be used to violate a causal inequality is known to be physically implementable. There is therefore considerable interest in determining which processes are physical and which are just mathematical artefacts of the framework. Here we make the first step in this direction, by proposing a purification postulate: processes are physical only if they are purifiable. We derive necessary conditions for a process to be purifiable, and show that several known processes do not satisfy them.

4.1 Introduction

It is widely believed that any future theory that unifies quantum mechanics and gravity will feature quantum uncertainty in the metric tensor [31], which should produce indefinite causal structures. We currently lack, however, good understanding of what an indefinite causal structure even is. To investigate that, one approach is to consider simple, concrete models that are compatible with the laws of quantum information processing. One such model – the process matrix formalism – was introduced by Oreshkov *et al.* as the most general causal structure that can connect

local laboratories where standard quantum mechanics is valid without creating paradoxical causal loops [3].

These processes have been shown to enable the realization of tasks that are otherwise impossible: they allow for the violation of causal inequalities [3, 37, 47–49, 78, 79], they can be detected by causal witnesses [80], they provide an advantage in quantum computation [1, 6, 28], and they enable a reduction in communication complexity [81, 82]. But even though one such process – the quantum switch – has been implemented experimentally [2, 4], in general it is not known if all process are physical or some of them are just mathematical artefacts of the formalism. They were, after all, defined only from the requirement of not generating logical contradictions, and the processes realisable in nature are likely to fulfil additional physical constraints.

One can, therefore, look for requirements other than mere logical consistency in order to shed light on the physicality of these processes. As seen from the search of physical principles to determine the set of quantum correlations, such meta-theoretical principles can provide nontrivial constraints on the possible theories [83–87]. The principle we choose to investigate here is reversibility of the transformations between quantum states: it is, after all, valid in all fundamental physical theories, and was a central ingredient in all of the reconstructions of quantum mechanics [27, 88–93].

To define what reversibility means for processes we first needed to generalise their definition: we extend them with a global past and a global future, so that they can be seen as inducing a transformation from quantum states in the past to quantum states in the future, after passing through the causally indefinite region of the local laboratories. We can then define pure processes as those that preserve the reversibility of the underlying operations, i.e., those that induce a unitary transformation from the past to the future whenever unitary transformations are also applied in the local laboratories.

With these definitions in hand, we can propose the purification postulate: *processes are physical only if they are purifiable*, i.e., if they can be expressed as a part of a pure process in a larger space. It turns out to be rather difficult to determine if a given process is purifiable. We derive, then, a necessary but not sufficient condition for purifiability, and show that several known processes fail to satisfy this condition, among which the process from Ref. [3] that was the first one shown to violate a causal inequality. In fact, no bipartite process that was able to violate a causal inequality was found to be purifiable. There exists, however, a tripartite process which is purifiable and able to violate causal inequalities. This implies that the purification postulate is necessary but not sufficient for physicality of a process, if one takes the view that violations of causal inequalities is impossible in Nature [94].

The paper is organized as follows: in section 4.2 we generalise the definition of a process and introduce the notion of pure processes. In section 4.3 we propose the purification postulate. In section 4.4 we derive necessary and sufficient conditions for a process to be purifiable, and in section 4.5 we derive a necessary but not sufficient condition for purifiability that can be easily tested. In section 4.6 we apply our condition to several known processes, and in section 4.7 we present the tripartite pure process that violates causal inequalities.

4.2 Pure processes

In Ref. [3], the process matrix was introduced as a (multi)linear function that takes completely positive (CP) maps to probabilities. To be able to talk about purification, we need to extend this definition to take CP maps not to probabilities, but to other CP maps. This view of a process as a higher-order transformation is much in the spirit of quantum combs [32], and can be recovered from the multipartite definition of process presented for example in Ref. [80]. We shall, nevertheless, explicitly define processes as higher-order transformations for clarity and to make this paper self-contained.

To this end, let $\mathcal{A}$ be a completely positive trace preserving (CPTP) map that takes the Hilbert spaces A_I, A'_I to A_O, A'_O, and $\mathcal{B}$ a CPTP map that takes B_I, B'_I to B_O, B'_O. Then a process is defined as the most general linear transformation that acts trivially on A'_I, B'_I, A'_O, B'_O and takes $\mathcal{A}$ and $\mathcal{B}$ to a CPTP map $\mathcal{G}_{\mathcal{A},\mathcal{B}}$ from A'_I, B'_I, P to A'_O, B'_O, F, as represented in Fig. 4.1.

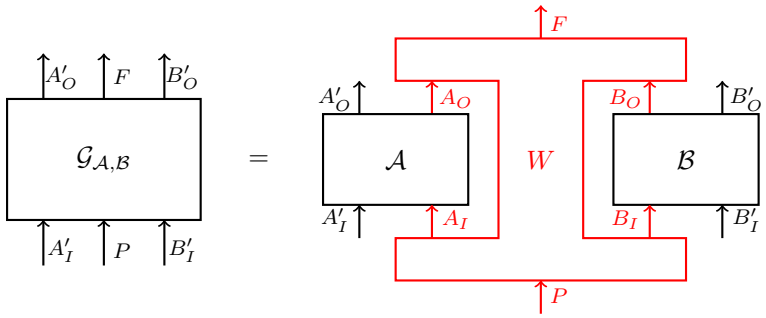


Figure 4.1: A process W is a bilinear function from the CPTP maps $\mathcal{A}, \mathcal{B}$ to a CPTP map $\mathcal{G}_{\mathcal{A},\mathcal{B}}$.

To find a matrix representation for a process, we make use of the Choi-Jamiołkowski (CJ) isomorphism, which we recap in appendix R. Let then $A = \mathfrak{C}(\mathcal{A})$, $B = \mathfrak{C}(\mathcal{B})$, and $G_{A,B} = \mathfrak{C}(\mathcal{G}_{\mathcal{A},\mathcal{B}})$ be the CJ representations of the CPTP maps $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{G}_{\mathcal{A},\mathcal{B}}$. By linearity we can represent the mapping from A, B to $G_{A,B}$ as

$$G_{A,B} = \mathrm{tr}_{A_I A_O B_I B_O}[W^{T_{A_I A_O B_I B_O}}(A \otimes B)], \tag{4.1}$$

where $W \in P \otimes F \otimes A_I \otimes A_O \otimes B_I \otimes B_O$ is the process matrix we are defining, and identity matrices on the subsystems P, F, A'_I, A'_O, B'_I, and B'_O are left implicit. This formula can be rewritten in a more convenient way using the link product, which was designed to conveniently express the CJ representation of a composition of quantum operations [32], and we recap in appendix S:

$$G_{A,B} = W * (A \otimes B) \tag{4.2}$$

We need $G_{A,B}$ to be a valid CPTP map for all valid CPTP maps A and B. This imposes the following restrictions on W, which we derive in appendix T:

$$W \geq 0 \tag{4.3}$$

$$\mathrm{tr}\, W = d_{A_O} d_{B_O} d_P \tag{4.4}$$

$$W = L_V(W), \tag{4.5}$$

where L_V is a projector on the linear subspace of valid process matrices defined in equation (T.20).

As an explicit example of a process defined as a higher-order transformation consider a situation where a state in $P = P_1 \otimes P_2$ is sent to Alice and Bob, and the resulting state from Alice and Bob is sent to $F = F_1 \otimes F_2$. The process matrix is given by $|w_{\text{state}}\rangle\langle w_{\text{state}}|$, where

$$|w_{\text{state}}\rangle = |\mathbb{1}\rangle\!\rangle^{P_1 A_I} |\mathbb{1}\rangle\!\rangle^{P_2 B_I} |\mathbb{1}\rangle\!\rangle^{A_O F_1} |\mathbb{1}\rangle\!\rangle^{B_O F_2}, \tag{4.6}$$

and $|\mathbb{1}\rangle\!\rangle$ is the "pure" CJ representation of the identity map between the appropriate Hilbert spaces, as defined in Appendix R.

A less trivial example is the process where a state in P goes first to Alice, then to Bob, and finally to F. Using again the vector representation the process is

$$|w_{\text{channel}}\rangle = |\mathbb{1}\rangle\!\rangle^{P A_I} |\mathbb{1}\rangle\!\rangle^{A_O B_I} |\mathbb{1}\rangle\!\rangle^{B_O F}. \tag{4.7}$$

Finally, an example of a process that encodes an indefinite causal order is the quantum switch [1, 80], which in this representation is:

$$\begin{aligned} |w_{\text{switch}}\rangle = |0\rangle^{P_1} |\mathbb{1}\rangle\!\rangle^{P_2 A_I} |\mathbb{1}\rangle\!\rangle^{A_O B_I} |\mathbb{1}\rangle\!\rangle^{B_O F_2} |0\rangle^{F_1} \\ + |1\rangle^{P_1} |\mathbb{1}\rangle\!\rangle^{P_2 B_I} |\mathbb{1}\rangle\!\rangle^{B_O A_I} |\mathbb{1}\rangle\!\rangle^{A_O F_2} |1\rangle^{F_1} \end{aligned} \tag{4.8}$$

To connect this version of the quantum switch with the one in Refs. [1, 80], one should prepare the control qubit and send it to P_1, and prepare the target qubit and send it to P_2. After the process is done the control qubit will be found in F_1 (unchanged), and the target qubit in F_2.

This definition of processes allows a natural definition of what it means for a process to be "pure". Just as we can define unitaries as the most general linear transformations that map pure states to pure states of the same dimension[1], we can define pure processes to be most general linear transformations that map unitaries to unitaries. More precisely:

Definition 5. *A process W is pure if for all unitaries $\mathcal{A}$, $\mathcal{B}$ the resulting transformation $\mathcal{G}_{\mathcal{A},\mathcal{B}}$ is also a unitary.*

It turns out that pure processes are unitary transformations from A_O, B_O, P to A_I, B_I, F and can be conveniently represented as vectors, as shown in the following theorem:

Theorem 6. *A process W is pure if and only if $W = |U_w\rangle\!\rangle\langle\!\langle U_w|$ for some unitary U_w.*

Proof. If W is pure, then $\mathcal{G}_{\mathcal{A},\mathcal{B}}$ must be a unitary in particular when $\mathcal{A}$ and $\mathcal{B}$ are SWAPs that map A'_I, A_I to A_O, A'_O and B'_I, B_I to B_O, B'_O. Then

$$G_{A,B} = W * \left(|\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|^{A'_I A_O} |\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|^{A_I A'_O} \otimes |\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|^{B'_I B_O} |\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|^{B_I B'_O} \right) = W, \tag{4.9}$$

so the resulting transformation $\mathcal{G}_{\mathcal{A},\mathcal{B}} = \mathfrak{C}^{-1}(G_{A,B}) = \mathfrak{C}^{-1}(W) =: \mathcal{W}$ is just the process itself, with the relabelling $A_I \mapsto A'_O$, $A_O \mapsto A'_I$, $B_I \mapsto B'_O$, and $B_O \mapsto B'_I$,

[1] More precisely, a linear map $\mathcal{E}$ is a unitary iff for all pure states $|\psi\rangle$ the transformed state $\mathcal{I} \otimes \mathcal{E}(|\psi\rangle\langle\psi|)$ is a pure state of the same dimension.

so $\mathcal{W}$ must be a unitary transformation from A_O, B_O, P to A_I, B_I, F. Writing $\mathcal{W}(\rho) = U_w \rho U_w^\dagger$, we have that its CJ representation is $W = \mathfrak{C}(\mathcal{W}) = |U_w\rangle\!\rangle\langle\!\langle U_w|$.

Conversely, if $W = |U_w\rangle\!\rangle\langle\!\langle U_w|$, $A = |U_a\rangle\!\rangle\langle\!\langle U_a|$, and $B = |U_b\rangle\!\rangle\langle\!\langle U_b|$, then

$$G_{A,B} = |U_w\rangle\!\rangle\langle\!\langle U_w| * (|U_a\rangle\!\rangle\langle\!\langle U_a| \otimes |U_b\rangle\!\rangle\langle\!\langle U_b|) = |U_g\rangle\!\rangle\langle\!\langle U_g|, \tag{4.10}$$

where

$$U_g^{A'_I B'_I P} = \mathrm{tr}_{A_I B_I}\left[\left(U_w^{P A_I B_I} \otimes \mathbb{1}^{A'_I B'_I}\right)\left(\mathbb{1}^P \otimes U_a^{A_I A'_I} \otimes U_b^{B_I B'_I}\right)\right]. \tag{4.11}$$

Since by assumption $\mathcal{G}_{\mathcal{A},\mathcal{B}}$ is trace preserving, we have that $\mathrm{tr}_{A'_O B'_O F}\, G_{A,B} = \mathbb{1}^{A'_I B'_I P}$. Substituting $G_{A,B} = |U_g\rangle\!\rangle\langle\!\langle U_g|$ we get

$$\mathrm{tr}_{A'_O B'_O F}\, |U_g\rangle\!\rangle\langle\!\langle U_g| = U_g^\dagger U_g = \mathbb{1}^{A'_I B'_I P}. \tag{4.12}$$

Since, furthermore, U_g is a square finite-dimensional matrix, it has a right inverse which is equal to its left inverse, so U_g is a unitary, and so is $\mathcal{G}_{\mathcal{A},\mathcal{B}}$. □

4.3 A purification postulate

With the definition of a pure process in hand, we can now define what it means to purify a process: we say that a process W with trivial P and F is purifiable if you can recover it from a pure process S by inputting the state[2] $|0\rangle$ in P and tracing out F, i.e., if

$$W = S * (|0\rangle\langle 0|^P \otimes \mathbb{1}^F), \tag{4.13}$$

as shown in Fig. 4.2.

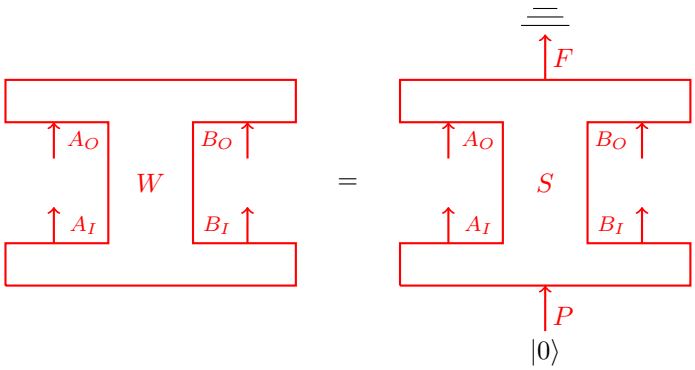


Figure 4.2: A process W with trivial P and F is purifiable if it can be recovered from a pure process S by inputting the state $|0\rangle$ in P and tracing out F.

We propose then the purification postulate: *a process is physical only if it is purifiable*. This postulate is motivated by the fact that a non-purifiable process would fundamentally map unitaries onto isometries or non-unitary CPTP maps, destroying the reversibility of the theory. Reversibility, on its turn, is a cherished

[2] We can fix the state to be $|0\rangle$ instead of allowing an arbitrary state without loss of generality.

principle: all fundamental physical theories are time reversible[3] and the hint that the formation and evaporation of a black hole might be an irreversible evolution is one of the major problems in modern physics [98, 99]. Moreover, in all reconstructions of quantum mechanics [27, 88–93] reversibility was used as a central ingredient. This supports the idea that it is indeed a fundamental part of quantum mechanics, and it should not be done away with lightly. Furthermore, although there is speculation that in a full theory of quantum gravity new degrees of freedom are created by the expansion of the Universe [100–102], in concrete models these degrees of freedom are born in the vacuum state [103], making the time evolution actually be reversible. In any case, current inflationary models using quantum field theory on a curved expanding spacetime do not feature creation of new degrees of freedom [104].

As a sanity check, note that all processes known to be physical – all processes where the order between the parties is fixed, e.g. $|w_{\text{channel}}\rangle$ (equation (4.7)), controlled by a quantum system, e.g. $|w_{\text{switch}}\rangle$ (equation (4.8)), or incoherently mixed – are purifiable.

4.4 Necessary and sufficient conditions for purification

We shall now proceed to finding necessary and sufficient conditions for a process W with trivial P and F to be purifiable, i.e., to be recoverable via equation (4.13) from a process S that is pure and valid. From Theorem 6, we know that S is pure iff $S = |U_s\rangle\!\rangle\langle\!\langle U_s|$ for some unitary U_s. Defining

$$|w_\psi\rangle := \langle\psi^*|^P |U_s\rangle\!\rangle \tag{4.14}$$

and noting that

$$|U_s\rangle\!\rangle\langle\!\langle U_s| * (|\psi\rangle\langle\psi|^P \otimes \mathbb{1}^F) = |w_\psi\rangle\langle w_\psi| * \mathbb{1}^F = \mathrm{tr}_F |w_\psi\rangle\langle w_\psi|, \tag{4.15}$$

we can rewrite equation (4.13) as

$$W = \mathrm{tr}_F |w_0\rangle\langle w_0|, \tag{4.16}$$

where $|w_0\rangle = \langle 0|^P |U_s\rangle\!\rangle$. We can also state the condition that S is valid purely in terms of the vectors $|w_\psi\rangle$. To do that, remember that S is valid iff $S * (A \otimes B)$ is a CPTP map for all CPTP maps A and B. In its turn, $S * (A \otimes B)$ is a CPTP map iff for all input states $|\psi\rangle$ its output $[S * (A \otimes B)] * |\psi\rangle\langle\psi|^P$ is a valid quantum state. Note, however, that

$$[S * (A \otimes B)] * |\psi\rangle\langle\psi|^P = (S * |\psi\rangle\langle\psi|^P) * (A \otimes B),$$

and the condition that the righ hand side is a valid quantum state for all $|\psi\rangle$, A, and B is precisely the condition that the process with trivial P given by $(S * |\psi\rangle\langle\psi|^P)$ produces a valid CPTP (with trivial P) when linked with the CPTP maps A and B, which means that the process $(S * |\psi\rangle\langle\psi|^P) = |w_\psi\rangle\langle w_\psi|$ is a valid process. Writing out the validity conditions explicitly, we have

$$\forall |\psi\rangle \quad |w_\psi\rangle\langle w_\psi| = L_V(|w_\psi\rangle\langle w_\psi|) \quad \text{and} \quad \mathrm{tr}|w_\psi\rangle\langle w_\psi| = d_{A_O} d_{B_O} \tag{4.17}$$

[3] Note that to treat the collapse of the wavefunction during a measurement as an irreversible process one needs to use objective collapse models [95–97] instead of standard quantum mechanics. We are taking the view that collapse is not a physical process.

where L_V is the projector onto the valid subspace defined in equation (T.20) particularized for $d_P = 1$.

With this, we reduce the problem of purifying a process W to finding a set of vectors $|w_\psi\rangle$ such that equations (4.16) and (4.17) are satisfied. We can simplify it further by noting that equation (4.16) is just the purification of a positive matrix, which admits a simple solution. Let then the eigendecomposition of W be

$$W = \sum_{i=0}^{r-1} \lambda_i |\lambda_i\rangle\langle\lambda_i|, \tag{4.18}$$

where r is the rank of W, and $\lambda_i, |\lambda_i\rangle$ its nonzero eigenvalues and corresponding eigenvectors. Then a valid purification for it is

$$|w_0\rangle = \sum_{i=0}^{r-1} \sqrt{\lambda_i} |\lambda_i\rangle |i\rangle, \tag{4.19}$$

which is unique modulo some isometry V on the purifying system. But this isometry has no effect on the validity of $|w_\psi\rangle\langle w_\psi|$, as it only affects the state output by the process, and an isometry maps valid quantum states to valid quantum states. This implies that we can choose without loss of generality the isometry to be the identity, and the dimension of the purifying system $d_F = d_P$ to be the rank of W. This allows us to use a r^2-dimensional basis for P to simplify condition (4.17). Choosing $\{|i\rangle\langle j|\}_{i,j=0}^{r-1}$ as this basis, and rewritting the equations explicitly in terms of the vectors $|w_i\rangle = \langle i|^P |U_s\rangle\!\rangle$, the condition translates to

$$\forall i, j \quad L_V^\perp(|w_i\rangle\langle w_j|) = 0 \quad \text{and} \quad \langle w_i|w_j\rangle = d_{A_O} d_{B_O} \delta_{ij},$$

where for brevity we are using the projector $L_V^\perp := \mathbb{1} - L_V$.

We summarize the results of this section in the following theorem:

Theorem 7. *A process matrix W of rank r and eigendecomposition*

$$W = \sum_{i=0}^{r-1} \lambda_i |\lambda_i\rangle\langle\lambda_i| \tag{4.20}$$

is purifiable if and only if there exists a set of vectors $\{|w_i\rangle\}_{i=0}^{r-1}$ such that

$$|w_0\rangle = \sum_{i=0}^{r-1} \sqrt{\lambda_i} |\lambda_i\rangle |i\rangle \tag{4.21}$$

and

$$\forall i, j \quad L_V^\perp(|w_i\rangle\langle w_j|) = 0 \quad \textit{and} \quad \langle w_i|w_j\rangle = d_{A_O} d_{B_O} \delta_{ij} \tag{4.22}$$

If they exist, a process $S = |U_s\rangle\!\rangle\langle\!\langle U_s|$ that purifies W is given by

$$\langle i|^P |U_s\rangle\!\rangle := |w_i\rangle \tag{4.23}$$

4.5 Necessary conditions for purification

It is not easy to find a set of vectors $\{|v_i\rangle\}$ that satisfies the conditions of Theorem 7, as the constraint $L_V^\perp(|v_i\rangle\langle v_j|) = 0$ is quadratic on them. We shall not solve this

problem in full, but rather we shall prove an upper bound on the number of vectors that can satisfy those conditions for a given process. If this upper bound is smaller than the rank of the process, this is a proof that this process is not purifiable.

We shall start by focussing in the simplest of the conditions, namely that $L_V^\perp(|v_i\rangle\langle w_0|) = 0$. Since this condition is linear on $|v_i\rangle$ it is straightforward to characterize the vector space V_W formed by the vectors that satisfy it. Restricting our attention to this (hopefully small) subspace makes it easier to consider the other, non-linear, conditions.

To construct an orthonormal basis for V_W, first we upscale the projector $L_V^\perp$ to act on the "double-kets" of process matrices, i.e., we define the matrix $\Pi_{L_V^\perp}$ such that[4]

$$\forall i,j \quad \Pi_{L_V^\perp}\Big||i\rangle\langle j|\Big\rangle\Big\rangle = \Big|L_V^\perp(|i\rangle\langle j|)\Big\rangle\Big\rangle$$

We have then

$$L_V^\perp(|v\rangle\langle w_0|) = 0 \iff \Pi_{L_V^\perp}|w_0^*\rangle|v\rangle = 0,$$

so if we define

$$O_W := \big(\langle w_0^*| \otimes \mathbb{1}\big)\Pi_{L_V^\perp}\big(|w_0^*\rangle \otimes \mathbb{1}\big)$$

we have that

$$O_W|v\rangle = 0 \iff \Pi_{L_V^\perp}|w_0^*\rangle|v\rangle = 0,$$

so the null eigenvectors of O_W form an orthonormal basis for V_W. With it in hand we can now restrict the projector $\Pi_{L_V^\perp}$ to the subspace[5] $V_W^* \otimes V_W$. Let $\{|r_i\rangle\}$ be an orthonormal basis for V_W, and $R = \sum_i |i\rangle\langle r_i|$. The restriction $\Pi_{L_V^\perp|V_W}$ is then given by

$$\Pi_{L_V^\perp|V_W} = (R^* \otimes R)\,\Pi_{L_V^\perp}\,(R^{*\dagger} \otimes R^\dagger)$$

We can now consider the condition that $L_V^\perp(|v_i\rangle\langle v_j|) = 0$. Let $\{|m_k\rangle\}_{k=0}^{d_m-1}$ be the set of eigenvectors of $\Pi_{L_V^\perp|V_W}$ with eigenvalue different than[6] 0. Then

$$\Pi_{L_V^\perp|V_W}|v_j^*\rangle|v_i\rangle = 0 \iff \forall k \quad \langle v_j^*|\langle v_i||m_k\rangle = 0$$

We want to rewrite the inner product $\langle v_j^*|\langle v_i||m_k\rangle$ in a more convenient way. For that, note that $\langle\psi|A|\psi\rangle = \langle\psi^*|\langle\psi||A^T\rangle\rangle$, so $\langle v_j^*|\langle v_i||m_k\rangle = \langle v_i|M_k|v_j\rangle$ for matrices M_k such that $|M_k^T\rangle\rangle = |m_k\rangle$. These matrices are in general not Hermitian, so for convenience we define $C_k = M_k + M_k^\dagger$ and $C_{k+d_m} = i(M_k - M_k^\dagger)$. Then

$$\Pi_{L_V^\perp|V_W}|v_j^*\rangle|v_i\rangle = 0 \iff \forall k \quad \langle v_i|C_k|v_j\rangle = 0$$

Let now d_{C_k} be the dimension of the largest subspace such that all $|v\rangle$ in this subspace respect $\langle v|C_k|v\rangle = 0$. We can easily calculate d_{C_k} using the the null-square lemma proved in Appendix U: let n_{k+}, n_{k-}, and n_{k0} be the number of positive, negative, and null eigenvalues of C_k. Then

$$d_{C_k} = n_{k0} + \min\{n_{k+}, n_{k-}\}$$

[4] $\Pi_{L_V^\perp}$ can be explicitly written as $\Pi_{L_V^\perp} = \mathfrak{C}(L_V^\perp)^R$, where R is the reshuffling operation [35].

[5] One can restrict instead to $V_W \setminus |w_0\rangle$ for a slight computational advantage.

[6] It is possible that this set is empty. In this case all vectors in V_W respect the conditions of Theorem 7, and the process is purifiable iff $\dim(V_W) \geq \mathrm{rank}(W)$.

A simple upper bound on the dimension of the largest subspace of vectors that respect the conditions of Theorem 7 is then

$$d_{\max}(W) := \min_i d_{C_i} \tag{4.24}$$

Therefore if $d_{\max}(W) < \operatorname{rank}(W)$ the process W is not purifiable.

4.6 Examples

In this section we shall omit superscripts that identify subsystems and tensor products to avoid clutter. The expression (e.g.) $Z\mathbb{1}XZ$ should be understood as $Z^{A_I} \otimes \mathbb{1}^{A_O} \otimes X^{B_I} \otimes Z^{B_O}$.

Let

$$W_{\text{OCB}} = \frac{1}{4}\left[\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1} + \frac{\mathbb{1}ZZ\mathbb{1} + Z\mathbb{1}XZ}{\sqrt{2}}\right] \tag{4.25}$$

be the process introduced in Ref. [3]. It was proven, under a restriction on the allowed operations, to produce the maximal violation of the original causal inequality for any dimension [46]. Since $\operatorname{rank}(W_{\text{OCB}}) = 8$ and $d_{\max}(W_{\text{OCB}}) = 5$ it is not purifiable.

Let

$$\begin{aligned} W_{\max} = \frac{1}{4}\Big[&\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1} + a_0\, Z\mathbb{1}Z\mathbb{1} - a_1\big(Z\mathbb{1}\mathbb{1}\mathbb{1} + \mathbb{1}\mathbb{1}Z\mathbb{1}\big) \\ &- a_2\big(Z\mathbb{1}\mathbb{1}Z + \mathbb{1}ZZ\mathbb{1}\big) + a_3\big(Z\mathbb{1}ZZ + ZZZ\mathbb{1}\big) \\ &+ a_4\big(Z\mathbb{1}XX - Z\mathbb{1}YY + XXZ\mathbb{1} - YYZ\mathbb{1}\big)\Big], \end{aligned} \tag{4.26}$$

where $a_0 \approx 0.2744$, $a_1 \approx 0.2178$, $a_2 \approx 0.3628$, $a_3 \approx 0.3114$, and $a_4 \approx 0.2097$. This process was introduced in Appendix C of Ref. [49] and conjectured to produced the maximal violation of GYNI and LGYNI inequalities for its dimension. Since $\operatorname{rank}(W_{\max}) = 11$ and $d_{\min}(W_{\max}) = 10$ it is not purifiable.

Let

$$W_{\text{opt}} = \frac{1}{4}\left[\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1} + \frac{1}{\sqrt{3}}\, Z\mathbb{1}XZ + \frac{\sqrt{3}-1}{3}\big(\mathbb{1}X\mathbb{1}X + \mathbb{1}Y\mathbb{1}Y + \mathbb{1}Z\mathbb{1}Z\big)\right] \tag{4.27}$$

be the process introduced in Ref. [94]. By itself it cannot violate any causal inequalities, but it becomes able to do so when it is extended with an entangled ancilla. However, when admixed with a small amount of noise, the known violations disappear, so this noisy version was conjectured to be unable to violate any causal inequalities. That is, W_{opt} was conjectured to be "extensibly causal". Since extensibly causal processes seem physically reasonable, but we have no physical interpretation for W_{opt} or its noisy version, it would very interesting to test its purifiability to gather evidence about its physicality. Unfortunately, since $\operatorname{rank}(W_{\text{opt}}) = 12$ and $d_{\max}(W_{\text{opt}}) = 17$ our test is inconclusive. This is also the case for the noisy version, which has rank 16 and $d_{\max}$ equal to 41.

4.7 Counter-example

Since the processes examined in the previous section either are not purifiable or have unknown status, it might be tempting to conjecture that all processes that

can violate causal inequalities are not purifiable. While this might be true for the bipartite case, we have a counter-example for the general case: a tripartite process which can violate causal inequalities and is purifiable.

This process, which was first published in Ref. [78], can be concisely described as a function from A_O, B_O, C_O to A_I, B_I, C_I that incoherently maps the basis states $|a, b, c\rangle$ to the basis states $|\neg b \wedge c, \neg c \wedge a, \neg a \wedge b\rangle$, where $\neg$ and $\wedge$ represent logical negation and logical conjunction. It can be represented in a more cumbersome way as the process matrix

$$W_{\text{det}} = \sum_{abc} |a, b, c\rangle\langle a, b, c| \otimes |\neg b \wedge c, \neg c \wedge a, \neg a \wedge b\rangle\langle \neg b \wedge c, \neg c \wedge a, \neg a \wedge b| \tag{4.28}$$

where unlike in the other processes we wrote the subsystems in the order A_O, B_O, C_O, A_I, B_I, C_I to match the function described above. To purify this process [105] we use the standard trick to turn a irreversible function into a reversible one, that is, changing the function from $|x\rangle \mapsto |f(x)\rangle$ to $|x\rangle|y\rangle \mapsto |x\rangle|y \oplus f(x)\rangle$. In terms of our function this means mapping $|a, b, c\rangle|i, j, k\rangle$ to $|a, b, c\rangle|i \oplus \neg b \wedge c, j \oplus \neg c \wedge a, k \oplus \neg a \wedge b\rangle$. This reversible function is then the purification of W_{det}, and can be written as the process vector

$$|w_{\text{det}}\rangle = \sum_{\substack{abc \\ ijk}} |a, b, c\rangle|i, j, k\rangle \otimes |a, b, c\rangle|i \oplus \neg b \wedge c, j \oplus \neg c \wedge a, k \oplus \neg a \wedge b\rangle \tag{4.29}$$

where the subsystems are written in the order A_O,B_O,C_O,P,F,A_I,B_I,C_I. Note that unlike the other subsystems P and F consist of three qubits each.

4.8 Conclusion

We have proposed a purification postulate, motivated by considerations that are independent of this formalism, that allows us for the first time to declare a large class of processes to be unphysical. This can be made part of a axiomatic, top-down approach to determining the set of physical processes. Ideally, one should marry it with a bottom-up approach, that shows processes to be physical by actually constructing laboratory implementations for them. As of yet, however, there are still processes which are neither excluded by the purification principle nor known to be implementable, showing that more work must be done in order to consumate this marriage.

In this gray zone we find, for example, $|w_{\text{det}}\rangle$ (equation (4.29)), showing that the set of purifiable processes is rather rich, and that the purification principle does not rule out the violation of causal inequalities in Nature. It might still be the case, however, that processes that cannot violate causal inequalities, i.e., extensibly causal processes, are always purifiable. In order to decide this it would be interesting to know whether the noisy of version of W_{opt} (equation (4.27)) is purifiable.

Chapter 5

Quantum computation with linear closed timelike curves

In this chapter we consider how to use indefinite causal structures as a computational resource. The goal is twofold: if it turns out that a indefinite causal structure that is known to be physical allows for the solution of a practical problem asymptotically more efficiently than ordered causal structures, that would be a huge discovery. On the other hand, if some indefinite causal structure of unknown physicality is shown to be extremely powerful, for example by allowing the efficient solution of NP-complete problems, this would be evidence against its physical plausibility.

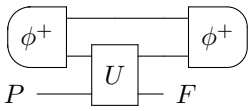
So far, the only known result about this subject is that the quantum switch reduces the query complexity of a specific problem from $O(n^2)$ queries to $O(n)$ queries [28]. To go beyond such isolated cases and investigate the power of the process matrix formalism as a whole one needs a well-defined computational model. Here we are going to propose such a model, using a simulation of arbitrary process matrices with closed timelike curves. This model allows us to prove the first known upper bound on the computational power of process matrices, showing it to be upper bounded by PostBQP [106].

A different computational model also based on the process matrix formalism was proposed in [107, 108]. Note, however, that this model restricts which operations one is allowed to perform inside the indefinite causal structure, and is therefore more contrived.

5.1 The Bennett-Schumacher-Svetlichny model of closed timelike curves

The model of closed timelike curves we use was proposed independently by Bennett and Schumacher [109] and Svetlichny [110] (see also [111–113]). It is motivated by the observation that, in the quantum teleportation protocol, one does not need to do any correction on the target state when the outcome of the Bell measurement is $|\phi^+\rangle$. In this case one can write the target state as already being there before the measure was made, and even before the state to be teleported was even created. This post-selected teleportation protocol is then used to teleport the output state of a part of the output of a unitary matrix to a part of its input, whereas the rest of the unitary simply maps a quantum system from the past P to the future F, as shown

in the circuit (5.1).



(5.1)

Naming the subsystems as $1, 2$, and 3 (from bottom to top), the operator that maps the chronology-respecting system from the past to the future is then

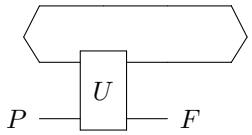
$$K = \langle\phi^+|^{23} U^{12} |\phi^+\rangle^{23} \propto \mathrm{tr}_2\, U^{12}. \tag{5.2}$$

Since a partial trace of a unitary is in general not a unitary, K will not necessarily preserve the norm of a input state, so for the evolution to be valid we will need to renormalize it. The evolution of a state $|\psi\rangle$ input in P is therefore defined as

$$|\psi\rangle \mapsto \frac{1}{\|K|\psi\rangle\|_2} K|\psi\rangle, \tag{5.3}$$

which of course is undefined when $\|K|\psi\rangle\|_2 = 0$, illustrating how pathological this model is.

This post-selected closed timelike curve (known as P-CTC) is usually represented as



(5.4)

with the renormalization implicitly included.

Note that although we used pure states and unitary transformations, this formalism of P-CTCs can be easily extended to mixed states and CPTP maps.

5.2 Processes as linear closed timelike curves

Now we proceed to show how having access to P-CTCs allows one to implement any process, or equivalently how to simulate any process using post-selection. For that, we shall use the processes as defined in equation (4.1), which says that an operator $W \in P \otimes F \otimes A_I \otimes A_O \otimes B_I \otimes B_O$ is a process matrix if and only if for all CPTP maps $\mathcal{A} : A_I \to A_O$ and $\mathcal{B} : B_I \to B_O$ the operator[1]

$$G_{A,B} = \mathrm{tr}_{A_I A_O B_I B_O}\left[W^{T_{A_I A_O B_I B_O}} (A \otimes B)\right] \tag{5.5}$$

is a valid CPTP map, where

$$\begin{aligned} A &= \mathcal{I} \otimes \mathcal{A}(|\mathbb{1}^{A_I}\rangle\!\rangle\langle\!\langle \mathbb{1}^{A_I}|) \in A_I \otimes A_O \\ B &= \mathcal{I} \otimes \mathcal{B}(|\mathbb{1}^{B_I}\rangle\!\rangle\langle\!\langle \mathbb{1}^{B_I}|) \in B_I \otimes B_O \\ G_{A,B} &= \mathcal{I} \otimes \mathcal{G}_{\mathcal{A},\mathcal{B}}(|\mathbb{1}^{P}\rangle\!\rangle\langle\!\langle \mathbb{1}^{P}|) \in P \otimes F \end{aligned}$$

are the Choi-Jamiołkowski (CJ) representations of the CPTP maps $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{G}_{\mathcal{A},\mathcal{B}}$, which we review in Appendix R.

[1] We shall restrict ourselves to the bipartite case and omit the auxilliary Hilbert spaces A'_I, A'_O, B'_I, and B'_O to reduce clutter, which is already excessive. Taking them into account doesn't change anything fundamental in the argument.

Using the fact that the trace can be seen as the Hilbert-Schmidt inner product, under which transposition is a self-adjoint map, we can rewrite equation (5.5) as

$$G_{A,B} = \mathrm{tr}_{A_I A_O B_I B_O}\left[W(A^T \otimes B^T)\right]. \tag{5.6}$$

Substituting the definition of the CJ isomorphism for A and B and noting that

$$\left(\mathcal{I} \otimes \mathcal{A}(|\mathbb{1}^{A_I}\rangle\!\rangle\langle\!\langle \mathbb{1}^{A_I}|)\right)^T = \mathcal{A}^* \otimes I(|\mathbb{1}^{A_O}\rangle\!\rangle\langle\!\langle \mathbb{1}^{A_O}|), \tag{5.7}$$

where $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$, we arrive at

$$G_{A,B} = \mathrm{tr}_{A_I A_O B_I B_O}\left[W\left(\mathcal{A}^* \otimes \mathcal{I}(|\mathbb{1}^{A_O}\rangle\!\rangle\langle\!\langle \mathbb{1}^{A_O}|)\right) \otimes \left(\mathcal{B}^* \otimes \mathcal{I}(|\mathbb{1}^{B_O}\rangle\!\rangle\langle\!\langle \mathbb{1}^{B_O}|)\right)\right], \tag{5.8}$$

which after collecting terms becomes

$$G_{A,B} = \mathrm{tr}_{A_I A_O B_I B_O}\left[W\left((\mathcal{A}^* \otimes \mathcal{B}^*) \otimes \mathcal{I}(|\mathbb{1}^{A_O B_O}\rangle\!\rangle\langle\!\langle \mathbb{1}^{A_O B_O}|)\right)\right]. \tag{5.9}$$

As noted in Appendix T, W must be itself the CJ representation of a CPTP map $\mathcal{W}$ from $P \otimes A_O \otimes B_O$ to $F \otimes A_I \otimes B_I$, so defining it via

$$W = \mathcal{W} \otimes \mathcal{I}(|\mathbb{1}^{A_O B_O P}\rangle\!\rangle\langle\!\langle \mathbb{1}^{A_O B_O P}|) \tag{5.10}$$

and substituting this in equation 5.9 shows us that $G_{A,B}$ is equal to

$$\mathrm{tr}_{A_I A_O B_I B_O}\left[\left(\mathcal{W} \otimes \mathcal{I}(|\mathbb{1}^{A_O B_O P}\rangle\!\rangle\langle\!\langle \mathbb{1}^{A_O B_O P}|)\right)\left((\mathcal{A}^* \otimes \mathcal{B}^*) \otimes \mathcal{I}(|\mathbb{1}^{A_O B_O}\rangle\!\rangle\langle\!\langle \mathbb{1}^{A_O B_O}|)\right)\right] \tag{5.11}$$

which, after taking the adjoint of the map $\mathcal{A}^* \otimes \mathcal{B}^*$ to the other side of the inner product, becomes

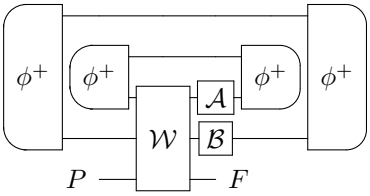
$$\mathrm{tr}_{A_I A_O B_I B_O}\left[\left(\left((\mathcal{A} \otimes \mathcal{B}) \circ \mathcal{W}\right) \otimes \mathcal{I}(|\mathbb{1}^{A_O B_O P}\rangle\!\rangle\langle\!\langle \mathbb{1}^{A_O B_O P}|)\right)|\mathbb{1}^{A_O B_O}\rangle\!\rangle\langle\!\langle \mathbb{1}^{A_O B_O}|\right] \tag{5.12}$$

Finally, we substitute $|\mathbb{1}^{A_O B_O}\rangle\!\rangle / \sqrt{d_{A_O} d_{B_O}}$ with the maximally entangled state $|\phi^+\rangle^{A_O B_O}$, which gives us

$$\begin{aligned} G_{A,B} = d_{A_O}^2 d_{B_O}^2\, \mathrm{tr}_{A_I A_O B_I B_O}\Big[&|\phi^+\rangle\langle\phi^+|^{A_O B_O} \\ &\times \left(\left((\mathcal{A} \otimes \mathcal{B}) \circ \mathcal{W}\right) \otimes \mathcal{I}\left(|\mathbb{1}^P\rangle\!\rangle\langle\!\langle \mathbb{1}^P| \otimes |\phi^+\rangle\langle\phi^+|^{A_O B_O}\right)\right)\Big] \end{aligned} \tag{5.13}$$

This identity can be stated more clearly as the circuit equation:

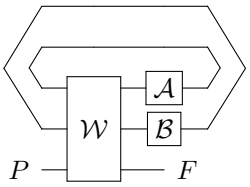
$$\mathcal{G}_{\mathcal{A},\mathcal{B}} = d_{A_O}^2 d_{B_O}^2$$



(5.14)

which shows how to simulate the map $\mathcal{G}_{\mathcal{A},\mathcal{B}}$ via post-selection: one prepares the state $|\phi^+\rangle^{A_O B_O} = |\phi^+\rangle^{A_O} \otimes |\phi^+\rangle^{B_O}$, applies the CPTP map $\mathcal{W}$ to it, applies the CPTP maps $\mathcal{A}$ and $\mathcal{B}$ to the ouput of $\mathcal{W}$, and measures the resulting state in a basis that includes $|\phi^+\rangle^{A_O B_O}$, post-selecting on obtaining it.

In the language of P-CTCs, one can also understood this as applying the CPTP maps $\mathcal{A}$ and $\mathcal{B}$ to the output of $\mathcal{W}$, and using a P-CTC to teleport the output of$\mathcal{A}$ and $\mathcal{B}$ to the input of $\mathcal{W}$. It can be represented more compactly as

$\mathcal{G}_{\mathcal{A},\mathcal{B}} =$ 

(5.15)

The crucial point to notice is that the probability of post-selection is always $1/d_{A_O}^2 d_{B_O}^2$, independently of the CPTP maps $\mathcal{A}$ and $\mathcal{B}$ or the state ρ input in P, which makes the renormalization of this circuit, and therefore the mapping $\mathcal{A}, \mathcal{B}, \rho \mapsto \mathcal{G}_{\mathcal{A},\mathcal{B}}(\rho)$ a *linear* transformation. One could, instead, let $\mathcal{W}$ be an arbitrary CPTP map, as in a general P-CTC, and define the transformation from P to F as done in equation (5.3):

$$\rho \mapsto \frac{\mathcal{G}_{\mathcal{A},\mathcal{B}}(\rho)}{\operatorname{tr} \mathcal{G}_{\mathcal{A},\mathcal{B}}(\rho)}, \tag{5.16}$$

This transformation is linear only if the denominator is a constant, that does not depend on $\mathcal{A}$, $\mathcal{B}$, and ρ. But this is the case only if $\mathcal{G}_{\mathcal{A},\mathcal{B}}$ is proportional to a CPTP map, for all $\mathcal{A}$ and $\mathcal{B}$, which is simply the definition of a process, modulo an irrelevant multiplicative constant.

This shows that process matrices are exactly the linear particular case of P-CTCs, when those are defined with the variable gates $\mathcal{A}$ and $\mathcal{B}$. Since their nonlinearity is the source of major pathologies, such as the ability to solve NP-complete problems and to distinguish non-orthogonal states, one can expect the process matrices to be exactly the non-pathological P-CTCs, that however still allow for indefinite causal order.

5.3 The computational model

Now we are in a position to define the complexity class of the problems that can be efficiently solved with process matrices. We shall give two equivalent definitions, one based directly on the definition of a process matrix, and the other based on its interpretation in terms of P-CTCs. In both of them we shall use as basic element the operator

$$W \in P \otimes F \otimes A_I^1 \otimes A_O^1 \otimes A_I^2 \otimes A_O^2 \otimes \ldots \otimes A_I^k \otimes A_O^k \tag{5.17}$$

which, as above, shall be understood as the CJ representation of a CPTP map from $P, A_O^1, A_O^2 \ldots, A_O^k$ to $F, A_I^1, A_I^2 \ldots, A_I^k$, where P, F correspond to the normal past and future of the circuit, and the other k subsystems are where one can apply an arbitrary CPTP map $A_i \in A_I^i \otimes A_O^i$ in a causally "suspect" way (that is, by fiat as in equation (5.5) or by using a P-CTC as in equation (5.15)). The output of the circuit is then determined by a measurement on one of the qubits in F. If the result of the measurement is 1 we say that the circuit accepts input x, and if the result is 0 we say that input x was rejected.

The complexity class is then:

Definition 8. BQP$_{\ell\text{CTC}}$ *is the class of languages* L *for which there exists a uniform*[2] *family of polynomial-size quantum circuits* $W = \{W_n\}_{n\in\mathbb{N}}$ *such that for all inputs* x *of size* n

1. $G_n = \mathrm{tr}_{1,2,\ldots,k}\left[W_n^{T_{1,2,\ldots,k}}(A_1 \otimes A_2 \otimes \ldots A_k)\right]$ *is a CPTP map for all CPTP maps* A_i*.*
2. *If* $x \in L$ *then* $P[W$ *accepts* $x] \geq \frac{2}{3}$*.*
3. *If* $x \notin L$ *then* $P[W$ *rejects* $x] \geq \frac{2}{3}$*.*

It should be clear, given the interpretation of processes in terms of P-CTCs shown in the previous section, that this definition is equivalent to:

Definition 9. BQP$_{\ell\text{CTC}}$ *is the class of languages* L *for which there exists a uniform family of polynomial-size quantum circuits* $W = \{W_n\}_{n\in\mathbb{N}}$ *such that for all inputs* x *of size* n*.*

1. *For all CPTP maps* $A_1, A_2, \ldots, A_k$ *the probability that all post-selections succeed is* $1/d^{2k}$*, where* d *is the dimension of every subsystem.*
2. *If* $x \in L$ *then given that all post-selections succeed* $P[W$ *accepts* $x] \geq \frac{2}{3}$*.*
3. *If* $x \notin L$ *then given that all post-selections succeed* $P[W$ *rejects* $x] \geq \frac{2}{3}$*.*

Note that although this computational model is motivated by the pure processes defined in [114], we are not demanding W to be a unitary (corresponding to a pure process). It would be interesting to know whether demanding this would change the computational power of the model. It is vital, however, to allow the maps $A_1, A_2, \ldots, A_k$ to be general CPTP maps and not only unitaries, as we want the quantum circuit W to generate consistent probabilities also when the local operations A_i are subject to noise, which we model precisely by making them be CPTP maps, or, equivalently, by extending them with an ancilla.

It should be immediately clear that BQP $\subseteq$ BQP$_{\ell\text{CTC}}$, as one obtains a normal BQP circuit simply by setting $k = 0$ or by simulating a causally ordered circuit using the P-CTCs.

Another consequence of this definition is that for BQP$_{\ell\text{CTC}}$ to have any advantage over BQP, the number of P-CTCs k must increase faster than logarithmically with the size of the input n. This is because one can simply repeat the circuit until all post-selections succeed, and if the number of repetitions is small enough then the repeated circuit will still be in BQP. More precisely, the probability that all post-selections succeed simultaneously is 2^{-2k}, so repeating the circuit a number of times greater than $\log(3)2^{2k}$ guarantees that the probability that at all post-selections succeed simultaneously at least once is greater than $2/3$. Since this is polynomial in n if k grows logarithmically with n or slower, the repeated circuit is simply a valid BQP circuit.

This stands in stark contrast with the case of general P-CTCs, where having access to a single one is already enough to access the power of the whole complexity class, as noted by Lloyd *et al.* [112] and proven in more detail by Brun and Wilde [113]. The argument is simply that the power of quantum computation with P-CTCs is the same as the power of quantum computation with post-selection,

[2] That is, whose description can be generated by a Python program in time polynomial in n.

and that a single P-CTC is sufficient to simulate any desired post-selection. Since post-selection is also sufficient to simulate the whole $\mathsf{BQP}_{\ell\mathsf{CTC}}$, it is clear that the complexity class of quantum computation with unrestricted post-selection – called PostBQP – contains $\mathsf{BQP}_{\ell\mathsf{CTC}}$. In its turn, PostBQP was shown by Aaronson [106] to be equal to the rather powerful complexity class PP [115], which can be seen as a pathological version of BPP, that does not require the probabilities of acceptance and rejection to be bounded away from $1/2$.

Summarizing the conclusions of the previous paragraph, we have that[3]

$$\mathsf{BQP}_{\ell\mathsf{CTC}} \subseteq \mathsf{PostBQP} = \mathsf{PP},$$

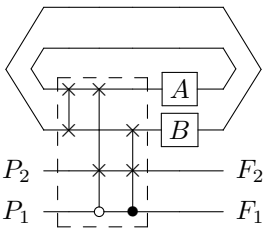
which is a rather unsatisfactory upper bound, as PP allows for example the solution of NP-complete problems, which we don't expect to be possible in a physically realizable computational model. A somewhat tighter upper bound is the class $\mathsf{PostBQP}_{\mathsf{size}}$ defined by Morimae and Nishimura [117], which is PostBQP with the restriction that the probability of successful post-selection can only depend on the size of the input. Since this is the case for $\mathsf{BQP}_{\ell\mathsf{CTC}}$, we have that

$$\mathsf{BQP}_{\ell\mathsf{CTC}} \subseteq \mathsf{PostBQP}_{\mathsf{size}} \subseteq \mathsf{PostBQP}.$$

This containment is not very useful, however, as little is known about the class $\mathsf{PostBQP}_{\mathsf{size}}$.

5.4 Examples

It might be enlightening to explicitly write a couple of processes as a circuit that includes P-CTCs. As a first example, consider the quantum switch, as defined in equation (4.8). Its CTC representation is

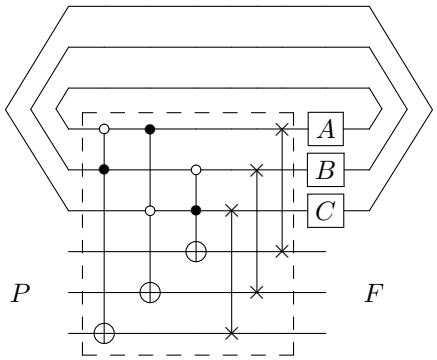


where P_1 is the control qubit, and P_2 is the target system. The dashed box singles out the unitary that is identified with the process matrix of the quantum switch.

It should be clear that using the strategy defined in Appendix E one can build a family of circuits that can efficiently create superpositions of n causal orders, by using n P-CTCs. This family of circuits, in its turn, can be used to solve the computational problem proposed in [28]. For $n = 2$, the solution is obtained by applying a Hadamard to the qubit F_1 and then measuring it in the computational basis. For general n, the solution is obtained by doing a Fourier transform on the control qudit and measuring it in the computational basis.

[3] A technical detail is that for the equality PostBQP $=$ PP to hold one needs to restrict the post-selection probabilities to be exponentially small or greater (when nonzero) [116]. Since this restriction is automatically obeyed by $\mathsf{BQP}_{\ell\mathsf{CTC}}$ we don't need to worry about it.

The second example is a pure tripartite process capable of violating causal inequalities, defined in equation (4.29). Its CTC representation is



It would be interesting to generalise this circuit to n P-CTCs and find a quantum algorithm for which it is useful.

5.5 Conclusion

We have defined a complexity class for quantum computation with process matrices, and shown that its power is upperbounded by that of PostBQP. Since PostBQP is an extremely powerful complexity class, we expect this upper bound to be rather loose. One possible way to tighten it would be by optimizing the scheme for implementing processes matrices via post-selection. The scheme we showed has probability of success $1/d^{2k}$ independently of the process, which is clearly non-optimal, since there are process matrices that can be implemented without any post-selection at all, such as causally ordered ones. If one would find, for example, a scheme to implement any process with a probability of post-selection success that is polynomially small in k, this would be a proof that process matrices are no more powerful than causally ordered circuits, up to a polynomial factor.

Conclusion

The study of causal relations in quantum mechanics is a rather new field. By the time I started my PhD there existed only a few articles written about it, and the formalism used to describe causal structures was not well-developed. For this reason, a big part of my PhD was dedicated to develop the basic technical tools to investigate it, such as the causal witnesses and the causal inequalities. This means also that several basic results were still not known, such as the computational advantage allowed by the quantum switch, the connection with P-CTCs, and the fact that several process matrices are not compatible with reversiblity.

I see this thesis, therefore, not as giving a final word about indefinite causal structures, but rather as laying the groundwork for future research. We have seen that the set of process matrices is extraordinarily rich, and though many answers about it were given, there are certainly many interesting questions that still haven't been even asked. The two specific questions that motivated the thesis – which indefinite causal structures exist in reality, and what is their computational power – are still not answered, but I believe that with the tools presented here they will be decided sooner rather than later.

In the absence of more evidence, we can speculate that the computational power of process matrices will turn out to be similar to what has been shown in Chapter 1: allowing for a polynomial speed-up over BQP, but no more than this. Such a speed up would not create a new computational class, but still might be of practical interest: remember that Grover's algorithm only reduces the number of queries necessary to do unstructured search from $O(n)$ to $O(\sqrt{n})$, but it was nevertheless revolutionary. Furthermore, if process matrices do turn out to allow only this polynomial speed up, it would be a good argument in favour of their plausibility. An obvious avenue for future research is to use the precise computational model defined in chapter 5 to settle this question.

Another, not completely independent, avenue for future research is to pursue the approach in Chapter 4 in a manner analogous to the reconstructions of quantum mechanics: to find physical principles which characterize the set of process matrices which correspond to actual, physical causal structures. Of course, one can only be sure to have found the set of physical causal structures if one has a concrete physical theory that can produce them, such as a theory of quantum gravity. This is a very ambitious goal, but still it must be done eventually.

A final question that is left open is to find is to find Tsirelson bounds for causal inequalities, or more generally to characterize the set of correlations that can be produced by process matrices. The analogous question for the case of Bell inequalities has proven to have even practical application; even if this might turn out not to be the case for process matrices, it is still an elegant mathematical puzzle that has repeatedly resisted attempts at a full solution.

Appendix A

Existence proof of the sets of unitaries

We need to show that for every y there exists a set of unitaries that satisfies property $\mathbf{P}_y$, otherwise the problem becomes trivial. More specifically, we will show that for every integer $n \geq 1$ and every $y \in \{0, \ldots, n! - 1\}$ it is possible to find a set of n unitary matrices $\{U_i\}_0^{n-1}$ such that

$$\Pi_x = \omega^{xy} \Pi_0 \tag{A.1}$$

is true for all x for some choice of y, where

$$\Pi_x = U_{\sigma_x(n-1)} \ldots U_{\sigma_x(1)} U_{\sigma_x(0)}, \tag{A.2}$$

and $\omega = e^{i\frac{2\pi}{n!}}$.

We first present the construction for $y = 1$. In this case, to each permutation of the matrices corresponds a different phase ω^x. We start by proving that the pairwise relations

$$U_k U_j = \omega^{k!} U_j U_k \quad \text{for} \quad k > j \geq 0 \tag{A.3}$$

generate all the $n!$ required phases. In order to prove this, it is convenient to introduce an explicit labeling of the permutations. A generic permutation of the n matrices $\{U_{n-1}, \ldots, U_0\}$ is labeled by a sequence of integers $(a_{n-1}, \ldots, a_1)$, with $0 \leq a_k \leq k$, and is obtained by shifting the matrix U_k to the right a_k times, starting from $k = 1$. As an example, let us construct the four-element permutation labeled by $(3, 1, 1)$. First, we swap U_0 and U_1 (i.e., we shift U_1 one position to the right), obtaining $U_3 U_2 U_0 U_1$, then we shift U_2 one position to the right and obtain $U_3 U_0 U_2 U_1$. Finally, we shift U_3 three times to the right and get the desired permutation, $U_0 U_2 U_1 U_3$. In this procedure, each matrix U_k is swapped a_k times with matrices U_j having $j < k$. Thus, if the relations (A.3) hold, the permutation labeled by $(a_{n-1}, \ldots, a_1)$ produces a phase ω^x, where

$$x = \sum_{k=1}^{n-1} a_k k! \tag{A.4}$$

is the expansion of the integer x in the *factorial basis* (or *factoradic* expansion). For the given example, $x = 3 \times 3! + 1 \times 2! + 1 \times 1! = 21$.

We construct now a set of n unitary matrices that satisfy the pairwise relations (A.3). They are constructed from the generalized X and Z matrices

$$X = \sum_{j=0}^{n!-1} |j \oplus 1\rangle\langle j|, \tag{A.5a}$$

$$Z = \sum_{j=0}^{n!-1} \omega^j |j\rangle\langle j|, \tag{A.5b}$$

where $\oplus$ denotes the sum modulo $n!$. Note that they satisfy the commutation relation

$$ZX = \omega XZ. \tag{A.6}$$

Then we define

$$\begin{aligned} U_k &= \left(Z^{k!}\right)^{\otimes k} \otimes X \otimes \mathbb{1}^{\otimes n-k-1} \quad \text{for} \quad k < n-1, \\ U_{n-1} &= \left(Z^{(n-1)!}\right)^{\otimes n}, \end{aligned} \tag{A.7}$$

where we are using the convention that $A^{\otimes 0} = 1$. For $n = 4$, the set of unitaries defined by (A.7) reads

$$\begin{aligned} U_0 &= X \otimes \mathbb{1} \otimes \mathbb{1}, \\ U_1 &= Z \otimes X \otimes \mathbb{1}, \\ U_2 &= Z^2 \otimes Z^2 \otimes X, \\ U_3 &= Z^6 \otimes Z^6 \otimes Z^6. \end{aligned} \tag{A.8}$$

It is easy to check that the matrices generated according to these rules satisfy property (A.3). A set of n unitary matrices that satisfies property (A.1) for an arbitrary y can be found by repeating this construction with ω^y in place of ω.

Note that this construction is enough to show that there is an infinity of sets of unitaries satisfying property (A.1), since the choice of basis in the definition of the matrices X and Z in equation (A.5) is arbitrary.

This construction generates matrices with dimension $d = n!^{n-1}$, which therefore must act on $O(n^2 \log n)$ qubits. One can also see that they can be implemented with a polynomial amount of elementary gates, since they are tensor products of a linear amount of X and Z^k matrices, that can themselves be implemented with a polynomial amount of elementary gates.

One can also get constructions with the correct phases in lower dimension, which might be interesting for an experimental implementation. For example, for $n = 3$, the following 6-dimensional matrices also work:

$$\begin{aligned} U_0 &= X, \\ U_1 &= ZX, \\ U_2 &= Z^2. \end{aligned} \tag{A.9}$$

This construction saturates the lower bound $d \geq n!$ we presented in the main text. It is an interesting puzzle to find out whether there exists a construction with $d = n!$ for every n.

Appendix B

Tolerating experimental error

In an experimental implementation, no property $\mathbf{P}_y$ can be satisfied exactly, so to run this algorithm one needs to be able to tolerate some deviations from it. To do this, notice that property $\mathbf{P}_y$ is satisfied if and only if

$$\frac{1}{n!^2 d}\left\|\sum_{x=0}^{n!-1}\omega^{-xy}\Pi_x\right\|_{HS}^2 = 1, \tag{B.1}$$

where $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm. This is a simple consequence of the fact that this norm is defined through an inner product. We then reformulate the problem such that a set of unitaries satisfies the modified property $\mathbf{P}'_y$ if

$$\frac{2}{3} \leq \frac{1}{n!^2 d}\left\|\sum_{x=0}^{n!-1}\omega^{-xy}\Pi_x\right\|_{HS}^2 \leq 1 \tag{B.2}$$

for a given $y \in \{0, \ldots, n!-1\}$. The problem is still to decide which property $\mathbf{P}'_y$ the set of unitaries have, given the promise that they have one of them.

Note that, in this version of the problem, we cannot anymore use an arbitrary pure state $|\psi\rangle$ to perform the protocol, since it is not true anymore that the probabilities

$$p_s = \frac{1}{n!^2}\left\|\sum_{x=0}^{n!-1}\omega^{-xs}\Pi_x|\psi\rangle\right\|^2$$

are independent of $|\psi\rangle$. In fact, it is possible to have a set of matrices such that the *lhs* of Eq. (B.1) is arbitrarily close to 1, while p_y for some state $|\psi\rangle$ is equal to 0.

Instead, we can use the maximally mixed state, since then the outcome probabilities p_s become directly related to the Hilbert-Schmidt norm, as

$$p_s = \frac{1}{n!^2 d}\left\|\sum_{x=0}^{n!-1}\omega^{xs}\Pi_x\right\|_{HS}^2 \tag{B.3}$$

and therefore the promise implies that $p_s \geq 2/3$ for $s = y$, and $p_s < 1/3$ for $s \neq y$. These two conditions are the ones necessary for a probabilistic algorithm to work; they imply that if we run the algorithm k times and take as our guess for y the most common answer, the probability of making a mistake goes down exponentially with k.

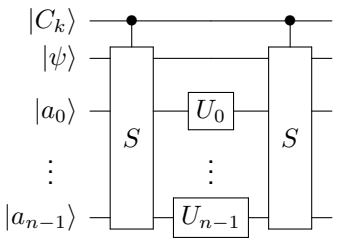
Appendix C

Implementing the n-SWITCH gate in the quantum circuit model

Here we describe how to simulate the n-SWITCH gate in the quantum circuit model. The simulation is based on the circuit presented on Ref. [8], with the difference that our scheme can be used for quantum (and not only classical) control of the order.

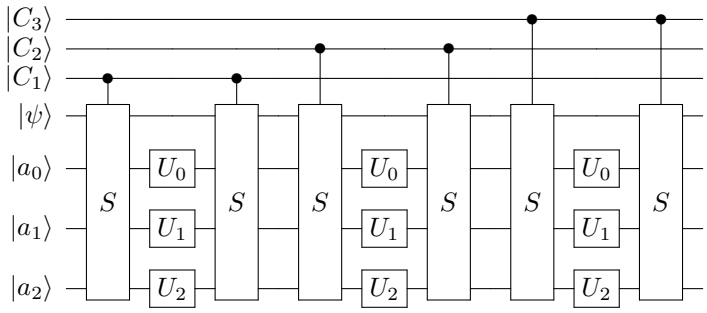
In the main text we described the control system $|C\rangle$ as encoding the permutation to be applied simply as a state running from $|0\rangle$ to $|n!-1\rangle$. Here we shall use a more convenient representation, expressing $|C\rangle = |C_1\rangle \dots |C_n\rangle$, where each $|C_k\rangle$ runs from $|0\rangle$ to $|n-1\rangle$, and indicates the unitary to be applied in the position k. For example, to encode the permutation that first applies U_1, then U_0, and then U_2, we shall write $|C\rangle = |102\rangle$. This representation requires $n\lceil \log_2 n \rceil$ qubits, a small overhead over the $\lceil \log_2 n! \rceil$ qubits that are necessary to encode a permutation of n elements. We also remark that the conversion between these two representations can be done on a classical computer in polynomial time, and therefore we shall note it no further.

The circuit consists of a composition of n instances of the following element, where k goes from 1 to n:



The $|a_i\rangle$ are ancillæ, and S is a gate that swaps $|\psi\rangle$ with the ancilla $|a_i\rangle$ controlled on $|C_k\rangle = |i\rangle$, leaving the other ancillæ invariant. This gate clearly can be implemented with a linear amount of elementary gates, so we shall not discuss its implementation. Note that each element contains n unitaries; therefore, since we need n of these elements to implement the n-SWITCH, the total number of queries in this implementation is n^2.

For example, using this construction for $n = 3$ gives us the circuit



Note that when each unitary is applied only once to the target system, i.e. when $|C\rangle$ encodes a permutation or a superposition of permutations, then the ancillæ disentangle from the control system at the end of the circuit (since U_i is applied $n-1$ times on $|a_i\rangle$, independently of $|C\rangle$). This is not the case when $|C\rangle$ encodes a more general sequence of unitaries, thus this circuit cannot be used to implement quantum control of arbitrary sequences of n unitaries[1].

If the time used by the swap gates is neglected, this circuit has running time n. It is easy to see that a circuit with this running time cannot use less than n^2 queries: in order to be possible to implement all permutations, each of the n unitaries must be available in each of the n time steps. It is however possible to reduce the number of queries at the expense of the running time (for the 2-SWITCH, an implementation with one query to U_0 and two queries to U_1 is possible, see Ref. [6].) As shown in the main text, no implementation with less than $\Omega(n^2)$ queries is possible. Note that this coincides with the definition of "gate uses" introduced in the main text.

[1] The ancillæ also disentangle for sequences in which each unitary is applied to the target a fixed number of times. For example, when $|C\rangle$ is in a superposition of $|002\rangle$, $|020\rangle$, and $|200\rangle$ the ancillæ still disentangle (in each case U_0 is applied twice and U_2 once), but not when it is a superposition of $|002\rangle$ and $|021\rangle$.

Appendix D

Relationship with other quantum algorithms

We would like to understand the relationship of our algorithm with other known classes of quantum algorithms; in particular, there are some similarities between our algorithm and those that solve the (abelian) hidden subgroup problem, and we wanted to explore how deep they are.

In the hidden subgroup problem one is given a group G and needs to find (a set of generators for) a hidden subgroup H, by using a function $f(g)$ that is constant in each coset of H, and different on different cosets. One is given access to f via a black-box unitary that does the mapping

$$U|x\rangle|y\rangle = |x\rangle|y + f(x)\rangle.$$

In our problem, one is given as input a set of unitaries $\{U_j\}_{j=0}^{n-1}$, with the promise that its permutations Π_x act as

$$S_n|x\rangle|\psi\rangle = |x\rangle\Pi_x|\psi\rangle = \omega^{xy}|x\rangle\Pi_0|\psi\rangle.$$

A first difference is that our unitaries act by applying a phase, instead of shifting the state of the ancilla. But if one nevertheless considers the function $f(x) = \omega^{xy}$ one can see that it obeys the promise of the hidden subgroup problem: it is a periodic function with period $r = n!/\gcd(y_0, n!)$, and if one takes the group as $G = Z_{n!}$ with addition modulo $n!$, and $H = \{0, r, 2r, \ldots\}$, then $f(x)$ is constant in each coset of H and distinct for distinct cosets. Moreover, our algorithm finds y and therefore r, solving the hidden subgroup problem for this function.

Can we push this analogy further? More specifically, can we use our algorithm to find the period of any function $f'(x)$ that is constant on each coset of the hidden subgroup and different for different cosets? We are going to show that this is not the case. Let then $f'(x) = \omega^{g(x)}$ be such a function. We start in the state

$$|C\rangle|\psi\rangle = \frac{1}{\sqrt{n!}} \sum_{x=0}^{n!-1} |x\rangle|\psi\rangle,$$

and apply the switch:

$$S_n|C\rangle|\psi\rangle = \frac{1}{\sqrt{n!}} \sum_{x=0}^{n!-1} |x\rangle\Pi_x|\psi\rangle = \frac{1}{\sqrt{n!}} \sum_{x=0}^{n!-1} \omega^{g(x)}|x\rangle\Pi_0|\psi\rangle.$$

Now we split the sum into the subgroup H and its cosets:

$$\begin{aligned} S_n|C\rangle|\psi\rangle &= \frac{1}{\sqrt{n!}} \sum_{x=0}^{n!-1} \omega^{g(x)}|x\rangle\Pi_0|\psi\rangle \\ &= \frac{1}{\sqrt{n!}} \sum_{l=0}^{r-1} \sum_{a=0}^{\frac{N}{r}-1} \omega^{g(l+ar)}|l+ar\rangle\Pi_0|\psi\rangle. \end{aligned}$$

Now we use the fact that $g(x+r) = g(x)$,

$$S_n|C\rangle|\psi\rangle = \frac{1}{\sqrt{n!}} \sum_{l=0}^{r-1} \sum_{a=0}^{\frac{N}{r}-1} \omega^{g(l)}|l+ar\rangle\Pi_0|\psi\rangle,$$

and apply the inverse Fourier transform on the first register:

$$\begin{aligned} \mathcal{F}_{n!}S_n|C\rangle|\psi\rangle &= \frac{1}{n!} \sum_{s=0}^{n!-1} \sum_{l=0}^{r-1} \sum_{a=0}^{\frac{n!}{r}-1} \omega^{g(l)-y(l+ar)}|s\rangle\Pi_0|\psi\rangle \\ &= \frac{1}{r} \sum_{k=0}^{r-1} \left(\sum_{l=0}^{r-1} \omega^{g(l)-k\frac{n!}{r}l} \right) |k\tfrac{n!}{r}\rangle\Pi_0|\psi\rangle \end{aligned}$$

The algorithm ends by measuring the first register in the computational basis. At this point, it is useful to compare this state to the one you would get if you were applying the period-finding algorithm. With it, the state of the first register would be

$$\frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} |k\tfrac{n!}{r}\rangle,$$

a *uniform* superposition over the multiples of $\frac{n!}{r}$, and a constant amount of repetitions of the algorithm would be enough to determine r with high probability.

Instead, because in our case the superposition is not uniform, one needs an exponential amount of repetitions in the worst case. To see that, consider the case when $g(l) = l \mod n!$. Then the probability of obtaining $|0\rangle$ when measuring the first register is

$$\frac{\sin^2\left(\frac{\pi r}{n!}\right)}{r^2 \sin^2\left(\frac{\pi}{n!}\right)},$$

and this means that, for constant r, one would need to make $O(n!^2)$ measurements to get an outcome different than $|0\rangle$ and any information at all about the period r.

For this reason, it seems that the analogy between our algorithm and the period-finding algorithm (and, therefore, those solving the hidden subgroup problem) cannot be pushed far. We think the true connection is only at the more basic level that they are both applications of the quantum Fourier transform.

Appendix E

Decomposition of the n-ROUTER

We show how an n-ROUTER can be constructed using a polynomial number of elementary resources, each equivalent to a polarizing beam splitter (PBS). Let $|x\rangle$ be a basis element of the control system, and $|j\rangle_{\text{in}}$ denote the j-th input mode to the router, with $j = 0, \dots, n-1$ (for simplicity, we don't write explicitly the target system). Then the n-ROUTER performs the transformation $|x\rangle|j\rangle_{\text{in}} \mapsto |x\rangle|\sigma_x(j)\rangle_{\text{out}}$, where $|\sigma_x(j)\rangle_{\text{out}}$ is the $\sigma_x(j)$-th output mode. In order to decompose this gate in elementary ones, we first express the control variable in terms of its factoradic representation, introduced in Section A: $x \leftrightarrow (a_{n-1}, \dots, a_1)$, with $0 \leq a_k \leq k$. Then, we represent each coefficient a_k using k bits[1] $b_{k\,1}, \dots, b_{k\,k}$, with $b_{k\,j} = 1$ for $j < a_k$ and $b_{k\,j} = 0$ for $j \geq a_k$. For example, the values of a_3 are represented as

$a_3 =$	0	1	2	3
$b_{3\,1} =$	0	1	1	1
$b_{3\,2} =$	0	0	1	1
$b_{3\,3} =$	0	0	0	1

In this way, we can encode the control quantum system in $\frac{n(n-1)}{2}$ qubits, $|x\rangle \to \bigotimes_{k=1}^{n} \bigotimes_{j=1}^{k} |b_{k\,j}\rangle$. Now we can construct the n-ROUTER using a controlled binary swap of modes for each control qubit $|b_{k\,j}\rangle$. Recall that an arbitrary permutation $|j\rangle_{\text{in}} \mapsto |\sigma_x(j)\rangle_{\text{out}}$ is obtained shifting each mode k "to the right" by a number a_k of positions, i.e. applying the unitary $|k\rangle_{\text{in}} \mapsto |k - a_k\rangle_{\text{out}}$, $|j\rangle_{\text{in}} \mapsto |j+1\rangle_{\text{out}}$ for $k - a_k \leq j < k$. The idea is to decompose this shift in swaps between neighboring modes, with each swap controlled by one control qubit. Explicitly, the controlled mode-swaps are defined as

$$\begin{aligned} |b_{k\,j}\rangle|k-j\rangle_{\text{in}} &\mapsto |b_{k\,j}\rangle|k-j+b_{k\,j}\rangle_{\text{out}} \\ |b_{k\,j}\rangle|k-j+1\rangle_{\text{in}} &\mapsto |b_{k\,j}\rangle|k-j+1-b_{k\,j}\rangle_{\text{out}}, \end{aligned} \tag{E.1}$$

and identity for the other modes. The n-ROUTER is then obtained by first applying the mode-swap controlled by $|b_{1\,1}\rangle$, then the one controlled by $|b_{2\,1}\rangle$ followed by the one controlled by $|b_{2\,2}\rangle$ and so on, applying successively each mode-swap controlled by $|b_{k\,j}\rangle$ increasing k and, for each k, increasing j.

[1] This encoding is clearly not optimal, since it requires $O(n^2)$ bits to encode the $n!$ permutations, instead of the required $O(n \log n)$. We will however not invest more time to optimize this aspect of the problem.

As an example, consider a permutation identified by the single non-vanishing factoradic coefficient $a_3 = 2$. The corresponding control qubits are then in the states $|b_{3\,1} = 1\rangle|b_{3\,2} = 1\rangle|b_{3\,3} = 0\rangle$. First we apply the swap between modes 3 and 2 controlled on $|b_{3\,1}\rangle$. Since $b_{3\,1} = 1$ this results in the mode-swap $|3\rangle \leftrightarrow |2\rangle$. Then, since $b_{3\,2} = 1$, the modes 2 and 1 are swapped. Composing the two swaps, we obtain the transformation $|3\rangle \rightarrow |1\rangle$, $|1\rangle \rightarrow |2\rangle$, $|2\rangle \rightarrow |3\rangle$. Since $b_{3\,3} = 0$, the last mode-swap is not applied, the result of the two mode-swaps is therefore the shift of mode 3 by $a_3 = 2$ positions to the right. By composing this procedure for an arbitrary set of coefficients $(a_{n-1}, \ldots, a_1)$, the corresponding permutation of modes is realized.

The building block of this construction is the controlled swap of two modes, which is essentially equivalent to a PBS (for a PBS, the control qubit is the photon polarization). We stress that this element does not correspond to any traditional elementary gate of a quantum circuit, and it should thus be considered as a new elementary resource.

In Ref. [8], a different approach was proposed for the scalable realization of the n-SWITCH: a construction was proposed that realizes the n-SWITCH using $O(n^2)$ 2-SWITCHes. This construction, however, is not applicable to our case, since our definitions differ slightly[2]. Furthermore, the construction based on the n-ROUTER element is more directly related to the interferometric implementation proposed in the main text.

[2] According to the definition from [8], the n-SWITCH orders the n unitaries according to the prescribed permutation, but it does not directly compose them to each other. It is thus possible to plug additional elements between any pairs of unitaries, making the construction of the n-SWITCH from 2-SWITCHes possible. The 2-SWITCH of [8] can be reproduced by a 3-SWITCH as defined here, with the prescription that only the first and last position are swapped, while the position in the middle is fixed and can be used either as an identity or to plug input and output of another switch.

Appendix F

Details of the formalism

Here we explore in more details the properties of the Choi-Jamiołkowski (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 [3] and allows a direct identification of non-signaling processes with quantum states.

F.1 Choi-Jamiołkowski isomorphism

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 [118, 119]. For a linear operator $A : \mathcal{H}^{A_I} \to \mathcal{H}^{A_O}$, we define[1]

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

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

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

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.

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 entangled state by applying a local unitary to one subsystem of a reference maximally entangled state. Therefore, the CJ vector $|U^*\rangle\!\rangle^{A_I A_O} = \mathbb{1} \otimes$

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

$U^*|\mathbb{1}\rangle\!\rangle$ *is maximally entangled if and only if* U *is a unitary*. More explicitly, an operator

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

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

$$|U^*\rangle\!\rangle^{A_I A_O} = \sum_{jk} u^*_{jk} |k\rangle^{A_I} |j\rangle^{A_O} \tag{F.4}$$

is maximally entangled.

Measurement-preparation. Another useful linear operator is $|\psi\rangle\langle\psi|\phi$, 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 (F.1), we find the CJ representation

$$\big|\, (|\psi\rangle\langle\psi|\phi\,)^*\big\rangle\!\big\rangle^{A_I A_O} = |\phi\rangle^{A_I} \otimes \big(|\psi\rangle^*\big)^{A_O}. \tag{F.5}$$

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:

$$\big|\, (|\phi\rangle\langle\phi|\phi\,)^*\big\rangle\!\big\rangle^{A_I A_O} = |\phi\rangle^{A_I} \otimes \big(|\phi\rangle^*\big)^{A_O}. \tag{F.6}$$

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

$$M^{A_I A_O} := \big[\mathcal{I} \otimes \mathcal{M}^A(|\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|)\big]^T. \tag{F.7}$$

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

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

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

According to Choi's theorem [71], a linear map $\mathcal{M}^A : A_I \to A_O$ is CP if and only if its CJ matrix is positive semidefinite, $M^{A_I A_O} \geq 0$. A characterization of the trace-preserving condition can be found using the inverse CJ isomorphism,

$$\mathcal{M}^A(\rho) = \big[\,\mathrm{tr}_{A_I}[\rho^{A_I} \otimes \mathbb{1}^{A_O} \cdot M^{A_I A_O}]\big]^T. \tag{F.9}$$

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

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

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

$$\mathbb{1}^{A_I} - \mathrm{tr}_{A_O} M^{A_I A_O} \geq 0\,. \tag{F.11}$$

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

$$M^{A_I A_O} = \mathbb{1}^{A_I} \otimes (\rho^T)^{A_O} . \tag{F.12}$$

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

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

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 $\mathcal{M}^A(\sigma) = \rho \operatorname{tr}[E\sigma]$, which has CJ representation

$$M^{A_I A_O} = E^{A_I} \otimes (\rho^T)^{A_O} . \tag{F.14}$$

F.2 Process matrices

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

Quantum states. Consider a bipartite process matrix of the form

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

According to the generalized Born rule, Eq. (2.3), 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

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

where $E^{A_I} := \operatorname{tr}_{A_O} M^{A_I A_O}$ and $E^{B_I} := \operatorname{tr}_{B_O} M^{B_I B_O}$. These operators are positive semidefinite and, because of Eq. (F.11), they can be completed to form a POVM. Thus Eq. (F.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 (F.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 (F.15) and the CJ representation of state preparation, Eq. (F.12).

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

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

where we used (F.12) and (F.13) for the local operations. This is equivalent to saying that B measures E in the state $\mathrm{tr}_{A_O}\left[\big(\rho^T\big)^{A_O} \otimes \mathbb{1}^{B_I} \cdot W^{A_O B_I}\right] = \Big[\mathrm{tr}_{A_O}\left[\rho^{A_O} \otimes \mathbb{1}^{B_I} \cdot \big(W^T\big)^{A_O B_I}\right]\Big]^T$. Comparing this with the inverse CJ transformation (F.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

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

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

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 j-th 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

$$\begin{aligned} &\overline{W}(M^{A_I^j A_O^j}) \\ &:= \mathrm{tr}_{A_I^j A_O^j}\Big[\big(\mathbb{1}^{A_I^1 A_O^1} \otimes \ldots M^{A_I^j A_O^j} \otimes \ldots \mathbb{1}^{A_I^N A_O^N}\big) \cdot W\Big]. \end{aligned} \tag{F.19}$$

With the usual generalized Born rule (2.3), the reduced process matrix gives the probability for the remaining $N-1$ parties to measure arbitrary CP maps, given that the j-th 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 (F.18). If A prepares a state ρ, the reduced process matrix for B is

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

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.

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

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

The probability is then obtained as the modulus square of the amplitude and conforms to the general expression (2.3). Given that party j performs the unitary U_j, the reduced process is clearly given by the partial scalar product

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

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

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

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

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

Appendix G

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 Ref. [3], where they were formulated in a basis-dependent way. Here we derive the equivalent characterization of valid process matrices given in Eqs. (2.4)–(2.6); 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.

G.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 (2.3), 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 completely positive 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[1]

$$\mathrm{tr}\Big[\big(M^{A'_I A_I A_O} \otimes M^{B'_I B_I B_O}\big) \cdot \big(\rho^{A'_I B'_I} \otimes W^{A_I A_O B_I B_O}\big)\Big] \geq 0$$
$$\forall\ M^{A'_I A_I A_O} \geq 0\,,\ M^{B'_I B_I B_O} \geq 0\,,\ \rho^{A'_I B'_I} \geq 0\,. \tag{G.1}$$

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 (G.1) is equal to $\mathrm{tr}[M^{B'_I B_I B_O} \cdot W^{A_I A_O B_I B_O}]$. Requiring that its value is non-negative for all $M^{B'_I B_I B_O} \geq 0$ implies that W must be positive semidefinite.

Reciprocally, $W \geq 0$ clearly implies that (G.1) is satisfied. Hence, the non-negativity of probabilities is equivalent to W being positive semidefinite, Eq. (2.4).

[1] 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 [120].

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 $\mathrm{tr}_{A_O} M^{A_I A_O} = \mathbb{1}^{A_I}$ – or equivalently, using the notation of Eq. (2.12), ${}_{A_O}M^{A_I A_O} = \mathbb{1}^{A_I A_O}/d_{A_O}$. Ignoring here for simplicity the possible use of an ancillary system (which leads to the same conclusion[2]), the normalization of probabilities is thus equivalent to

$$\begin{aligned} &\mathrm{tr}\big[\big(M^{A_I A_O} \otimes M^{B_I B_O}\big) \cdot W^{A_I A_O B_I B_O}\big] = 1 \\ &\qquad \forall\ M^{A_I A_O} \geq 0\,,\ M^{B_I B_O} \geq 0\,, \\ \text{s.t. }\ &{}_{A_O}M^{A_I A_O} = \mathbb{1}^{A_I A_O}/d_{A_O}\ ,\ {}_{B_O}M^{B_I B_O} = \mathbb{1}^{B_I B_O}/d_{B_O}\,. \end{aligned} \tag{G.2}$$

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[3]. 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 Eq. (G.2) is equivalent to

$$\begin{aligned} &\mathrm{tr}\left[\Big({}_{[1-A_O]}x + \mathbb{1}/d_{A_O}\Big) \otimes \Big({}_{[1-B_O]}y + \mathbb{1}/d_{B_O}\Big) W\right] = 1 \\ &\qquad \forall\ x,\ y\,. \end{aligned} \tag{G.3}$$

For $x = y = 0$, this yields the normalization condition of Eq. (2.5),

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

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

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

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

---

[2] Taking into account a possible ancillary state $\rho^{A'_I B'_I}$, the same reasoning as below leads to Eqs. (G.4) and (G.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 $\mathrm{tr}\,\rho^{A'_I B'_I} = 1$. One can easily check that these are indeed equivalent to (G.4) and (G.15) in their original form.

[3] More precisely: for any hermitian operator $M^{A_I A_O}$, there always exists $\alpha > 0$ such that $M^{A_I A_O} + \alpha\, \mathbb{1}^{A_I A_O}/d_{A_O} \geq 0$. 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 $M_+^{A_I A_O} = \frac{1}{\alpha+1}(M^{A_I A_O} + \alpha\, \mathbb{1}^{A_I A_O}/d_{A_O}) \geq 0$ and $M_-^{A_I A_O} = \mathbb{1}^{A_I A_O}/d_{A_O} \geq 0$ 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 $M_+^{B_I B_O} \geq 0$, $M_-^{B_I B_O} \geq 0$ 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 Eq. (G.2), and therefore $\mathrm{tr}\big[\big(M_\pm^{A_I A_O} \otimes M_\pm^{B_I B_O}\big) \cdot W^{A_I A_O B_I B_O}\big] = 1$. Expanding $\mathrm{tr}\big[\big(M^{A_I A_O} \otimes M^{B_I B_O}\big) \cdot W^{A_I A_O B_I B_O}\big]$ using the decomposition just constructed, we find that its value is also 1. Hence, if Eq. (G.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.

which then imply

$$\begin{gathered}\operatorname{tr}\big[\big({}_{[1-A_O]}x \otimes {}_{[1-B_O]}y\big)W\big] = 0 \\ \forall\, x,\ y\,.\end{gathered} \tag{G.7}$$

Reciprocally, Eqs. (G.4)–(G.7) clearly imply (G.3), so that these are equivalent to Eq. (G.2).

Thinking of the trace as the Hilbert-Schmidt inner product

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

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

$${}_{[1-A_O]}\big(\operatorname{tr}_{B_I B_O} W\big) = 0, \tag{G.8}$$
$${}_{[1-B_O]}\big(\operatorname{tr}_{A_I A_O} W\big) = 0, \tag{G.9}$$
$${}_{[1-A_O][1-B_O]}W = 0, \tag{G.10}$$

which we can rewrite as

$${}_{B_I B_O}W = {}_{A_O B_I B_O}W\,, \tag{G.11}$$
$${}_{A_I A_O}W = {}_{A_I A_O B_O}W\,, \tag{G.12}$$
$$W = {}_{B_O}W + {}_{A_O}W - {}_{A_O B_O}W\,, \tag{G.13}$$

which are conditions (2.9)–(2.11).

Note that each condition (G.8)–(G.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[4]

$$\mathcal{L}_V = \big\{W \in A_I \otimes A_O \otimes B_I \otimes B_O \,|\, W = L_V(W)\big\}, \tag{G.14}$$

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 (G.11)–(G.13) explicitly as projections onto subspaces, i.e., as

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

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

$$L_A(W) = W - {}_{B_I B_O}W + {}_{A_O B_I B_O}W\,, \tag{G.16}$$
$$L_B(W) = W - {}_{A_I A_O}W + {}_{A_I A_O B_O}W\,, \tag{G.17}$$
$$L_{AB}(W) = {}_{B_O}W + {}_{A_O}W - {}_{A_O B_O}W\,, \tag{G.18}$$

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.,

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

[4] 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.

which, after simplification, can be written as

$$L_V(W) = {}_{A_O}W + {}_{B_O}W - {}_{A_OB_O}W \\ - {}_{B_IB_O}W + {}_{A_OB_IB_O}W \\ - {}_{A_IA_O}W + {}_{A_IA_OB_O}W \quad \text{(G.20)}$$

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 $W \geq 0$, $\operatorname{tr} W = d_{A_O} d_{B_O}$, and $W = L_V(W)$, as in Eqs (2.4)–(2.6).

G.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 $W \geq 0$, $\operatorname{tr} W = d_{A_O} d_{B_O} d_{C_O}$, and

$$W = L_A(W)\,, \quad W = L_B(W)\,, \quad W = L_C(W)\,, \\ W = L_{AB}(W)\,, \quad W = L_{AC}(W)\,, \quad W = L_{BC}(W)\,, \\ W = L_{ABC}(W)\,, \quad \text{(G.21)}$$

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

$$L_A(W) = {}_{[1-(1-A_O)B_IB_OC_IC_O]}W\,, \\ L_B(W) = {}_{[1-(1-B_O)A_IA_OC_IC_O]}W\,, \\ L_C(W) = {}_{[1-(1-C_O)A_IA_OB_IB_O]}W\,, \\ L_{AB}(W) = {}_{[1-(1-A_O)(1-B_O)C_IC_O]}W\,, \\ L_{AC}(W) = {}_{[1-(1-A_O)(1-C_O)B_IB_O]}W\,, \\ L_{BC}(W) = {}_{[1-(1-B_O)(1-C_O)A_IA_O]}W\,, \\ L_{ABC}(W) = {}_{[1-(1-A_O)(1-B_O)(1-C_O)]}W\,,$$

where we used the shorthand notation

$${}_{[\sum_X \alpha_X X]}W = \sum_X \alpha_X \cdot {}_X W \quad \text{(G.22)}$$

for a sum over products of subsystems X with coefficients α_X (and with ${}_1W := W$).

The constraints in (G.21) are equivalent to

$$W = L_V(W)\,, \quad \text{(G.23)}$$

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,

$$L_V(W) = {}_{\left[1-(1-A_O+A_IA_O)(1-B_O+B_IB_O)(1-C_O+C_IC_O)+\ A_IA_OB_IB_OC_IC_O\right]}W\,, \quad \text{(G.24)}$$

which defines a projector onto the linear subspace

$$\mathcal{L}_V = \left\{W \in A_I \otimes A_O \otimes B_I \otimes B_O \otimes C_I \otimes C_O \,|\, W = L_V(W)\right\}. \quad \text{(G.25)}$$

G.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 $W \geq 0$, $\mathrm{tr}\, W = d_{A_O^1} \ldots d_{A_O^N}$, and for all $2^N - 1$ non-empty subsets $\mathcal{X}$ of $\{1, \ldots, N\}$,

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

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

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

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

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

which again defines a projector onto the linear subspace

$$\mathcal{L}_V = \left\{W \in A_I^1 \otimes A_O^1 \otimes \ldots \otimes A_I^N \otimes A_O^N \mid W = L_V(W)\right\}. \tag{G.29}$$

Proof. For any subset $\mathcal{X}$ of $\{1, \ldots, N\}$, define $P_{\mathcal{X}} = \prod_{i \in \mathcal{X}} (1 - A_O^i) \prod_{i \notin \mathcal{X}} A_I^i A_O^i$. 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, $\left[(1 - A_O^{i_0}) A_O^{i_0}\right] W = 0$, so that $\left[P_{\mathcal{X}} P_{\mathcal{X}'}\right] W = 0$.

Developing L_V, we thus find

$$L_V(W) = \left[\prod_{\mathcal{X} \neq \emptyset} (1 - P_{\mathcal{X}})\right] W = \left[1 - \sum_{\mathcal{X} \neq \emptyset} P_{\mathcal{X}}\right] W = \left[1 - \sum_{\mathcal{X}} P_{\mathcal{X}} + \prod_i A_I^i A_O^i\right] W.$$

Now, one can write

$$\sum_{\mathcal{X}} P_{\mathcal{X}} = \sum_{k_1=0}^{1} \ldots \sum_{k_N=0}^{1} (1 - A_O^1)^{k_1} (A_I^1 A_O^1)^{1-k_1} \ldots (1 - A_O^N)^{k_N} (A_I^N A_O^N)^{1-k_N} \tag{G.30}$$

$$= \prod_i \sum_{k_i=0}^{1} (1 - A_O^i)^{k_i} (A_I^i A_O^i)^{1-k_i} = \prod_i (1 - A_O^i + A_I^i A_O^i), \tag{G.31}$$

from which Eq. (G.28) follows. □

Appendix H

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 Ref. [39][1].

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

$$\mathcal{K}^* = \{x^* \in E' \mid \langle x^*, x \rangle \geq 0 \quad \forall x \in \mathcal{K}\}\,. \tag{H.1}$$

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

$$\mathcal{L}^* = \mathcal{L}^\perp = \{x^* \in E' \mid \langle x^*, x \rangle = 0 \quad \forall x \in \mathcal{L}\}. \tag{H.2}$$

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

$$[\mathrm{conv}(\mathcal{K}_1 \cup \mathcal{K}_2)]^* = \mathcal{K}_1^* \cap \mathcal{K}_2^*\,, \tag{H.3a}$$
$$(\mathcal{K}_1 \cap \mathcal{K}_2)^* = \mathrm{conv}(\mathcal{K}_1^* \cup \mathcal{K}_2^*)\,, \tag{H.3b}$$

   where conv denotes the convex hull.

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

[1] 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 10. *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 $\operatorname{tr} W = d_O$, $W \in \mathcal{L}_V$ (i.e., it satisfies Eqs. (2.9)–(2.11)), and it can be written as*

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

$$\textit{with} \quad W^{A\prec B} \geq 0\,, \quad W^{A\prec B} = {}_{B_O}W^{A\prec B}\,, \tag{H.5}$$

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

Proof. The "only if" direction is straightforward (simply replace $q\,W^{A\prec B} \to W^{A\prec B}$ and $(1-q)\,W^{B\prec A} \to W^{B\prec A}$ to go from (2.18) to (H.4), so that $W^{A\prec B}$ and $W^{B\prec A}$ in (H.4) are not normalized).

To see that the converse also holds, first note that $W^{A\prec B} \geq 0$ and $W^{B\prec A} \geq 0$ imply that $W \geq 0$, 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 (2.9). Since $W \in \mathcal{L}_V$ also satisfies (2.9), 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 (2.10). 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 (2.11). 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 (2.18)), 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

$$\operatorname{tr}[S W^{\text{sep}}] \geq 0 \tag{H.7}$$

for all causally separable process matrices W^{sep}. Note that the condition $\operatorname{tr}[W^{\text{sep}}] = d_O$ 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 10:

$$\mathcal{W}^{\text{sep}} = \operatorname{conv}\big[(\mathcal{P} \cap \mathcal{L}_{B_O}) \cup (\mathcal{P} \cap \mathcal{L}_{A_O})\big] \cap \mathcal{L}_V, \tag{H.8}$$

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

$$\mathcal{L}_{B_O} = \{W \mid W = {}_{B_O}W\}, \tag{H.9}$$

$$\mathcal{L}_{A_O} = \{W \mid W = {}_{A_O}W\}, \tag{H.10}$$

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

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

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

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

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

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

Focusing on $(\mathcal{P} \cap \mathcal{L}_{B_O})^*$, using again the duality relations (H.2)–(H.3), we see that

$$\begin{aligned}(\mathcal{P} \cap \mathcal{L}_{B_O})^* &= \text{conv}(\mathcal{P} \cup \mathcal{L}_{B_O}^\perp) \\ &= \{S_+ + S_0 \mid S_+ \geq 0,\ {}_{B_O}S_0 = 0\} \\ &= \{S \mid {}_{B_O}S \geq 0\},\end{aligned} \tag{H.15}$$

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

Similarly,

$$(\mathcal{P} \cap \mathcal{L}_{A_O})^* = \{S \mid {}_{A_O}S \geq 0\}. \tag{H.16}$$

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

Appendix I

Explicit positive semidefinite constraints

For the convenience of the reader, we present the SDP problems (2.38) and (2.39) with all conic constraints rewritten in terms of the positive semidefinite cone, to facilitate implementation. To rewrite (2.38), we need a characterisation 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

$$\begin{aligned} \mathcal{W}_V^* &= \mathcal{W}^* \cap \mathcal{L}_V && \text{(I.1)} \\ &= (\mathcal{P} \cap \mathcal{L}_V)^* \cap \mathcal{L}_V && \text{(I.2)} \\ &= \mathrm{conv}(\mathcal{P} \cup \mathcal{L}_V^\perp) \cap \mathcal{L}_V && \text{(I.3)} \\ &= \{L_V(\Sigma_P) | \Sigma_P \geq 0\}\,, && \text{(I.4)} \end{aligned}$$

where Equation (I.4) follows from an argument analogous to the one used to derive Corollary 2.

With this characterisation, the SDP problem (2.38) then becomes

$$\begin{aligned} &\min \mathrm{tr}(SW) \\ \text{s.t.} \quad &S = L_V(S_P)\,,\ {}_{A_O}S_P \geq 0\,,\ {}_{B_O}S_P \geq 0\,, \\ &\mathbb{1}/d_O - S = L_V(\Sigma_P)\,,\ \Sigma_P \geq 0\,. \end{aligned} \tag{I.5}$$

To rewrite the SDP problem (2.39), we use the characterisation of $\mathcal{W}^{\text{sep}}$ given in Lemma 10:

$$\begin{aligned} &\min \mathrm{tr}(\Omega)/d_O \\ \text{s.t.} \quad &W + \Omega = W^{A\prec B} + W^{B\prec A}\,, \\ &W^{A\prec B} \geq 0\,, \quad W^{A\prec B} = {}_{B_O}W^{A\prec B}\,, \\ &W^{B\prec A} \geq 0\,, \quad W^{B\prec A} = {}_{A_O}W^{B\prec A}\,, \\ &\Omega \geq 0\,, \quad \Omega = L_V(\Omega)\,. \end{aligned} \tag{I.6}$$

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

Appendix J

Duality for conic problems

In this appendix we show that the two problems defined in Section 2.3.2 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 (2.40) holds.

Let us first recall the definitions of primal and dual conic problems (Definition 4.2.1 in [40]), of which SDP problems are a particular case:

Definition 11. *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*

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

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 $\lambda \geq 0$ for which

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

is a causally separable process, optimized over all valid (normalised) 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 $\operatorname{tr}\Omega/d_O$. Remembering also that the normalisation $1/(1+\lambda)$ is irrelevant for conic constraints, the problem reduces to minimizing $\operatorname{tr}\Omega/d_O$ such that

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

To translate this SDP problem into the language of Definition 11, let us define

$$E = \mathcal{L}_V \times \mathcal{L}_V\,, \tag{J.3}$$
$$\mathcal{K} = \mathcal{W}^{\text{sep}} \times \mathcal{W}\,, \tag{J.4}$$
$$\mathcal{L} = \{(\Omega, \Omega)\,|\,\Omega \in \mathcal{L}_V\}\,, \tag{J.5}$$
$$b = (W, 0)\,, \tag{J.6}$$
$$c = (0, \mathbb{1}/d_O)\,, \tag{J.7}$$

and the inner product

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

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

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

which indeed corresponds to the SDP problem (2.39).

To construct the dual SDP problem, first note that

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

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

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

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

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

which corresponds to the SDP problem (2.38).

Let us emphasize that here the duals of $\mathcal{W}^{\text{sep}}$ 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 $\mathcal{W}^{\text{sep}}$ and $\mathcal{W}$ (and therefore $\mathcal{K}=\mathcal{W}^{\text{sep}}\times\mathcal{W}$) have empty interiors, and therefore these cones would not satisfy the requirements of Definition 11. 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., $\mathcal{W}^{\text{sep}}\times\mathcal{W}\subseteq\mathcal{L}_V\times\mathcal{L}_V$ 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 $\mathcal{W}^{\text{sep}}\times\mathcal{W}$, as both cones require their elements to be positive semidefinite, and $W\geq 0$ and $-W\geq 0$ imply that $W=0$. To show that $\mathcal{W}^{\text{sep}}\times\mathcal{W}$ has nonempty interior, it is enough[1] to find an operator that belongs to $\mathrm{int}\,\mathcal{W}^{\text{sep}}$. This is done through the following lemma:

Lemma 12. *$\mathbb{1}^\circ + \Omega \in \mathrm{int}\,\mathcal{W}^{\text{sep}}$ for any $\Omega\in\mathcal{L}_V$ such that $\|\Omega\|_2 < \frac{1}{2d_I}$, where $d_I = d_{A_I}d_{B_I}$ and $\|\cdot\|_2$ is the Hilbert-Schmidt norm.*

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

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

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

are causally ordered (in the sense that they satisfy Eq. (2.17) and the analogous relation for the order $B \prec A$, respectively), and so are the operators

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

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

Since, furthermore,

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

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

$$\left\|\omega^{A\prec B}\right\| \leq \frac{1}{2d_I} \quad \text{and} \quad \left\|\omega^{B\prec A}\right\| \leq \frac{1}{2d_I}\,, \tag{J.18}$$

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

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

which implies that

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

and therefore

$$\|\Omega\|_2 \leq \frac{1}{2d_I} \tag{J.21}$$

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 (J.21) belongs to the interior of $\mathcal{W}^{\text{sep}}$, i.e., $\|\Omega\|_2 < \frac{1}{2d_I}$ implies that $\mathbb{1}^\circ + \Omega \in \operatorname{int} \mathcal{W}^{\text{sep}}$. □

This concludes the proof that problems (2.38) and (2.39) are SDP problems dual to each other. We shall now proceed to show that the Duality Theorem (Theorem 4.2.1 in [40]) applies to them:

Theorem 13. *Let (P), (D) be a primal-dual pair of conic problems as defined above, and let the pair be such that*

leftmirgin=5mm The set of primal solutions $\mathcal{K} \cap (\mathcal{L} + b)$ *intersects* $\operatorname{int} \mathcal{K}$*;*

leftmiirgiin=5mm The set of dual solutions $\mathcal{K}^* \cap (\mathcal{L}^\perp + c)$ *intersects* $\operatorname{int} \mathcal{K}^*$*;*

leftmiiirgiiin=5mm $\langle c, x\rangle$ *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*

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

Let us check that for the SDP problems (J.9) and (J.13), the three assumptions of the Duality Theorem are indeed satisfied.

To see that *1.* is satisfied, we need to find $\Omega \in \operatorname{int} \mathcal{W}$ such that $W + \Omega \in \operatorname{int} \mathcal{W}^{\text{sep}}$. Take $\Omega = \lambda \mathbb{1}^\circ$; then $W + \lambda \mathbb{1}^\circ \in \operatorname{int} \mathcal{W}^{\text{sep}}$ iff $\frac{1}{\lambda} W + \mathbb{1}^\circ \in \operatorname{int} \mathcal{W}^{\text{sep}}$. Using Lemma 12, we conclude that this is true if

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

Since

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

it is enough to choose

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

to satisfy inequality (J.23), and we're done.

To see that *2.* is satisfied, we need to exhibit a witness S such that $S \in \operatorname{int} \mathcal{S}_V$ and $\mathbb{1}/d_O - S \in \operatorname{int} \mathcal{W}_V^*$. 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 $S = \mathbb{1}/(2 d_O)$.

To see that *3.* is satisfied, note that $\Omega \geq 0$ implies that $\operatorname{tr} \Omega / d_O \geq 0$.

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

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

as claimed in Eq. (2.40). As discussed in Sec. 2.3.2, a value $\operatorname{tr}[\Omega^*]/d_O = -\operatorname{tr}[S^* W] > 0$ guarantees that the process matrix W is causally nonseparable, and the solution S^* of the dual problem provides an explicit causal witness; a value $\operatorname{tr}[\Omega^*]/d_O = 0$ 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 (I.6)).

Appendix K

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 [121]. 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 $\mathcal{N}(W) \geq 0$ for every process matrix W, with $\mathcal{N}(W) = 0$ if and only if W is causally separable.

Convexity $\mathcal{N}(\sum_i p_i W_i) \leq \sum_i p_i \mathcal{N}(W_i)$ for any process matrices W_i and any $p_i \geq 0$, with $\sum_i p_i = 1$.

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

Now we shall prove that both $R_{\mathsf{g}}(W)$ and $R_{\mathsf{r}}(W)$ as defined in equations (2.42) and (2.46) respect the properties of **Discrimination** and **Convexity**, whereas $R_{\mathsf{g}}(W)$ respects **Monotony** but $R_{\mathsf{r}}(W)$ does not.

Discrimination follows from the definition of the SDP problems (2.38)–(2.39) and (2.45)–(2.44). Note that since they satisfy the assumptions of the Duality Theorem (13), 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 $p_i \geq 0$,

$$\operatorname{tr}[S_{W_j} W_j] \leq \operatorname{tr}\Big[\Big(S_{\sum_i p_i W_i}\Big) W_j\Big] \tag{K.1}$$

and therefore

$$-\operatorname{tr}\Big[\Big(S_{\sum_i p_i W_i}\Big) \sum_i p_i W_i\Big] \leq -\sum_i p_i \operatorname{tr}[S_{W_i} W_i], \tag{K.2}$$

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

Now we show that **Monotony** does hold for $R_{\mathsf{g}}(W)$. 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

$$\mathrm{tr}[(C_2^A \otimes C_2^B) \cdot \$(W)] = \mathrm{tr}[(C_{123}^A \otimes C_{123}^B)W], \tag{K.3}$$

where

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

is the Choi-Jamiołkowski operator of the composition of the each party's operations. The processes W and $\$(W)$ are illustrated in Figure K.1.

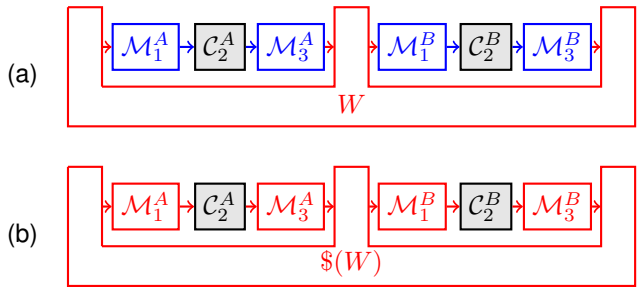


Figure K.1: (a) The situation where the parties share a bipartite process W (in red) and apply the CPTP maps $\mathcal{M}_1^X$ and $\mathcal{M}_3^X$ (in blue) to their inputs and outputs can be equivalently described by (b) a single bipartite process $\$(W)$ (in red).

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

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

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

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

and the arguments in Appendix G show that requiring the probabilities $P(\mathcal{C}_2^A, \mathcal{C}_2^B)$ to be positive and normalised 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 [32].

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,

$$R_{\mathsf{g}}(\$(W)) \leq R_{\mathsf{g}}(W), \tag{K.7}$$

or equivalently that

$$-\mathrm{tr}\left[\left(S_{\$(W)}\right)\$(W)\right] \leq -\mathrm{tr}[(S_W)W]\,. \tag{K.8}$$

By duality, this is equivalent to

$$-\mathrm{tr}\left[\$^*\left(S_{\$(W)}\right)W\right] \leq -\mathrm{tr}[(S_W)W] \tag{K.9}$$

(where $\$^*$ is the dual map of $\$$), which follows from the optimality of S_W if $\$^*\left(S_{\$(W)}\right)$ is a valid causal witness that respects the normalisation condition for generalised

robustness (as defined in SDP problem (2.38)). Therefore, we need to show it has the two following properties:

$$\mathrm{tr}\big[\$^*\big(S_{\$(W)}\big)W^{\mathsf{sep}}\big] \geq 0 \quad \forall W^{\mathsf{sep}}\,, \tag{K.10}$$

$$\mathbb{1}/d_O - \$^*\big(S_{\$(W)}\big) \in \mathcal{W}^*\,. \tag{K.11}$$

The first one follows from duality

$$\mathrm{tr}\big[\$^*\big(S_{\$(W)}\big)W^{\mathsf{sep}}\big] = \mathrm{tr}\big[S_{\$(W)}\$(W^{\mathsf{sep}})\big] \tag{K.12}$$

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

The second one is equivalent to

$$\mathrm{tr}\big[\big(\mathbb{1}/d_O - \$^*\big(S_{\$(W)}\big)\big)\Omega\big] \geq 0 \tag{K.13}$$

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

$$\mathrm{tr}\big[S_{\$(W)}\,\$(\Omega)\big] \leq \mathrm{tr}(\Omega)/d_O\,, \tag{K.14}$$

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 (2.38)).

An analogous proof fails for random robustness, as the dual map $\$^*(\cdot)$ can increase the trace of a witness, and therefore make it fail to satisfy the normalisation condition for SDP problem (2.44). To show that $R_{\mathsf{r}}(W)$ does not in fact satisfy **Monotony**, it is enough to find a process and local operations such that $R_{\mathsf{r}}\big(\$(W)\big) > R_{\mathsf{r}}(W)$.

A concrete counterexample can be obtained by considering W_{OCB} and S_{OCB} from section 2.3.4. Let

$$W_1 = W_{\mathsf{OCB}} \otimes \frac{\mathbb{1}^{A'_I}}{2} \tag{K.15}$$

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)

$$R_{\mathsf{r}}(W_1) = -\,\mathrm{tr}\,S_{W_1}\,W_1 = \sqrt{2} - 1, \tag{K.16}$$

where

$$S_{W_1} = 2\,S_{\mathsf{OCB}} \otimes |0\rangle\langle 0|^{A'_I} \tag{K.17}$$

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

$$\$(W_1) = W_{\mathsf{OCB}} \otimes |0\rangle\langle 0|^{A'_I} \tag{K.18}$$

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)

$$R_{\mathsf{r}}(\$(W_1)) = -\,\mathrm{tr}\,S_{\$(W_1)}\,\$(W_1) = 2(\sqrt{2} - 1), \tag{K.19}$$

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

$$R_{\mathsf{r}}(\$(W_1)) > R_{\mathsf{r}}(W_1), \tag{K.20}$$

so random robustness is not monotonous under local operations.

Appendix L

Characterisation of tripartite causal witnesses

Proof of Theorem 3. As defined in section 2.2.2, the cone of tripartite causally separable processes with $d_{C_O} = 1$ is

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

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

$$\mathcal{S}_{3C} = \mathcal{W}_{3C}^{\text{sep}\,*}\,. \tag{L.2}$$

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

$$\begin{aligned} \mathcal{S}_{3C} &= (\mathcal{P} \cap \mathcal{L}_{A\prec B\prec C})^* \cap (\mathcal{P} \cap \mathcal{L}_{B\prec A\prec C})^* && \text{(L.3)}\\ &= \left[\text{conv}\big(\mathcal{P} \cup \mathcal{L}^{\perp}_{A\prec B\prec C}\big)\right] \cap \left[\text{conv}\big(\mathcal{P} \cup \mathcal{L}^{\perp}_{B\prec A\prec C}\big)\right], && \text{(L.4)} \end{aligned}$$

with

$$\text{conv}\big(\mathcal{P} \cup \mathcal{L}^{\perp}_{A\prec B\prec C}\big) = \{S^P_{ABC} + S^{\perp}_{ABC} \mid S^P_{ABC} \geq 0,\ L_{A\prec B\prec C}(S^{\perp}_{ABC}) = 0\} \tag{L.5}$$

and

$$\text{conv}\big(\mathcal{P} \cup \mathcal{L}^{\perp}_{B\prec A\prec C}\big) = \{S^P_{BAC} + S^{\perp}_{BAC} \mid S^P_{BAC} \geq 0,\ L_{B\prec A\prec C}(S^{\perp}_{BAC}) = 0\}\,. \tag{L.6}$$

□

Appendix M

Optimizing Chiribella's task

We want to optimize the weights $q_{ij}^{[,]}, q_{ij}^{\{,\}}$ so as to minimize the maximal probability of success[1] for causally separable processes $p_{\text{succ}}^{\text{sep}}$, i.e., we want to minimize the upper bound

$$\text{tr}(G_{\text{finite}} W^{\text{sep}}) \leq p_{\text{succ}}^{\text{sep}}. \tag{M.1}$$

This is relevant because, according to equation (2.91), a lower $p_{\text{succ}}^{\text{sep}}$ corresponds to a larger resistance to worst-case noise.

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

We are thus led to define the following SDP problem:

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

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

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

(and we omit the optimal $q_{ij}^{[,]}, q_{ij}^{\{,\}}$ for brevity).

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

Appendix N

Characterization of the causal polytope

N.1 Vertices of the causal polytope

As explained in the main text, the causal polytope, defined as the set of causal correlations of the form (3.3), is the convex hull of the polytope of correlations compatible with the causal order $A \prec B$ (which cannot signal to Alice) and the polytope of correlations compatible with the order $B \prec A$ (which cannot signal to Bob), so that its vertices are vertices of at least one of these two polytopes.

Let us characterize the vertices of the polytope of correlations $p^{A\prec B}$. Using Bayes' rule, we can write

$$p^{A\prec B}(a,b|x,y) = p(a|x,y)\,p(b|x,y,a) = p(a|x)\,p(b|x,y,a),$$

where the last equality follows from the no-signaling-to-Alice constraint (3.1). No further constraint is imposed on $p(a|x)$ and $p(b|x,y,a)$ (except that they must be valid probability distributions); these can be written as convex combinations of deterministic distributions, in the form

$$\begin{aligned} p(a|x) &= \sum_{\alpha} q_\alpha\, \delta_{a,\,\alpha(x)}\,, \\ p(b|x,y,a) &= \sum_{\beta} q_\beta\, \delta_{b,\,\beta(x,y,a)}\,, \end{aligned}$$

with $q_\alpha, q_\beta \geq 0$, $\sum_\alpha q_\alpha = \sum_\beta q_\beta = 1$, where the α's denote deterministic functions of Alice's input x, and the β's denote deterministic functions of both Alice and Bob's inputs x, y and of Alice's output a. Combining them, we get

$$\begin{aligned} p^{A\prec B}(a,b|x,y) &= \sum_{\alpha,\beta} q_\alpha\, q_\beta\; \delta_{a,\,\alpha(x)}\; \delta_{b,\,\beta(x,y,a)} \\ &= \sum_{\alpha,\beta} q_\alpha\, q_\beta\; \delta_{a,\,\alpha(x)}\; \delta_{b,\,\beta_\alpha(x,y)} \\ &= \sum_{\alpha,\beta'} q_{\alpha,\beta'}\; \delta_{a,\,\alpha(x)}\; \delta_{b,\,\beta'(x,y)}\,, \end{aligned}$$

where $\beta_\alpha(x,y) = \beta(x,y,\alpha(x))$ and $q_{\alpha,\beta'} = q_\alpha\, q_{\beta'|\alpha}$ with $q_{\beta'|\alpha} = \sum_\beta \delta_{\beta_\alpha,\beta'}\, q_\beta$, such that $q_{\alpha,\beta'} \geq 0$ and $\sum_{\alpha,\beta'} q_{\alpha,\beta'} = 1$. Hence, any correlation $p^{A\prec B}$ can be written as a convex combination of deterministic correlations compatible with the order $A \prec B$—which thus correspond to the vertices of the corresponding polytope of correlations $p^{A\prec B}$.

If Alice and Bob's inputs can take m_A and m_B different values, respectively, and their outputs can take k_A and k_B values, resp., then there are $k_A^{m_A}$ different deterministic functions $\alpha(x)$ and $k_B^{m_A m_B}$ functions $\beta'(x,y)$, so that the polytope of correlations $p^{A\prec B}$ has $k_A^{m_A} \times k_B^{m_A m_B}$ vertices. Note that $k_B^{m_B}$ of the functions $\beta'(x,y)$ do not depend on x; hence, out of all the vertices, $k_A^{m_A} k_B^{m_B}$ are non-signaling, while the other $k_A^{m_A}(k_B^{m_A m_B} - k_B^{m_B})$ are signaling to Bob.

Similarly, the vertices of the polytope of correlations $p^{B\prec A}$ are the $k_A^{m_A m_B}\, k_B^{m_B}$ deterministic correlations compatible with the order $B \prec A$, of which $k_A^{m_A} k_B^{m_B}$ are non-signaling and are thus common to the previous polytope of correlations $p^{A\prec B}$, while the other $k_B^{m_B}(k_A^{m_A m_B} - k_A^{m_A})$ are signaling to Alice. The vertices of the causal polytope are all the deterministic correlations[1] compatible with either the order $A \prec B$, or the order $B \prec A$, or both—which makes a total of $k_A^{m_A}\, k_B^{m_A m_B} + k_A^{m_A m_B}\, k_B^{m_B} - k_A^{m_A} k_B^{m_B}$ vertices.

N.2 Dimensions

Because of the $m_A m_B$ normalization constraints $\sum_{a,b} p(a,b|x,y) = 1$, the probability space of correlations $p(a,b|x,y)$ is of dimension $m_A m_B(k_A k_B - 1)$. With the no-signaling-to-Alice and no-signaling-to-Bob constraints (3.1, 3.2), the dimensions of the polytopes of correlations $p^{A\prec B}$ and $p^{B\prec A}$ are reduced to $m_A m_B(k_A k_B - 1) - m_A(m_B - 1)(k_A - 1)$ and $m_A m_B(k_A k_B - 1) - (m_A - 1)m_B(k_B - 1)$, respectively. However, the causal polytope—i.e., their convex hull—remains of the same dimension as the full probability space.

N.3 For binary inputs and outputs

In the case where both Alice and Bob's inputs and outputs take binary values, the 10-dimensional polytopes of correlations $p^{A\prec B}$ and $p^{B\prec A}$ both have 64 vertices, among which 16 are non-signaling vertices common to both polytopes. The 12-dimensional causal polytope thus has $64 + 64 - 16 = 112$ different vertices.

We enumerated the facets of the causal polytope for binary inputs and outputs by solving the convex hull problem using the software `lrs` [68]. As described in the main text, we obtained 48 facets, which can be grouped into 3 families of equivalent facets (up to relabelings of inputs and outputs). Explicitly, these are

- 16 trivial facets of the form $p(a,b|x,y) \geq 0$ for all $x, y, a, b = 0, 1$;
- 16 facets of the GYNI type, which can be written (in the same form as (3.6), implicitly assuming uniform input bits) as

$$p(a \oplus \alpha_1 x \oplus \alpha_0 = y, b \oplus \beta_1 y \oplus \beta_0 = x) \;\leq\; \frac{1}{2},$$

[1] Note that the fact that the causal polytope has deterministic vertices was not trivial *a priori*; this contrasts for instance with the no-signaling polytope, which has non-deterministic vertices like the PR-box [76].

for all $\alpha_0, \alpha_1, \beta_0, \beta_1 = 0, 1$;

- 16 facets of the LGYNI type, which can be written (in the same form as (3.7), implicitly assuming uniform input bits) as

$$p\big((x \oplus \alpha_1)(a \oplus \alpha_0 \oplus y) = 0, (y \oplus \beta_1)(b \oplus \beta_0 \oplus x) = 0\big) \leq \frac{3}{4},$$

for all $\alpha_0, \alpha_1, \beta_0, \beta_1 = 0, 1$.

Note that this causal polytope for binary inputs and outputs coincides with the polytope of correlations obtained from a local model augmented with one bit of communication, as described in Ref. [122]. This is because the use of just one bit of (one-way) communication is of course compatible with a definite causal order, either $A \prec B$ or $B \prec A$, and for binary inputs one bit is enough for one party to send all the information about her input to the other party. In general however, the polytopes described in Ref. [122] are different from causal polytopes.

Appendix O

See-Saw algorithm

Maximizing the violation of a causal inequality over the process matrix and the instruments is a nonlinear problem, which makes it intractable directly. To address this problem, we used an approach inspired by the See-Saw algorithm of Werner and Wolf [74]. The idea is that if Alice and Bob's instruments are fixed, then the combination of probabilities that enters the causal inequality is a linear function of the W matrix, and maximizing it is a semidefinite programming (SDP) problem [40] that can be solved efficiently. In the same spirit, if the W matrix and the instruments of one party are fixed, then the value of interest is a linear function of the instruments of the other party, and again its optimization is a SDP problem. Hence, one can try to approach the maximum violation of a causal inequality by optimizing over the process matrix and the parties' instruments in an iterative manner.

More specifically, let $\omega(W, \mathcal{A}, \mathcal{B})$ be the value taken by the combination of probabilities in the causal inequality, considered as a function of the process matrix W and the sets of instruments $\mathcal{A} = \{M_{a|x}^{A_I A_O}\}_a$ and $\mathcal{B} = \{M_{b|y}^{B_I B_O}\}_b$ (in their CJ representation). We start the algorithm by generating random sets of instruments $\mathcal{A}_0$ and $\mathcal{B}_0$, and for these fixed instruments we maximize ω considered as a function of W, via the following SDP problem:

$$\begin{gathered}\text{maximize} \quad \omega(W, \mathcal{A}_0, \mathcal{B}_0) \\ \text{subject to} \\ W \geq 0\,, \quad \operatorname{tr} W = d_{A_O} d_{B_O}\,, \\ {}_{B_I B_O}W = {}_{A_O B_I B_O}W\,, \quad {}_{A_I A_O}W = {}_{A_I A_O B_O}W\,, \\ W = {}_{B_O}W + {}_{A_O}W - {}_{A_O B_O}W\,.\end{gathered}$$

With the optimal process matrix W_0 thus obtained and the fixed set of instruments $\mathcal{B}_0$ for Bob, we proceed to optimize ω as a function of Alice's instruments, via the SDP problem

$$\begin{gathered}\text{maximize} \quad \omega(W_0, \mathcal{A}, \mathcal{B}_0) \\ \text{subject to} \\ \forall\, x, a, \quad M_{a|x}^{A_I A_O} \geq 0\,, \quad \operatorname{tr}_{A_O} \sum_a M_{a|x}^{A_I A_O} = \mathbb{1}^{A_I}\,.\end{gathered}$$

With the optimal set of instruments $\mathcal{A}_0$ obtained now and the previously obtained process matrix W_0, we do the analogous optimization over Bob's set of instruments $\mathcal{B}$, and iterate the three optimization steps of the algorithm until it converges.

One can see that at each step of the algorithm the value of ω can only increase, so it is guaranteed to converge to a local maximum. One does not, however, always get the global maximum, and in practice one must repeat the algorithm several times to get a good lower bound on the maximal value of ω.

Note that this See-Saw algorithm can of course straightforwardly be adapted to more than two parties.

Appendix P

Maximal violations for qubits

The best violations of our GYNI and LGYNI causal inequalities that we found, for local dimensions $d = d_{A_I} = d_{A_O} = d_{B_I} = d_{B_O} = 2$ (i.e., for "qubits"), is reached by any convex combination

$$q\, W_{\mathsf{max},d=2} + (1-q)\, W'_{\mathsf{max},d=2}$$

(with $0 \leq q \leq 1$) of the two process matrices $W_{\mathsf{max},d=2}$ and $W'_{\mathsf{max},d=2}$ defined as

$$\begin{aligned} W^{(\prime)}_{\mathsf{max},d=2} = \frac{1}{4}\Big[\mathbb{1}^{\otimes 4} + a_0^{(\prime)}\, Z\mathbb{1}Z\mathbb{1} - a_1^{(\prime)}\big(Z\mathbb{1}\mathbb{1}\mathbb{1}+\mathbb{1}\mathbb{1}Z\mathbb{1}\big) - a_2^{(\prime)}\big(Z\mathbb{1}\mathbb{1}Z+\mathbb{1}ZZ\mathbb{1}\big) \\ + a_3^{(\prime)}\big(Z\mathbb{1}ZZ+ZZZ\mathbb{1}\big) + a_4^{(\prime)}\big(Z\mathbb{1}XX-Z\mathbb{1}YY+XXZ\mathbb{1}-YYZ\mathbb{1}\big)\Big] \end{aligned}$$

(with implicit tensor products and implicit superscripts), where the coefficients a_0, a_1, a_2, a_3, and a_4 are, respectively, real roots of the polynomials

$$\begin{gathered} 4\,608\,x^4 - 1\,575\,x^3 + 525\,x^2 - 117\,x - 1, \\ 221\,184\,x^4 + 142\,479\,x^3 - 19\,701\,x^2 - 15\,603\,x + 2\,363, \\ 9\,216\,x^4 - 16\,857\,x^3 + 11\,724\,x^2 - 3\,660\,x + 430, \\ 221\,184\,x^4 - 50\,895\,x^3 - 16\,200\,x^2 + 1\,368\,x + 602, \\ 221\,184\,x^4 + 16\,335\,x^3 - 37\,008\,x^2 - 11400\,x + 3\,440, \end{gathered}$$

and the primed coefficients a'_0, a'_1, a'_2, a'_3, and a'_4 are, respectively, real roots of the polynomials

$$\begin{gathered} 4\,608\,x^4 + 8\,595\,x^3 + 5\,583\,x^2 + 873\,x - 43, \\ 221\,184\,x^4 - 101\,601\,x^3 - 1\,701\,x^2 + 2\,745\,x + 305, \\ 3\,072\,x^4 - 2\,229\,x^3 + 540\,x^2 - 60\,x + 4, \\ 221\,184\,x^4 - 294\,975\,x^3 + 145\,080\,x^2 - 31\,224\,x + 2\,492, \\ 221\,184\,x^4 + 16\,335\,x^3 - 37\,008\,x^2 - 11\,400\,x + 3\,440. \end{gathered}$$

Numerically, their values are

$$\begin{aligned} a_0 &= 0.2744, & a'_0 &= 0.0390, \\ a_1 &= 0.2178, & a'_1 &= 0.3355, \\ a_2 &= 0.3628, & a'_2 &= 0.2451, \\ a_3 &= 0.3114, & a'_3 &= 0.4291, \\ & a_4 = a'_4 = 0.2097. \end{aligned}$$

Using the same instruments for Alice and Bob as those defined in Eqs. (3.13–3.16), our maximal probability $p_{\text{GYNI}}^{\text{max},d=2}$ of winning the GYNI game with qubits is then found to be the smallest real root of the polynomial

$$1\,769\,472\,x^4 - 2\,884\,032\,x^3 + 1\,630\,800\,x^2 - 380\,052\,x + 34\,087,$$

and our maximal probability $p_{\text{LGYNI}}^{\text{max},d=2}$ of winning the LGYNI game with qubits is $p_{\text{LGYNI}}^{\text{max},d=2} = p_{\text{GYNI}}^{\text{max},d=2} + 1/4$. Numerically, we obtain

$$p_{\text{GYNI}}^{\text{max},d=2} \approx 0.5694 > \frac{1}{2}\,, \quad p_{\text{LGYNI}}^{\text{max},d=2} \approx 0.8194 > \frac{3}{4}\,.$$

It is somewhat surprising that these maximal violations of the GYNI and LGYNI inequalities with qubits have such complicated expressions—contrary for instance to the case of Oreshkov *et al.*'s original causal inequality (for which the violation exhibited in Ref. [3] was proven, under certain constraints, to be optimal [46]), or to the case of the well known CHSH Bell inequality [67, 75]. Note also that, as mentioned in the main text, we could find higher violations of the GYNI inequality using higher-dimensional quantum systems (see Table 3.1), while we couldn't find any higher violations of the LGYNI inequality.

Appendix Q

The set of process matrix correlations is convex

In this Appendix we show that the set of process matrix correlations is convex.

Let $p_0(a,b|x,y)$ and $p_1(a,b|x,y)$ be two correlations realized by the (valid) process matrices and instruments $\{W_0, M_{0;\,a|x}^{A_I A_O}, M_{0;\,b|y}^{B_I B_O}\}$ and $\{W_1, M_{1;\,a|x}^{A_I A_O}, M_{1;\,b|y}^{B_I B_O}\}$, respectively, so that

$$\begin{aligned} p_0(a,b|x,y) &= \mathrm{tr}\left[(M_{0;\,a|x}^{A_I A_O} \otimes M_{0;\,b|y}^{B_I B_O}) \cdot W_0\right], \\ p_1(a,b|x,y) &= \mathrm{tr}\left[(M_{1;\,a|x}^{A_I A_O} \otimes M_{1;\,b|y}^{B_I B_O}) \cdot W_1\right]. \end{aligned}$$

Without loss of generality we assume that Alice and Bob's input and output systems in $\{W_0, M_{0;\,a|x}^{A_I A_O}, M_{0;\,b|y}^{B_I B_O}\}$ and in $\{W_1, M_{1;\,a|x}^{A_I A_O}, M_{1;\,b|y}^{B_I B_O}\}$ have the same dimensions (one can indeed always embed lower-dimensional systems into larger-dimensional ones). Let us now introduce some ancillary 2-dimensional input systems with Hilbert spaces $\mathcal{H}^{A'_I}$ and $\mathcal{H}^{B'_I}$ and define, for any $q \in [0,1]$,

$$\begin{aligned} W &= q\,|0\rangle\langle 0|^{A'_I} \otimes |0\rangle\langle 0|^{B'_I} \otimes W_0 \\ &\quad + (1-q)\,|1\rangle\langle 1|^{A'_I} \otimes |1\rangle\langle 1|^{B'_I} \otimes W_1\,, \end{aligned}$$

$$M_{a|x}^{A'_I A_I A_O} = |0\rangle\langle 0|^{A'_I} \otimes M_{0;\,a|x}^{A_I A_O} + |1\rangle\langle 1|^{A'_I} \otimes M_{1;\,a|x}^{A_I A_O}\,,$$

$$M_{b|y}^{B'_I B_I B_O} = |0\rangle\langle 0|^{B'_I} \otimes M_{0;\,b|y}^{B_I B_O} + |1\rangle\langle 1|^{B'_I} \otimes M_{1;\,b|y}^{B_I B_O}\,.$$

One can easily check that W thus defined is a valid process matrix, and that $M_{a|x}^{A'_I A_I A_O}$ and $M_{b|y}^{B'_I B_I B_O}$ define valid instruments. Furthermore, a straightforward calculation shows that for these process matrix and instruments,

$$\begin{aligned} p(a,b|x,y) &= \mathrm{tr}\left[(M_{a|x}^{A'_I A_I A_O} \otimes M_{b|y}^{B'_I B_I B_O}) \cdot W\right] \\ &= q\ \mathrm{tr}\left[(M_{0;\,a|x}^{A_I A_O} \otimes M_{0;\,b|y}^{B_I B_O}) \cdot W_0\right] \\ &\quad + (1-q)\ \mathrm{tr}\left[(M_{1;\,a|x}^{A_I A_O} \otimes M_{1;\,b|y}^{B_I B_O}) \cdot W_1\right] \\ &= q\,p_0(a,b|x,y) + (1-q)\,p_1(a,b|x,y)\,. \end{aligned}$$

Thus, any convex combination $q\,p_0 + (1-q)\,p_1$ of process matrix correlations can also be realized with a process matrix and suitable instruments, which shows that

the set of process matrix correlations is indeed convex. It is straightforward to generalize the proof to a scenario with more parties.

Note that our construction only shows that the convex hull of the set of correlations produced by process matrices of dimensions $d_{A_I} \times d_{A_O} \times d_{B_I} \times d_{B_O}$ is contained in the set of correlations produced by process matrices of dimensions $2d_{A_I} \times d_{A_O} \times 2d_{B_I} \times d_{B_O}$, opening up the possibility that the set of process matrix correlations is not convex for any fixed input dimension (see Figure 3.2), analogously to the set of (nonsignaling) quantum correlations for fixed dimensions [123].

Appendix R

The Choi-Jamiołkowski isomorphism

We use two levels of the Choi-Jamiołkowski isomorphism. The first level is the "pure" CJ isomorphism, used to represent matrices as vectors. The CJ operator of a matrix $A : \mathcal{H}_I \to \mathcal{H}_O$ is its "double-ket" [118, 119]:

$$|A\rangle\!\rangle := \mathbb{1} \otimes A|\mathbb{1}\rangle\!\rangle = \sum_{i=0}^{d_{\mathcal{H}_I}-1} |i\rangle A|i\rangle \tag{R.1}$$

The second level is the "mixed" CJ isomorphism, used to represent linear operators that act on matrices as matrices themselves. The CJ operator of a map $\mathcal{M} : \mathcal{L}(\mathcal{H}_I) \to \mathcal{L}(\mathcal{H}_O)$ is

$$\mathfrak{C}(\mathcal{M}) := \mathcal{I} \otimes \mathcal{M}(|\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|) = \sum_{i,j=0}^{d_{\mathcal{H}_I}-1} |i\rangle\langle j| \otimes \mathcal{M}(|i\rangle\langle j|) \tag{R.2}$$

Note that this convention differs from the one in [3, 80] by a transposition. This is necessary to avoid representing completely positive maps with a non-positive matrix, as would happen with the convention from [3, 80] when, for example, we take in equation (4.1) the process as being identity from P to A_I and identity from A_O to F, and the map $\mathcal{A}$ to be a SWAP.

Appendix S

The link product

The motivation for defining the link product is to express the CJ operator of a composition of two linear maps $\mathcal{M}$ and $\mathcal{N}$ directly as a function of the CJ operators of the individual maps [32], as

$$\mathfrak{C}(\mathcal{N} \circ \mathcal{M}) = \mathfrak{C}(\mathcal{N}) * \mathfrak{C}(\mathcal{M}).$$

We do not need, however, to explicitly use the CJ isomorphism, as the link product can be more simply defined as an abstract operation on two matrices. Namely, the link product between two operators $M \in \bigotimes_{i\in I} A^i$ and $N \in \bigotimes_{i\in J} A^i$ is

$$N * M := \mathrm{tr}_{I\cap J}[(\mathbb{1}^{J\setminus I} \otimes M^{T_{I\cap J}})(N \otimes \mathbb{1}^{I\setminus J})] \tag{S.1}$$

Note that when $I \cap J = \emptyset$ the link product reduces to $N * M = N \otimes M$ and, conversely, when $I \cap J = I \cup J$, it reduces to $N * M = \mathrm{tr}\, M^T N$.

Appendix T

Conditions for validity

In this appendix we derive the constraints on W such that the map

$$G_{A,B} = W * (A \otimes B)$$

is a valid CPTP map for all valid CPTP maps A and B. Recall that a linear map $\mathcal{A} : A_I \to A_O$ is completely positive and trace preserving if and only if its CJ operator $A = \mathfrak{C}(\mathcal{A})$ is positive and $\mathrm{tr}_{A_O} A = \mathbb{1}^{A_I}$. Therefore we need that

$$\mathrm{tr}_{A'_O B'_O F}\, W * (A \otimes B) = \mathbb{1}^{A'_I B'_I P} \quad \text{and} \quad W * (A \otimes B) \geq 0 \tag{T.1}$$

$$\forall A, B \quad \text{s.t.} \quad A \geq 0, \quad \mathrm{tr}_{A_O A'_O} A = \mathbb{1}^{A_I A'_I} \tag{T.2}$$

$$B \geq 0, \quad \mathrm{tr}_{B_O B'_O} B = \mathbb{1}^{B_I B'_I} \tag{T.3}$$

First we show that $W * (A \otimes B) \geq 0$ iff $W \geq 0$. To see that, note that when $\mathcal{A}$ and $\mathcal{B}$ are swaps we have $W * (A \otimes B) = W$, so $W \geq 0$ is a necessary condition for $G_{A,B} \geq 0$. To see that it is also sufficient, just note that the link product of two positive operators is always positive.

To get the constraints on W such that $\mathrm{tr}_{A'_O B'_O F}\, W * (A \otimes B) = \mathbb{1}^{A'_I B'_I P}$, first note that the positivity of A, B, and W is not relevant for that, since the set of positive operators is a full dimensional subset of the space of hermitian operators. Therefore we only need the normalisation condition

$$\mathrm{tr}_{A'_O B'_O F}\, W * (A \otimes B) = \mathbb{1}^{A'_I B'_I P} \tag{T.4}$$

$$\forall A, B \quad \text{s.t.} \quad \mathrm{tr}_{A_O A'_O} A = \mathbb{1}^{A_I A'_I}, \quad \mathrm{tr}_{B_O B'_O} B = \mathbb{1}^{B_I B'_I} \tag{T.5}$$

Note also that the dimension of A'_I and A'_O does not affect which set of operators $\mathrm{tr}_{A'_I A'_O} A$ is[1], so we can without loss of generality set $d_{A'_I} = d_{A'_O} = d_{B'_I} = d_{B'_O} = 1$, simplifying this condition to

$$\mathrm{tr}_F\, W * (A \otimes B) = \mathbb{1}^P \tag{T.6}$$

$$\forall A, B \quad \text{s.t.} \quad \mathrm{tr}_{A_O} A = \mathbb{1}^{A_I}, \quad \mathrm{tr}_{B_O} B = \mathbb{1}^{B_I} \tag{T.7}$$

Note also that

$$\mathrm{tr}_{A_O} A = \mathbb{1}^{A_I} \iff A = a - {}_{A_O}a + \frac{\mathbb{1}^{A_I A_O}}{d_{A_O}} \tag{T.8}$$

[1] It would if the CPTP maps A were restricted to be unitaries.

for some Hermitian operator a, where we are using the shorthand notation

$$ {}_X W = \frac{\mathbb{1}^X}{d_X} \otimes \mathrm{tr}_X W. \tag{T.9} $$

Using this in Alice and Bob's sides allows us to reduce the normalization condition to

$$ \forall a, b \quad (\mathrm{tr}_F W) * \left[\left(a - {}_{A_O} a + \frac{\mathbb{1}^{A_I A_O}}{d_{A_O}} \right) \otimes \left(b - {}_{B_O} b + \frac{\mathbb{1}^{B_I B_O}}{d_{B_O}} \right) \right] = \mathbb{1}^P \tag{T.10} $$

For $a = b = 0$ this yields the condition

$$ (\mathrm{tr}_F W) * \left(\frac{\mathbb{1}^{A_I A_O}}{d_{A_O}} \otimes \frac{\mathbb{1}^{B_I B_O}}{d_{B_O}} \right) = \mathbb{1}^P \tag{T.11} $$

For $b = 0$ and $a = 0$, respectively, equation (T.11) together with (T.10) imply that

$$ \forall a \quad (\mathrm{tr}_F W) * \left[(a - {}_{A_O} a) \otimes \frac{\mathbb{1}^{B_I B_O}}{d_{B_O}} \right] = 0 \tag{T.12} $$

$$ \forall b \quad (\mathrm{tr}_F W) * \left[\frac{\mathbb{1}^{A_I A_O}}{d_{A_O}} \otimes (b - {}_{B_O} b) \right] = 0 \tag{T.13} $$

which then imply that

$$ \forall a, b \quad (\mathrm{tr}_F W) * [(a - {}_{A_O} a) \otimes (b - {}_{B_O} b)] = 0 \tag{T.14} $$

Using the fact that the maps ${}_{A_O}$ and ${}_{B_O}$ are self-adjoint under the Hilbert-Schmidt inner product, we see that conditions (T.12)-(T.14) are equivalent to

$$ \mathrm{tr}_{B_I B_O F} W = {}_{A_O} \mathrm{tr}_{B_I B_O F} W \tag{T.15} $$

$$ \mathrm{tr}_{A_I A_O F} W = {}_{B_O} \mathrm{tr}_{A_I A_O F} W \tag{T.16} $$

$$ \mathrm{tr}_F W = {}_{A_O} \mathrm{tr}_F W + {}_{B_O} \mathrm{tr}_F W - {}_{A_O B_O} \mathrm{tr}_F W \tag{T.17} $$

These conditions together with (T.11) characterize completely the affine subspace to which valid process matrices belong. It is, however, convenient to rewrite them so that we characterize this affine subspace as a linear subspace together with a single constraint on the trace of W. To do that, note that conditions (T.15)-(T.17) define linear subspaces, so we only need to rewrite condition (T.11). It is easy to see that it is equivalent to:

$$ \mathrm{tr}_{A_I A_O B_I B_O F} W = \mathrm{tr}_{P A_I A_O B_I B_O F} W \tag{T.18} $$

$$ \mathrm{tr}\, W = d_P d_{A_O} d_{B_O} \tag{T.19} $$

So the linear subspace of valid process matrices is the intersection of the linear subspaces defined in the equations (T.15)-(T.18). Since the projectors onto these subspaces commute, we can obtain the projector onto the subspace of valid process matrices L_V simply by composing them. Elementary manipulations give us then

$$ \begin{aligned} L_V(W) = W - {}_F W + {}_{A_O F} W + {}_{B_O F} W - {}_{A_O B_O F} W - {}_{A_I A_O F} W + {}_{A_I A_O B_O F} W \\ - {}_{B_I B_O F} W + {}_{A_O B_I B_O F} W - {}_{A_I A_O B_I B_O F} W + {}_{P A_I A_O B_I B_O F} W \end{aligned} \tag{T.20} $$

To summarize the results of this section, W is a valid process matrix iff

$$W \geq 0 \tag{T.21}$$

$$\operatorname{tr} W = d_P d_{A_O} d_{B_O} \tag{T.22}$$

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

Appendix U

The null-square lemma

The purpose of this section is to prove that, for any Hermitian matrix C, the greatest null square submatrix of C achievable via a change of basis has size $d_C \equiv \min(n_+, n_-) + n_0$, where n_+, n_-, n_0 denote, respectively, the dimensionality of the subspaces spanned by the eigenvectors of C with positive, negative and zero eigenvalues.

We will first establish that the above quantity is an upper bound for the greatest zero-square submatrix. Let then S be a subspace with the property that $\langle v|C|w\rangle = 0$, for $|v\rangle, |w\rangle \in S$, and denote by Π_0 the projector onto the null space of C.

Note that $\mathrm{dim}(S) \leq \mathrm{dim}(S_0) + \mathrm{dim}(S_{+-})$, where $S_0 \equiv \Pi_0 S$ and $S_{+-} \equiv (\mathbb{1} - \Pi_0)S$. Notice as well that $\langle v|C|w\rangle = 0$, for $|v\rangle, |w\rangle \in S_{+-}$. Clearly, the dimension of S_0 is upper bounded by n_0, hence all we need to do is show that the dimension of S_{+-} is upper bounded by $\min(n_+, n_-)$.

Let $\{|u^{(i)}\rangle\}_i$ be a basis for S_{+-}, and let $\{\lambda_k\}_{k=1}^{n_+}$ ($\{-\mu_k\}_{k=1}^{n_-}$) be the set of positive (minus negative) eigenvalues of C, with eigenvectors $\{|\psi_k\rangle\}_{k=1}^{n_+}$ ($\{|\phi_k\rangle\}_{k=1}^{n_-}$). Then,

$$|u^{(i)}\rangle = \sum_{k=1}^{n_+} v_k^{(i)} |\psi_k\rangle + \sum_{k=1}^{n_-} w_k^{(i)} |\phi_k\rangle.$$

From the condition $\langle u^{(i)}|C|u^{(j)}\rangle = 0$ we have that

$$\sum_{k=1}^{n_+} (v_k^{(i)})^* v_k^{(j)} \lambda_k - \sum_{k=1}^{n_-} (w_k^{(i)})^* w_k^{(j)} \mu_k = 0,$$

or, in other words, $\langle \tilde{v}^{(i)} | \tilde{v}^{(j)} \rangle = \langle \tilde{w}^{(i)} | \tilde{w}^{(j)} \rangle$, with $\tilde{v}_k^{(i)} \equiv \sqrt{\lambda_k} v_k^{(i)}$, $\tilde{w}_k^{(i)} \equiv \sqrt{\mu_k} w_k^{(i)}$. This implies that there exists an isometry U such that $|\tilde{v}^{(i)}\rangle = U|\tilde{w}^{(i)}\rangle$. The vectors $|\tilde{v}^{(i)}\rangle \oplus |\tilde{w}^{(i)}\rangle$ (and thus $|u^{(i)}\rangle$) are generated by applying a linear transformation on $|\tilde{w}^{(i)}\rangle$ and so the space they span cannot be greater than n_-. An identical argument shows that the space spanned by $\{|u^{(i)}\rangle\}_{i=1}^m$ cannot be greater than n_+ either. We have just proven that d_C is an upper bound for the size of the null square.

To prove that it is also a lower bound, note that the $\min(n_+, n_-)$ linearly independent vectors $\{|u_i\rangle \equiv \lambda_k^{-1/2} |\psi_k\rangle + \mu_k^{-1/2} |\phi_k\rangle : i = 1, ..., \min(n_+, n_-)\}$ satisfy $\langle u_i|C|u_j\rangle = 0$ for all i, j. Adding all n_0 vectors from the null space of C, we obtain a basis $\{|u_i\rangle\}_{i=1}^{d_C}$ with the property that $\langle u_i|A|u_j\rangle = 0$.

Bibliography

[1] G. Chiribella, G. M. D'Ariano, P. Perinotti and B. Valiron. 'Quantum computations without definite causal structure'. *Phys. Rev. A* **88** 022318 (2013). arXiv:0912.0195 [quant-ph].

[2] L. M. Procopio, A. Moqanaki, M. Araújo, F. Costa, I. A. Calafell, E. G. Dowd, D. R. Hamel, L. A. Rozema, Č. Brukner and P. Walther. 'Experimental superposition of orders of quantum gates'. *Nat. Commun.* **6** 7913 (2015). arXiv:1412.4006 [quant-ph].

[3] O. Oreshkov, F. Costa and Č. Brukner. 'Quantum correlations with no causal order'. *Nat. Commun.* **3** 1092 (2012). arXiv:1105.4464 [quant-ph].

[4] G. Rubino, L. A. Rozema, A. Feix, M. Araújo, J. M. Zeuner, L. M. Procopio, Č. Brukner and P. Walther. 'Experimental Verification of an Indefinite Causal Order' (2016). arXiv:1608.01683 [quant-ph].

[5] D. Deutsch. 'Quantum Computational Networks'. *Proc. R. Soc. Lond. A* **425** 73–90 (1989).

[6] G. Chiribella. 'Perfect discrimination of no-signalling channels via quantum superposition of causal structures'. *Phys. Rev. A* **86** 040301 (2012). arXiv:1109.5154 [quant-ph].

[7] E. Andersson, J. Bergou and I. Jex. 'Comparison of unitary transforms using Franson interferometry'. *J. Mod. Opt.* **52** 1485–1494 (2005).

[8] T. Colnaghi, G. M. D'Ariano, S. Facchini and P. Perinotti. 'Quantum computation with programmable connections between gates'. *Phys. Lett. A* **376** 2940–2943 (2012). arXiv:1109.5987 [quant-ph].

[9] D. Kleitman and D. Kwiatkowski. 'A lower bound on the length of a sequence containing all permutations as subsequences'. *Journal of Combinatorial Theory, Series A* **21** 129 –136 (1976).

[10] S. Radomirović. 'A Construction of Short Sequences Containing All Permutations of a Set as Subsequences'. *The Electronic Journal of Combinatorics* **19** P31 (2012). URL: http://www.combinatorics.org/ojs/index.php/eljc/article/view/v19i4p31.

[11] A. Y. Kitaev. 'Quantum measurements and the Abelian Stabilizer Problem' (1995). arXiv:quant-ph/9511026.

[12] X.-Q. Zhou, T. C. Ralph, P. Kalasuwan, M. Zhang, A. Peruzzo, B. P. Lanyon and J. L. O'Brien. 'Adding control to arbitrary unknown quantum operations'. *Nat. Commun.* **2** (2011). arXiv:1006.2670 [quant-ph].

[13] X.-Q. Zhou, P. Kalasuwan, T. C. Ralph and J. L. O'Brien. 'Calculating unknown eigenvalues with a quantum algorithm'. *Nature Photonics* **7** 223–228 (2013). arXiv:1110.4276 [quant-ph].

[14] M. Araújo, A. Feix, F. Costa and Č. Brukner. 'Quantum circuits cannot control unknown operations'. *New J. Phys.* **16** 093026 (2014). arXiv:1309.7976 [quant-ph].

[15] G. Chiribella, R. Ionicioiu, T. Jennewein and D. Terno (). Private communication.

[16] M. Mirhosseini, M. Malik, Z. Shi and R. W. Boyd. 'Efficient separation of the orbital angular momentum eigenstates of light'. *Nat. Commun.* **4** 2781 (2013). arXiv:1306.0849 [quant-ph].

[17] G. C. G. Berkhout, M. P. J. Lavery, J. Courtial, M. W. Beijersbergen and M. J. Padgett. 'Efficient Sorting of Orbital Angular Momentum States of Light'. *Phys. Rev. Lett.* **105** 153601 (2010).

[18] M. Arndt, O. Nairz, J. Voss-Andreae, C. Keller, G. van der Zouw and Z. A. 'Wave-particle duality of C_{60} molecules'. *Nature* **401** 680–682 (1999).

[19] S. Eibenberger, S. Gerlich, M. Arndt, M. Mayor and J. Tuxen. 'Matter-wave interference with particles selected from a molecular library with masses exceeding 10 000 amu'. *Phys. Chem. Chem. Phys* **15** 14696–14700 (2013). arXiv:1310.8343 [quant-ph].

[20] K. Nakago, M. Hajdušek, S. Nakayama and M. Murao. 'Parallelized adiabatic gate teleportation'. *Phys. Rev. A* **92** 062316 (2015). arXiv:1310.4061 [quant-ph].

[21] A. Soeda. 'Limitations on quantum subroutine designing due to the linear structure of quantum operators' (). Talk at ICQIT2013.

[22] J. Thompson, M. Gu, K. Modi and V. Vedral. 'Quantum Computing with black-box Subroutines' (2013). arXiv:1310.2927 [quant-ph].

[23] D. Deutsch. 'Quantum Computational Networks'. *Proc. R. Soc. Lond. A* **425** 73–90 (1989).

[24] L. Hardy. 'Foliable operational structures for general probabilistic theories'. *Deep Beauty: Understanding the Quantum World through Mathematical Innovation; Halvorson, H., Ed* 409 (2009). arXiv:0912.4740 [quant-ph].

[25] B. Coecke. 'Quantum picturalism'. *Contemporary physics* **51** 59–83 (2010).

[26] G. Chiribella, G. M. D'Ariano and P. Perinotti. 'Probabilistic theories with purification'. *Phys. Rev. A* **81** 062348 (2010). arXiv:0908.1583 [quant-ph].

[27] G. Chiribella, G. M. D'Ariano and P. Perinotti. 'Informational derivation of quantum theory'. *Phys. Rev. A* **84** 012311 (2011). arXiv:1011.6451 [quant-ph].

[28] M. Araújo, F. Costa and Č. Brukner. 'Computational Advantage from Quantum-Controlled Ordering of Gates'. *Phys. Rev. Lett.* **113** 250402 (2014). arXiv:1401.8127 [quant-ph].

[29] L. Hardy. 'Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure'. *J. Phys. A: Math. Gen.* **40** 3081–3099 (2007). arXiv:gr-qc/0608043.

[30] C. Rovelli. 'Quantum mechanics without time: A model'. *Phys. Rev. D* **42** 2638–2646 (1990).

[31] A. Ashtekar. 'Large Quantum Gravity Effects: Unforeseen Limitations of the Classical Theory'. *Phys. Rev. Lett.* **77** 4864–4867 (1996). arXiv:gr-qc/9610008.

[32] G. Chiribella, G. M. D'Ariano and P. Perinotti. 'Theoretical framework for quantum networks'. *Phys. Rev. A* **80** 022339 (2009). arXiv:0904.4483 [quant-ph].

[33] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[34] E. Davies and J. Lewis. 'An operational approach to quantum probability'. English. *Comm. Math. Phys.* **17** 239–260 (1970).

[35] K. Życzkowski and I. Bengtsson. *Geometry of Quantum States*. Cambridge University Press, 2006.

[36] G. Gutoski and J. Watrous. 'Toward a general theory of quantum games'. *In Proceedings of 39th ACM STOC*. 2006, pp. 565–574. arXiv:quant-ph/0611234.

[37] O. Oreshkov and C. Giarmatzi. 'Causal and causally separable processes'. *New J. Phys.* **18** 093020 (2015). arXiv:1506.05449 [quant-ph].

[38] O. Gühne and G. Tóth. 'Entanglement detection'. *Phys. Rep.* **474** 1–75 (2009). arXiv:0811.2803 [quant-ph].

[39] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[40] Y. Nesterov and A. Nemirovskii. *Interior Point Polynomial Algorithms in Convex Programming*. Studies in Applied Mathematics. Society for Industrial and Applied Mathematics, 1987.

[41] M. Steiner. 'Generalized robustness of entanglement'. *Phys. Rev. A* **67** 054305 (2003). arXiv:quant-ph/0304009.

[42] G. Vidal and R. Tarrach. 'Robustness of entanglement'. *Phys. Rev. A* **59** 141 (1999). arXiv:quant-ph/9806094.

[43] J. Lofberg. 'YALMIP : A Toolbox for Modeling and Optimization in MATLAB'. *Proceedings of the CACSD Conference*. Taipei, Taiwan, 2004. URL: http://users.isy.liu.se/johanl/yalmip.

[44] *The MOSEK optimization toolbox for MATLAB*. URL: http://www.mosek.com/.

[45] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani and S. Wehner. 'Bell nonlocality'. *Rev. Mod. Phys.* **86** 419–478 (2014). arXiv:1303.2849 [quant-ph].

[46] Č. Brukner. 'Bounding quantum correlations with indefinite causal order'. *New J. Phys.* **17** 083034 (2015). arXiv:1404.0721 [quant-ph].

[47] Ä. Baumeler and S. Wolf. 'Perfect signaling among three parties violating predefined causal order'. *Information Theory (ISIT), 2014 IEEE International Symposium on*. 2014, pp. 526–530. arXiv:1312.5916 [quant-ph].

[48] Ä. Baumeler, A. Feix and S. Wolf. 'Maximal incompatibility of locally classical behavior and global causal order in multi-party scenarios'. *Phys. Rev. A* **90** 042106 (2014). arXiv:1403.7333 [quant-ph].

[49] C. Branciard, M. Araújo, A. Feix, F. Costa and Č. Brukner. 'The simplest causal inequalities and their violation'. *New J. Phys.* **18** 013008 (2015). arXiv:1508.01704 [quant-ph].

[50] M. Zych, F. Costa, I. Pikovski and Č. Brukner. 'Bell Inequalities for Temporal Order'. (in preparation).

[51] L. Gurvits. 'Classical deterministic complexity of Edmonds' Problem and quantum entanglement'. *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. ACM. 2003, pp. 10–19.

[52] J. S. Bell. 'On the Einstein-Poldolsky-Rosen paradox'. *Physics* **1** 195–200 (1964).

[53] H Reichenbach. *The Direction of Time*. University of Califonia Press, 1956.

[54] F. Arntzenius. 'Reichenbach's Common Cause Principle'. *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2010.

[55] K. B. Laskey. 'Quantum Causal Networks' (2007). arXiv:0710.1200 [quant-ph].

[56] M. S. Leifer and R. W. Spekkens. 'Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference'. *Phys. Rev. A* **88** 052130 (2013). arXiv:1107.5849 [quant-ph].

[57] C. J. Wood and R. W. Spekkens. 'The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning'. *New J. Phys.* **17** 033002 (2015). arXiv:1208.4119 [quant-ph].

[58] E. G. Cavalcanti and R. Lal. 'On modifications of Reichenbach's principle of common cause in light of Bell's theorem'. *J. Phys. A: Math. Theor.* **47** 424018 (2014). arXiv:1311.6852 [quant-ph].

[59] J. Henson, R. Lal and M. F. Pusey. 'Theory-independent limits on correlations from generalized Bayesian networks'. *New J. Phys.* **16** 113043 (2014). arXiv:1405.2572 [quant-ph].

[60] J. Pienaar and Č. Brukner. 'A graph-separation theorem for quantum causal models'. en. *New J. Phys.* **17** 073020 (2015). arXiv:1406.0430 [quant-ph].

[61] L. Hardy. 'Probability theories with dynamic causal structure: a new framework for quantum gravity' (2005). arXiv:gr-qc/0509120.

[62] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi and Č. Brukner. 'Witnessing causal nonseparability'. *New J. Phys.* **17** 102001 (2015). arXiv:1506.03776 [quant-ph].

[63] Č. Brukner. 'Quantum causality'. *Nat. Phys.* **10** 259–263 (2014).

[64] G. Ziegler. *Lectures on Polytopes*. Springer-Verlag, 1995.

[65] I. Pitowsky. *Quantum Probability – Quantum Logic*. Lecture Notes in Physics. Berlin: Springer-Verlag, 1989.

[66] S. Pironio. 'Lifting Bell inequalities'. *J. Math. Phys.* **46** 062112 (2005). arXiv:quant-ph/0503179 [quant-ph].

[67] J. F. Clauser, M. A. Horne, A. Shimony and R. A. Holt. 'Proposed experiment to test local hidden-variable theories'. *Phys. Rev. Lett.* **23** 880–884 (1969).

[68] D. Avis. *lrs*. URL: http://cgm.cs.mcgill.ca/~avis/C/lrs.html.

[69] M. L. Almeida, J.-D. Bancal, N. Brunner, A. Acín, N. Gisin and S. Pironio. 'Guess Your Neighbor's Input: A Multipartite Nonlocal Game with No Quantum Advantage'. *Phys. Rev. Lett.* **104** 230404 (2010). arXiv:1003.3844 [quant-ph].

[70] A. Jamiołkowski. 'Linear transformations which preserve trace and positive semidefiniteness of operators'. *Rep. Math. Phys* **3** 275–278 (1972).

[71] M.-D. Choi. 'Completely positive linear maps on complex matrices'. en. *Linear Algebra Appl.* **10** 285–290 (1975).

[72] N. Schachner. "Ancestral Voices," in *Astounding Stories Volume XII, Number 4*. Street & Smith Publications, 1933.

[73] R. Barjavel. *Le voyageur imprudent ("The imprudent traveller")*. Denoël, 1943.

[74] R. F. Werner and M. M. Wolf. 'Bell inequalities and entanglement'. *Quant. Inf. and Comp.* **3** 1 (2001). arXiv:quant-ph/0107093.

[75] B. S. Cirel'son. 'Quantum generalizations of Bell's inequality'. *Letters in Mathematical Physics* **4** 93–100 (1980).

[76] S. Popescu and D. Rohrlich. 'Quantum nonlocality as an axiom'. *Found. Phys.* **24** 379–385 (1994).

[77] I. Ibnouhsein and A. Grinbaum. 'Information-theoretic constraints on correlations with indefinite causal order'. *Phys. Rev. A* **92** 042124 (2015). arXiv:1408.6172 [quant-ph].

[78] Ä. Baumeler and S. Wolf. 'The space of logically consistent classical processes without causal order'. *New J. Phys.* **18** 013036 (2016). arXiv:1507.01714 [quant-ph].

[79] A. A. Abbott, C. Giarmatzi, F. Costa and C. Branciard. 'Multipartite Causal Correlations: Polytopes and Inequalities' (2016). arXiv:1608.01528 [quant-ph].

[80] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi and Časlav Brukner. 'Witnessing causal nonseparability'. *New J. Phys.* **17** 102001 (2015). arXiv:1506.03776 [quant-ph].

[81] A. Feix, M. Araújo and Č. Brukner. 'Quantum superposition of the order of parties as a communication resource'. *Phys. Rev. A* **92** 052326 (2015). arXiv:1508.07840 [quant-ph].

[82] P. Allard Guérin, A. Feix, M. Araújo and Č. Brukner. 'Exponential communication complexity advantage from quantum superposition of the direction of communication'. *Phys. Rev. Lett.* **117** 100502 (2016). arXiv:1605.07372 [quant-ph].

[83] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp and F. Unger. 'Limit on Nonlocality in Any World in Which Communication Complexity Is Not Trivial'. *Phys. Rev. Lett.* **96** 250401 (2006). eprint: quant-ph/0508042.

[84] N. Linden, S. Popescu, A. J. Short and A. Winter. 'Quantum Nonlocality and Beyond: Limits from Nonlocal Computation'. *Phys. Rev. Lett.* **99** 180502 (2007). arXiv:quant-ph/0610097.

[85] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter and M. Żukowski. 'Information causality as a physical principle'. *Nature* **461** 1101–1104 (2009). arXiv:0905.2292 [quant-ph].

[86] M. Navascués and H. Wunderlich. 'A glance beyond the quantum model'. *Proc. Royal Soc. A* **466** 881–890 (2009). arXiv:0907.0372 [quant-ph].

[87] T. Fritz, A. B. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier and A. Acín. 'Local orthogonality as a multipartite principle for quantum correlations'. *Nat. Commun.* **4** 2263 (2013). arXiv:1210.3018 [quant-ph].

[88] L. Hardy. 'Quantum Theory From Five Reasonable Axioms' (2001). arXiv:quant-ph/0101012.

[89] B. Dakić and Č. Brukner. 'Deep Beauty: Understanding the Quantum World through Mathematical Innovation'. Cambridge University Press, 2011. Chap. Quantum Theory and Beyond: Is Entanglement Special?, pp. 365–392. arXiv:0911.0695 [quant-ph].

[90] L. Masanes and M. P. Müller. 'A derivation of quantum theory from physical requirements'. *New J. Phys.* **13** 063001 (2011). arXiv:1004.1483 [quant-ph].

[91] H. Barnum, M. P. Müller and C. Ududec. 'Higher-order interference and single-system postulates characterizing quantum theory'. *New J. Phys.* **16** 123029 (2014). arXiv:1403.4147 [quant-ph].

[92] P. A Höhn. 'Toolbox for reconstructing quantum theory from rules on information acquisition' (2014). arXiv:1412.8323 [quant-ph].

[93] P. A Höhn and C. Wever. 'Quantum theory from questions' (2015). arXiv:1511.01130 [quant-ph].

[94] A. Feix, M. Araújo and Č. Brukner. 'Causally nonseparable processes admitting a causal model'. *New J. Phys.* **18** 083040 (2016). arXiv:1604.03391 [quant-ph].

[95] G. C. Ghirardi, A. Rimini and T. Weber. 'Unified dynamics for microscopic and macroscopic systems'. *Phys. Rev. D* **34** 470–491 (1986).

[96] G. C. Ghirardi, P. Pearle and A. Rimini. 'Markov processes in Hilbert space and continuous spontaneous localization of systems of identical particles'. *Phys. Rev. A* **42** 78–89 (1990).

[97] A. Bassi, K. Lochan, S. Satin, T. P. Singh and H. Ulbricht. 'Models of wavefunction collapse, underlying theories, and experimental tests'. *Rev. Mod. Phys.* **85** 471–527 (2013). arXiv:1204.4325 [quant-ph].

[98] S. W. Hawking. 'Breakdown of predictability in gravitational collapse'. *Phys. Rev. D* **14** 2460–2473 (1976).

[99] D. Harlow. 'Jerusalem Lectures on Black Holes and Quantum Information'. *Rev. Mod. Phys.* **88** 015002 (2016). arXiv:1409.1231 [hep-th].

[100] T. Jacobson. 'Trans-Planckian Redshifts andthe Substance of the Space-Time River'. *Progress of Theoretical Physics Supplement* **136** 1–17 (1999). arXiv:hep-th/0001085.

[101] M. Bojowald, D. Cartin and G. Khanna. 'Lattice refining loop quantum cosmology, anisotropic models, and stability'. *Phys. Rev. D* **76** 064018 (2007). arXiv:0704.1137 [gr-qc].

[102] S. Gielen and L. Sindoni. 'Quantum Cosmology from Group Field Theory Condensates: a Review'. *SIGMA* **12** 082 (2016). arXiv:1602.08104 [gr-qc].

[103] P. A. Höhn. 'Quantization of systems with temporally varying discretization. I. Evolving Hilbert spaces'. *J. Math. Phys.* **55** 083508 (2014). arXiv:1401.6062 [gr-qc].

[104] V. Mukhanov. *Physical Foundations of Cosmology*. Cambridge University Press, 2005.

[105] Ä. Baumeler and S. Wolf (2015). Private communication.

[106] S. Aaronson. 'Quantum Computing, Postselection, and Probabilistic Polynomial-Time'. *Proc. R. Soc. A* **461** 3473–3482 (2005). arXiv:quant-ph/0412187.

[107] Ä. Baumeler and S. Wolf. 'Device-independent test of causal order and relations to fixed-points'. *New J. Phys.* **18** 035014 (2016). arXiv:1511.05444 [quant-ph].

[108] Ä. Baumeler and S. Wolf. 'Non-causal computation avoiding the grandfather and information antinomies' (2016). arXiv:1601.06522 [quant-ph].

[109] C. H. Bennett and B. Schumacher. *Teleportation, Simulated Time Travel, and How to Flirt with Someone Who Has Fallen into a Black Hole*. Talk in QUPON. 2005. URL: http://web.archive.org/web/20060207123032/http://www.research.ibm.com/people/b/bennetc/QUPONBshort.pdf.

[110] G. Svetlichny. 'Time Travel: Deutsch vs. Teleportation'. *Int. J. Theor. Phys.* **50** 3903–3914 (2011). arXiv:0902.4898 [quant-ph].

[111] R. Dias da Silva, E. F. Galvão and E. Kashefi. 'Closed timelike curves in measurement-based quantum computation'. *Phys. Rev. A* **83** 012316 (2011). arXiv:1003.4971 [quant-ph].

[112] S. Lloyd et al. 'Closed Timelike Curves via Postselection: Theory and Experimental Test of Consistency'. *Phys. Rev. Lett.* **106** 040403 (2011). arXiv:1005.2219 [quant-ph].

[113] T. A. Brun and M. M. Wilde. 'Perfect State Distinguishability and Computational Speedups with Postselected Closed Timelike Curves'. *Found. Phys.* **42** 341–361 (2012). arXiv:1008.0433 [quant-ph].

[114] M. Araújo, A. Feix, M. Navascués and Č. Brukner. 'A purification postulate for quantum mechanics with indefinite causal order' (2016). arXiv:1401.8127 [quant-ph].

[115] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

[116] S. Aaronson. *PostBQP Postscripts: A Confession of Mathematical Errors*. 2014. URL: http://www.scottaaronson.com/blog/?p=2072.

[117] T. Morimae and H. Nishimura. 'Quantum interpretations of AWPP and APP'. *Quant. Inf. and Comp.* **16** 0498–0514 (2016). arXiv:1502.00067 [quant-ph].

[118] A. Royer. 'Wigner function in Liouville space: A canonical formalism'. *Phys. Rev. A* **43** 44–56 (1991).

[119] S. L. Braunstein, G. M. D'Ariano, G. J. Milburn and M. F. Sacchi. 'Universal Teleportation with a Twist'. *Phys. Rev. Lett.* **84** 3486–3489 (2000). arXiv:quant-ph/9908036.

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

[121] F. G. S. L. Brandão. 'Quantifying entanglement with witness operators'. *Phys. Rev. A* **72** 022310 (2005). arXiv:quant-ph/0503152.

[122] D. Bacon and B. F. Toner. 'Bell Inequalities with Auxiliary Communication'. *Phys. Rev. Lett.* **90** 157904 (2003). arXiv:quant-ph/0208057.

[123] K. F. Pál and T. Vértesi. 'Concavity of the set of quantum probabilities for any given dimension'. *Phys. Rev. A* **80** 042114 (2009). arXiv:0905.2915 [quant-ph].

Mateus Araújo

Curriculum vitæ

Address: Boltzmanngasse 5
1090 Wien, Austria
☎ *+4301427772583*
✉ *mateus.santos@univie.ac.at*
Birth: 1988-08-13

Research Interests

Foundations of Quantum Mechanics, Causal Structures, Quantum Information, Computational Complexity

Education

2012– **Ph.D in Physics**, *University of Vienna*, Vienna, (in progress).
2010–2012 **M.Sc in Physics**, *UFMG*, Belo Horizonte.
2007–2010 **B.Sc in Physics**, *UFMG*, Belo Horizonte.

Master's thesis

title *Quantum realism and quantum surrealism*
supervisor Marcelo Terra Cunha

Languages

Portuguese Native
English Fluent
Spanish Fluent
German Intermediate
French Basic

Conferences

Invited talks

- *Conditions of possible experience* at FQXi Workshop on Quantum Contextuality and Sequential Measurements – 2013-11-06; Sevilla, Spain

Contributed talks

- *A purification postulate for quantum mechanics with no causal order* at CEQIP – 2016-06-19. Valtice, Czech Republic
- *A purification postulate for quantum mechanics with no causal order* at Quantum Networks – 2016-04-01. Barcelona, Spain
- *Witnessing causal nonseparability* at V Quantum Information School and Workshop – 2015-08-12. Paraty, Brazil
- *Witnessing causal nonseparability* at Quantum Information – 2015-06. Benasque, Spain

- *Witnessing causal nonseparability* at IQOQI breakfast – 2015-03-24. Wien, Austria
- *Computational advantage from quantum-controlled ordering of gates* at SFB Meeting – 2014-12-11. Wien, Austria
- *Decrease in query complexity for quantum computers with superposition of circuits* at CoQuS student seminar – 2014-04-07. Wien, Austria
- *Decrease in query complexity for quantum computers with superposition of circuits* at 78. Jahrestagung der DPG und DPG-Frühjahrstagung – 21-03-2014, Berlin, Germany
- *Decrease in query complexity for quantum computers with superposition of circuits* at ÖPG/SPS Annual Meeting 2013 – 2013-09-05. Linz, Austria
- *(Im)possibility of controlling a black box quantum gate* at IV Quantum Information School and Workshop – 2013-08-12. Paraty, Brazil
- *Decrease in query complexity for quantum computers with superposition of circuits* at Quantum Information – 2013-06. Benasque, Spain
- *From the detection loophole to the transmission loophole* at III Encontro temático do INCT-IQ – 2011. Natal, Brazil
- *Maximal CHSH violations with low efficiency photodetection and homodyne measurements* at III Quantum Information School and Workshop – 2011-08-09. Paraty, Brazil
- *Noções de computação quântica* at XXII Escola de Inverno – Física – UFMG – Computação Quântica e Informação Quântica. 2011. Belo Horizonte, Brazil

Talks in scientific visits

- *Quantum computation with linear closed timelike curves* at University of Oxford – 2016-05-10. Oxford, United Kingdom
- *A purification postulate for quantum mechanics with no causal order* at University of Siegen – 2016-04-08. Siegen, Germany
- *Quantum computation with linear closed timelike curves* at Institute of Photonic Sciences – 2016-04-04. Castelldefels, Spain
- *Quantum circuits cannot control unknown operations* at National Quantum Information Centre in Gdansk – 2013-11-29. Gdańsk, Poland
- *Quantum circuits cannot control unknown operations* at University of Bristol – 2013-10-09. Bristol, United Kingdom
- *(Im)possibility of controlling an arbitrary gate* at Slovak Academy of Sciences – 2013-07-29. Bratislava, Slovakia
- *Contextualidade e negatividade são noções equivalentes de não-classicalidade* at Universidade Federal Fluminense – 2012-08-22. Niterói, Brasil

Posters

- *Witnessing causal nonseparability* at QuPoN 2015 – 2015-05-19. Wien, Austria
- *Computational advantage from quantum-controlled ordering of gates* at QIP2015 – 2015-01-13. Sydney, Australia
- *Decrease in query complexity for quantum computers with superposition of circuits*, at Quantum Contextuality, Non-Locality, and the Foundations of Quantum Mechanics – 2014-07-02. Bad Honnef, Germany

- *Quantum circuits cannot control unknown operations*, at Quantum [Un]Speakables II: 50 Years of Bell's Theorem – 2014-19-06. Wien, Austria
- *Quantum circuits cannot control unknown operations* at QIP2014 – 2014-02-04. Barcelona, Spain
- *All noncontextuality inequalities for the n-cylce* at QIP2013 – 2013-02. Beijing, China
- *General bounds for detection efficiency in Bell tests* at TQC2012 – 2012-05. Tokyo, Japan
- *Mapas positivos e testemunhas de emaranhamento* at WECIQ2010 – 2010. Petrópolis, Brazil

Participation

- 28º Colóquio Brasileiro de Matemática – 2011; Teaching Assistant: Mecânica Quântica para Matemáticos em Formação. Rio de Janeiro, Brazil
- Clay Mathematics Institute 2010 Summer School – Probability and Statistical Physics in Two and more Dimensions – 2010. Búzios, Brazil
- XIV Escola Brasileira de Probabilidade – 2010. Búzios, Brazil

Publications

Pre-prints

[1] G. Rubino, L. A. Rozema, A. Feix, M. Araújo, J. M. Zeuner, L. M. Procopio, Č. Brukner, and P. Walther. "Experimental Verification of an Indefinite Causal Order" (2016). arXiv:1608.01683 [quant-ph].

Journal Articles

[2] P. Allard Guérin, A. Feix, M. Araújo, and Č. Brukner. "Exponential communication complexity advantage from quantum superposition of the direction of communication". *Phys. Rev. Lett.* **117** 100502 (2016). arXiv:1605.07372 [quant-ph].

[3] A. Feix, M. Araújo, and Č. Brukner. "Causally nonseparable processes admitting a causal model". *New J. Phys.* **18** 083040 (2016). arXiv:1604.03391 [quant-ph].

[4] C. Branciard, M. Araújo, A. Feix, F. Costa, and Č. Brukner. "The simplest causal inequalities and their violation". *New J. Phys.* **18** 013008 (2015). arXiv:1508.01704 [quant-ph].

[5] A. Feix, M. Araújo, and Č. Brukner. "Quantum superposition of the order of parties as a communication resource". *Phys. Rev. A* **92** 052326 (2015). arXiv:1508.07840 [quant-ph].

[6] M. Navascués, A. Feix, M. Araújo, and T. Vértesi. "Characterizing finite-dimensional quantum behavior". *Phys. Rev. A* **92**, 042117 042117 (2015). arXiv:1507.07521 [quant-ph].

[7] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, and Č. Brukner. "Witnessing causal nonseparability". *New J. Phys.* **17** 102001 (2015). arXiv:1506.03776 [quant-ph].

[8] L. M. Procopio, A. Moqanaki, M. Araújo, F. Costa, I. A. Calafell, E. G. Dowd, D. R. Hamel, L. A. Rozema, Č. Brukner, and P. Walther. "Experimental superposition of orders of quantum gates". *Nat. Commun.* **6** 7913 (2015). arXiv:1412.4006 [quant-ph].

[9] M. Araújo, F. Costa, and Č. Brukner. "Computational Advantage from Quantum-Controlled Ordering of Gates". *Phys. Rev. Lett.* **113** 250402 (2014). arXiv:1401.8127 [quant-ph].

[10] M. Araújo, A. Feix, F. Costa, and Č. Brukner. "Quantum circuits cannot control unknown operations". *New J. Phys.* **16** 093026 (2014). arXiv:1309.7976 [quant-ph].

[11] G. Borges, M. Carvalho, P.-L. de Assis, J. Ferraz, M. Araújo, A. Cabello, M. T. Cunha, and S. Pádua. "Quantum contextuality in a Young-type interference experiment". *Phys. Rev. A* **89** 052106 (2014). arXiv:1304.4512 [quant-ph].

[12] M. Araújo, M. T. Quintino, C. Budroni, M. Terra Cunha, and A. Cabello. "All noncontextuality inequalities for the n-cycle scenario". *Phys. Rev. A.* **88** 022118 (2013). arXiv:1206.3212 [quant-ph].

[13] C. Teo, M. Araújo, M. T. Quintino, J. Minář, D. Cavalcanti, V. Scarani, M. Terra Cunha, and M. França Santos. "Realistic loophole-free Bell test with atom-photon entanglement". *Nature Comm.* **4**, 2104 (2013). arXiv:1206.0074 [quant-ph].

[14] M. Araújo, M. T. Quintino, D. Cavalcanti, M. França Santos, A. Cabello, and M. Terra Cunha. "Tests of Bell inequality with arbitrarily low photodetection efficiency and homodyne measurements". *Phys. Rev. A* **86** 030101 (2012). arXiv:1112.1719 [quant-ph].

[15] M. T. Quintino, M. Araújo, D. Cavalcanti, M. França Santos, and M. Terra Cunha. "Maximal violations and efficiency requirements for Bell tests with photodetection and homodyne measurements". *J. Phys. A: Math. Theor.* **45** 215308 (2012). arXiv:1106.2486 [quant-ph].

Other

[16] M. Araújo. "Quantum realism and quantum surrealism". Master's thesis. 2012. arXiv:1208.6283 [quant-ph].

[17] M. Araújo. "Fundamentos matemáticos da separabilidade quântica". Presented at IMPA. Monografia. 2010. URL: http://www.mat.ufmg.br/~tcunha/MonografiaMateus.pdf.