DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Indefinite causal structures: physical implementability and applications“

verfasst von / submitted by

Adrien Feix

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

Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2018 / Vienna 2018

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. Caslav Brukner

Abstract

In quantum theory, it is usually assumed that events are embedded in a global causal order. In this thesis, we examine the consequences of lifting this assumption from a quantum information perspective, building upon the recently developed "processes matrix" formalism. We first investigate why and how certain processes can violate "causal inequalities"—constraints on the probability distributions that all causally ordered processes satisfy. This sheds light on possible criteria to distinguish physically relevant processes from those that appear to be mere mathematical artefacts of the formalism. Second, we study a specific class of physically implementable processes, in which the order in which two parties, Alice and Bob, apply their operations is put in a coherent superposition of "Alice being before Bob" and "Bob being before Alice". We demonstrate that these resources allow for a reduction of the communication required to complete a certain task, and then prove that this advantage scales exponentially with the length of the parties' inputs. Third, we apply the formalism to a fixed causal order with three parties "Alice being before Bob" where the causal relationship between Alice's and Bob's event is in a quantum superposition of a direct causal link and a shared common cause. We develop a criterion to distinguish such a situation from classical mixtures of causal structures and propose a physical implementation combining a coherent spatial superposition of a mass with general relativistic time dilation, two features that are expected to be present in any quantum gravity theory.

Zusammenfassung

Die Quantentheorie beruht üblicherweise auf der Annahme, dass Ereignisse in eine globale Kausalordnung eingebettet werden können. Diese Dissertation setzt sich, aus der Perspektive der Quanteninformation, mit den Konsequenzen des Verzichts auf diese Annahme auseinander, aufbauend auf dem kürzlich entwickelten "Prozessmatrix-Formalismus". Wir untersuchen zunächst inwieweit gewisse Prozesse "kausale Ungleichungen" – Bedingungen auf der Ebene der Wahrscheinlichkeitsditribution, die für alle Prozesse mit kausaler Ordnung gelten – verletzen können. Dies wirft ein neues Licht auf mögliche Kriterien um physikalisch sinnvolle Prozesse von denen, die bloße mathematische Artefakte des Formalismus sind, zu unterscheiden. Daraufhin betrachten wir eine spezifische Klasse von Prozessen die physikalisch implementiert werden können, bei denen die kausale Beziehung zwischen den zwei Parteien, Alice und Bob, in einer kohärenten Superposition von "Alice implementiert ihre Operation vor Bob" und "Bob implementiert seine Operation vor Alice" ist. Wir zeigen, dass diese Ressourcen für eine bestimmte Aufgabe die Möglichkeit eröffnet, das Ausmaß der benötigten Kommunikation zu reduzieren und beweisen, dass diese Reduktion exponentiell mit der Länge der Inputs der Parteien skaliert. Schließlich wenden wir den Formalismus auf ein Szenario mit drei Parteien und fixierter kausaler Ordnung an ("Alice vor Bob") bei dem die kausale Beziehung zwischen Alice's und Bob's Ereignissen in einer kohärenten Superposition zwischen einem direkten kausalen Einfluss und einer gemeinsamen Ursache ist. Wir entwickeln eine Methode, um eine solche Situation von einer inkohärenten Mischung zu unterscheiden und schlagen als physikalische Implementierung ein Gedankenexperiment vor, das eine räumliche Superposition einer Masse mit der allgemein relativistischen Zeitdilatation kombiniert – zwei Aspekte, die in jede Theorie der Quantengravitation berücksichtigen muss.

Contents

Chapter 1

Introduction

Archibald Wheeler's famous intuition that physical quantities ultimately derive their significance from answers to binary questions—epitomized in the slogan "it from bit" [Whe90]—has two facets. Its ontological aspect is beyond the grasp of physics, but the related *methodological* question is meaningful: can and should physics study phenomena from the point of view of information science?

The only way to answer the question is to explore this possibility, in particular in areas of physics where other approaches have not always been satisfactory, such as in the field of quantum theory. Indeed, *quantum information*, the study of the way information is carried and transformed by quantum systems has proven to be wildly successful. It has contributed to significant advances in quantum computing, quantum cryptography, quantum many-body physics and understanding Bell's inequalities and their violation.

In this dissertation, we will apply the quantum information paradigm—without making any ontological assumptions or claims—to the study of phenomena arising within and beyond standard quantum theory. The main toolbox that we make use of is the recently developed *process matrix formalism* [OCB12, ABC$^+$15], proposed to study causality from a quantum information perspective. This introduction will give a brief overview over the formalism and some of the open questions arising in its development, putting the contributions in the thesis in an overarching context.

1.1 Higher-order operations and process matrices

The process matrix formalism has proved to be a convenient tool because it combines two useful features. First, it provides a unified representation of quantum states, quantum operations on states and "higher-order" operations on quantum operations. Second, it allows for arbitrary causal structures for those higher-order operations on operations. We will briefly describe these two characteristics in this section; for a more thorough treatment of the subject, we refer the reader to Appendices D.1 and D.2.

Choi-Jamiołkowski isomorphism and quantum combs

Quantum systems are usually described by states in Hilbert spaces. For a pure state in the Hilbert space $\mathcal{H}^I$, a more general *mixed state* is represented by a density matrix $\rho \in I$, where I is the space of linear operators on $\mathcal{H}^I$. The evolution of a quantum system and its associated quantum state can be described in very general terms by quantum operations on the quantum state, that is, a completely positive trace preserving (CPTP) map or quantum channel $\mathcal{M} : I \to O$, where O is the space of density matrices corresponding to the output of the map [NC00]. In a graphical representation of a circuit, such an operation can be represented by a "box" acting on a "wire" that stands for the quantum system (see Fig. 1.1 (a)).

An equivalent representation of the linear map $\mathcal{M}$, which puts input and output Hilbert spaces on an equal footing, is obtained by applying $\mathcal{M}$ to one half of a maximally entangled state, leaving the other half untouched. The resulting positive matrix $M \in I \otimes O$ contains information about the output of the map for every possible input state. The *isomorphism* between $\mathcal{M}$ and the matrix M is called Choi-Jamiołkowski (CJ) isomorphism [Cho75, Jam72]. In the convention that will use throughout the thesis,[1] it is given by:

$$M := [(\mathcal{I} \otimes \mathcal{M})(|I\rangle\rangle\langle\langle I|)]^{\mathrm{T}} \in I \otimes O, \tag{1.1}$$

where $\mathcal{I}$ is the identity map, $|I\rangle\rangle := \sum_{j=1}^{d_{\mathcal{H}_I}} |jj\rangle \in \mathcal{H}_I \otimes \mathcal{H}_I$ is a non-normalized maximally entangled state and $^{\mathrm{T}}$ denotes matrix transposition in the computational basis.

Besides quantum operations on states, one can also consider linear operations on quantum operations. This can for instance be relevant to investigate "black-box" or "functional" quantum computing, in which algorithm takes unknown quantum operations as an input [CDP08a, AFCB14]. Consider, for instance, a supermap transforming a quantum operation $\mathcal{W}(\mathcal{M}) = \mathcal{M}'$, where $\mathcal{M}' : I' \to O'$ is a quantum operation acting on a state in I' (see Fig. 1.1 (b)). A natural normalization condition in this case is that W map CPTP maps to CPTP maps. Making use of the CJ-isomorphism of (1.1) for quantum maps again, one can think of $\mathcal{W}$ not as a supermap acting on maps, but rather as a map acting on the CJ-matrix $M \in I \otimes O$ or, by applying (1.1) again, as a CJ-matrix itself: $W \in I \otimes O \otimes I' \otimes O'$.

When considering supermaps of a several different quantum maps $\mathcal{W}(\mathcal{M}_A, \mathcal{M}_B, \mathcal{M}_C, \dots) = \mathcal{W}'$, things get more complicated. *One* type of supermaps corresponds to "quantum circuit with open wires" or "quantum comb" [CDP08a, CDP08b, CDP09, BCDP11], into which the maps or systems are plugged in, resulting in another (super)map (see, for instance, the supermap acting on four maps depicted in Fig. 1.1 (c)). Quantum combs have a clear physical interpretation, since they can be implemented by a sequence of quantum channels with memory [CDP09].

Each open wire of the quantum comb corresponds to a Hilbert space,[2] and the full quantum comb's CJ-matrix is an element of the tensor product of the Hilbert spaces corresponding to all open wires. Applying a supermap $\mathcal{W}$ to its inputs $\mathcal{M}_i$ ("plugging the operations into" the circuit) in terms of CJ-matrices W and M_i is simple. The resulting (super)map's CJ representation W' is the partial trace over the Hilbert spaces corresponding to connected wires of the product of W and the tensor product of maps $\bigotimes_i M_i$; it is an element of the tensor product of Hilbert spaces corresponding to all open wires *remaining* after connecting the operations.

$$\mathrm{tr}_{\text{connected Hilbert spaces}}\left[W \cdot \bigotimes_i M_i\right] = W' \tag{1.2}$$

The quantum comb formalism provides a unified treatment of states, maps, and supermaps of any order—all of them are quantum combs. This feature simplifies a number of optimization problems over parts of quantum circuits [BCD+10, BDPS14, BDPC11]. However, it is important to realise that the assumption that all supermaps can be implemented as quantum circuits is nontrivial. In particular, it implies that the supermap has a *fixed causal order* of applying those operations—the order in which they feature in the

[1] Other conventions differ by a complex conjugation or a transpose.

[2] This guarantees that the supermap is linear in its inputs, i.e., that each operation acted upon is "plugged into" the quantum circuit exactly once.

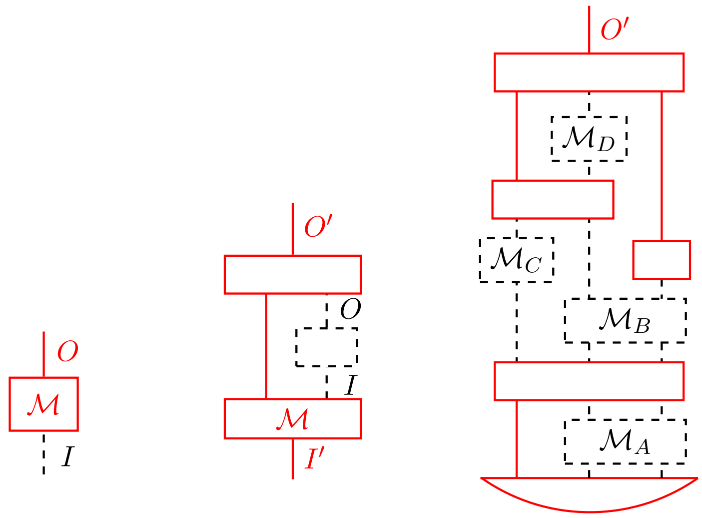

Figure 1.1: Circuit representation of different types of quantum (super)maps with a fixed causal order. The elements belonging to the (super)maps are depicted in solid red, the systems or operations acted upon (to be "plugged in") in dashed black: (a) an operation mapping a state to a state ($\in O$), (b) an operation mapping a quantum map to a quantum map ($\in I' \to O'$), (c) an operation mapping four quantum maps to a state ($\in O'$).

corresponding quantum circuit. In the next section, we will see that this assumption can be lifted to describe a wider range of linear supermaps.

Dropping causality: process matrices

Dropping the assumption that linear quantum supermaps apply quantum operations they act upon only according to a fixed order is not completely straightforward. This is related to *normalization*, which was touched upon in the previous section. We call a general supermap normalized if it maps normalized (super)maps to normalized (super)maps, for instance normalized density matrices or to CPTP maps. For quantum combs, this condition is automatically satisfied because quantum comb is built from a sequence of CPTP maps, i.e., from quantum channels with memory. If we drop the assumption of causal order for the supermap, this is not the case anymore and additional normalization conditions have to be imposed "by hand". They guarantee that the supermap is well-behaved (or "valid")—a precondition for developing a physical interpretation and implementation of the supermap. In order to distinguish supermaps with a fixed causal order from the broader class of valid quantum supermaps, we will call the latter "processes", and their corresponding CJ-matrix "process matrices" following the standard terminology used introduced in Ref. [OCB12].

The conditions for the validity of process matrices are conveniently formulated on the process matrix level. Eq. (1.2) also applies to process matrices without a fixed causal order. Requiring normalization simply means that the process matrix W applied to any collection of CJ-matrices of CPTP maps (M_i^{CPTP}) should be a valid (normalized) process matrix itself [OCB12], which means that taking the trace over the *remaining open wires* with any collection of CPTP maps (N_i^{CPTP}) should be equal to one:[3]

[3]Which can be interpreted as a "trivial" valid process matrix.

$$\operatorname{tr}\left[W \cdot \bigotimes_i M_i^{\mathrm{CPTP}} \bigotimes_j N_j^{\mathrm{CPTP}}\right] = 1, \quad \forall M_i^{\mathrm{CPTP}}, N_j^{\mathrm{CPTP}} \tag{1.3}$$

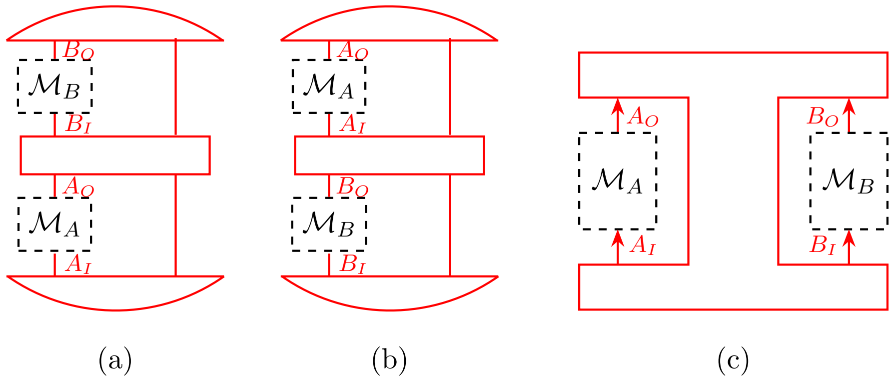

Figure 1.2: Supermaps mapping two quantum maps $\mathcal{M}_A, \mathcal{M}_B$ to a probability. For quantum combs, either (a) $\mathcal{M}_A$ is applied before $\mathcal{M}_B$ or (b) $\mathcal{M}_B$ before $\mathcal{M}_A$. A general quantum process (c) has no such *a priori* structure besides being normalized.

At this point, an interesting question is whether validity of process matrices is not ultimately equivalent to a fixed causal order, which applies to quantum combs, which would make the class of process matrices equivalent to the class of quantum combs. The first interesting case[4] to test this hypothesis is that of process matrices mapping two CP maps $\mathcal{M}_A, \mathcal{M}_B$ to a probability, graphically depicted in Fig. 1.2. A quantum comb can either apply $\mathcal{M}_A$ before $\mathcal{M}_B$ or $\mathcal{M}_B$ before $\mathcal{M}_A$, while a general process has no such restriction.

Oreshkov, Costa, and Brukner [OCB12] provided an explicit example of a valid process matrix (fulfilling the constraint of Eq. (1.3)) which is neither a quantum comb nor a *causally separable* process, that could be decomposed as a convex combination of quantum combs. This shows that there are mathematically well-defined quantum supermaps which cannot be represented as quantum circuits with open wires. We will examine some previously studied mathematical results, physical interpretations, and applications of such processes in the following section.

1.2 Interpretation and applications of the process matrix formalism

To study interpretations and applications of such "causally nonseparable" processes, it will prove useful to interpret the CP maps the process acts upon as laboratories belonging to *parties*: Alice implements $\mathcal{M}_A$, Bob $\mathcal{M}_B$, Charlie $\mathcal{M}_C$, etc. In these laboratories, standard quantum theory applies, that is, the the parties apply standard quantum experiments, including measurement and preparation of states. In this context, the process encodes the causal relations between the different parties and the presence or absence of a causal order in the process has a physical significance. As a simple example, take a process with a fixed causal order where Alice is after Bob (i.e., $\mathcal{M}_A$ is applied after $\mathcal{M}_B$ in the corresponding circuit); in this case, there can be no signalling from Alice to Bob. From this point of

[4] For maps acting on a single operation or only on states, the normalization condition is equivalent to causality [OCB12].

view, causally nonseparable processes, such as the one found in Ref. [OCB12] are resources allowing the parties to complete certain tasks or "games", potentially with an advantage over processes with a definite causal order and convex combinations thereof.[5]

Causal inequalities

The most general type of task is as follows: each party is given a (possibly empty) *input* and the goal is that each party (or only a subset thereof) provide a *correct output*, which depends on the inputs of one or more other parties. Let us first consider games where the inputs and outputs are *classical bit strings*. Each party will implement a *quantum instrument* [DL70]—a set of CP maps corresponding to the outcomes of a generalized measurement and repreparation—corresponding to their classical input. The output is given by the result of this generalized measurement.

In the bipartite case, two parties are each given an input bit: x for Alice and y for Bob. We denote Alice's CP maps (in their CJ representation) by $M_A^{a|x}$ and Bob's CP maps by $M_B^{b|y}$. Since Alice and Bob implement a valid quantum instrument, the sum over the outcomes of the CP maps is a CPTP map, that is:

$$\mathrm{tr}_{A_O} \sum_a M_A^{a|x} = \mathbb{1}^{A_I}, \mathrm{tr}_{B_O} \sum_b M_B^{b|y} = \mathbb{1}^{B_I}, \tag{1.4}$$

where $\mathbb{1}$ is the identity matrix.

For a given process matrix W given as a resource to the two parties (see Fig. 1.2 (c)), the conditional probability distribution is given by applying Eq. (1.2):

$$p(a,b|x,y) = \mathrm{tr}\Big[W \cdot M_A^{a|x} \otimes M_B^{b|y}\Big]. \tag{1.5}$$

A very simple task is for Alice to output Bob's input ($a = y$) and for Bob to output Alice input ($b = x$), the "Guess Your Neighbour's Input" game [BAF+16]. The probability of success can be expressed as:

$$p_{\text{succ}} = p(a = y, b = x) = \frac{1}{4} \sum_{x,y,a,b} \delta_{a,y}\delta_{b,x} \,\mathrm{tr}\Big[W M_A^{a|x} \otimes M_B^{b|y}\Big]. \tag{1.6}$$

Any causally ordered process is of the type where Alice's operation is applied first ($A \prec B$) or Bob's operation is applied first ($B \prec A$), as graphically depicted in Fig. 1.2 (a) and (b). To identify which probability distributions $p(a,b|x,y)$ can be obtained with such processes takes four steps [BAF+16] that are very similar to the ones used to characterize probability distributions compatible with local realism in the study of Bell inequalities [Bel64, BCP+13].

First, one enumerates all *deterministic* probability distributions (i.e., mappings from x, y to a, b) for the given scenario, such as the one where Alice and Bob have an input and an output bit (in this case, there are $4^4 = 256$ such deterministic probability distributions). Second, one eliminates those probability distributions that are not compatible with either $A \prec B$ (the condition being $p^{A\prec B}(a,b|x,y) = p(a|x)p(b|x,y,a)$) or $B \prec A$ (the condition being $p^{B\prec A} = p(b|y)p(a|x,y,b)$), leaving only deterministic probability distributions compatible with a fixed causal order (in the example with one bit of input and output per party, 112 of them). Third, one takes these probability distributions as *vertices of a polytope* in the space of probability distributions (its dimension is the product of the dimensions of

[5]In the game interpretation, such convex combinations can allow for "two-way signalling" in the sense that the probability distribution of Alice's output depends on Bob's input and vice-versa.

a, b, x, y, i.e., $2^4 = 16$ in our example). The polytope represents all the strategies resulting from access to *shared randomness* among Alice and Bob, which includes randomly choosing between the causal orders, i.e., all correlations arising from *causally separable* processes. Finally, one can compute an alternative representation of the polytope, in terms of its *facets*. Some of these facets describe trivial constraints (such as normalization of probabilities); the nontrivial ones can be interpreted as "causal inequalities".

For the simple example of one bit of input and output for Alice and Bob, there are two nontrivial causal inequalities [BAF+16], one of which is a bound on the success probability of the GYNI task that holds for all causally separable process matrices W_{sep}:

$$p_{\text{succ}}^{\text{sep}} = p^{\text{sep}}(a = y, b = x) = \frac{1}{4} \sum_{x,y,a,b} \delta_{a,y} \delta_{b,x} \operatorname{tr}\left[W_{\text{sep}} M_A^{a|x} \otimes M_B^{b|y}\right] \leq \frac{1}{2} \tag{1.7}$$

Any probability distribution that lies *outside* this causal polytope, and thus violates one or more causal inequalities, is the signature of an underlying causally nonseparable process. We emphasize that this conclusion can be drawn without any assumptions about the dimension Alice's and Bob's input and output space or on the maps that Alice and Bob implement—it is "device-independent" [ABG+07]. Causal inequalities are analogous to Bell inequalities, whose violation rules out that the underlying state is separable without requiring additional assumptions about its dimension or on the operations used to produce the probability distribution.

A number of examples of processes that enable the violation of different types of causal inequalities when combined with adequate local maps for Alice and Bob were found in Refs. [OCB12, BAF+16]. In these bipartite examples, it is interesting to note that the causal inequalities are apparently *not violated maximally.* For instance, the logical bound of GYNI inequality of Eq. (1.6) is $p_{\text{succ}} = 1$, yet numerical optimization indicates that process matrices seem to be able to reach only $p_{\text{succ}} \approx 0.62$ [BAF+16]. For a slightly more complicated causal inequality and for a restricted class of operations, it was proven that valid process matrices cannot reach the logical bound and the actual bound (analogously to the "Tsirelson bound" [Tsi80] for Bell inequalities) was derived [Bru15]. It was also shown that for three parties, there are causal inequalities that be maximally violated by valid process matrices [BW13], adding to the similarity between Bell inequality violation and the violation of causal inequalities.

Yet, a number of interesting questions remained open. First, what type of operations do the parties need to implement to violate causal inequalities? In the two-party case, there is a proof that when Alice and Bob are limited to *classical operations*[6], no violation of causal inequalities is possible [OCB12, BB16], prompting the hypothesis that quantum mechanics and the violation of causal inequalities are intricately linked. We will come back to this hypothesis in Chapter 2. Second, the *physical meaning* of processes that violate causal inequalities was poorly understood. Their most straightforward interpretation—in terms of "channels going back in time"—seems to be physically implausible. One possibility would be that respecting causal inequalities could itself be a fundamental principle, and only processes that do not allow for their violation are physical. We will explore this possibility in detail in Chapter 3.

Causal witnesses

Since there is a correspondence between Bell inequalities to device-independently certify entanglement and causal inequalities to device-independently certify causal nonseparability,

[6] That is, all their CP maps can be diagonalized in the same basis.

it is natural to explore the analogy further. There are entangled states that do *not* allow for the violation of any Bell inequalities so they admit a local hidden variable model [Wer89]. While their entanglement cannot be certified only on the basis of probability distributions, it can in principle be inferred *device-dependently*, from a full state tomography. Actually, a *partial tomography* is often sufficient to compute the expectation value of an "entanglement witness" [CS14], an operator R which has positive eigenvalues for all separable states, but negative ones for (some) entangled states: $\mathrm{tr}[\rho_{\mathrm{sep}} \cdot R] \geq 0 \ \ \forall \rho_{\mathrm{sep}}$.

In the same way, one can define "causal witnesses"—operators S that has positive eigenvalues for causally separable processes but negative ones for (some) causally nonseparable ones [ABC+15]:

$$\mathrm{tr}[W_{\mathrm{sep}} \cdot S] \geq 0 \quad \forall W_{\mathrm{sep}}. \tag{1.8}$$

Interestingly, and in contrast to entanglement witnesses, the optimal causal witness can be computed efficiently for any causally separable process, since the corresponding conditions are *semidefinite constraints* [ABC+15].

The parties can measure the expectation value of the causal witness S by decomposing it in a basis of CP maps, performing a partial process tomography. The correlations Alice and Bob, *together* with the knowledge of what operators they are implementing, allow for the device-dependent certification of causal nonseparability, for every causally nonseparable process. In between causal inequalities—without any additional assumptions—and causal witnesses—relying on complete knowledge of the operations implemented—there are a number of weaker assumptions to certify causal nonseparability. One of them is to rely only on an assumption about the *dimension* of the quantum systems acted upon. The significance of this "semi-device independent" [LVB11, PB11] certification of causal nonseparability is shown in Chapter 4.

Applications

Another way of interpreting the certification of causal nonseparability through causal witnesses is as a task that the parties have to achieve. In this case, the inputs are no longer bit strings, but *unknown quantum operations*. For instance, take three parties: Alice, Bob, and Charlie. Alice and Bob are each given unitaries (U_A, U_B) that either commute $([U_A, U_B] = 0)$ or anticommute $(\{U_A, U_B\} = 0)$. They are "black boxes" that cannot be used more than once (for instance to tomograph them). Charlie then has to determine whether the unitaries commute or anticommute.

As first shown in Ref. [CDPV13], this task cannot be successfully completed using processes with a definite causal order between Alice, Bob and Charlie, yet there is a process matrix that allows for it: the "quantum switch". In this process, the order (Alice before Bob or Bob before Alice) in which the operations are applied to a target system ρ_T is controlled by the value of a quantum control system σ_C. It is causally non-separable and can be thought of as a *coherent superposition* of circuits (or of directions of communication) controlled by a control qubit, see Fig. 1.3 for a graphical representation.

Using the quantum switch with the control state $|+\rangle_C = (|0\rangle_C + |1\rangle_C)/\sqrt{2}$, Alice and Bob simply apply their unitaries U_A and U_B to the target state $|\psi_T\rangle$. Charlie then receives a state with that has an amplitude proportional to the commutator $[U_A, U_B]\,|\psi_T\rangle$ and to the anticommutator $\{U_A, U_B\}\,|\psi_T\rangle$ in the superposition basis $(|0\rangle_C \pm |1\rangle_C)/\sqrt{2}$, such that a measurement can distinguish between the commuting and the anticommuting case. The quantum switch thus allows for the discrimination to be successfully completed with just one use of the unknown unitaries U_A, U_B, while an ordered circuit would require either Alice or Bob to apply their unitary at least twice. This advantage in the number of queries to blackbox unitaries can be extended to n unitaries with an adaptation of the task and of

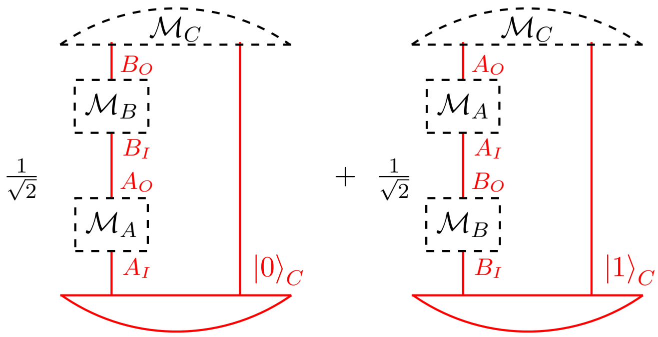

Figure 1.3: The "quantum switch": a coherent superposition of circuits involving three parties. When the control qubit is in the state $|0\rangle_C$, Alice's operation is applied before Bob'; when the control qubit is $|1\rangle_C$, Bob's operation is applied before Alice's. A coherent superposition of the control qubit leads to entanglement between the control qubit and the causal order. Charlie can use this property, for example to determine whether Alice's and Bob's unitaries commute or anticommute.

the process, resulting in a *reduction in query complexity* from $O(n^2)$ for causally separable processes to $O(n)$ for causally nonseparable processes [ACB14].

These advantages are significant because the generalized quantum switches have a clear physical interpretation and can in principle be implemented in the lab. For instance, the proof-of-principle of the quantum switch superposing the order of two parties was first implemented in Ref. [PMA+15] and subsequently extended in Ref. [RRF+17] to show causal nonseparability by direct measurement of a causal witness. It is therefore also interesting to explore other similar processes in terms of information processing power. Since the query complexity advantage was proven for a task derived from a causal witness [ABC+15], one can examine tasks connected to semi-device independent certification of causal nonseparability. In Chapter 4, we show that the associated advantage is a reduction of the communication necessary to achieve a specific task; in Chapter 5, we study the way this advantage scales with the input length.

Physical relevance

The quantum switch and similar processes, which consist of the quantum superposition of causally ordered processes, are also interesting from a foundational point of view. In any theory of quantum gravity, the causal structure (determined by general relativity) should be dynamic and quantized, and it is reasonable to expect that quantum superposition of causal structures will be one of its features. While not being in any way a theory of quantum gravity, the process matrix formalism can nevertheless be used to describe such situations.

One effect that can be expected from the interplay of general relativity and quantum theory concerns the gravitational time-dilation arising from a massive body in a quantum superposition of positions. It should lead to entanglement between proper time of a clock close to the mass and the position of the mass. Take Alice and Bob to have initially synchronized clocks and the events in their laboratories to be defined with respect to their local clocks: Bob's operation being applied at local time τ_B and Alice's at local time τ_A. If a massive body is put in a superposition between being close to Alice and close to Bob, it can also put the *causal relationship* between the two events defined by Alice's and Bob's local clocks in a superposition. If the difference is sufficiently large, one can achieve a situation in which, for one position of the mass, Alice's event can be before Bob, while, for the other position, Bob's event is before Alice's. In this way, the quantum switch can be *implemented gravitationally*, giving rise to a number of foundational questions [ZCPB17, CRGB17]. In Chapter 6, we study a conceptually similar, but more realistic scenario, in which the gravitational time dilation can be arbitrarily small, only moving Bob's event in or out of Alice's event's lightcone.

1.3 Outline of the thesis

The thesis consists of five chapters, each corresponding to the main text of a published paper. Supplementary information to each chapter is contained in the corresponding appendixes.

Maximal incompatibility with a global causal order

Chapter 2 deals with causal inequalities for multipartite scenarios and makes progress on two major open questions of the process formalism. The first result is about the *maximal violation* of causal inequalities by process matrices. Extending a previous result for the three-party case [BW13], we show that process matrices can maximally violate some causal inequalities for three or more parties. The second result concerns the role of quantum mechanics for the violation of causal inequalities. We mentioned in the previous section that in the two-party case, Alice and Bob have to implement genuinely quantum operations in order to violate causal inequalities. Surprisingly, for three and more parties, this is not the case. The process matrices and the operations of the parties can be chosen to be diagonal in the parties' computational basis, so that they effectively are only operations on classical bits. The corresponding "classical process matrices" can therefore produce just as intriguing violations of causal inequalities as genuinely quantum ones.

Nonseparable processes with a causal model

Chapter 3 investigates the relationship between causal inequalities and causal witnesses in the bipartite case. The previously known two-party causally nonseparable processes [OCB12, BAF^{+}16] capable of violating causal inequalities with the appropriate strategies have no clear physical interpretation, while the "quantum switch", a three-party process that cannot violate causal inequalities [ABC^{+}15] has a physical interpretation. This prompts the conjecture that the violation of causal inequalities is somehow forbidden by nature, while causal nonseparability is not. We show that there are two-party processes that are nonseparable and cannot violate any causal inequality, yet also lack any physical interpretation. Additionally, we give numerical evidence to show that a more sophisticated criterion, related to the capability of a process to violate causal inequalities when the parties additionally share entanglement [OG16] appears to be similarly inadequate in distinguishing between physically implementable processes and mere mathematical artefacts.

A novel quantum communication resource

In Chapter 4, we develop a criterion for causal nonseparability that lies between the device-independent violation of causal inequalities and the device-independent violation of causal witnesses. A *semi-device independent* [LVB11, PB11] certification of causal nonseparability is based merely on the conditional probability distribution and an assumption on the dimension of the quantum systems that leave the parties' laboratories, but makes no assumptions on the operations the parties implement. Inspired by a quantum causal witness, we introduce a three-party task that cannot be completed for any process with a fixed causal order if the quantum systems exchanged are restricted to qubits. We then show that, using the quantum switch and putting the *order of communication in superposition*, the parties can succeed with unit probability, even if when restricted to sending out qubits. This shows that the quantum switch can reduce the amount of communication necessary for certain tasks and that its causal nonseparability can be certified with very restricted assumptions.

Exponential communication complexity advantage

Chapter 5 builds upon the previous result by examining the way the communication advantage of the quantum switch scales with the length of the inputs given to the parties. We first show that when requiring *deterministic* success for a suitable generalization of the task, using the quantum switch as a resource can reduce the amount of communication required (the "communication complexity") exponentially. We then show that this communication complexity advantage remains even when allowing for a *bounded probability of failing*, showing how powerful a resource the superposition of the direction of communication can be.

Quantum superpositions of causal structures

In Chapter 6, we apply the process matrix formalism to a slightly different scenario. We consider a fixed underlying causal structure with three events A, B, C. A is before B and C and B is before C, but the *causal relationship* between A directly influencing B and A and B sharing a common cause. We develop the tools to distinguish such a quantum causal structure from a merely classical mixture of direct and common cause processes and define "witnesses of causal nonclassicality". We then propose an implementation of the process involving quantum superposition of a massive particle and general relativistic time dilation. The entanglement between the position of the mass and the proper time of Alice and Bob leads allows for the causal relationship between events defined with respect to Alice's and Bob's local clocks to be put in a superposition, for arbitrarily small superpositions.

Chapter 2

Maximal incompatibility of locally classical behavior and global causal order in multi-party scenarios

Abstract

Quantum theory in a global space-time gives rise to non-local correlations, which cannot be explained causally in a satisfactory way; this motivates the study of theories with reduced global assumptions. Oreshkov, Costa, and Brukner [OCB12] proposed a framework in which quantum theory is valid locally but where, at the same time, no global space-time, i.e., predefined causal order, is assumed beyond the absence of logical paradoxes. It was shown for the two-party case, however, that a global causal order always emerges in the classical limit. Quite naturally, it has been conjectured that the same also holds in the multi-party setting. We show that counter to this belief, classical correlations locally compatible with classical probability theory exist that allow for deterministic signaling between three or more parties incompatible with any predefined causal order.

Published. Ä. Baumeler, A. Feix, and S. Wolf. Maximal incompatibility of locally classical behavior and global causal order in multiparty scenarios. *Phys. Rev. A*, 90(4):042106, 2014.

Contribution to the main proof establishing the maximal violation of causal inequalities for any dimension and to the writing of the manuscript.

2.1 Motivation and main result

According to Bell [Bel81], correlations cry out for explanation. In such a spirit, already Einstein, Podolsky, and Rosen (EPR) [EPR35] had asked for an extension of quantum theory that incorporates a *causal explanation* [Rei56, Wis14] of the correlations arising when two parts of an entangled quantum state are measured. Such an explanation can describe the emergence of the correlations either through *pre-shared information* or through *influences*. Because of relativity, EPR argued further, the latter cannot be the cause of such correlations. Later, (finite-speed) influences were ruled out by theory [WS15, BBLG13] and experiments [Asp82, SS97, WJS+98, Asp99, RKM+01, SBv+08, GMR+13]. Therefore, still according to EPR's reasoning, physical quantities need to be *predefined*. This, however, had been rejected by Bell [Bel64] under the assumption that spatially separated settings can be chosen (at least partially [Hal11, BG11, PRB+14]) freely and independently; such correlations are called *non-local*. Remarkably, this means that there are *not predefined yet correlated* physical quantities emerging in a space-like separated way. However, although the EPR program as such may have failed, it seems natural to continue to ask for a causal explanation of non-local correlations. A possible approach is to refrain from considering space-time as fundamental, treating it as emerging (potentially along with other macroscopic quantities) from a deeper fundament [Par03, Woo84, BLMS87, D'A11] — comparably to temperature. A step in this direction was taken by Hardy [Har05, Har07] with his program of merging general relativity with quantum theory, in which he proposes to extend the latter to *dynamical causal orders*, a feature of relativity (see [Bru14] for a recent review on quantum theory and causality). Chiribella, D'Ariano, and Perinotti [CDP09, CDPV13] studied quantum supermaps called "quantum combs" that allow for superpositions of causal orders. Based on Hardy's idea, Oreshkov, Costa, and Brukner [OCB12] developed a framework for quantum correlations without predefined causal order by dropping the assumption of a *global background time* while keeping the assumptions that *locally*, nature is described by quantum theory and that *no logical paradoxes* arise. Some causal structures emerging from this framework cannot be predefined [OCB12, BW14, Bru15] — just like physical quantities exhibiting non-local correlations [Bel64, Tsi80, GHZ89, GHSZ90]. If, in the *two-party* case, we consider the classical limit of the quantum systems, *i.e.*, enforce both parties' physics to be described by *classical probability theory* (instead of quantum theory), then a predefined causal order *always emerges* [OCB12]. This is in accordance with our experience and, hence, natural and unsurprising; it strongly indicates that the same may hold in the *multi-party* case [Cos13]. This, however, fails to be true, as we show in the present work.

2.2 Input-output systems, and causal order

By definition, measurement settings and outcomes are classical, *i.e.*, perfectly distinguishable. Therefore, we think of physical systems as black boxes which we probe with classical inputs and that respond with classical outputs. When taking this perspective, we describe all *physical quantities*, *i.e.*, *outputs*, as functions of *inputs*. The party S is described by a set of *inputs* $V(S) = \{A_i\}_I$ chosen *freely* by S, and by a set of *outputs* $Q(S) = \{X_j\}_J$ the party can access (instantiations of the inputs and outputs are denoted by the same letters but in lowercase). Since we refrain from assuming global *space-time* as given *a priori*, we cannot define *free randomness* based on such a causal structure, as done elsewhere [CR11, GR13]. Instead, we take the concept of *free randomness* as fundamental — in accordance with a recent trend to derive properties of quantum theory from information-theoretic principles [CFS02, BZ03, CBH03, Bra05, PPK+09, CDP11, MM13, PW13] — and *postulate*

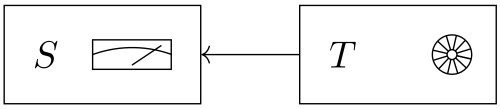

Figure 2.1: If party T can freely choose an input (here, visualized by a knob), and party S can read off an output that depends on T's input, then T can *signal* to S, which implies that S is in the *causal future* of T ($S \succeq T$).

inputs as being *free*.

Outputs are functions of inputs. Based on this relationship, we define *causal order*. If an output X_j is a function of A_i, we say that X_j *causally depends* on A_i and is in the *causal future* of A_i or, equivalently, that A_i is in the *causal past* of X_j, denoted by $X_j \succeq A_i$ or $A_i \preceq X_j$. The negations of these relations are denoted by $\not\succeq$ and $\not\preceq$. This definition does neither induce a causal order between outputs nor between inputs nor between any output and the input it does not depend on.

Let us introduce a second party T described by the set of inputs $V(T) = \{B_k\}_K$ and with access to the outputs $Q(T) = \{Y_\ell\}_L$. Outputs can depend on inputs of both parties. If party S has an output that depends on an inputs of T, then we say that T can *signal* to S (see Figure 2.1). In the following, we will assume *unidirectional* signaling: If S can signal to T, then T cannot signal to S. This enables us to causally order parties. If at least one output of S depends on an input of T, but no output of T depends on any input of S (which is the condition for unidirectional signaling), then S is in the causal future of T. Formally, if there exist $X \in Q(S)$ and $B \in V(T)$ fulfilling $X \succeq B$ and if for all $Y \in Q(T)$ and for all $A \in V(S)$, the relation $Y \not\succeq A$ holds, then we have $S \succeq T$.

Consider a two-party scenario with parties S, T, each having a single input A, B, a single output X, Y, respectively, and a shared random variable Λ. We call a theory compatible with *predefined causal order* if all achievable probability distributions $P(x, y|a, b)$ can be written as a convex mixture of possible causal orders, *i.e.*,

$$
\begin{aligned}
P(x, y|a, b) = \Pr(\alpha) \sum_\lambda \Pr(\lambda|\alpha) \Pr(x|a, \lambda, \alpha) \Pr(y|a, b, \lambda, \alpha) \\
+ \Pr(\neg\alpha) \sum_\lambda \Pr(\lambda|\neg\alpha) \Pr(x|a, b, \lambda, \neg\alpha) \Pr(y|b, \lambda, \neg\alpha)\,,
\end{aligned}
$$

where α is the event $S \preceq T$, and λ is an instantiation of Λ that depends on a input not in either of the sets $V(S)$ or $V(T)$. For more than two parties, the definition of predefined causal order becomes more subtle. Suppose we have three parties S, T, and U, where S is in the causal past of both T and U. We call a causal order *predefined* even if S is free to choose the causal order between T and U [Cos, Gia]. In general, in a predefined causal order, a party is allowed to determine the causal order between all parties in her causal future. Hence, a theory with the parties $S_0, \ldots, S_{n-1}$, inputs $A_0, \ldots, A_{n-1}$ (shorthand $\mathbf{A}$), and outputs $X_0, \ldots, X_{n-1}$ ($\mathbf{X}$), respectively, is compatible with *predefined causal order* if all achievable probability distributions $P(\mathbf{x}|\mathbf{a})$ can be written as

$$
P(\mathbf{x}|\mathbf{a}) = \sum_{i=0}^{n-1} \Pr(\alpha_i \wedge \neg\alpha_0 \wedge \cdots \wedge \neg\alpha_{i-1}) \Pr(\mathbf{x}|\mathbf{a}, S_i \text{ is first})\,,
$$

where α_i is the event that each party $S_{j(\neq i)}$ either is in the causal future of S_i ($S_i \preceq S_j$) or has no causal relation with S_i ($S_i \not\preceq S_j$ and $S_i \not\succeq S_j$). The term $\Pr(\mathbf{x}|\mathbf{a}, S_i \text{ is first})$ is a convex mixture of distributions compatible with the causal structures in which S_i is first and chooses the causal order between the remaining parties.

2.3 Game

The following multi-party game cannot be won in a scenario with predefined causal order. Denote by $S_0, \dots, S_{n-1}$ the parties that participate in the game. Each party S_i has a uniformly distributed binary input A_i as well as a binary output X_i and access to the shared random variable M uniformly distributed in the range $\{0, \dots, n-1\}$. The random variable M belongs to a dummy party (we need her as a source of shared randomness). For given $M = m$, the game is won whenever S_m's output X_m is the parity of the inputs to all other parties, *i.e.*, $X_m = \bigoplus_{i \neq m} A_i$. Therefore, the success probability for winning the game is

$$p_{\text{succ}} = \frac{1}{n} \sum_{m=0}^{n-1} \Pr\left(X_m = \bigoplus_{i \neq m} A_i \,\middle|\, M = m \right) . \tag{2.1}$$

In a setup with predefined causal order, this success probability is upper bounded by $1 - 1/(2n)$. To see this, note that if, without loss of generality, S_0 is first, she will remain first. For $n > 2$, the last party can be specified by S_0. Thus, all the terms of the sum in expression (2.1) are 1 except for the first summand, reflecting the fact that S_0 herself has to guess the parity of the other's inputs, which is 1/2. By repeating the experiment $\omega(n)$ times, one can bring the winning probability arbitrarily closely to 0.

2.4 Framework for classical correlations without causal order

Instead of assuming that locally, nature is described by *quantum theory* [OCB12], we take the *classical limit* of the systems and thus assume that locally, nature is described by *classical probability theory*. In addition to this, we require the probabilities of the outcomes to be non-negative and to sum up to 1; this excludes logical paradoxes [OCB12, CDPV13]. We suppose that each party has a closed laboratory that can be opened once — which is when the only interaction with the environment happens. When a laboratory is opened, the party receives, manipulates, and outputs a state. Thus, in the setting of local validity of classical probability theory, such a laboratory is described by a conditional probability distribution $P_{O|I}$, where I is the input to and O is the output from the laboratory.

Let us consider the parties as described in the game. We denote the input to S_i by I_i and the output from S_i by O_i. Therefore, the ith local laboratory is described by the distribution $P_{X_i,O_i|A_i,I_i}$. As we do not make global assumptions other than that the overall picture should describe a probability distribution, we describe everything outside the laboratories by the distribution W (see Figure 2.2) satisfying the condition that for any choice of $a_0, \dots, a_{n-1}$, *i.e.*, for any probability distribution $P_{X_0,O_0|I_0}, \dots, P_{X_{n-1},O_{n-1}|I_{n-1}}$, the values of the product with W, *i.e.*, the values of

$$P_{X_0,O_0|I_0} \cdot \ldots \cdot P_{X_{n-1},O_{n-1}|I_{n-1}} \cdot P_{I_0,\dots,I_{n-1}|O_0,\dots,O_{n-1}} , \tag{2.2}$$

and in general of

$$P_{O_0|I_0} \cdot \ldots \cdot P_{O_{n-1}|I_{n-1}} \cdot P_{I_0,\dots,I_{n-1}|O_0,\dots,O_{n-1}} , \tag{2.3}$$

are non-negative and sum up to 1. For tackling this condition formally, we represent a probability distribution P_X as a real positive diagonal matrix $\hat{P}_X$ with trace 1 and diagonal entries $P_X(x)$. A conditional probability distribution $P_{X|Y}$ is a collection of (unconditional) probability distributions $P_{X|Y=y}$ for each value of y. Thus, we represent $P_{X|Y}$ similarly,

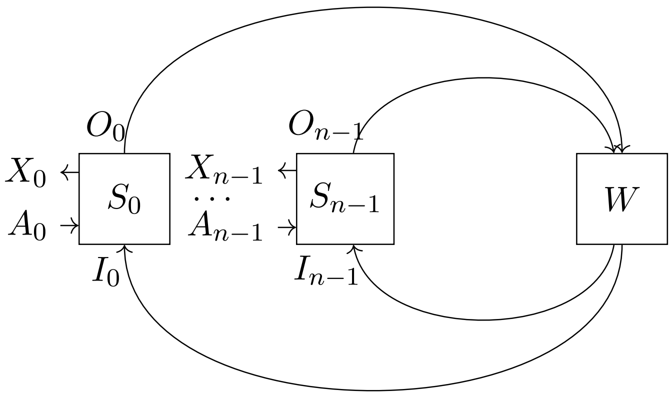

Figure 2.2: Party S_i is described by $P_{X_i,O_i|A_i,I_i}$. Her output O_i is fed to W, which describes *everything* outside the laboratories. Therefore, W also provides the input I_i and is described by $W = P_{I_0,\ldots,I_{n-1}|O_0,\ldots,O_{n-1}}$.

yet with trace $|\mathcal{Y}|$, where $\mathcal{Y}$ is the set of values y can take, and we use the symbol $\hat{P}_{X|Y}$. The values $P_{X|Y=y}(x)$ are ordered with respect to the ordering of the subscripts of $P_{X|Y}$, *e.g.*, for binary X and Y the matrix $\hat{P}_{X|Y}$ is

$$\begin{pmatrix} P_{X|Y=0}(0) & 0 & 0 & 0 \\ 0 & P_{X|Y=1}(0) & 0 & 0 \\ 0 & 0 & P_{X|Y=0}(1) & 0 \\ 0 & 0 & 0 & P_{X|Y=1}(1) \end{pmatrix} . \tag{2.4}$$

The condition that the probabilities $P_{X|Y=y}(x)$ sum up to 1 for fixed y is reflected by the condition that if we trace out X from the matrix $\hat{P}_{X|Y}$ (denoted by $\operatorname{tr}_X \hat{P}_{X|Y}$), we are left with the identity. The product of two distributions $\hat{P}_X$ and $\hat{P}_Y$ in the matrix representation corresponds to the tensor product denoted by $\hat{P}_X \otimes \hat{P}_Y$. To obtain the marginal distribution from a joint distribution, we use the partial trace. This implies that the output state of a laboratory $P_{O_i|I_i}$, given the input state P_{I_i}, is $\operatorname{tr}_{I_i}(\hat{P}_{O_i|I_i} \cdot (\mathbb{1}_{O_i} \otimes \hat{P}_{I_i}))$, where $\mathbb{1}_{O_i}$ is the identity matrix with the same dimension as $\hat{P}_{O_i}$. This allows us to use the framework of Oreshkov, Costa, and Brukner [OCB12], where we restrict ourselves to diagonal matrices, *i.e.*, all objects (W and local operations) are simultaneously diagonalizable in the computational basis and can, hence, be expressed using the identity $\mathbb{1}$ and the Pauli matrix σ_z. We know from their framework [OCB12] that if we express $P_{I_0,\ldots,I_{n-1}|O_0,\ldots,O_{n-1}}$ as a matrix $W = c \sum_i g_i$, where c is a normalization constant and $g_i = R_{i,0} \otimes \cdots \otimes R_{i,n-1} \otimes T_{i,0} \otimes \cdots \otimes T_{i,n-1}$. For every i, the summand g_i represents a channel from all S_j with $\operatorname{tr} T_{i,j} = 0$ to all S_k with $\operatorname{tr} R_{i,k} = 0$. In order to avoid logical paradoxes, g_i must describe a channel where at least one party is a recipient without being a sender [OCB12]. In other words, g_i must either be the identity or there exists j such that $T_{i,j} = \mathbb{1}$ and $\operatorname{tr} R_{i,j} = 0$.

2.5 Winning the game perfectly

To win the game using this framework, we need to provide the distribution $P_{I_0,\ldots,I_{n-1}|O_0,\ldots,O_{n-1}}$ and all distributions describing the laboratories. For that purpose, we use the fact that if a set $\{g_i\}_I$ of matrices with all eigenvalues in $\{-1,1\}$ forms an Abelian group with respect to matrix multiplication, then $\sum_{i\in I} g_i$ is a positive semi-definite matrix. To prove this, take the eigenvector $\mathbf{v}$ which has the smallest eigenvalue $\lambda_{\min}$, *i.e.*,

$$\sum_{i\in I} g_i \mathbf{v} = \sum_{i\in I} \lambda_i^{\mathbf{v}} \mathbf{v} = \lambda_{\min} \mathbf{v}\,, \tag{2.5}$$

where $\lambda_i^{\mathbf{v}}$ is the eigenvalue of g_i with respect to the eigenvector $\mathbf{v}$. Let g_{i_0} be an element contributing negatively to $\lambda_{\min}$, *i.e.*, $g_{i_0}\mathbf{v} = -\mathbf{v}$. As the set forms a group, for every j there exists a $k \neq j$ such that $g_{i_0} \cdot g_j = g_k$. This implies $-\lambda_j^{\mathbf{v}} = \lambda_k^{\mathbf{v}}$ and $\sum_{i\in I} \lambda_i^{\mathbf{v}} = 0$.

Construction of W_n for odd n

We construct the distribution $P_{I_0,\dots,I_{n-1}|O_0,\dots,O_{n-1}}$ for odd $n > 2$. Let $\{g_i\}_I$ be the set of matrices that can be written as $g_i = g_{i,1} \otimes g_{i,2} \otimes \cdots \otimes g_{i,n}$, with the objects $g_{i,k} \in \{\mathbb{1}, \sigma_z\}$, and with an even number of σ_z's for each $i \in I$. We use the notation $g_{i,j:k}$ to denote the matrix $g_{i,j} \otimes g_{i,j+1} \otimes \cdots \otimes g_{j,k}$. The fact $\sigma_z^2 = \mathbb{1}$ implies that the product $g_i \cdot g_j$, for every $i,j \in I$, is a tensor product of $\mathbb{1}$ and σ_z with an even number of σ_z's, and is thus an element of $\{g_i\}_I$. Furthermore, all elements mutually commute, have all eigenvalues in $\{-1,1\}$ and, hence, each element is an involution. Therefore, their sum is a positive semi-definite matrix. The distribution $P_{I_0,\dots,I_{n-1}|O_0,\dots,O_{n-1}}$ as a matrix W_n is built by taking the sum over all group elements, where the matrix $g_{i,k}$ of the group element g_i contributes to the input I_k of party S_k, and to the output $O_{k-1 \bmod n}$ of the party labeled by $(k-1 \bmod n)$,

$$W_n = \hat{P}_{I_0,\dots,I_{n-1}|O_0,\dots,O_{n-1}} = \frac{1}{2^n} \sum_{i\in I} g_i \otimes g_{i,2:n} \otimes g_{i,1}\,. \tag{2.6}$$

By construction, W_n is positive semi-definite, *i.e.*, all probabilities are positive. Because n is odd, there exists for each group element g_i $(\neq \mathbb{1})$ at least one position k such that $g_{i,k} \otimes g_{i,k+1 \bmod n} = \sigma_z \otimes \mathbb{1}$, which excludes logical paradoxes. Furthermore, for every $i \in \{0,1,\dots,n-1\}$ the object W_n contains the channel from all parties $S_{j(\neq i)}$ to S_i — a condition to perfectly win the game.

Example: W_3

For illustration, we construct W_3. The group from which W_3 is constructed is $\{g_0, g_1, g_2, g_3\}$ with the group elements

$$g_0 = \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1}\,, \tag{2.7}$$
$$g_1 = \mathbb{1} \otimes \sigma_z \otimes \sigma_z\,, \tag{2.8}$$
$$g_2 = \sigma_z \otimes \mathbb{1} \otimes \sigma_z\,, \tag{2.9}$$
$$g_3 = \sigma_z \otimes \sigma_z \otimes \mathbb{1}\,. \tag{2.10}$$

The matrix W_3 is thus

$$\begin{aligned} W_3 = \frac{1}{8}\big(&\mathbb{1}^{\otimes 6} + \mathbb{1} \otimes \sigma_z \otimes \sigma_z \otimes \sigma_z \otimes \sigma_z \otimes \mathbb{1} \\ &+ \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \sigma_z \\ &+ \sigma_z \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z\big)\,. \end{aligned} \tag{2.11}$$

The second summand of W_3 represents a channel from S_0, S_1 to S_2, the third summand represents a channel from S_1, S_2 to S_0, and finally, the last summand represents a channel from S_0, S_2 to S_1.

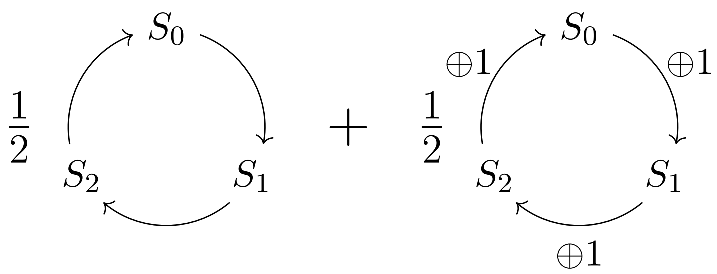

Figure 2.3: The conditional probability distribution W_3 is a mixture of a circular identity channel with a circular bit-flip channel.

It can easily be verified that if the three parties S_0, S_1, and S_2 input $P_{O_0,O_1,O_2}(o_0, o_1, o_2) = 1$ into W_3, then W_3 outputs the distribution

$$P_{I_0,I_1,I_2}(i_0, i_1, i_2) = \begin{cases} \frac{1}{2}, & i_0 = o_2,\ i_1 = o_0,\ i_2 = o_1, \\ \frac{1}{2}, & i_0 = \bar{o}_2,\ i_1 = \bar{o}_0,\ i_2 = \bar{o}_1, \\ 0, & \text{otherwise}, \end{cases} \tag{2.12}$$

where $\bar{o}_i = o_i \oplus 1$. Therefore, W_3 implements a uniform mixture of the loops where the input of party $S_{i \bmod 3}$ is sent to party $S_{i+1 \bmod 3}$, and where the input of party $S_{i \bmod 3}$ is flipped and sent to $S_{i+1 \bmod 3}$ (see Figure 2.3) [Bru]. It is evident from Figure 2.3 that logical paradoxes are not possible. If all intermediate parties forward what they receive (by applying any reversible transformation), both loops (see Figure 2.3) cancel each other out, *i.e.*, the correlations interfere destructively. Then again, if one intermediate party does not forward what she receives, the loop is broken. Conversely, any party can signal to her predecessor in the loop, because then an even number of bit-flips are applied, and thus the correlations interfere constructively. The same reasoning holds for any W_n for odd $n > 2$.

Construction of W_n for even n

The above construction works for odd $n > 2$. For even n, the group contains the element $\sigma_z^{\otimes n}$, which leads to a logical paradox since all inputs are correlated to all outputs [OCB12]. This can also be seen in Figure 2.3, where for even n, the sum of both channels leads to a logical paradox. In this case (n even), we double the dimensions of the output of the second-to-last party and of the input of the last party, and construct the distribution based on the group of matrices for the case of $n-1$. Let $\{g_i\}_I$ be the group used to construct W_{n-1}. The set for n parties is the Abelian subgroup $\{g_i \otimes g_i'\}_I \cup \{\bar{g}_i \otimes g_i'\}_I$, where $g_i' = g_{i,1} \otimes g_{i,2}$ and $\bar{g}_i = g_i \cdot \sigma_z^{\otimes(n-1)}$.

The distribution $P_{I_0,\ldots,I_{n-1}|O_0,\ldots,O_{n-1}}$ as a matrix W_n is constructed as before, with the exception that g_i' is considered a *single* submatrix,

$$W_n = \frac{1}{2^{n+1}} \sum_{i \in I} \Big(g_i \otimes g_i' \otimes g_{i,2:n-1} \otimes g_i' \otimes g_{i,1} + \bar{g}_i \otimes g_i' \otimes \bar{g}_{i,2:n-1} \otimes g_i' \otimes \bar{g}_{i,1} \Big) . \tag{2.13}$$

Again, by construction, W_n fulfills all requirements and contains all channels required to perfectly win the game.

Example: W_4

As an example, we construct the matrix W_4. The group $\{h_0, h_1, h_2, h_3, h_4, h_5, h_6, h_7\}$ for W_4 is constructed from the group $\{g_0, g_1, g_2, g_3\}$ and has the elements

$$h_0 = g_0 \otimes g_0' = (\mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1}) \otimes (\mathbb{1} \otimes \mathbb{1})\,, \tag{2.14}$$
$$h_1 = g_1 \otimes g_1' = (\mathbb{1} \otimes \sigma_z \otimes \sigma_z) \otimes (\mathbb{1} \otimes \sigma_z)\,, \tag{2.15}$$
$$h_2 = g_2 \otimes g_2' = (\sigma_z \otimes \mathbb{1} \otimes \sigma_z) \otimes (\sigma_z \otimes \mathbb{1})\,, \tag{2.16}$$
$$h_3 = g_3 \otimes g_3' = (\sigma_z \otimes \sigma_z \otimes \mathbb{1}) \otimes (\sigma_z \otimes \sigma_z)\,, \tag{2.17}$$
$$h_4 = \bar{g}_0 \otimes g_0' = (\sigma_z \otimes \sigma_z \otimes \sigma_z) \otimes (\mathbb{1} \otimes \mathbb{1})\,, \tag{2.18}$$
$$h_5 = \bar{g}_1 \otimes g_1' = (\sigma_z \otimes \mathbb{1} \otimes \mathbb{1}) \otimes (\mathbb{1} \otimes \sigma_z)\,, \tag{2.19}$$
$$h_6 = \bar{g}_2 \otimes g_2' = (\mathbb{1} \otimes \sigma_z \otimes \mathbb{1}) \otimes (\sigma_z \otimes \mathbb{1})\,, \tag{2.20}$$
$$h_7 = \bar{g}_3 \otimes g_3' = (\mathbb{1} \otimes \mathbb{1} \otimes \sigma_z) \otimes (\sigma_z \otimes \sigma_z)\,. \tag{2.21}$$

The matrix W_4 is thus

$$\begin{aligned} W_4 = \frac{1}{32}\big(&\mathbb{1}^{\otimes 10} \\ &+ \mathbb{1} \otimes \sigma_z \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \sigma_z \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \mathbb{1} \\ &+ \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \sigma_z \otimes \mathbb{1} \otimes \mathbb{1} \otimes \sigma_z \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z \\ &+ \sigma_z \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \sigma_z \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \sigma_z \otimes \sigma_z \\ &+ \sigma_z \otimes \sigma_z \otimes \sigma_z \otimes \mathbb{1} \otimes \mathbb{1} \otimes \sigma_z \otimes \sigma_z \otimes \mathbb{1} \otimes \mathbb{1} \otimes \sigma_z \\ &+ \sigma_z \otimes \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1} \otimes \sigma_z \otimes \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1} \otimes \sigma_z \otimes \sigma_z \\ &+ \mathbb{1} \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \mathbb{1} \otimes \mathbb{1} \\ &+ \mathbb{1} \otimes \mathbb{1} \otimes \sigma_z \otimes \sigma_z \otimes \sigma_z \otimes \mathbb{1} \otimes \sigma_z \otimes \sigma_z \otimes \sigma_z \otimes \mathbb{1}\big)\,. \end{aligned} \tag{2.22}$$

The second to the fifth summands represent the channels that are used to perfectly win the game.

The conditional probability distribution W_4 responds to input $P_{O_0,O_1,O_2,O_3}(o_0, o_1, o_{2,1}, o_{2,2}, o_3) = 1$ with the following output

$$P_{I_0,I_1,I_2,I_3}(i_0, i_1, i_2, i_{3,1}, i_{3,2}) = \tag{2.23}$$
$$\begin{cases} \frac{1}{4}, & i_0 = o_3,\ i_1 = o_0,\ i_2 = o_1, \\ & i_{3,1} = o_{2,1},\ i_{3,2} = o_{2,2}, \\ \frac{1}{4}, & i_0 = \bar{o}_3,\ i_1 = o_0,\ i_2 = \bar{o}_1, \\ & i_{3,1} = o_{2,1},\ i_{3,2} = \bar{o}_{2,2}, \\ \frac{1}{4}, & i_0 = o_3,\ i_1 = \bar{o}_0,\ i_2 = \bar{o}_1, \\ & i_{3,1} = \bar{o}_{2,1},\ i_{3,2} = o_{2,2}, \\ \frac{1}{4}, & i_0 = \bar{o}_3,\ i_1 = \bar{o}_0,\ i_2 = o_1, \\ & i_{3,1} = \bar{o}_{2,1},\ i_{3,2} = \bar{o}_{2,2}, \\ 0, & \text{otherwise}, \end{cases}$$

where $o_{2,1}, o_{2,2}$ are both bits of the random variable O_2, $i_{3,1}, i_{3,2}$ are both bits of the random variable I_3, and where $\bar{o}_i = o_i \oplus 1$. Therefore, W_4 implements a uniform distribution of four circular channels (see Figure 2.4).

By construction of W_4, no logical paradox arises. More intuitively, in any strategy that does not break any of the four circular channels of Figure 2.4 (*i.e.*, every party's output depends on its input), parties S_2 and S_3 use the first bit, the second, or both bits to communicate. If they use the first bit, then the correlations arising from the first two

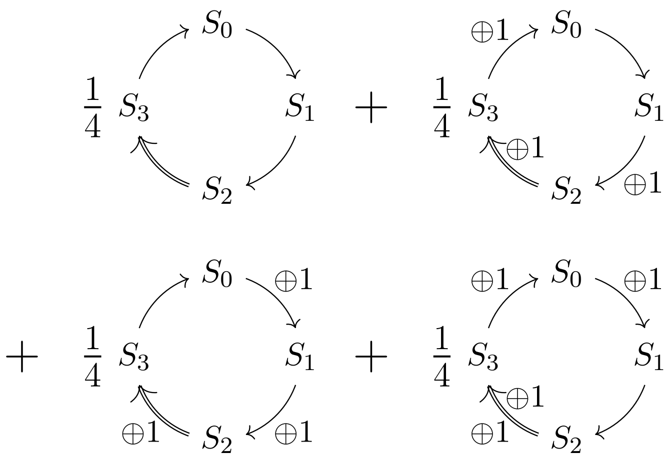

Figure 2.4: The conditional probability distribution W_4 is a mixture of four circular channels. The channel from S_2 to S_3 is a two-bit channel (double line). If the $\oplus 1$ operation for the channel from S_2 to S_3 is outside the circle, then the $\oplus 1$ operation is applied to the first channel, *i.e.*, the first bit is flipped, if it is inside the circle, then the $\oplus 1$ operation is applied to the second channel, *i.e.*, the second bit is flipped.

loops and the last two loops of Figure 2.4 interfere constructively. Both pairs together, however, break the loop. If they use the second bit, then the first and the third loop, and the second with the last loop yield the same output in every cycle. In total, all loops cancel each other out. For the last case as well, where both bits are used for communication, the correlations from the first and the last loop interfere constructively, and so do the second and the third. Ultimately, again, all loops cancel each other out, and no logical paradox can be created.

For larger even n, the conditional probability distribution W_n as well is constructed out of four loops, as in Figure 2.4, that cancel each other out when one tries to build a logical paradox. For $n = 2$, the same construction does not work, because the two-bit channel cannot be used to signal from its source to its destination — it can only be used when combined with other channels. In a two-party scenario, however, in order to win the game, each party needs to signal to the other.

Winning strategy

For odd n, the strategy $Q_i^m = \hat{P}_{X_i=x_i,O_i|A_i=a_i,I_i}$ for party S_i to win the game is

$$Q_i^m = Q_{i,O}^m \otimes Q_{i,I}^m = \left(\frac{\mathbb{1} + (-1)^{a_i'} \sigma_z}{2} \right) \otimes \left(\frac{\mathbb{1} + (-1)^{x_i} \sigma_z}{2} \right), \tag{2.24}$$

where $a_i' = a_i$ for $i \equiv m+1 \pmod n$, and $a_i' = a_i + x_i$ otherwise. The strategies for even n are equivalent to the strategies for odd n, except that S_{n-2} has a two-bit output and S_{n-1} has a two-bit input. Depending on M, they use the first, second, or both bit(s) to receive or send the desired bit. All local operations are classical since they are diagonal, *i.e.*, consist only of measuring and preparing states in the σ_z basis.

The distribution $P(x_m|a_0,\ldots,a_{n-1},M=m)$ is

$$\sum_{\substack{x_i\in\{0,1\}\\ i\neq m}} P(\mathbf{x}|\mathbf{a},M=m) \tag{2.25}$$

$$= \sum_{\substack{x_i\in\{0,1\}\\ i\neq m}} \mathrm{tr}\big[\big(Q^m_{0,I}\otimes\cdots\otimes Q^m_{n-1,I}\otimes Q^m_{0,O}\otimes\cdots\otimes Q^m_{n-1,O}\big)\cdot W_n\big] \tag{2.26}$$

$$= \frac{1}{2}\Big(1+(-1)^{X_m+\sum_{i\neq m}A_i}\Big)\,, \tag{2.27}$$

where we rearranged the submatrices of Q_i^m in the trace expression such that the ordering of the conditional probabilities in W_n match. This result is obtained because after taking the trace, each term except 1 and $(-1)^{X_m+\sum_{i\neq m}A_i}$ is either zero or depends on a variable $X_{i(\neq m)}$ which, in the process of marginalization over $X_{i(\neq m)}$, cancels out. For each m, the winning probability is

$$\Pr\left(X_m = \bigoplus_{i\neq m} A_i \,\middle|\, M=m\right) = 1\,.$$

Therefore, the game is won with certainty.

Example: $n=3$

The probability of obtaining x_0 in the case $M=0$ is

$$P(x_0|a_0,a_1,a_2,M=0) \tag{2.28}$$

$$= \sum_{x_1,x_2\in\{0,1\}} \mathrm{tr}\big[\big(Q^0_{0,I}\otimes Q^0_{1,I}\otimes Q^0_{2,I}\otimes Q^0_{0,O}\otimes Q^0_{1,O}\otimes Q^0_{2,O}\big)\cdot W_3\big]$$

$$= \frac{1}{8}\frac{\mathrm{tr}\big[\mathbb{1}^{\otimes 6}\big]}{2^6} \sum_{x_1,x_2\in\{0,1\}} \big(1+(-1)^{x_0+x_1+x_2+a_0+a_1} + (-1)^{x_0+a_1+a_2} + (-1)^{x_1+x_2+a_0+a_2}\big)$$

$$= \frac{1}{2}\big(1+(-1)^{x_0+a_1+a_2}\big)\,. \tag{2.29}$$

Therefore, the probability of the event $X_0 = A_1\oplus A_2$ is 1. The distribution of X_1 in the case $M=1$ is

$$P(x_1|a_0,a_1,a_2,M=1) \tag{2.30}$$

$$= \frac{1}{8}\sum_{x_0,x_2\in\{0,1\}} \big(1+(-1)^{x_0+x_2+a_0+a_1} + (-1)^{x_0+x_1+x_2+a_1+a_2} + (-1)^{x_1+a_0+a_2}\big) \tag{2.31}$$

$$= \frac{1}{2}\big(1+(-1)^{x_1+a_0+a_2}\big)\,, \tag{2.32}$$

and, finally, the distribution of X_2 in the case $M=2$ is

$$P(x_2|a_0,a_1,a_2,M=2) \tag{2.33}$$

$$= \frac{1}{8}\sum_{x_0,x_1\in\{0,1\}} \big(1+(-1)^{x_2+a_0+a_1} + (-1)^{x_0+x_1+a_1+a_2} + (-1)^{x_0+x_1+x_2+a_0+a_2}\big)$$

$$= \frac{1}{2}\big(1+(-1)^{x_2+a_0+a_1}\big)\,. \tag{2.34}$$

The probabilities of the events $X_1 = A_0 \oplus A_2$ and $X_2 = A_0 \oplus A_1$ are both 1. Therefore, the game is won with certainty.

Intuitively, in the case $M = m$, party $S_{m+1 \bmod 3}$ sends $O_{m+1 \bmod 3} = a_{m+1 \bmod 3}$ on both circular channels of Figure 2.3. Thus, party $S_{m+2 \bmod 3}$ receives the uniform mixture of $I_{m+2 \bmod 3} = a_{m+1 \bmod 3}$ (left channel of Figure 2.3) and $I_{m+2 \bmod 3} = a_{m+1 \bmod 3} \oplus 1$ (right channel of Figure 2.3). Party $S_{m+2 \bmod 3}$ thereafter sends $O_{m+2 \bmod 3} = I_{m+2 \bmod 3} \oplus a_{m+2 \bmod 3}$, *i.e.*, the uniform mixture of $O_{m+2 \bmod 3} = a_{m+1 \bmod 3} \oplus a_{m+2 \bmod 3}$ and $O_{m+2 \bmod 3} = a_{m+1 \bmod 3} \oplus a_{m+2 \bmod 3} \oplus 1$, on both circular channels, yielding the deterministic input $I_m = a_{m+1 \bmod 3} \oplus a_{m+2 \bmod 3}$ to party S_m.

Example: $n = 4$

In the example $n = 4$, we explicitly write the local operations for the third and fourth parties, as the third party has a two-bit output, and the fourth party has a two-bit input. The local operations for the third party (S_2) are

$$
\begin{aligned}
Q_2^{m=0} &= \left(\frac{\mathbb{1} \otimes \mathbb{1} + (-1)^{a_2+x_2}\sigma_z \otimes \mathbb{1}}{4}\right) \otimes Q_2' \\
Q_2^{m=1} &= \left(\frac{\mathbb{1} \otimes \mathbb{1} + (-1)^{a_2}\sigma_z \otimes \sigma_z}{4}\right) \otimes Q_2' \\
Q_2^{m=2} &= \left(\frac{\mathbb{1} \otimes \mathbb{1}}{4}\right) \otimes Q_2' \\
Q_2^{m=3} &= \left(\frac{\mathbb{1} \otimes \mathbb{1} + (-1)^{a_2+x_2}\mathbb{1} \otimes \sigma_z}{4}\right) \otimes Q_2' \,,
\end{aligned}
$$

with

$$
Q_2' = \left(\frac{\mathbb{1} + (-1)^{x_2}\sigma_z}{2}\right). \tag{2.35}
$$

Party 4 (S_3) uses

$$
\begin{aligned}
Q_3^{m=0} &= Q_3' \otimes \left(\frac{\mathbb{1} \otimes \mathbb{1} + (-1)^{x_3}\sigma_z \otimes \mathbb{1}}{2}\right) \\
Q_3^{m=1} &= Q_3' \otimes \left(\frac{\mathbb{1} \otimes \mathbb{1} + (-1)^{x_3}\sigma_z \otimes \sigma_z}{2}\right) \\
Q_3^{m=2} &= \left(\frac{\mathbb{1} + (-1)^{a_3}\sigma_z}{2}\right) \otimes \left(\frac{\mathbb{1} \otimes \mathbb{1}}{2}\right) \\
Q_3^{m=3} &= Q_3' \otimes \left(\frac{\mathbb{1} \otimes \mathbb{1} + (-1)^{x_3}\mathbb{1} \otimes \sigma_z}{2}\right),
\end{aligned}
$$

where we use shorthand Q_3' for

$$
Q_3' = \left(\frac{\mathbb{1} + (-1)^{a_3+x_3}\sigma_z}{2}\right). \tag{2.36}
$$

The distributions of X_0, X_1, X_2, X_3, under the condition $M = 0$, $M = 1$, $M = 2$, $M =$

3, respectively, are

$$\begin{aligned}
P(x_0|a_0,a_1,a_2,a_3,M=0) &= \frac{1}{2}\left(1+(-1)^{x_0+a_1+a_2+a_3}\right), \\
P(x_1|a_0,a_1,a_2,a_3,M=1) &= \frac{1}{2}\left(1+(-1)^{x_1+a_0+a_2+a_3}\right), \\
P(x_2|a_0,a_1,a_2,a_3,M=2) &= \frac{1}{2}\left(1+(-1)^{x_2+a_0+a_1+a_3}\right), \\
P(x_3|a_0,a_1,a_2,a_3,M=3) &= \frac{1}{2}\left(1+(-1)^{x_3+a_0+a_1+a_2}\right).
\end{aligned}$$

Therefore, the event $X_0 = A_1 \oplus A_2 \oplus A_3$, given $M = 0$, the event $X_1 = A_0 \oplus A_2 \oplus A_3$, in the case $M = 1$, the event $X_2 = A_0 \oplus A_1 \oplus A_3$, if $M = 2$, and the event $X_3 = A_0 \oplus A_1 \oplus A_2$ in the case $M = 3$ have probability 1. Which implies that the game is won with certainty.

By consulting Figure 2.4, we can describe the strategy in the following way. If $M = m$, then party $S_{m+1 \bmod 4}$ sends $a_{m+1 \bmod 4}$ to the next party by using all four channels of Figure 2.4. Each of the next two parties in clockwise orientation, *i.e.*, party $S_{m+2 \bmod 4}$ and party $S_{m+3 \bmod 4}$, sends the parity of what she receives from the previous party and her input ($a_{m+2 \bmod 4}$, $a_{m+3 \bmod 4}$, respectively). Depending on M, parties S_2 and S_3 use the first, the second, or both single-bit channels. In particular, if $M = 0$, then S_2 uses the first channel to communicate to S_3 — the second channel is ignored. For $M = 1$ they use both channels, *i.e.*, the parity of the inputs to both channels is equal to the bit S_2 sends. For $M = 2$, the two-bit channel between S_2 and S_3 is ignored. Finally, for $M = 3$ they use the second channel. By doing so, S_m obtains $a_{m+1 \bmod 4} + a_{m+2 \bmod 4} + a_{m+3 \bmod 4}$, as the introduced bit-flips from the four channels (see Figure 2.4) cancel each other out.

2.6 Conclusion

In an attempt to construct a theory that combines aspects of general relativity and quantum theory, Oreshkov, Costa, and Brukner [OCB12] proposed a framework for quantum correlations without causal order. They proved that some correlations are incompatible with any *a priori* causal order and, therefore, are *not compatible with predefined causal order* although they satisfy quantum theory *locally*. We consider the classical limit of this framework and show that in sharp contrast to the two-party scenario [OCB12], *classical and logically consistent multi-party correlations can be incompatible with any predefined causal order*. To show this, we propose a game that cannot be won in a scenario with predefined causal order, but is won with certainty when no causal order is fixed.

Recently, the ideas of indefinite causal order and of superpositions of causal orders were applied to quantum computation [Har09, Chi12, CDFP12, CDPV13, ACB14, Mor14]. Furthermore, Aaronson and Watrous [AW09] showed that closed timelike curves render classical and quantum computing equivalent. Our result is similar in the sense that the winning probability of the game is the same for the quantum and for the classical framework. Since the W object in Figure 2.2 can be thought of as a *channel back in time*, closed timelike curves can be interpreted as being part of the framework. Closed timelike curves *per se* are consistent with general relativity [Göd49]. However, Aaronson and Watrous take Deutsch's approach [Deu91] to closed timelike curves which, as opposed to the framework studied here, is a *non-linear* extension of quantum theory — such extensions are known to allow for communication faster than at the speed of light [Gis90].

Acknowledgments

We thank Časlav Brukner, Fabio Costa, Christina Giarmatzi, Issam Ibnouhsein, Ognyan Oreshkov, and Jibran Rashid for helpful discussions. We thank Časlav Brukner for the interpretation of W_3 as two loops. We thank Ognyan Oreshkov for pointing out that whenever the matrix W is diagonal in the computational basis, all local operations can be reduced to objects diagonal in the same basis. We thank three anonymous reviewers for helpful comments on the presentation of the results—especially the last reviewer for the detailed comments. The present work was supported by the Swiss National Science Foundation (SNF), the National Centre of Competence in Research "Quantum Science and Technology" (QSIT), the COST action on Fundamental Problems in Quantum Physics, the European Commission Project RAQUEL, the John Templeton Foundation, FQXi, and the Austrian Science Fund (FWF) through CoQuS, SFB FoQuS, and the Individual Project 24621.

Chapter 3

Causally nonseparable processes admitting a causal model

Abstract

A recent framework of quantum theory with no global causal order predicts the existence of "causally nonseparable" processes. Some of these processes produce correlations incompatible with any causal order (they violate so-called "causal inequalities" analogous to *Bell inequalities*) while others do not (they admit a "causal model" analogous to a *local model*). Here we show for the first time that bipartite causally nonseparable processes with a causal model exist, and give evidence that they have no clear physical interpretation. We also provide an algorithm to generate processes of this kind and show that they have nonzero measure in the set of all processes. We demonstrate the existence of processes which stop violating causal inequalities but are still causally nonseparable when mixed with a certain amount of "white noise". This is reminiscent of the behavior of Werner states in the context of entanglement and nonlocality. Finally, we provide numerical evidence for the existence of causally nonseparable processes which have a causal model even when extended with an entangled state shared among the parties.

Published. A. Feix, M. Araújo, and Č. Brukner. Causally nonseparable processes admitting a causal model. *New J. Phys.*, 18(8):083040, 2016.

Contribution in conceiving the research project, deriving the main results and proofs, and writing the manuscript.

3.1 Introduction

It is well-known that quantum mechanics is at odds with naive notions of reality and locality as predicted by Bell's celebrated theorem [Bel64, CHSH69]. One might wonder whether peculiar quantum features could challenge other fundamental notions, like the concept of *causality*, as well. The *process matrix formalism* of Oreshkov, Costa and Brukner [OCB12] was developed to explore this question—studying the most general causal structures compatible with local quantum mechanics for two parties A and B.

Surprisingly, the formalism predicts causal structures which are "causally nonseparable": they correspond neither to A being before B nor to B being before A, nor to a probabilistic mixture thereof. These causal structures can produce correlations incompatible with any definite causal order, violating so-called "causal inequalities" [OCB12, BAF⁺16].[1]

However, the empirical relevance of these results is still completely unclear. Do they appear in some physical situations or are they merely a mathematical artifact of the process matrix formalism?

For three and more parties, there are causally nonseparable processes whose physical realization is known—one instance is the "quantum switch" [CDPV13] where the causal order between two parties A and B is controlled by a quantum system belonging to a third party C. Processes of this kind, however, *cannot violate* causal inequalities [ABC⁺15, OG16]: they admit a "causal model", i.e., a causally separable process is capable of reproducing their correlations. Their causal nonseparability can only be certified through device-dependent "causal tomography" or "causal witnesses" [ABC⁺15]. This is analogous to states which are entangled but cannot violate Bell inequalities, i.e., for which a "local model" exists [Wer89].

Since the only causally nonseparable processes known to be physically implementable have a causal model, it is tempting to conjecture that the inability to violate causal inequalities without [Bru14] or with [OG16] operations extended to shared entangled states by all parties *singles out* the physical causal structures from unphysical ones. Ref. [OG16] contains an example of a tripartite process matrix with a causal model but which does not remain causal under extensions, i.e., is not "extensibly causal", demonstrating the difference between the two notions for more than two parties.

In this paper, we provide an example of a bipartite causally nonseparable process with a causal model. Furthermore, we give numerical evidence that *bipartite* nonseparable processes exist which do not violate causal inequalities, even when extended with entanglement. No physical interpretation is known for these processes.

The paper is organized as follows: Sec. 3.2 introduces the process matrix formalism and the definitions of causal nonseparability and causal inequalities. In Sec. 3.3, we define a class of two-party causally nonseparable processes and construct a causal model for them. This shows that the sets of causally nonseparable and causal inequality violating processes are distinct also in the bipartite case. Since the causally nonseparable processes with a causal model can be interpreted as the mixture of physically implementable process with an unphysical process, this gives evidence that they are not implementable in nature.

In Sec. 3.4, we provide an algorithm to construct nonseparable processes with a causal model by composing a random causally separable process with a non-completely positive map on one party. Using a random sample generated by a "hit-and-run" Markov chain [Smi80, Smi84], we also show that nonseparable processes with a causal model have nonzero measure in the space of all processes.

[1] For two parties, the operations implemented to violate causal inequalities have to be quantum [OCB12]; surprisingly, for three or more parties, classical operations are sufficient [BFW14].

In Sec. 3.5, we construct a family of "Werner processes", which mimic the behavior of Werner states [Wer89] with respect to nonlocality and entanglement. The Werner processes' causal nonseparability is more resistant to the introduction of "white noise" than its ability to violate causal inequalities. This shows that the analogy between causal nonseparability and causal inequalities on the one hand, and entanglement and Bell inequalities on the other, extends beyond what was previously known [OCB12, Bru15, ABC+15, OG16].

Finally, in Sec. 3.6, we examine the behavior of the processes with added shared entanglement between the parties. While some of the processes we constructed *do violate causal inequalities* when extended in this way (are not "extensibly causal"), numerical calculations indicate that the ability to violate causal inequalities disappears when adding a little white noise, at which the causal nonseparability is preserved. We conjecture that there are processes which are extensibly causal, and yet not physically implementable.

3.2 Causal nonseparability and causal inequalities

Quantum circuits can be thought of as a formalization of causal structures with a definite causal order. They consist of *wires*, representing quantum systems, which connect boxes, representing *quantum operations*. While for quantum circuits, the order of the operations is fixed [CDP09], situations where the order of operations is not well-defined are readily represented in the *process matrix formalism* [OCB12], which can be thought of as a generalization of the quantum circuit formalism. We will briefly introduce the main elements of the formalism; a more detailed introduction to it can be found in Ref. [ABC+15].

A quantum operation maps a density matrix $\rho_{A_I} \in A_I$ to a density matrix $\rho_{A_O} \in A_O$ (where A_I (A_O) denotes the space of linear operators on the Hilbert space $\mathcal{H}^{A_I}$ ($\mathcal{H}^{A_O}$)). The most general operations within the quantum formalism are *completely positive* (CP) maps $\mathcal{M}_A : A_I \to A_O$. Using the Choi-Jamiołkowski [Cho75, Jam72] (CJ) isomorphism, one can represent every CP map as an operator acting on the tensor product of the input and output Hilbert spaces:

$$M_A := [(\mathcal{I} \otimes \mathcal{M}_A)(|I\rangle\rangle\langle\langle I|)]^{\mathrm{T}} \in A_I \otimes A_O, \tag{3.1}$$

where $\mathcal{I}$ is the identity map and $|I\rangle\rangle := \sum_{j=1}^{d_{\mathcal{H}_I}} |jj\rangle \in \mathcal{H}_I \otimes \mathcal{H}_I$ is a non-normalized maximally entangled state; $^{\mathrm{T}}$ denotes matrix transposition in the computational basis.

The CJ-isomorphism can also be used to represent "superoperators" or "processes" which map *quantum maps* to quantum maps, quantum states or probabilities [GW07, CDP08b, CDP09, LS13, OCB12]. In this paper, we will focus on processes mapping *two quantum operations* ξ_x^a and η_y^b—corresponding to the Choi-Jamiołkowski representation of Alice's and Bob's CP maps—to a *probability* (see Fig. 3.1). Requiring linearity of probabilities in the operations, we can represent it as

$$p(\xi_x^a, \eta_y^b) := \mathrm{tr}[W \cdot \xi_x^a \otimes \eta_y^b], \tag{3.2}$$

$$W \in A_I \otimes A_O \otimes B_I \otimes B_O. \tag{3.3}$$

To ensure positivity of probabilities for all pairs of possible CP maps (as well as for extended operations where the parties share additional entanglement) the "process matrix" W has to be positive semidefinite $W \geq 0$ [OCB12]. The normalization of probabilities implies that $\mathrm{tr}[W \cdot \xi^{\mathrm{CPTP}} \otimes \eta^{\mathrm{CPTP}}] = 1$ for all CJ-representations of completely positive *trace-preserving maps* ξ^{CPTP} and η^{CPTP} [OCB12].

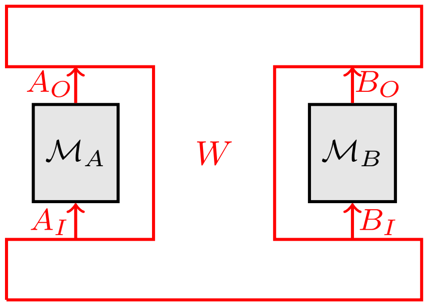

Figure 3.1: Representation of a bipartite process W, which linearly maps Alice's and Bob's CP maps $\mathcal{M}_A, \mathcal{M}_B$ to a probability. A_I (B_I) represents Alice's (Bob's) input Hilbert space and A_O (B_O) Alice's (Bob's) output Hilbert space.

We call a process $W^{A\prec B}$ ($W^{B\prec A}$) "causally ordered" if it does not allow for signalling from Bob to Alice (Alice to Bob), which is equivalent to the conditions [ABC+15]

$$W^{A\prec B} = \mathrm{tr}_{B_O}[W^{A\prec B}] \otimes \mathbb{1}^{B_0}/d_{B_O}, \tag{3.4}$$
$$W^{B\prec A} = \mathrm{tr}_{A_O}[W^{B\prec A}] \otimes \mathbb{1}^{A_0}/d_{A_O}. \tag{3.5}$$

A process matrix $W^{\mathrm{sep}} \in \mathcal{W}_{\mathrm{sep}}$ that can be decomposed into a convex combination ($0 \leq q \leq 1$) of causally ordered processes is "causally separable":

$$W_{\mathrm{sep}} = qW^{A\prec B} + (1-q)W^{B\prec A}. \tag{3.6}$$

It was recently shown that one can *efficiently* determine whether a process is causally (non)separable using a *semidefinite program* (SDP). Here we will use the SDP for "random robustness" [ABC+15]

$$\begin{aligned} &\min \lambda \\ \text{s.t.}\quad & W = W^{A\prec B} + W^{B\prec A} - \lambda \mathbb{1}^\circ, \\ & W^{A\prec B} = \mathrm{tr}_{B_O}[W^{A\prec B}] \otimes \mathbb{1}^{B_0}/d_{B_O}, \\ & W^{B\prec A} = \mathrm{tr}_{A_O}[W^{B\prec A}] \otimes \mathbb{1}^{A_0}/d_{A_O}, \end{aligned} \tag{3.7}$$

where $\mathbb{1}^\circ := \mathbb{1}^{A_I A_O B_I B_O}/(d_{A_I} d_{B_I})$. The random robustness $R_r(W)$ is defined as the result of the optimization $R_r(W) := \lambda_{\mathrm{opt}}$.

If $R_r(W) \leq 0$, the SDP gives an explicit decomposition of W into $W^{A\prec B}$ and $W^{B\prec A}$; if $R_r(W) > 0$, the process is not causally separable. The value of $R_r(W)$ is also an operational measure of "causal nonseparability". It is related to the minimal amount of "white noise" $\mathbb{1}^\circ$ that needs to be mixed with the process to make it causally separable. That is, for $\gamma \geq R_r(W)$, the process $(\gamma \mathbb{1}^\circ + W)/(1+\gamma)$ is causally separable.

A so-called "dual SDP" to (3.7) can then provide the optimal "causal witness", i.e., a hermitian operator S such that $\mathrm{tr}[W_{\mathrm{sep}}\, S] \geq 0$ for all causally separable processes W_{sep} [ABC+15]. The property $\mathrm{tr}[W\, S] < 0$ can be verified experimentally by measuring a set of operators for Alice and Bob, certifying that the process W is not causally separable. Note that, in analogy to entanglement witnesses [CS14], this certification of causal nonseparability relies on a partial tomography of the process and thus requires trust in Alice's and Bob's local operations: it is "device-dependent".

It is well-known that the entanglement of a quantum state can be certified *device-independently* (without requiring trust Alice's and Bob's operations) if the probability distribution resulting from a set of measurements violates a Bell inequality [ABG+07]. In an analogous way, causal nonseparability can be device-independently confirmed using *causal inequalities* [OCB12, BAF$^+$16, ABC$^+$15, OG16, BFW14], where the "non-causal" correlations between Alice and Bob alone suffice to show that the process they use is not causally separable, without additional trust in their local operations.

The condition for a probability distribution to be "causal", i.e., *not* to violate any causal inequality, is simply that it can be decomposed into a convex combination of a probability distribution which is no-signaling from Bob to Alice[2] ($p_{A\prec B}$) and a probability distribution which is no-signaling from Alice to Bob ($p_{B\prec A}$) [BAF$^+$16]:

$$p_{\text{causal}} = q p_{A\prec B} + (1-q) p_{B\prec A}. \tag{3.8}$$

Note that the correlations generated by a causally ordered process cannot violate any causal inequality.

For the scenario where Alice (Bob) has an input bit x (y) and outputs one bit a (b), one causal inequality is a bound on the probability of success of the "guess your neighbor's input" (GYNI) [BAF$^+$16]:

$$p_{\text{GYNI}} := \frac{1}{4}\sum_{x,y} p(a=y, b=x|x,y) \leq \frac{1}{2}. \tag{3.9}$$

Some valid processes and local strategies which result in correlations violating (3.9) are described in Ref. [BAF$^+$16].

The relation between causally nonseparable processes and the violation of causal inequalities is not yet fully understood. On the one hand, there exist causally nonseparable processes that can be physically implemented [PMA$^+$15] but have a *causal model*—they do not violate causal inequalities [ABC$^+$15, OG16]. An example of such a process is the "quantum switch" [CDPV13]. On the other hand, there are processes that can violate causal inequalities, but it is not known if they can be realized in nature, prompting the conjecture that only processes with a causal model are physically implementable [Bru14].

Another natural feature to investigate is the (in)ability for a process to violate causal inequalities, even when extended with an entangled state shared by all parties. We will call processes which do not allow for such a violation "extensibly causal" [OG16] and come back to this concept in detail in Sec. 3.6.

In the bipartite case, all previously known nonseparable processes violate causal inequalities and it is not clear if nonseparable processes with a causal model even exist [BAF$^+$16]. In the next section, we explicitly provide a class of nonseparable bipartite processes that allow for a causal model.

3.3 Causally nonseparable processes with a causal model

We will consider the following class of processes (all the operators are understood to act on qubits, $d_{A_I} = d_{A_O} = d_{B_I} = d_{B_O} = 2$):

$$\begin{aligned} W^{A\prec B} &:= \mathbb{1}^\circ + \frac{1}{12}(\mathbb{1}ZZ\mathbb{1} + \mathbb{1}XX\mathbb{1} + \mathbb{1}YY\mathbb{1}), \\ W^{B\prec A} &:= \mathbb{1}^\circ + \frac{1}{4}(Z\mathbb{1}XZ), \\ W &:= qW^{A\prec B} + (1-q+\epsilon)W^{B\prec A} - \epsilon\mathbb{1}^\circ. \end{aligned} \tag{3.10}$$

[2]When Alice is given the input x and outputs a (Bob is given an input y and outputs b), no-signaling from Bob to Alice implies that the marginal probability on Alice's side does not depend on Bob's input: $\sum_b p_{A\prec B}(ab|xy) = \sum_b p_{A\prec B}(ab|xy'), \forall y, y'$.

Here $\mathbb{1}, X, Y, Z$ are the Pauli matrices and the tensor products between the Hilbert spaces A_I, A_O, B_I, B_O are implicit, as in the remainder of the paper. The process matrix (3.10) is positive semidefinite for $\epsilon \leq q - 1 + \sqrt{\frac{(1-q)(q+3)}{3}}$ and causally nonseparable for $\epsilon > 0$. As shown in Appendix A.1, its random robustness is $R_r(W) = \epsilon$. It is maximal for $q = \sqrt{3} - 1 \approx 0.732$, where $\epsilon = \frac{4}{\sqrt{3}} - 2 \approx 0.309$.

The proof that the process (3.10) cannot be used to violate any causal inequalities, for any local strategy[3] consists of two steps: (i) we show that the set of correlations compatible with W is the same as the set of correlations achievable with W^{T_B} (where $^{\mathrm{T}_B}$ denotes the partial transpose of the systems $B_I B_O$ with respect to the computational basis); (ii) we verify that W^{T_B} is valid and causally separable, hence cannot violate causal inequalities. Taken together, this establishes that W cannot violate any causal inequalities either and therefore admits a causal model.

The first part of the proof is simple. Using definition (3.2) and the self-duality of transposition, we rewrite the probability distribution:

$$p(ab|xy) = \mathrm{tr}[W \xi_x^a \otimes \eta_y^b] = \mathrm{tr}[W^{\mathrm{T}_B} \xi_x^a \otimes (\eta_y^b)^{\mathrm{T}}]. \tag{3.11}$$

Additionally, for any quantum instrument [DL70] $\{\eta_y^b\}$, the instrument $\{\eta_y^{b\,\mathrm{T}}\}$ is also valid, since transposition maps completely positive maps to completely positive maps and trace-preserving maps to trace-preserving maps[4]. This establishes (i), namely that the correlations achievable with W^{T_B} are the same as those compatible with W—note that this holds for *any* process, even when W^{T_B} is *not positive semidefinite*, and therefore not a valid process matrix. In such a case, the probability distribution will be well-defined for local measurements, but not when extending the process with an entangled state between Alice and Bob, an extension which is physically meaningful and to which we will come back in Sec. 3.6.

For the class of process matrices given in Eq. (3.10), W^{T_B} *is* always positive semidefinite. We will now explicitly decompose W^{T_B} as a convex combination of causally ordered process matrices, proving that it is causally separable (formally, this implies that $R_r(W^{\mathrm{T}_B}) \leq 0$).

First, one should notice the similarity of $W^{A \prec B}$ with the process matrix $D_{2/3}^{A \prec B}$ of a depolarizing channel (with $\frac{2}{3}$ probability of depolarizing and $\frac{1}{3}$ probability of perfectly transmitting the state) from Alice to Bob

$$D_{2/3}^{A \prec B} := \mathbb{1}^\circ + \frac{1}{12}(\mathbb{1}ZZ\mathbb{1} + \mathbb{1}XX\mathbb{1} - \mathbb{1}YY\mathbb{1}), \tag{3.12}$$

where only the sign of the term $\mathbb{1}YY\mathbb{1}$ differs compared to $W^{A \prec B}$ of (3.10). This exactly corresponds to a partial transpose of the systems $B_I B_O$, such that $W^{A \prec B} = (D_{2/3}^{A \prec B})^{\mathrm{T}_B}$. Using the definition of the depolarizing process $D_{2/3}^{A \prec B} = \frac{2}{3}\mathbb{1}^\circ + \frac{1}{3} I^{A \prec B}$, where $I^{A \prec B} = \mathbb{1}^{A_I} |I\rangle\rangle\langle\langle I|^{A_O B_I} \mathbb{1}^{B_O}/2$. Since $(W^{B \prec A})^{\mathrm{T}_B} = W^{B \prec A}$, we can write W^{T_B} as:

$$\begin{aligned} W^{\mathrm{T}_B} &= \frac{2q}{3}\mathbb{1}^\circ + \frac{q}{3} I^{A \prec B} + (1 - q + \epsilon) W^{B \prec A} - \epsilon \mathbb{1}^\circ \\ &= \frac{q}{3} I^{A \prec B} + (1 - q + \epsilon) W^{B \prec A} + \left(\frac{2q}{3} - \epsilon\right) \mathbb{1}^\circ, \end{aligned} \tag{3.13}$$

[3]Note that our proof guarantees the existence of a causal model, which means that the correlations belong to *every* causal polytope, without restriction on the number of inputs and outputs for each party.

[4]The condition on the CJ representation of a CPTP map is that $\mathrm{tr}_{B_O} M^{B_I B_O} = \mathbb{1}^{B_I}$ and it implies $\mathrm{tr}_{B_O} (M^{B_I B_O})^{\mathrm{T}} = \mathbb{1}^{B_I}$.

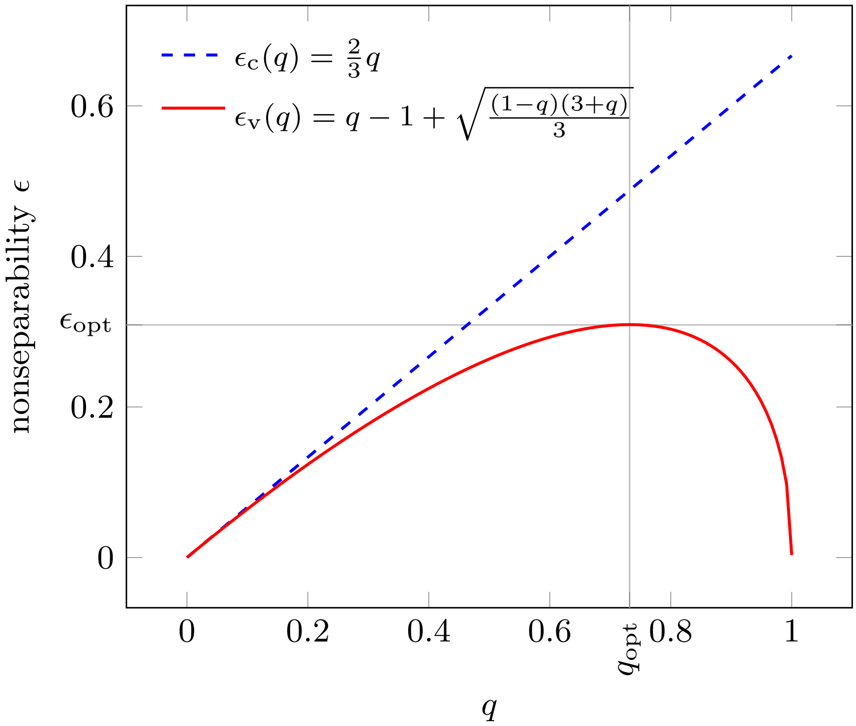

Figure 3.2: For W of Eq. (3.10) to be a valid process, $\epsilon \leq \epsilon_v(q)$ (in the region below the red curve). A process only generates causal correlations for $\epsilon \leq \epsilon_c(q)$ (the region below the dotted blue curve). Since $\epsilon_v \leq \epsilon_c$, every valid process of the form (3.10) allows for a "causal" probability distribution. ϵ_{opt} and q_{opt} are the parameters maximizing causal nonseparability.

which is a convex decomposition into causally ordered processes as long as $\epsilon \leq \epsilon_c(q) = \frac{2q}{3}$. Since $\epsilon \leq q - 1 + \sqrt{\frac{(1-q)(3+q)}{3}}$ for the process given in Eq. (3.10) to be valid, and $q - 1 + \sqrt{\frac{(1-q)(3+q)}{3}} \leq \epsilon_c(q)$, the whole class of processes defined in Eq. (3.10) cannot violate causal inequalities. For a graphical representation of this relationship, see Fig. 3.2.

This concludes the proof and provides an explicit causal model for the process W with the instruments $\{\xi_x^a\}$ and $\{\eta_y^b\}$: the process W^{T_B} with the instruments $\{\xi_x^a\}$ and $\{\eta_y^{b\mathrm{T}}\}$. Since W^{T_B} is causally separable, it can be interpreted as a probabilistic mixture of two causally ordered processes. There are infinitely many such decompositions since the term $\mathbb{1}^\circ$ in Eq. (3.13) can be split and added to $\frac{q}{3} I^{A \prec B}$ and $(1 - q + \epsilon) W^{B \prec A}$ in any proportion.

W can be taken to be W^{T_B} composed with a transpose map on B. The process matrix $\frac{q}{3}(I^{A \prec B})^{\mathrm{T}_B}$ becomes positive (meaning that the associated map is completely positive) when adding at least $\frac{2q}{3}$ of white noise. However, the maximal noise that can be admixed by transferring the $\mathbb{1}^\circ$ term in (3.13) is $\frac{2q}{3} - \epsilon$, which is strictly smaller. Therefore, the causal model for W suggests a natural interpretation for it as a convex combination of an *unphysical* channel from Alice to Bob (as it is not completely positive) with a *physical* channel from Bob to Alice. This provides some evidence—yet not a proof—that the process W is not physically implementable.

3.4 Random causally nonseparable processes with a causal model

In this section, we develop a method to construct a broad class of causally nonseparable processes with a causal model. Given a random causally separable process, one applies a positive, but not necessarily completely positive map $Q_B(\cdot)$ on Bob's side:

$$W_{\text{sep?}} = Q_B(W_{\text{sep}}). \tag{3.14}$$

If the resulting process has negative eigenvalues, it is discarded; otherwise, it is a valid process matrix. Using the same argument as in the preceding section, we know that the resulting process $W_{\text{sep?}}$ will have a causal model. Sometimes—and these are the interesting cases—the process will also be causally nonseparable. This can readily be checked this via SDP (3.7).

We generated causally separable process matrices W_{sep} (where $d_{A_I} = d_{B_I} = d_{A_O} = d_{B_O} = 2$) according to an asymptotically uniform distribution using the "hit-and-run" technique (see Appendix A.3 for details) and Eq. (3.14), using the transposition map $^{\mathrm{T}_B}$ for Q_B. We found that most (69%) of the resulting matrices were positive and hence valid process matrices. About half of these turn out to be causally nonseparable while—by construction—allowing for a causal model (we denote this set by $\mathcal{W}_{\text{nsep}}^{(\text{c})}$). The histogram of the resulting causal nonseparabilities is shown in Fig. 3.3. There is therefore a *finite* prob-

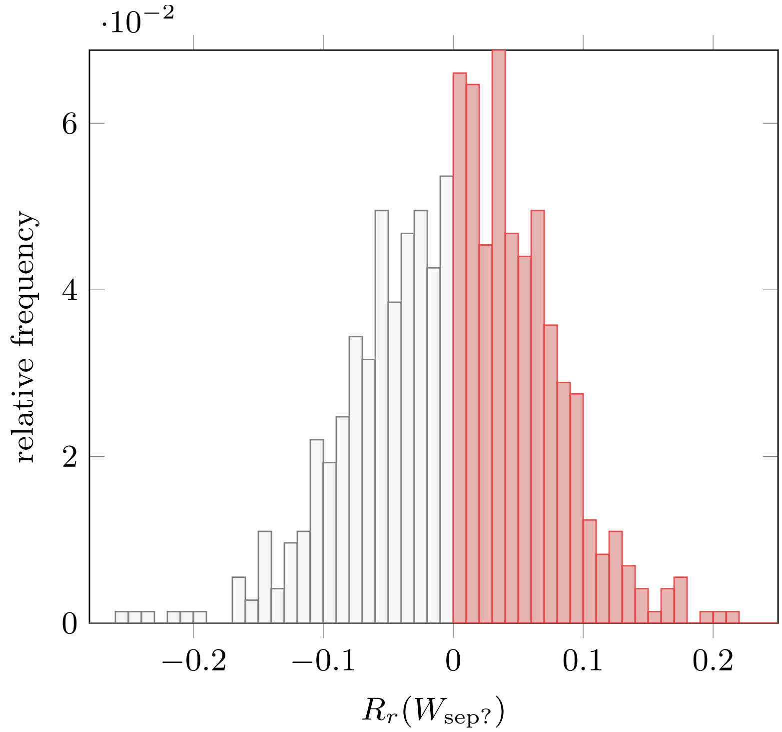

Figure 3.3: Histogram of the random robustness R_r of the subset of 690 (≈ 69%) valid processes $W_{\text{sep?}}$ generated from 1000 uniformly distributed causally separable processes (see Appendix A.3) and applying partial transposition on Bob's side. The 366 (≈ 53%) which are causally nonseparable, while admitting a causal model, are represented in red.

ability of generating a nonseparable process with a causal model starting from a random of causally separable process. Since the map $^{\mathrm{T}_B}$ is measure-preserving, the set of causally nonseparable processes which admit a causal model is of the same dimension as the set of valid processes $\mathcal{W}$ itself (see Appendix A.2 for details).

3.5 "Werner" causally nonseparable processes

We will denote the set of causally nonseparable processes as $\mathcal{W}_{\text{nsep}}$. It is composed of the set of processes with a causal model ($\mathcal{W}_{\text{nsep}}^{(\text{c})}$) and the set of processes that can violate causal inequalities ($\mathcal{W}_{\text{nsep}}^{(\text{nc})}$), which are graphically represented in Fig. 3.4.

We now construct bipartite processes violating a causal inequality, but which, mixed with some amount of "white noise" $\mathbb{1}^\circ$, turn into causally nonseparable processes with a causal model. This behavior is reminiscent of "Werner states", which violate Bell inequalities until a noise level of up to $\frac{1}{2}$ but are entangled when mixed with noise up to a level of $\frac{2}{3}$ [Wer89]. It shows that the analogy between causal inequalities and causal nonseparability, on the one hand, and Bell inequalities and entanglement, on the other hand, applies to the two-party case.

The idea is to use a convex combination of a process in $\mathcal{W}_{\text{nsep}}^{(\text{c})}$ and a process in $\mathcal{W}_{\text{nsep}}^{(\text{nc})}$, which is also invariant under partial transposition with respect to B[5]. In this way, one can generate a broad class of "Werner causally nonseparable processes".

We will use the process defined in Eq. (3.10) with the maximal causal nonseparability:

$$W_{\text{opt}} := (\sqrt{3}-1)W^{A\prec B} + \frac{1}{\sqrt{3}}W^{B\prec A} - \left(\frac{4}{\sqrt{3}} - 2\right)\mathbb{1}^\circ, \tag{3.15}$$

together with the process

$$W_{\text{OCB}} := \mathbb{1}^\circ + \frac{1}{4\sqrt{2}}(\mathbb{1}ZZ\mathbb{1} + Z\mathbb{1}XZ), \tag{3.16}$$

which was proposed and shown to violate causal inequalities in Ref. [OCB12]. In Appendix A.1, we show that the resulting mixture

$$W_{\text{mix}}(\alpha) := \alpha W_{\text{opt}} + (1-\alpha)W_{\text{OCB}} \tag{3.17}$$

has nonseparability $R_{\text{mix}}(\alpha) := R_r(W_{\text{mix}}(\alpha)) = \alpha R_r(W_{\text{opt}}) + (1-\alpha)R_r(W_{\text{OCB}})$, where $R_r(W_{\text{OCB}}) = \sqrt{2}-1$ is the random robustness of W_{OCB}.

Following the same argument as in the previous section, we can now examine the causal nonseparability of $W_{\text{mix}}^{\text{T}_B}$, since, by transferring the partial transpose onto Bob's CP maps, we know that it can produce exactly the same correlations as W_{mix}. Its nonseparability $R'_{\text{mix}}(\alpha) := R_r(W_{\text{mix}}^{\text{T}_B}(\alpha))$ is again the weighted average (see Appendix A.1) of the nonseparabilities of $W_{\text{opt}}^{\text{T}_B}$ and of $W_{\text{OCB}}^{\text{T}_B} = W_{\text{OCB}}$:

$$\begin{aligned} R'_{\text{mix}}(\alpha) &= \alpha R_r(W_{\text{opt}}^{\text{T}_B}) + (1-\alpha)R_r(W_{\text{OCB}}) \\ &= R_{\text{mix}}(\alpha) + \frac{2\alpha}{3}R_r(W_{\text{opt}}^{\text{T}_B}) < R_{\text{mix}}(\alpha), \end{aligned} \tag{3.18}$$

where we used $R_r(W_{\text{opt}}^{\text{T}_B}) = \frac{2\sqrt{3}-4}{3} < 0$ and $R_r(W_{\text{OCB}}) = \sqrt{2}-1$. This means that there is a finite gap between the nonseparability of W_{mix} and the nonseparability of $W_{\text{mix}}^{\text{T}_B}$.

Using a see-saw algorithm [BAF+16], we numerically verified that $W_{\text{mix}}(\alpha)$ indeed violates the causal inequalities of Ref. [BAF+16] as long as $W_{\text{mix}}^{\text{T}_B}$ is nonseparable ($R'_{\text{mix}}(\alpha) > 0$), i.e., when $\alpha < \frac{3(\sqrt{2}-1)}{1+3\sqrt{2}-2\sqrt{3}} \approx 0.6987$.

This gap translates into a gap between the level of white noise $\mathbb{1}^\circ$ that W_{mix} can tolerate before admitting a causal model and the level of noise at which it becomes nonseparable. We therefore define the "Werner process" as a convex combination of $\mathbb{1}^\circ$ and $W_{\text{mix}}(\alpha)$:

$$W_{\text{Wer}}(\gamma,\alpha) := (1-\gamma)W_{\text{mix}}(\alpha) + \gamma\mathbb{1}^\circ. \tag{3.19}$$

[5] Or, in the more general scenario described in the previous section, is invariant under the non-completely-positive operation $Q_B(\cdot)$.

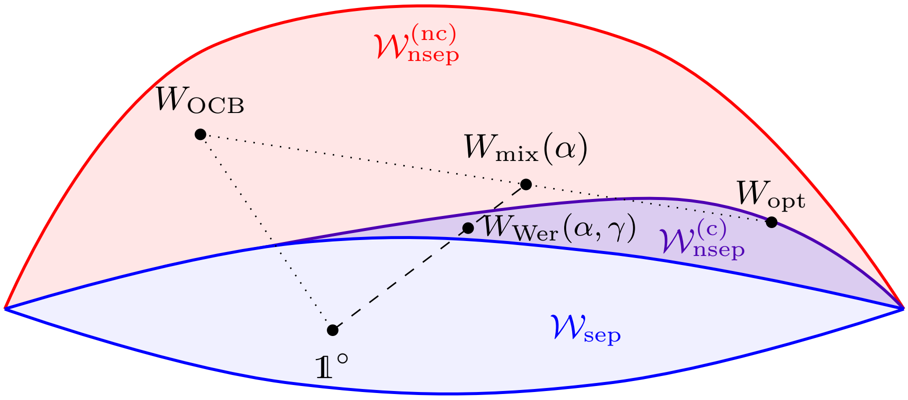

Figure 3.4: Schematic depiction of different types of process matrices. $\mathcal{W}_{\text{nsep}}^{(\text{nc})}$ is the set of process matrices which can violate causal inequalities and $W_{\text{nsep}}^{(\text{c})}$ is the set of nonseparable process matrices which admit a causal model and $\mathcal{W}_{\text{sep}}$ is the set of separable processes. The process $W_{\text{mix}} := \alpha W_{\text{opt}} + (1-\alpha) W_{\text{OCB}}$, where W_{opt} admits a causal model and W_{OCB} doesn't, gives rise to "Werner type" processes $W_{\text{Wer}}(\gamma,\alpha) := (1-\gamma)W_{\text{mix}}(\alpha) + \gamma\mathbb{1}^\circ$, which have a causal model but are causally nonseparable for a certain level of noise (3.21).

Using the definition of nonseparability (3.7), one can verify that the following relations hold (see Appendix A.1):

$$\begin{aligned} W_{\text{Wer}}\left(\gamma < \frac{R_{\text{mix}}(\alpha)}{1+R_{\text{mix}}(\alpha)}, \alpha\right) &\in \mathcal{W}_{\text{nsep}}, \\ W_{\text{Wer}}^{\text{T}_B}\left(\gamma \geq \frac{R'_{\text{mix}}(\alpha)}{1+R'_{\text{mix}}(\alpha)}, \alpha\right) &\in \mathcal{W}_{\text{sep}}. \end{aligned} \tag{3.20}$$

As W_{Wer} can violate causal inequalities only if $W_{\text{Wer}}^{\text{T}_B}$ is causally nonseparable (remember the proof of Sec. 3.3), we conclude from (3.20) that

$$W_{\text{Wer}}\left(\frac{R'_{\text{mix}}(\alpha)}{1+R'_{\text{mix}}(\alpha)} \leq \gamma < \frac{R_{\text{mix}}(\alpha)}{1+R_{\text{mix}}(\alpha)}, \alpha\right) \in \mathcal{W}_{\text{nsep}}^{(\text{c})}, \tag{3.21}$$

which mimics the behavior of Werner states. See Fig. 3.4 for a graphical representation of the location of W_{Wer} with respect to the different sets of processes.

3.6 Relationship to extensibly causal processes

In the context of the physical implementability of process matrices, it is natural to consider the *extension* of a process matrix with an *entangled state shared between the parties*. A process is "extensibly causal" if, even when extended with a shared entangled state, it *cannot violate causal inequalities* [OG16].

Extending a physically implementable process with an entangled state shared among the parties should also result in a physically implementable process. On this account, it is important to consider not only whether a process has a causal model, but rather whether it has such a model when extended with an entangled state. In Ref. [OG16], an example of a tripartite process with a causal model but which is not extensibly causal was presented, showing that both notions really differ and that the violation of causal inequalities can be "activated" by entanglement—and for which no physical implementation is known.

Note that the proof (Sec. 3.3) of the existence of causal model for W does not hold when the process is extended with an entangled state between Alice and Bob. It therefore cannot prove that W is extensibly causal. It crucially relies on the fact that the transpose $\{(\eta_y^b)^T\}$ of a valid instrument for Bob $\{\eta_y^b\}$ is still a valid instrument. However, taking the full transpose on Bob's instrument would lead to a "causal model" with a partial transpose of the shared entangled state, which can lead to negative probabilities. Conversely, the *partial transpose of Bob's instrument* (with no transposition on Bob's part of the entangled state) is not a valid instrument and does not yield positive probabilities in general.

To numerically study whether W_{opt} from Eq. (3.15) is extensibly causal, we extended it with a maximally entangled state of two ququarts ($|\phi\rangle^{A'_I B'_I} := \frac{1}{2}(|00\rangle + |11\rangle + |22\rangle + |33\rangle)$):

$$W_{\text{ext}} := W_{\text{opt}} \otimes |\phi\rangle \langle\phi|^{A'_I B'_I} . \tag{3.22}$$

We chose a maximally entangled ququart state because we believe that extending W_{opt} with a higher dimensional state would not improve its ability to violate causal inequalities.

Using the see-saw algorithm, we optimized W_{ext} for a violation of the simplest causal inequalities [BAF+16]. We found that W_{ext} *is* able to violate (by about $8 \cdot 10^{-5}$) the GYNI inequality (3.9), which proves that W_{opt} is *not* extensibly causal. Incidentally, it also shows that *in the bipartite case as well*, the violation of causal inequalities can be "activated" using entanglement.

If we adopt the view that extensively causal processes are physical, the activation of violation of causal inequalities, this suffices to exclude W_{opt} in the same way as the tripartite process with a causal model but which is not extensibly causal, given in Ref. [OG16], independently of the argument based on the decomposition of Sec. 3.3. However, this is not possible anymore when admixing a small amount of white noise: $(1-\kappa)W_{\text{ext}} + \kappa \mathbb{1}^\circ$. We ran the see-saw algorithm for different levels of white noise and found no violation of GYNI for $\kappa > 3.3 \cdot 10^{-4}$ (see Fig. 3.5 for a graphical representation of the relationship between noise and violation of GYNI). Similarly, neither the other causal inequality from Ref. [BAF+16] nor the "original" causal inequality from Ref. [OCB12] could be violated through see-saw optimization.

This gives reasonable evidence[6] that W_{opt} mixed with very little white noise is extensibly causal, while still being causally nonseparable (see Fig. 3.6 for a graphical representation). The argument for unphysicality of Sec. 3.3 still applies to it. This leads us to conjecture that some extensibly causal processes cannot be physically realized.

3.7 Conclusions

We studied the classification of causally nonseparable process matrices for two parties and found that composing a class of causally separable processes with a transpose map on one party's side results in nonseparable processes with causal models, i.e., that cannot violate causal inequalities.

Since the only interpretation we know relies on applying a non-completely positive map (which is itself is unphysical) to a valid process, the conjecture that processes which do not violate causal inequalities are physically implementable is undermined.

We also provided a simple algorithm to generate nonseparable processes with causal models—starting from a random separable process and composing it with a positive, but not completely positive map on one party's side. With a finite probability, this yields a nonseparable process with a causal model and shows that the measure of such processes is nonzero within the space of valid processes. The "hit-and-run" algorithm we used to generate random process matrices might be of independent interest.

[6] It falls short of being a proof because (i) the entangled state added in W_{ext} is finite-dimensional; (ii) the see-saw technique is not guaranteed to converge to the global optimum; (iii) only the three *known* bipartite causal inequalities were tested, inequalities with more settings might still be violated.

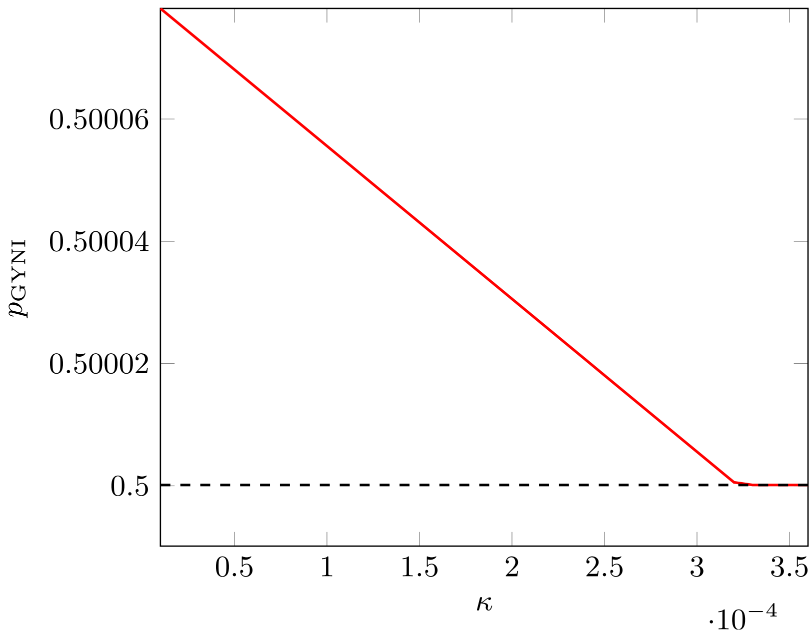

Figure 3.5: Numerically optimized (see-saw) violation of the GYNI inequality Eq. (3.9), using the noisy extended process $(1-\kappa)W_{\mathrm{ext}}+\kappa\mathbb{1}^{\circ}$ (see Eq. (3.22)) and causal bound (dashed). For a noise level of $\kappa > 3.3 \cdot 10^{-4}$, the algorithm fails to find a strategy violating the inequality, as well as for the other known bipartite causal inequalities [BAF+16, OCB12].

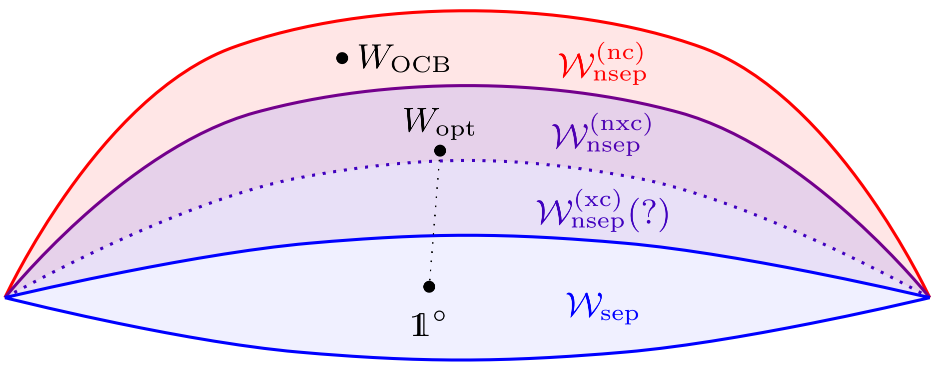

Figure 3.6: Schematic depiction of the sets process matrices with respect to extensible causal separability. $\mathcal{W}_{\mathrm{sep}}$ is the set of separable processes, $\mathcal{W}_{\mathrm{nsep}}^{(\mathrm{nc})}$ is the set of process matrices which can violate causal inequalities, $\mathcal{W}_{\mathrm{nsep}}^{(\mathrm{nxc})}$ the set of processes with a causal model but which can violate causal inequalities when extended with entanglement, such as W_{opt}. Based on our numerical evidence (see Fig. 3.5), we conjecture that the set $\mathcal{W}_{\mathrm{nsep}}^{(\mathrm{xc})}$ of processes which are causally nonseparable and "extensibly causal" is not empty. Note that the set of nonseparable processes with a causal model is $\mathcal{W}_{\mathrm{nsep}}^{(\mathrm{c})} = \mathcal{W}_{\mathrm{nsep}}^{(\mathrm{xc})} \cup \mathcal{W}_{\mathrm{nsep}}^{(\mathrm{nxc})}$.

We then developed the analogy between entanglement/nonlocality and causal nonseparability/noncausal correlations by providing a process analogous to a Werner state: it starts having a causal model when mixed with a certain amount of white noise, while still being strictly causally nonseparable.

Finally, we studied whether our processes still have a causal model when extended with an entangled state shared between Alice and Bob (whether they are "extensibly causal"). The numerical evidence prompted us to conjecture that some of the nonseparable processes we studied are extensibly causal, while not being physically implementable.

An important question remains open: if some processes which have an (extensible) causal model are nonphysical, which other criterion should be used to rule them out? One fairly natural approach is to postulate a "purification principle", according to which physically realizable processes can be recovered as part of a pure process in a larger space [AFNB17].

Acknowledgments

We thank Ognyan Oreshkov and Fabio Costa for pointing out the significance of "extensible causal separability" in the context of physical implementability, and Ralph Silva for a useful comment on an earlier version of this manuscript. We acknowledge support from the European Commission project RAQUEL (No. 323970); the Austrian Science Fund (FWF) through the Special Research Programme FoQuS, the Doctoral Programme CoQuS and Individual Project (No. 2462) and the John Templeton Foundation.

Chapter 4

Quantum superposition of the order of parties as a communication resource

Abstract

In a variant of communication complexity tasks, two or more separated parties cooperate to compute a function of their local data, using a limited amount of communication. It is known that communication of quantum systems and shared entanglement can increase the probability for the parties to arrive at the correct value of the function, compared to classical resources. Here we show that quantum superpositions of the direction of communication between parties can also serve as a resource to improve the probability of success. We present a tripartite task for which such a superposition provides an advantage compared to the case where the parties communicate in a fixed order. In a more general context, our result also provides the first semi-device-independent certification of the absence of a definite order of communication.

Published. A. Feix, M. Araújo, and Č. Brukner. Quantum superposition of the order of parties as a communication resource. *Phys. Rev. A*, 92(5):052326, 2015.

Contribution in conceiving the research project, deriving the main results and proofs, and writing the manuscript.

4.1 Introduction

In its short history, the field of quantum information has been very successful in discovering and explaining differences between classical and quantum information processing—in particular a variety of advantages that the use of quantum resources confers over the use of classical resources [NC00].

Quantum resources provide important benefits regarding *communication complexity tasks* [Yao79, Yao93, KN06] where two or more separated parties compute a function of their input strings, seeking to maximize the probability of success under the constraint of limited communication between them. Communicating quantum bits and sharing entanglement are two well-known resources that can be used to improve success probability in such scenarios [BCMd10].

A novel type of quantum resource—the *quantum switch*—, allows for the order in which quantum gates are applied to be in a quantum superposition, using an auxiliary quantum system that coherently controls the order in which the gates are applied [CDPV13]. The quantum switch has been shown to reduce the required number of queries to "blackbox" unitaries required to solve certain computational tasks [CDPV13, CDFP12, Chi12, ACB14, PMA+15].

Here we find that the *quantum control of the direction of communication between parties* is a novel, useful resource in communication complexity protocols. We demonstrate this by considering an explicit three-party communication task, in which Alice and Bob are each given input *trits* and Charlie has to determine whether they are equal or not. They are not allowed to share entanglement and the total communication is restricted to two qubits. We show that, when the order of communication between parties is fixed (or classically mixed), the success probability is bounded below one. However, using the quantum switch to superpose the direction of communication between Alice and Bob, there exists a protocol that always succeeds.

4.2 Process matrix formalism

Superpositions of the direction of communication are readily described in the *process matrix formalism*, first introduced in Ref. [OCB12]. We will briefly review some of its key aspects; for an extensive introduction to the subject, we refer the reader to Ref. [ABC+15].

The most general quantum operation, a completely positive (CP) map, maps a density operator $\rho_{A_I} \in A_I$ to a density operator $\rho_{A_O} \in A_O$. Here, A_I (A_O) denotes the space of linear operators on the Hilbert space $\mathcal{H}^{A_I}$ ($\mathcal{H}^{A_O}$); in general, the dimensions d_{A_I} and d_{A_O} of $\mathcal{H}^{A_I}$ and $\mathcal{H}^{A_O}$ do not have to be equal.

Using the Choi-Jamiołkowski [Cho75, Jam72] (CJ) isomorphism (where we follow the convention of Ref. [ABC+15]) one can represent a CP map $\mathcal{M}_A : A_I \to A_O$ as an operator

$$M_A := [(\mathcal{I} \otimes \mathcal{M}_A)(|I\rangle\langle I|)]^{\mathrm{T}} \in A_I \otimes A_O, \tag{4.1}$$

where $\mathcal{I}$ is the identity map and $|I\rangle := \sum_{j=1}^{d_{\mathcal{H}_I}} |jj\rangle \in \mathcal{H}_I \otimes \mathcal{H}_I$ is a non-normalized maximally entangled state and $^{\mathrm{T}}$ denotes transposition. The inverse transformation is

$$\mathcal{M}_A(\rho) = \mathrm{tr}_I[(\rho \otimes \mathbb{1})M_A]^{\mathrm{T}}. \tag{4.2}$$

Similarly, for two completely positive maps $\mathcal{M}_A : A_I \to A_O$ and $\mathcal{M}_B : B_I \to B_O$, the joint CJ-matrix is the tensor product of the CJ-matrix of the individual maps $\in A_I \otimes A_O \otimes B_I \otimes B_O$.

One can use this isomorphism to conveniently represent higher-order operations [GW07, CDP08b, CDP09, LS13, OCB12], which map quantum maps to quantum maps. These "superoperators" or "processes" can also be represented as *CJ-matrices* themselves, by applying the CJ-isomorphism repeatedly.

One can also meaningfully define operations acting *jointly on states and operations*. We will restrict our attention to the class of processes $\mathcal{W}$ mapping *two CP maps and two states* to *two states*:

$$\mathcal{W}(\mathcal{M}_A, \mathcal{M}_B, \sigma_C, \rho_T) = \mathrm{tr}_{A,B}\{W \cdot M_A \otimes M_B \otimes \sigma_C \otimes \rho_T\} = \rho'_{CT}. \tag{4.3}$$

4.3 Processes with and without a definite order of communication

Quantum circuits form a well-known class of processes in which gates corresponding to the operations $\mathcal{M}_A$ and $\mathcal{M}_B$ appear in a fixed order (as depicted in Fig. 4.1). Either $\mathcal{M}_A$ is

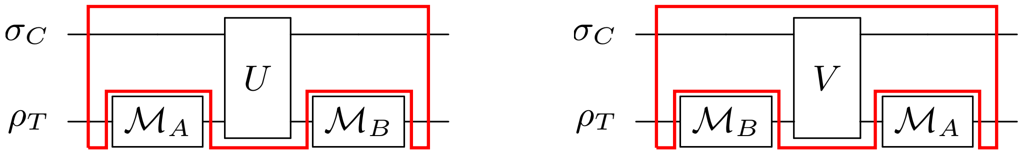

Figure 4.1: Examples of quantum circuits (in red) mapping two CPTP maps $\mathcal{M}_A$, $\mathcal{M}_B$ and two states σ_C, ρ_T to a state ρ'_{CT}. The *order of applying gates* is well-defined—$A \preceq B$ for the left circuit and $B \preceq A$ for the right one.

applied before $\mathcal{M}_B$ (corresponding to processes of the type $\mathcal{W}_{A\preceq B}$) or $\mathcal{M}_B$ is applied before $\mathcal{M}_A$ (corresponding to processes $\mathcal{W}_{B\preceq A}$) [CDP09]. Identifying $\mathcal{M}_A$ ($\mathcal{M}_B$) with Alice's (Bob's) operation, these "ordered processes" correspond to *a definite order of signaling between Alice and Bob*. More generally, we will also refer to classical mixtures thereof, which correspond to a classical random variable controlling the order of the process,

$$\mathcal{W}_{\text{ord.}} := p\mathcal{W}_{A\preceq B} + (1-p)\mathcal{W}_{B\preceq A}, \quad 0 \le p \le 1, \tag{4.4}$$

as "causally separable processes" [ABC+15, OG16].[1]

Not all physically implementable processes are causally separable: The *quantum switch*, first introduced by Chiribella et al. [CDPV13], corresponds to the process $\mathcal{W}_{\text{sw}}$, which applies two CP maps to a target system ρ_T in an order that is controlled by the value of a quantum control system σ_C. The quantum switch for pure target and control states $|\psi\rangle_T$, $|\phi\rangle_C$ and unitary operations U_A (U_B) on Alice's (Bob's) side is given by

$$\mathcal{W}_{\text{sw}}(U_A, U_B, |\phi\rangle_C, |\psi\rangle_T) = \langle 0|\phi\rangle\, |0\rangle_C\, U_B U_A\, |\psi\rangle_T + \langle 1|\phi\rangle\, |1\rangle_C\, U_A U_B\, |\psi\rangle_T \tag{4.5}$$

and can be extended by linearity to mixed states and general CP maps on Alice's and Bob's side [Chi12]. It is neither of the type $\mathcal{W}_{A\preceq B}$ nor of the type $\mathcal{W}_{B\preceq A}$. Since it is an extremal process, it also cannot be decomposed according to Eq. (4.4), which shows that there is no definite order of signaling for the quantum switch [ABC+15]. Rather, one should think of it as a *coherent superposition* of circuits or of directions of communication, controlled by a control qubit:

$$\frac{1}{\sqrt{2}}\Big(|0\rangle_C \,|\!\!-\!\boxed{U_A}\!-\!\boxed{U_B}\!-\!\rangle + |1\rangle_C \,|\!\!-\!\boxed{U_B}\!-\!\boxed{U_A}\!-\!\rangle\Big). \tag{4.6}$$

[1]Note that the definition of causal separability in Ref. [OG16] slightly differs from the one presented here.

It has been shown that using such a quantum control of circuits provides an advantage in query complexity for certain computational tasks [CDPV13, CDFP12, Chi12, ACB14]. It has also been implemented experimentally, using an interferometric setup [PMA+15].

4.4 The tripartite Hamming game

To demonstrate the relevance of the quantum switch in communication scenarios, we will introduce a communication game closely related to the distributed Deutsch-Josza promise problem [DJ92, BCW98, BCMd10] the *Simultaneous message passing model* (SMP) [Yao79, BCWd01] and *Random access codes* (RACs) [Wie83, Nay99, ANTSV02, HIN+06, ALMO08].

In our tripartite game—as for the SMP—, Alice and Bob receive input strings and Charlie computes a function of them. Communication between all the parties and shared (classical) randomness are also allowed. Charlie has to compute *the parity of the Hamming distance of Alice's and Bob's input strings*, generalizing the function of the distributed Deutsch-Josza promise problem (here, however, *no promise* on the Hamming distance of the inputs is required).

More precisely, Alice and Bob both are given n *trits* ($x \in \{0,1,2\}^n$ and $y \in \{0,1,2\}^n$ respectively), Charlie computes the Hamming parity $f(x,y)$ defined as

$$f(x,y) := \bigoplus_{i=1}^{n} \delta_{x_i y_i}. \tag{4.7}$$

In addition, the total length of the transcript communicated by Alice, Bob, and Charlie is restricted to be m bits (or qubits). This defines the $(n \log_2 3, m)$-Hamming game depicted in Fig. 4.2; the average success probability associated to it will be referred to as $p_{\text{succ.}}$.

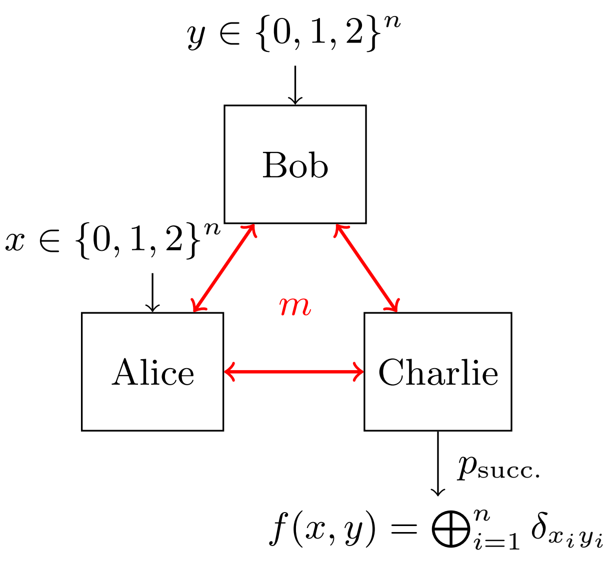

Figure 4.2: Tripartite $(n \log_2 3, m)$-Hamming game where Alice and Bob receive input strings of the length $n \log_2 3$ bits, and Charlie has to compute $f(x,y)$. The total communication is m bits or qubits; no entanglement is pre-shared.

Next we show that for the $(\log_2 3, 2)$-Hamming game (which is equivalent to the equality game for trits) the success probability is bounded below one when Alice, Bob and Charlie are restricted to using a causally separable process, i.e., when the direction of signaling is fixed or controlled by a classical random variable independent of the inputs. In contrast, using quantum control over the direction of signaling—the quantum switch—, Charlie can always compute $f(x,y)$. This demonstrates that causally nonseparable processes are useful resources for communication tasks.

Causally separable classical strategy

We will first consider the case where Alice, Bob and Charlie can only implement *classical* operations and use a process with a definite order of communication (or a mixture thereof). The optimal strategy involves Alice encoding her input trit x into a bit $a(x)$ and sending it to Bob, who sends the function $b(a,y)$ to Charlie, who finally outputs a function $g(b)$.

The deterministic strategies are the vertices of a convex polytope in the 9-dimensional (all possible combinations of x and y) space of probabilities $p(c|x,y)$. Given that Alice, Bob and Charlie share randomness, they can probabilistically combine determinstic strategies, reaching every point inside the convex polytope.

For equally distributed inputs, the probability of success for Charlie to output $f(x,y) = \delta_{x,y}$ is bounded by[2]:

$$p^{\mathcal{C}}_{\text{succ.}} := \frac{1}{9}\sum_{x,y} p(c = \delta_{x,y}|x,y) \leq \frac{7}{9}. \tag{4.8}$$

One deterministic strategy saturating this bound consists in Alice encoding whether her input is 0 or not ($a(x) = \delta_{x,0}$) and Bob answering 1 only if he is sure that Alice and he both have input 0 ($b(a,y) = \delta_{y,0}\delta_{a,1}$). Charlie simply returns Bob's answer. This strategy will fail only for input pairs $x = y = 1$ and $x = y = 2$.

Causally separable quantum strategy

We now turn to the case where Alice, Bob and Charlie use a causally separable process (consisting of quantum channels) and have access to quantum operations, as shown in Fig. 4.3. The parties are allowed to share randomness but not entanglement. In the optimal protocol with two qubits of communication in total, Alice encodes her input trit into a qubit $x \mapsto \rho_x$ and Bob applies a CPTP map $\mathcal{B}_y$ for each value of his input trit y onto the incoming qubit; Charlie then performs a two-outcome positive-operator valued measure (POVM) $\{C_m\}$ on the resulting state.

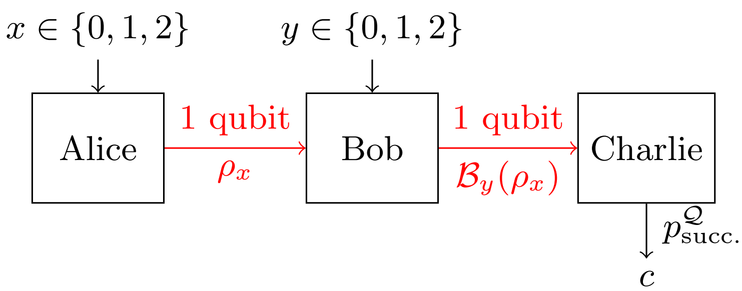

Figure 4.3: Optimal causally separable protocol for the equality game, where no entanglement is shared among the parties.

[2]Note that it is also a facet of the polytope, since it is saturated by vertices spanning an 8-dimensional affine subspace.

For equally distributed inputs, the probability of success for causally separable strategies is bounded by

$$p^{\mathcal{Q}}_{\text{succ.}} \leq \frac{1}{9} \max_{\rho_x, \{\mathcal{B}_y\}, \{C_m\}} \left(\sum_x \operatorname{tr}\{C_1 \mathcal{B}_x(\rho_x)\} + \sum_{x \neq y} \operatorname{tr}\{C_0 \mathcal{B}_y(\rho_x)\} \right), \tag{4.9}$$

which, in Appendix B, we prove to be

$$p^{\mathcal{Q}}_{\text{succ.}} \leq \frac{5}{6}. \tag{4.10}$$

Here, an optimal state preparation by Alice is

$$|a_0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \;\; |a_1\rangle = \sin\frac{\pi}{8}|0\rangle + e^{i\pi/4}\cos\frac{\pi}{8}|1\rangle, |a_2\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle + e^{-i\pi/4}|1\rangle\Big), \tag{4.11}$$

where $\rho_x = |a_x\rangle\langle a_x|$. Bob projectively measures in the basis $|a_y\rangle, |a_y^\perp\rangle$, where $|a_y^\perp\rangle$ is orthogonal to $|a_y\rangle$, and prepares the state $|x+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ or $|x-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, depending on the outcome.

Charlie simply applies a projective measurement in $|x\pm\rangle$-basis, the outcome of which constitues his guess c. The probability distribution arising from the optimal quantum strategy is shown in Table 4.1.

Table 4.1: Conditional probabilities of success with a causally separable process, for the optimal strategy (4.11), reaching $p^{\mathcal{Q}}_{\text{succ.}} = \frac{5}{6}$.

x, y	00	01	02	10	11	12	20	21	22
$p^{\mathcal{Q}}(c = \delta_{x,y}\|xy)$	1	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{3}{4}$	1	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{3}{4}$	1

Quantum superposition of the order of parties

We now show that when Alice, Bob, and Charlie can use the quantum switch to implement *quantum* control over the direction of communication between Alice and Bob, they can violate Eq. (4.10) maximally ($p^{\mathcal{Q}\text{-sw.}}_{\text{succ.}} = 1$).

Alice and Bob apply unitaries U_A^x, U_B^y to a target system and the quantum switch coherently superposes the order in which they are applied. Charlie receives the resulting state and applies a two-outcome projective measurement Π^+, Π^-. Since Alice and Bob only have access to a qubit subspace, they each only send one qubit out of their lab, while Charlie sends no system out. The total communication between Alice, Bob and Charlie is $m \leq \log_2(d_{A_O} \cdot d_{B_O} \cdot d_{C_O}) = 2$ qubits, in accordance with the assumptions of the $(\log_2 3, 2)$-Hamming game.

Alice and Bob choose a Pauli gate corresponding to their input trit $U_A^i = U_B^i = \sigma_i$ and the control state is $|\phi\rangle_C = |x+\rangle_C$ (the state $|\psi\rangle_T$ is irrelevant), see Fig. 4.4. Inserting this into Eq. (4.5), Charlie receives the state

$$\begin{aligned} &\mathcal{W}_{\text{sw}}(\sigma_x, \sigma_y, |x+\rangle_C, |\psi\rangle_T) \\ &= \frac{1}{\sqrt{2}}(|0\rangle_C \, \sigma_y\sigma_x |\psi\rangle_T + |1\rangle_C \, \sigma_x\sigma_y |\psi\rangle_T) \\ &= \frac{1}{2}(|x-\rangle_C \, [\sigma_y, \sigma_x] |\psi\rangle_T + |x+\rangle_C \, \{\sigma_y, \sigma_x\} |\psi\rangle_T), \end{aligned} \tag{4.12}$$

where $[\cdot, \cdot]$ is the commutator and $\{\cdot, \cdot\}$ the anticommutator.

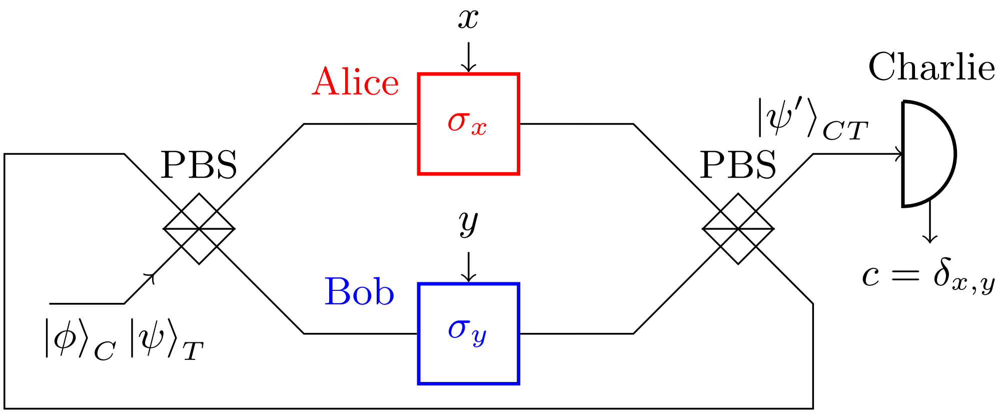

Figure 4.4: Linear optical implementation of the protocol using the quantum switch [CIJT, ACB14]. The control state $|\phi\rangle_C$ is encoded in polarization and the target state $|\psi\rangle_T$ in another photonic degree of freedom. Alice and Bob apply Pauli operators on the target system depending on their input x and y. Charlie performs a measurement in $|x\pm\rangle$ basis on the outcoming control system C and consequently outputs $\delta_{x,y}$. Note that in the experiment of Ref. [PMA+15], the control state was instead encoded in path.

If Charlie chooses a projective measurement on the resulting control system C, with $\Pi^+ = |x+\rangle\langle x+|_C$ and $\Pi^- = |x-\rangle\langle x-|_C$, he can determine whether $[\sigma_y, \sigma_x] = 0$ or $\{\sigma_y, \sigma_x\} = 0$ (because of the commutation relations of the Pauli matrices, one of them is always the case). If the former is true, Charlie deduces that $x = y$, otherwise, that $x \neq y$. Hence, he can compute $f(x, y) = \delta_{x,y}$ with unit probability, violating the bound (4.10).

Note that the protocol can be extended to any $(m \log_2 3, 2m)$ Hamming game (Alice and Bob each are given m trits and have access to an m-qubit system). Alice and Bob apply $\bigotimes_i \sigma_{x_i}$ and $\bigotimes_i \sigma_{y_i}$ respectively; Charlie, by measuring the control qubit in $|x\pm\rangle$-basis, can still determine whether $[\bigotimes_i \sigma_{x_i}, \bigotimes_i \sigma_{y_i}]$ or $\{\bigotimes_i \sigma_{x_i}, \bigotimes_i \sigma_{y_i}\}$ is zero. Since for each different trit, a factor of -1 appears when permuting the corresponding Pauli matrices, an even number of differences in the trit strings of Alice and Bob will result in a vanishing commutator, and an odd number of differences in a vanishing anticommutator. Using the quantum switch, Charlie can therefore always find the Hamming parity (4.7).

4.5 Conclusions

We demonstrated that a quantum superposition of the direction of communication between parties is a useful resource in communication complexity problems. This was explicitly shown for the $(\log_2 3, 2)$-Hamming game, where the probability of success for processes with a definite or classically mixed order of signaling is violated by using the quantum switch as a resource. The result points to the necessity for a general resource theory of communication to account for superpositions of the direction of communication. Note that having access to the quantum switch is *not* equivalent to sharing a maximally entangled state between Alice and Bob—for instance, the latter (through dense coding [BW92]) makes computing *any* binary function of two trits for Alice and Bob possible by exchanging just two qubits of communication, which is impossible with the quantum switch.

Our result also provides the first *semi-device-independent* [LVB11, PB11] way of certifying the causal nonseparability of a process, where Alice's and Bob's system is known to have (at most) a given dimension, but the operations themselves are not trusted. It lies between the stronger fully *device-independent* certification of causal nonseparability [OCB12, BAF$^+$16, OG16]—which was already shown to be impossible for the quantum switch [ABC$^+$15, OG16]—and the weaker *device-dependent* certification through *causal witnesses* [ABC$^+$15].

It would be interesting to improve the *scaling* (with the length of the inputs) of the reduction in communication achieved by using the quantum switch. To compute Hamming parity of two m-trit input strings, $2m$ qubits need to be exchanged using the quantum switch; making use of a process with a fixed order of communication, one can easily construct a protocol requiring only $m(1 + \log_2 3)$ qubits. Hence, both resources result in the same asymptotic scaling of communication for the Hamming game.

Acknowledgements

We would like to thank Fabio Costa, Michael Cuffaro, Sam Fletcher, Jacques Pienaar, Michal Sedlák, Christopher Timpson and Harald Weinfurter for useful discussions. We acknowledge support from the European Commission project RAQUEL (No. 323970); the Austrian Science Fund (FWF) through the Special Research Programme FoQuS, the Doctoral Programme CoQuS and Individual Project (No. 2462); FQXi and the John Templeton Foundation.

Chapter 5

Exponential communication complexity advantage from quantum superposition of the direction of communication

Abstract

In communication complexity, a number of distant parties have the task of calculating a distributed function of their inputs, while minimizing the amount of communication between them. It is known that with quantum resources, such as entanglement and quantum channels, one can obtain significant reductions in the communication complexity of some tasks. In this work, we study the role of the quantum superposition of the direction of communication as a resource for communication complexity. We present a tripartite communication task for which such a superposition allows for an exponential saving in communication, compared to one-way quantum (or classical) communication; the advantage also holds when we allow for protocols with bounded error probability.

Published. 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(10):100502, 2016.

Contribution in conceiving the research project, contributing to the main result, the proofs and participation in writing of the manuscript.

5.1 Introduction

Quantum resources make it possible to solve certain communication and computation problems more efficiently than what is classically possible. In communication complexity problems, a number of parties have to calculate a distributed function of their inputs while reducing the amount of communication between them [Yao79, KN06]. The minimal amount of communication is called the *complexity of the problem.* For some communication complexity tasks, the use of shared entanglement and quantum communication significantly reduces the complexity as compared to protocols exploiting shared classical randomness and classical communication [Yao93, BCMd10]. Important early examples for which quantum communication yields an exponential reduction in communication complexity over classical communication are the distributed Deutsch-Jozsa problem [BCW98] and Raz's problem [Raz99].

Quantum computation and communication are typically assumed to happen on a definite causal structure, where the order of the operations carried on a quantum system is fixed in advance. However, the interplay between general relativity and quantum mechanics might force us to consider more general situations in which the metric, and hence the causal structure, is indefinite. Recently, a quantum framework has been developed with no assumption of a global causal order [OCB12, ABC$^+$15, OG16]. This framework can also be used to study quantum computation beyond the circuit model, for instance using the "quantum switch" as a resource — a qubit coherently controlling the order of the gates in a quantum circuit [CDPV13]. It has recently been realized experimentally [PMA$^+$15].

It was shown that this new resource provides a reduction in complexity to n black-box queries in a problem for which the optimal quantum algorithm with fixed order between the gates requires a number of queries that scales as n^2 [ACB14]. The quantum switch is also useful in communication complexity; a task has been found for which the quantum switch yields an increase in the success probability, yet no advantage in the asymptotic scaling of the communication complexity was found [FAB15] (see Chapter 4). Most generally, no information processing task is known for which the quantum switch (or any other causally indefinite resource) would provide an exponential advantage over causal quantum (or classical) algorithms.

Here we find a tripartite communication complexity task for which there is an exponential separation in communication complexity between the protocol using the quantum switch and any causally ordered quantum communication scheme. The task requires no promise on inputs and is inspired by the problem of deciding whether a pair of unitary gates commute or anticommute, which can be solved by the quantum switch with only one query of each unitary [Chi12]. If the parties are causally ordered, the number of qubits that needs to be communicated to accomplish the task scales linearly with the number of input bits, whereas the protocol based on the quantum switch only requires logarithmically many communicated qubits. This shows that causally indefinite quantum resources can provide an exponential advantage over causally ordered quantum resources (i.e., entanglement and one-way quantum channels).

5.2 Communication scenario

The tripartite causally ordered communication scenario we consider in this paper is illustrated in Fig. 5.1. Alice and Bob are respectively given inputs $x \in X$ and $y \in Y$, taken from finite sets X, Y. There is a third party, Charlie, whose goal is to calculate a binary function $f(x, y)$ of Alice's and Bob's inputs, while minimizing the amount of communication between all three parties. We shall first assume that communication is one-way only:

from Alice to Bob and from Bob to Charlie. Furthermore, we grant the parties access to unrestricted local computational power and unrestricted shared entanglement. We will also consider bounded error communication, in which the protocol must succeed on all inputs with an error probability smaller than ϵ.

In quantum communication, the parties communicate with each other by sending quantum systems. Conditionally on their inputs, the parties may apply general quantum operations to the received system, and then send this system out. We require that the parties' local laboratories receive a system *only once* from the outside environment. We impose this requirement to exclude sequential communication, in which the parties communicate back and forth by sending quantum systems to each other at different time steps. Alice's laboratory has an input and output quantum state, consisting of N_{A_I} and N_{A_O} qubits, respectively; similar notation is used for Bob's and Charlie's systems. We seek to succeed at the communication task on all inputs with error probability lower than ϵ, while minimizing the number of communicated qubits $N := N_{A_O} + N_{B_O}$. The optimal causally ordered strategy is for Bob to calculate $f(x, y)$ and then communicate the result to Charlie using one bit of communication; in this case N_{A_O} is a good lower bound for N.

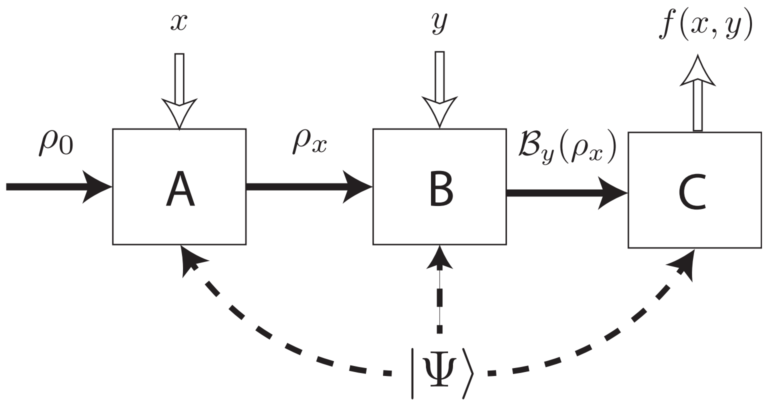

Figure 5.1: Causally ordered quantum communication complexity scenario. Conditionally on their inputs x and y, Alice sends a state ρ_x to Bob, who then applies a CP map $\mathcal{B}_y$ and sends the system to Charlie. The unlimited entanglement shared between the parties is represented by $|\Psi\rangle$. The optimal causally ordered protocol is the one that minimizes the number of qubits in ρ_x (which is a lower bound for the communication complexity of the task)

5.3 Causally ordered communication complexity

The communication complexity of any causally ordered tripartite communication complexity task can be bounded by considering the bipartite task obtained by identifying Bob and Charlie as a single party. Bearing this in mind, we prove a tight lower bound on the quantum communication complexity of an important family of one-way bipartite deterministic (error probability $\epsilon = 0$) communication tasks, which in turn implies a lower bound on the communication complexity of causally ordered tripartite tasks. This result appears in Theorem 5 of Ref. [Kla00], but we present a different proof here.

Lemma 1 *For deterministic one-way evaluation of any binary distributed function* $f : X \times Y \to \{0,1\}$ *such that* $\forall x_1, x_2 \in X$, *with* $x_1 \neq x_2$, $\exists y \in Y$ *for which* $f(x_1, y) \neq f(x_2, y)$, *the minimum Hilbert space dimension of the system sent between two parties sharing an arbitrary amount of entanglement is* $d = \lceil \sqrt{|X|} \rceil$. *Equivalently, the minimum number of communicated qubits is* $\lceil \log_2 d \rceil$.

Proof We recall a well-known result of quantum information [HJS+96], establishing that if Alice and Bob share unlimited entanglement, the largest number of orthogonal (perfectly distinguishable) states that Alice can transmit to Bob by sending a d-dimensional system is d^2. Therefore, they can deterministically compute f if Alice sends a system of Hilbert space dimension $\lceil \sqrt{|X|} \rceil$.

Suppose by way of contradiction that the Hilbert space dimension of the communicated system is only $(\lceil \sqrt{|X|} \rceil - 1)$. The maximal number of orthogonal states that can be transmitted by Alice to Bob is $(\lceil \sqrt{|X|} \rceil - 1)^2 < |X|$. Therefore, there exist inputs $x_1, x_2 \in X$ such that the corresponding states ρ_1, ρ_2 transmitted to Bob are not orthogonal, and thus not perfectly distinguishable [NC00]. By our assumption about the function f, there exists an input $y \in Y$ such that $f(x_1, y) \neq f(x_2, y)$. Therefore, if Bob receives the input y, he will need to distinguish between ρ_1 and ρ_2 in order to output the function correctly, but this cannot be done with zero error probability. ∎

The previous lemma establishes that for a very large class of deterministic communication complexity tasks, it is necessary for Alice to communicate all of her input to Bob. In these cases, the only advantage achieved by causal one-way quantum communication is a reduction by a constant factor of two due to dense coding [BW92]. An important example of this form is the Inner Product game [CvNT13, NS02]. Note that Lemma 1 does not apply to relational tasks such as the hidden matching problem [BYJK04], for which there is an exponential separation between quantum and classical communication complexity.

We now seek to establish a communication complexity task for which indefinite causal order can be used as a resource. In the following we assume that the parties have local laboratories, and that they receive a quantum system from the environment only once. They then perform a general quantum operation on their system, and send it out. An example of a noncausally ordered process is the quantum switch [CDPV13], whose use in the context of communication complexity is shown in Fig. 5.2. Charlie is in the causal future of both Alice and Bob, and an ancilla qubit coherently controls the causal ordering of Alice and Bob; both the target state and the control qubit are prepared externally. Assume that Alice and Bob apply unitary gates U_A and U_B to their respective input systems of N qubits. The global unitary describing the evolution of the system from Charlie's point of view is

$$V(U_A, U_B) = |0\rangle \langle 0|_c \otimes (U_B U_A)_t + |1\rangle \langle 1|_c \otimes (U_A U_B)_t, \tag{5.1}$$

where the index c denotes the control qubit, and the unitaries U_A and U_B act on the target Hilbert space of N qubits.

Using the quantum switch, one can determine whether two unitaries U_A, U_B commute or anticommute with a single query of each unitary, while at least one unitary must be queried twice in the causally ordered case [Chi12]. Explicitly, consider the quantum switch with the control qubit initially in state $|+\rangle_c = \frac{1}{\sqrt{2}}(|0\rangle_c + |1\rangle_c)$ and with initial target state $|\psi\rangle_t$. If A and B apply local unitaries U_A and U_B, the resulting state after applying $V(U_A, U_B)$ is

$$\frac{1}{\sqrt{2}}(|0\rangle_c \otimes U_B U_A |\psi\rangle_t + |1\rangle_c \otimes U_A U_B |\psi\rangle_t). \tag{5.2}$$

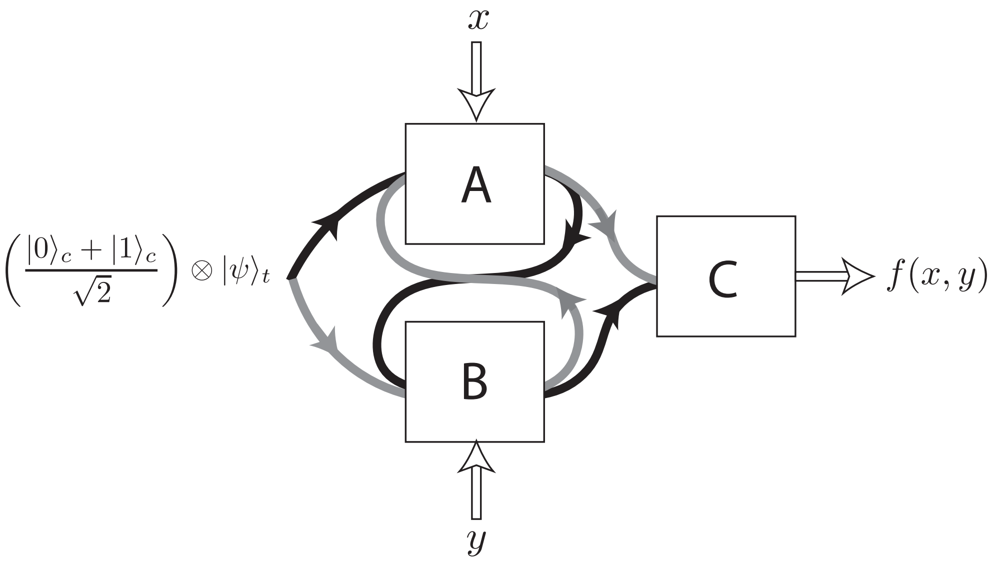

Figure 5.2: Communication complexity setup using the quantum switch. A qubit in the state $\frac{1}{\sqrt{2}}(|0\rangle_c + |1\rangle_c)$ coherently controls the path taken by a system of N qubits in initial state $|\psi\rangle_t$. One path goes first through Alice's lab and then Bob's, while the other path goes first through Bob's lab and then Alice's. Alice and Bob are given classical inputs $x \in X$, $y \in Y$, and Charlie (using the control qubit) computes a binary function of their inputs $f(x, y)$

.

If Charlie subsequently applies a Hadamard gate to the control qubit, the resulting state is

$$\frac{1}{2}(|0\rangle_c \otimes \{U_A, U_B\} |\psi\rangle_t - |1\rangle_c \otimes [U_A, U_B] |\psi\rangle_t). \tag{5.3}$$

Suppose that Alice and Bob randomly choose unitaries from a set $\mathcal{U}$ and that there exists a state $|\psi\rangle_t$ such that $\forall U, V \in \mathcal{U}$, either $[U, V] |\psi\rangle_t = 0$ or $\{U, V\} |\psi\rangle_t = 0$. Then Eq. (5.3) shows that the quantum switch with initial target state $|\psi\rangle_t$ and control qubit $|+\rangle_c$ as inputs allows Charlie to discriminate between these two possibilities with certainty by measuring the control qubit in the computational basis.

5.4 The Exchange Evaluation game

We now define a communication complexity task, the Exchange Evaluation game EE_n, for any integer n. In this game, Alice and Bob are respectively given inputs $(\mathbf{x}, f), (\mathbf{y}, g) \in \mathbb{Z}_2^n \times F_n$, where F_n is the set of functions over $\mathbb{Z}_2^n$ that evaluate to zero on the zero vector

$$F_n = \{f : \mathbb{Z}_2^n \to \mathbb{Z}_2 \,|\, f(\mathbf{0}) = 0\}. \tag{5.4}$$

Charlie must output

$$EE_n(\mathbf{x}, f, \mathbf{y}, g) = f(\mathbf{y}) \oplus g(\mathbf{x}), \tag{5.5}$$

where the symbol $\oplus$ denotes addition modulo 2. This game can be interpreted as the sum modulo 2 of two parallel random access codes [ANTSV99].

We first construct an encoding of the inputs $(\mathbf{x}, f), (\mathbf{y}, g)$ in terms of local n-qubit unitaries that all commute or anticommute; we then use the previous observation to conclude that the switch succeeds deterministically at this task with n qubits of communication.

We start with some definitions. The group of Pauli X operators on n qubits is defined as

$$X(\mathbf{x}) = X_1^{x_1} \otimes X_2^{x_2} \otimes \cdots \otimes X_n^{x_n}, \tag{5.6}$$

where x_i is the ith component of the binary vector $\mathbf{x} \in \mathbb{Z}_2^n$. Here, X_i is the single qubit Pauli X-operator acting on the ith qubit, and $X_i^0 = \mathbb{I}_i$ is the single qubit identity matrix.

We associate to every $f \in F_n$ a diagonal matrix

$$D(f) = \sum_{\mathbf{z}\in\mathbb{Z}_2^n} (-1)^{f(\mathbf{z})} \left|\mathbf{z}\right\rangle \left\langle\mathbf{z}\right|, \tag{5.7}$$

where $|\mathbf{z}\rangle$ is the state such that $Z_i |\mathbf{z}\rangle = (-1)^{z_i} |\mathbf{z}\rangle$, with Z_i the single qubit Pauli Z operator acting on qubit i. The set $\{D(f)\}_{f\in F_n}$ consists of all diagonal matrices with entries ± 1 in the computational basis, such that the first entry is $+1$.

We define the set of unitaries

$$\mathcal{U}_n = \{X(\mathbf{x})D(f) | (\mathbf{x}, f) \in \mathbb{Z}_2^n \times F_n\}, \tag{5.8}$$

which has dimension

$$|\mathcal{U}_n| = 2^{2^n+n-1}. \tag{5.9}$$

This superexponential scaling of $|\mathcal{U}_n|$ is essential to establish a communication advantage with the quantum switch. Also note that

$$X(\mathbf{x})D(f)X(\mathbf{y})D(g) \left|\mathbf{0}\right\rangle = (-1)^{f(\mathbf{y})} \left|\mathbf{x} \oplus \mathbf{y}\right\rangle. \tag{5.10}$$

Therefore, when acting on the n-qubit input state $|\mathbf{0}\rangle$, the elements of $\mathcal{U}_n$ all commute or anticommute with each other, and

$$\begin{aligned} [X(\mathbf{x})D(f), X(\mathbf{y})D(g)] \left|\mathbf{0}\right\rangle = 0\,, \text{if } (-1)^{f(\mathbf{y})} = (-1)^{g(\mathbf{x})} \\ \{X(\mathbf{x})D(f), X(\mathbf{y})D(g)\} \left|\mathbf{0}\right\rangle = 0\,, \text{if } (-1)^{f(\mathbf{y})} = (-1)^{g(\mathbf{x})+1}. \end{aligned}$$

Therefore, the game is equivalent to determining whether the corresponding unitaries $X(\mathbf{x})D(f)$ and $X(\mathbf{y})D(g)$ commute or anticommute *when applied to the state* $|\mathbf{0}\rangle$. By the discussion following Eq. (5.3), this problem can be solved deterministically by Charlie using the quantum switch with $O(n)$ qubits of communication from Alice to Bob, with a strategy consisting of applying the unitary corresponding to their input according to Eq. (5.8).

5.5 Exponential advantage

We now show that the Exchange Evaluation game satisfies the conditions of Lemma 1; this will allow us to conclude that for deterministic ($\epsilon = 0$) evaluation in the one-way causally ordered case, EE_n requires an amount of communicated qubits that grows exponentially with n.

Proposition 2 *For every $(\mathbf{x_1}, f_1), (\mathbf{x_2}, f_2) \in \mathbb{Z}_2^n \times F_n$, such that $(\mathbf{x_1}, f_1) \neq (\mathbf{x_2}, f_2)$, there exists $(\mathbf{y}, g) \in \mathbb{Z}_2^n \times F_n$ such that $EE_n(\mathbf{x_1}, f_1, \mathbf{y}, g) \neq EE_n(\mathbf{x_2}, f_2, \mathbf{y}, g)$.*

Proof First note that $EE_n(\mathbf{x_1}, f_1, \mathbf{y}, g) \neq EE_n(\mathbf{x_2}, f_2, \mathbf{y}, g)$ if and only if

$$f_1(\mathbf{y}) \oplus f_2(\mathbf{y}) \oplus g(\mathbf{x_1}) \oplus g(\mathbf{x_2}) = 1. \tag{5.11}$$

Then, since $(\mathbf{x_1}, f_1) \neq (\mathbf{x_2}, f_2)$, either $\mathbf{x_1} \neq \mathbf{x_2}$ or $f_1 \neq f_2$ holds. We check that the conditions of the lemma are satisfied in both cases.

(i) Case where $\mathbf{x_1} \neq \mathbf{x_2}$:

Suppose without loss of generality that $\mathbf{x_1} \neq \mathbf{0}$ and define g as the function such that $g(\mathbf{x_1}) = 1$ and $g(\mathbf{z}) = 0$, $\forall \mathbf{z} \neq \mathbf{x_1}$. Also, because $f_1, f_2 \in F_n$, $f_1(\mathbf{0}) = f_2(\mathbf{0}) = 0$. Therefore, the function g we just defined and $\mathbf{y} = \mathbf{0}$ satisfy Eq. (5.11).

(ii) Case where $f_1 \neq f_2$:

Let $\mathbf{y} \in \mathbb{Z}_2^n$ be a vector for which f_1 and f_2 differ, so that $f_1(\mathbf{y}) + f_2(\mathbf{y}) = 1$. Then this $\mathbf{y}$ and the zero function $g(\mathbf{x}) = 0 \,\forall \mathbf{x}$ satisfies Eq. (5.11). ∎

According to Eq. (5.9), the dimension of the set of inputs to EE_n is $|\mathcal{U}_n| = 2^{2^n+n-1}$. Direct application of Proposition 2 with Lemma 1 establishes that the number of qubits of communication required for deterministic success in the causally ordered case is $\frac{1}{2}\log_2 |\mathcal{U}_n| = \frac{1}{2}(2^n + n - 1) = \Omega(2^n)$, using dense coding. In comparison, we have seen that with the quantum switch as a resource, we need only n qubits of communication between Alice and Bob to calculate this function. We thus conclude that for the Exchange Evaluation game, there is an exponential separation in the deterministic communication complexity of EE_n.

Note that with two-way (classical) communication, it is possible to solve the Exchange Evaluation game with $2n + 2$ bits of communication, simply by having Alice and Bob send their vectors $\mathbf{x}$, $\mathbf{y}$ to the other party, followed by local evaluation of $f(\mathbf{y})$ and $g(\mathbf{x})$ by the parties and communication of the result to Charlie. We emphasize that once we allow two-way communication, the quantum advantage can also disappear in traditional quantum communication complexity (comparing causally ordered quantum communication with classical communication): this is the case for the distributed Deutsch-Jozsa problem [BCW98], but not for Raz's problem [RBK11].

For causally ordered communication complexity tasks, the exponential quantum-classical separation does not always continue to hold when allowing for protocols to have a small but nonzero error probability $\epsilon > 0$. Indeed, looking at early examples of tasks, the advantage disappears for the distributed Deutsch-Jozsa problem [BCW98], while it remains for Raz's problem [Raz99]. We prove in Appendix C that the one-way quantum communication complexity with bounded error for EE_n scales as $\Omega(2^n)$, and thus that the exponential separation in communication complexity due to superposition of causal ordering persists when allowing for a nonzero error probability.

To show that it is possible to operationally distinguish quantum control of causal order from two-way communication one could introduce counters at the output ports of Alice's and Bob's laboratories, whose role is to count the number of uses of the channels. Such an argument has already been made in Ref. [ACB14] to justify a computational advantage. We can model a counter as a qutrit initially in the state $|0\rangle$, whose evolution when a system exits the laboratory is $|i\rangle \to |i + 1 \bmod 3\rangle$, where $i \in \{0, 1, 2\}$. Then, for both one-way communication and the quantum switch, the counters of Alice and Bob will be in the state $|1\rangle$ at the end of the protocol; for genuine two-way communication, at least one of these counters will be in the final state $|2\rangle$. Therefore, the expectation value of the observables $N = \sum_{i=0}^{2} |i\rangle\langle i|$ for the counters allows us to distinguish realizations of the quantum switch, such as [PMA+15], from two-way quantum communication.

5.6 Conclusions

In conclusion, we have found a communication complexity task, the Exchange Evaluation game, for which a quantum superposition of the direction of communication — the quantum switch — results in an exponential saving in communication when compared to causally ordered quantum communication. An interesting feature of this game is that it is not a promise game, as are most known tasks for which quantum resources have an exponential advantage [BCMd10].

In future work, it would be interesting to explore other information processing tasks for which the quantum switch – or other causally indefinite processes – may yield interesting advantages. For example, one could look at the uses of the quantum switch for secure distributed computation [Yao82, Lo97, BCS12, LLWJ13]. Indeed, imagine that Alice and Bob both want to learn about the value of EE_n, in such a way that the other party does not learn about their inputs. They could achieve this goal by enlisting a third party and using the quantum switch with the EE_n protocol.

Acknowledgements

We thank Ashley Montanaro for pointing out Ref. [Kla00] to us, used to establish the bounded-error advantage. We acknowledge support from the European Commission project RAQUEL (No. 323970); the Austrian Science Fund (FWF) through the Special Research Programme FoQuS, the Doctoral Programme CoQuS and Individual Project (No. 2462), and the John Templeton Foundation. P. A. G. is also supported by FQRNT (Quebec).

Chapter 6

Quantum superpositions of "common-cause" and "direct-cause" causal structures

Abstract

The constraints arising for a general set of causal relations, both classically and quantumly, are still poorly understood. As a step in exploring this question, we consider a *coherently controlled superposition* of "direct-cause" and "common-cause" relationships between two events. We propose an implementation involving the spatial superposition of a mass and general relativistic time dilation. Finally, we develop a computationally efficient method to distinguish such genuinely quantum causal structures from classical (incoherent) mixtures of causal structures and show how to design experimental verifications of the nonclassicality of a causal structure.

Published. A. Feix and Č. Brukner. Quantum superpositions of "common-cause" and "direct-cause" causal structures. *New J. Phys.*, 19(12):123028, 2017.

Contribution in conceiving the research project, deriving the main results and proofs, and writing the manuscript.

6.1 Introduction

The deeply rooted intuition that the basic building blocks of the world are *cause-effect-relations* goes back over a thousand years [Ari33, Hum75, Rei56] and yet still puzzles philosophers and scientists alike.

In physics, general relativity provides a theoretic account of the causal relations that describe which events in spacetime can influence which other events. For two (infinitesimally close) events separated by a time-like or light-like interval, one event is in the future light cone of the other, such that there could be a direct cause-effect relationship between them. When a space-like interval separates two events, no event can influence the other. The causal relations in general relativity are *dynamical*, since they are imposed by the dynamical light cone structure [Bro09].

Incorporating the concept of causal structure in the quantum framework leads to novelties: it is expected that such a notion will be both *dynamical*, as in general relativity, as well as *indefinite*, due to quantum theory [Har07]. One might then expect indefiniteness with respect to the question of whether an interval between two events is time-like or space-like, or even whether event A is prior to or after event B for time-like separated events. Yet, finding a unified framework for the two theories is notoriously difficult and the candidate models still need to overcome technical and conceptual problems.

One possibility to separate conceptual from technical issues is to consider more general, *theory-independent* notions of causality. The *causal model* formalism [SGS93, Pea00] is such an approach, which has found applications in areas as diverse as medicine, social sciences and machine learning [IRW11]. The study of its quantum extension, allowing for non-local correlations [WS15, CL14, HLP14, Fri16] or including new information-theoretic principles [PB15, CMG15, CS16] might provide intuitions and insights that are currently missing from the theory-laden take at combining quantum mechanics with general relativity.

Recently, it was found that it is possible to formulate quantum mechanics without any reference to a global causal structure [OCB12]. The resulting framework—the *process matrix formalism*—allows for processes which are incompatible with any definite order between operations. One particular case of such a process is the "quantum switch", where an auxiliary quantum system can coherently control the order in which operations are applied [CDPV13]. This results in a quantum controlled superposition of the processes "A causing B" and "B causing A". The quantum switch can also be realized through a preparation of a massive system in a superposition of two distinct states, each yielding a different but definite causal structure for future events [Zyc15, ZCPB17]. Furthermore, it provides computational [ACB14] and communication [FAB15, AFAB16] advantages over standard protocols with a fixed order of events. The first experimental proof-of-principle demonstration of the switch has been reported recently [PMA$^+$15].

Given that one can implement superpositions of two different causal orders, one may ask if and how one could realize situations in which two events are in superpositions of being in "common-cause" (A does not cause B directly) and "direct-cause" (A and B share no common cause) relationships. Here we show that such superpositions exist and how to verify them.

We develop a framework for the computationally efficient verification of *coherent superpositions* of "direct-cause" and "common-cause" causal structures. We propose a natural physical realization of a quantum causal structure with the spatial superposition of a mass and general relativistic time dilation using the approach developed in Refs. [ZCPB17, Zyc15]. Finally, using the process matrix formalism, we define a degree of "nonclassicality of causal structures" and show how to design *experimental verifications* thereof using a semidefinite program [NN87].

6.2 Quantum causal models

To formalize the pre-theoretic notion of causality, the standard approach is to use *causal models* [SGS93, Pea00], consisting of (i) a causal network and (ii) model parameters. The *causal network* is represented by a *directed graph*, whose nodes are variables and whose directed edges represent causal influences between variables. The causal influence from A to B is identified with the possibility of *signaling* from A to B. To exclude the possibility of causal loops, one imposes the condition that the graph should be *acyclic* (a "DAG"), which induces a *partial order* ("causal order") over the variables. The *model parameters* then determine how the probability distribution of each variable or set of variables is to be computed as a function of the value of its parent nodes.

Fully characterizing the causal model requires information which is available only through "interventions", where the value of one or more variables is *set* to take a specific value, independently of the values of the rest of the variables. In the resulting causal network, the connections from all its parents are eliminated. Intervening on all relevant variables is sufficient to completely reconstruct the full causal model [Pea00]. Since this is often practically impossible, it is crucial to investigate the possibilities of causal inference from a *limited* set of interventions.

Moving to *quantum* causal models, we will define variables as results of generalized quantum operations applied to incoming quantum systems ("local operation"). Formally, a local operation $\mathcal{M}_A : A_I \to A_O$ is a map from a density matrix $\rho_{A_I} \in A_I$ to $\rho_{A_O} \in A_O$ (where A_I (A_O) denotes the space of linear operators on the Hilbert space $\mathcal{H}^{A_I}$ ($\mathcal{H}^{A_O}$)). The Choi-Jamiołkowski (CJ) isomorphism [Cho75, Jam72] provides a convenient representation of the local map as a positive operator $M_A \in A_I \otimes A_O$ (the explicit definition is given in Appendix D.1).

The *quantum causal structure*, which is the quantum analogue of the classical causal network, maps the aforementioned local operations to a probability distribution. It can be thought of as a *higher order operator* and can be formally represented in the "super-operator", "quantum comb" or "process matrix" formalisms [GW07, CDP08a, CDP09, BCDP11, LS13, OCB12].

We will focus on quantum causal structures with three laboratories (three nodes in the graph) A, B and C compatible with the causal order "A is not after B, which is not after C" ($A \prec B \prec C$). This means that there are no causal influences from B and C to A, nor from C to B (see Fig. 6.1). (Since C is last, C's output space C_O can be disregarded.)

In the process matrix formalism, the quantum causal structure is represented by the matrix $W \in A_I \otimes A_O \otimes B_I \otimes B_O \otimes C_I$ [OCB12, ABC+15]. The probabilities of observing the outcomes i, j, k at A, B, C (corresponding to implementing the completely positive (CP) maps M_A^i, M_B^j, M_C^k respectively) are given by the *generalized Born rule*:

$$p(A = i, B = j, C = k) = \mathrm{tr}[W\, M_A^i \otimes M_B^j \otimes M_C^k]. \tag{6.1}$$

The quantum causal structure and local operations should generate only meaningful (that is, *positive* and *normalized*) probability distributions. In addition, we require the probability distributions to be compatible with the causal order $A \prec B \prec C$. Note that both "common-cause" and "direct-cause" relationships between A and B are compatible with this causal order.

In terms of process matrices, these conditions are equivalent to requiring that W satisfies [ABC+15]:

$$W \geq 0, \quad W = \mathcal{L}_{A\prec B\prec C}(W) \tag{6.2}$$

$$\mathrm{tr}\, W = d_{A_O} d_{B_O}. \tag{6.3}$$

$\mathcal{L}_{A\prec B\prec C}(\cdot)$ is the projection onto processes compatible with the causal order $A \prec B \prec C$, defined in Appendix D.2. Eq. (6.2) defines a convex cone $\mathcal{W}$, eq. (6.3) a normalization constraint.

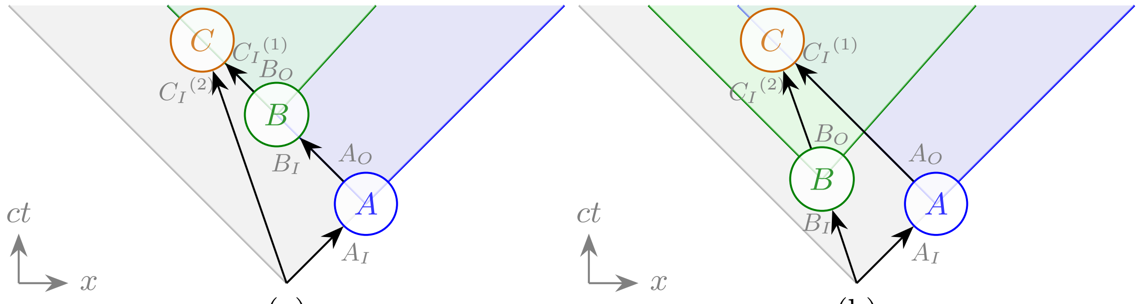

Figure 6.1: Space-time diagram of two causal structures compatible with the causal order $A \prec B \prec C$: (a) direct-cause process W^{dc} with a quantum channel between A_O and B_I; (b) common-cause process W^{cc} with a shared (possibly entangled state) between A_I and B_I, but no channel between A_O and B_I (A and B are space-like separated).

Following the standard DAG terminology, a purely "direct-cause" process W^{dc} contains only a *direct cause-effect relation between A and B*, excluding any form of *common cause* between A and B. Any correlation between A and B is therefore caused by A alone (Fig. 6.1 (a)) and Fig. 6.2 (a)). Tracing out C_I and B_O, the process matrix is a tensor product $\rho^{A_I} \otimes \tilde{W}^{A_O B_I}$. In our scenario, it will prove natural to *extend* this definition to include *convex mixtures* of direct-cause processes, i.e.,

$$\mathrm{tr}_{C_I B_O} W^{\mathrm{dc}} = \sum_i p_i \rho_i^{A_I} \otimes \tilde{W}_i^{A_O B_I}, \tag{6.4}$$

where $p_i \geq 0, \sum_i p_i = 1$, $\rho_i^{A_I}$ are arbitrary states and $\tilde{W}_i^{A_O B_I}$ arbitrary valid channels between Alice's output and Bob's input, representing to direct cause-effect links between A and B.

Such a process can be interpreted as a probability distribution over states entering A_I and corresponding channels from A_O to B_I. In the DAG framework, such probability distributions can be obtained from a graph with an additional latent node that acts as a common cause for all the observed nodes or simply ignorance of the graph that is implemented. Every channel from A to B with classical memory can be decomposed in this way; see Appendix D.7 for details.

On the other hand, a purely "common-cause" process W^{cc} does not include *any direct causal influence between A and B* (Fig. 6.1 (b)) and Fig. 6.2 (b)). This implies that there is no channel between A_O and B_I. Therefore, when B_O and C_I are traced out, the process factorizes as

$$\mathrm{tr}_{C_I B_O} W^{\mathrm{cc}} = \sigma^{A_I B_I} \otimes \mathbb{1}^{A_O}, \tag{6.5}$$

where $\sigma^{A_I B_I}$ is an arbitrary (possibly entangled, possibly mixed) state, representing the common-cause influencing A and B.

6.3 Classical and quantum superpositions of causal structures

One possibility of combining direct-cause and common-cause processes consists in allowing for *classical mixtures* thereof: imagine that flipping a (possibly biased) coin determines

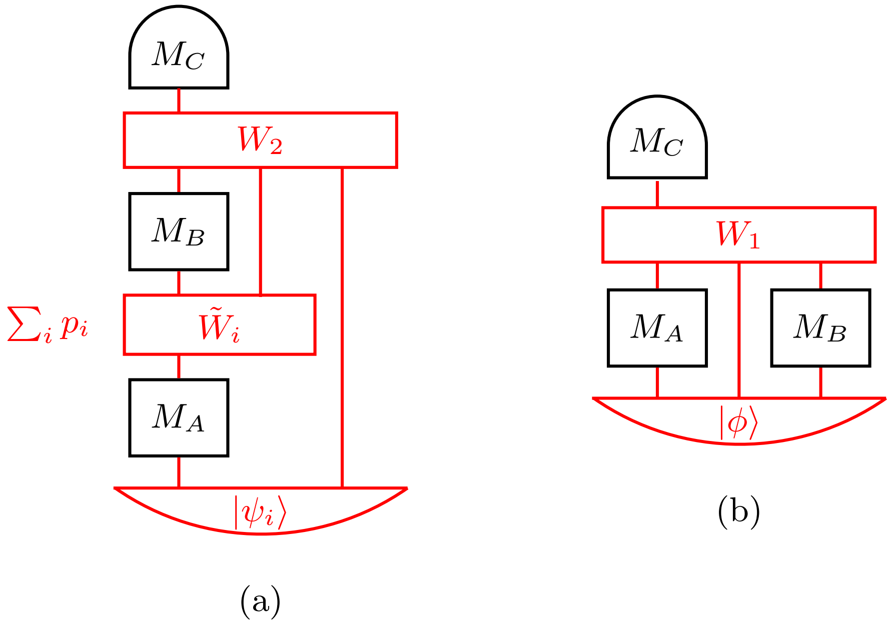

Figure 6.2: Circuit representation of the causal structures of Fig. 6.1, where $|\psi_i\rangle$ and $|\phi\rangle$ are states, $\tilde{W}_i, W_2$ and W_1 are CP trace preserving (CPTP) maps (lines can represent quantum systems of different dimensions). (a) The direct-cause process W^{dc} is the most general one satisfying (6.4); (b) the common-cause process W^{cc} is the most general one satisfying (6.5).

which process will be realized in an experimental run. Formally, this is described by a process W^{conv} which can be decomposed as a convex combination:

$$W^{\text{conv}} = qW^{\text{cc}} + (1-q)W^{\text{dc}}, \tag{6.6}$$

where $0 \leq q \leq 1$, W^{dc} satisfies (6.4) and W^{cc} satisfies (6.5). Note that such a classical mixture was experimentally implemented in Ref. [RAV$^+$15].

Can there be causal structures exhibiting *genuine quantum coherence*, i.e., that cannot be decomposed as a classical mixture of direct-cause and common-cause processes (while respecting the causal order $A \prec B \prec C$)?

We now give an example of such a coherent superposition. It is analogous to the "quantum switch" [CDPV13], which coherently superposes two causal orders $A \prec B \prec C$ and $B \prec A \prec C$, where the causal structure is entangled to a "control" system $C_I^{(0)}$ added to C's input space[1]. To keep the notation simple, we define it in the "pure" CJ-vector notation (see Appendix D.1):

$$\begin{aligned} |w\rangle = \frac{1}{\sqrt{2}}\Big(&|0\rangle^{C_I^{(0)}} |\psi\rangle^{A_I B_I} |I\rangle\rangle^{A_O C_I^{(1)}} |I\rangle\rangle^{B_O C_I^{(2)}} \\ &+ |1\rangle^{C_I^{(0)}} |\psi\rangle^{A_I C_I^{(2)}} |I\rangle\rangle^{A_O B_I} |I\rangle\rangle^{B_O C_I^{(1)}}\Big), \\ W^{\text{coherent}} &:= |w\rangle\langle w| \end{aligned} \tag{6.7}$$

where $|I\rangle\rangle := \sum_{j=1}^{d} |jj\rangle$ represents a non-normalized maximally entangled state—the CJ-representation of an identity channel. The corresponding superposition of circuits is shown

[1]See Ref. [MRSR17] for a different type of quantum causal structure proposed independently.

in Fig. 6.3. W^{coherent} satisfies neither the direct-cause condition (6.4) nor the common-cause condition (6.5) and is a projector on a pure vector, so it cannot be decomposed into *any nontrivial convex combination*, in particular not a mixture of direct-cause and common-cause processes. This proves that the process's causal structure is nonclassical.

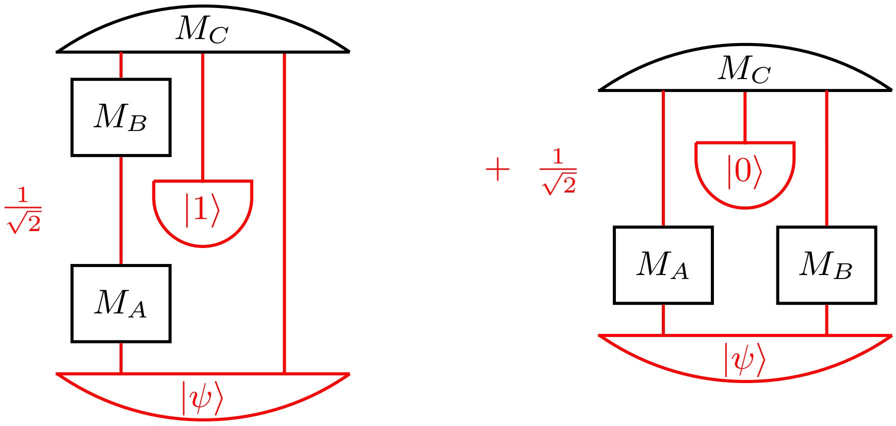

Figure 6.3: Coherent superposition of a direct-cause and a common-cause process, implementing the causal structure W^{coherent} of (6.7).

6.4 Physical implementation of the quantum causal structure

The causal structure W^{coherent} would not be of particular interest if it were a mere theoretical artifact. We now give an explicit and plausible physical scenario to realize the quantum causal structures in models which respect the principles of general relativistic time dilation and quantum superposition. We utilize the approach recently developed for the "gravitational quantum switch" to realize a superposition and entanglement of two different causal orders [Zyc15, ZCPB17].

Consider two observers, Alice and Bob, who have initially synchronized clocks. We *define* the events in the respective laboratories with respect to the *local clocks*. Bob's local operation will always be applied at his local time τ_B, while Alice's is applied at her local time τ_A. We will consider two configurations, which will be controlled by a quantum system. The state of the control system is given by the position of a massive body. In the first configuration, all masses are sufficiently far away such that the parties are in an approximately flat spacetime. The events in the two laboratories are chosen such that the event B is outside of A's light cone and the common-cause causal relationship is implemented. The coordinate times of the two events, as measured by a local clock of a distant observer, are $t_A \approx \tau_A$ and $t_B \approx \tau_B$. (Fig. 6.4 (a)). In the second configuration, a mass M is put closer to Bob's laboratory than to Alice's such that his clock runs slower with respect to hers due to gravitational time dilation. With a suitable choice of mass and distance between Alice and Bob, the event B, which is defined by his clock showing local time τ_B, will be inside A's future light cone. In terms of coordinate times one now has $t'_A = \tau_A/\sqrt{-g_{00}(A)}$ and $t'_B = \tau_B/\sqrt{-g_{00}(B)}$, where $g_{00}(A)$ and $g_{00}(B)$ are the "00" components of the metric tensor at the position of the laboratories. This configuration can implement the direct-cause relationship (Fig. 6.4 (b)).

If the mass M is initially in a coherent *spatial superposition* of a position close and a position far away from Bob, the quantum superposition of causal structures W_{coherent} is

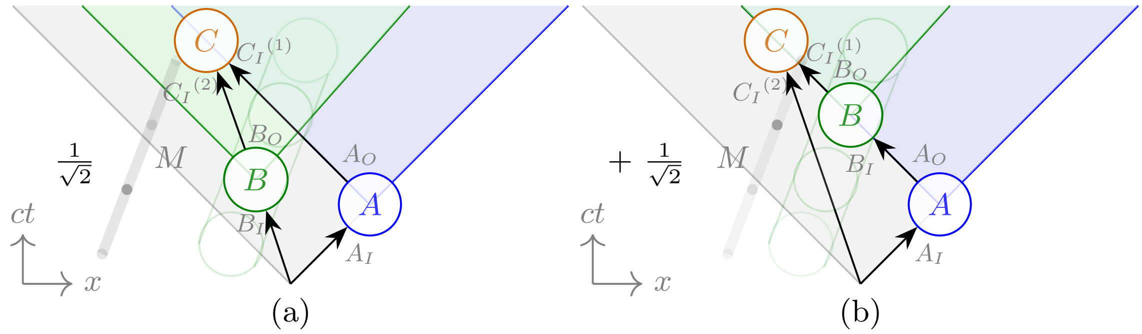

Figure 6.4: Space-time diagrams of events in a superposition of casual structures, as seen from a distant observer. Bob's laboratory is moving along a time-like curve, indicated by the circles showing his laboratory before and after τ_B. (a) If the mass M is far away from Bob, the event at his local time τ_B is space-like separated from A and a common-cause causal structure is realized. (b) If M is sufficiently close to B, because of time dilation, B's event at time τ_B, is in the future light cone of A, establishing a direct-cause structure between A, B and C. For a *coherent superposition* of the positions of M (the position of M being the control system $C_I^{(0)}$), the quantum causal structure will be described by W_{coherent}, as given in (6.7).

implemented. The *position of the mass* acts as the control system $C_I^{(0)}$;[2] it can be received by Charlie, who can manipulate it further (in particular, measure it in the superposition basis). Any possible information about the causal structure (direct cause or common cause) encoded in the degrees of freedom of the laboratories, such as for example in the clocks of the labs, must be erased, possibly using the methods of Ref [ZCPB17].

Note that, in contrast to the superposition of different causal orders [Zyc15, ZCPB17], the time dilation necessary to "move B in or out" of the light cone can, in principle, be made *arbitrarily small*, if Bob can define τ_B and thus the event B with a sufficiently precise clock[3].

To give an idea of the orders of magnitude involved: for a spatial superposition of the order of $\Delta x = 1\,\text{mm}$ and a mass of $M = 1\,\text{g}$, Bob's clock should resolve one part in 10^{27} to be able to certify the nonclassicality of the causal structure. This regime is still quite far from experimental implementation, since the best molecule interferometers [EGA+13] do not go beyond $M = 10^5\,\text{amu}$, $\Delta x = 10^{-6}\,\text{m}$, while the best atomic lattice clocks achieve a precision of one part in 10^{18} [NCH+15]. An additional difficulty consists in avoiding significant entanglement between the position of the mass and systems other than the local clocks. Nonetheless this regime is still far away from the Planck scale that is usually assumed to be relevant for quantum gravity effects.

We also stress that the process W_{coherent}, although it cannot be decomposed as a convex combination of a common cause and a direct cause process, is still compatible with the causal order $A \prec B \prec C$ and, as such [BDPC11], can be realized as a quantum circuit, as shown in Fig. D.1 (b) of Appendix D.2.

[2] The state $|0\rangle$ corresponding to the mass being far away from Bob and the state $|1\rangle$ corresponding to the mass being close to Bob.

[3] If Bob's clock cannot resolve the interval $\tau_B(1 - 1/\sqrt{-g_{00}(B)})$ within the time τ_B, the event B will be inside or outside A's light cone randomly and independently of the position of M, adding noise to the process.

6.5 Verifying the nonclassicality of causal structures

We now provide an *experimentally accessible* and *efficiently computable* measure of the nonclassicality of causality.

Let us first define the set $\mathcal{S}$ of operators which are positive on any convex combination W^{conv} of direct-cause and common-cause processes (i.e., processes satisfying (6.6)):

$$S \in \mathcal{S} \Rightarrow \operatorname{tr}[S\,W^{\text{conv}}] \geq 0 \quad \forall W^{\text{conv}}. \tag{6.8}$$

If S is positive on all convex combinations of direct-cause and common-cause process matrices, then it is also positive on all direct-cause ($\operatorname{tr}[S\,W^{\text{dc}}] \geq 0$) and common-cause ($\operatorname{tr}[S\,W^{\text{cc}}] \geq 0$) processes individually.

Since W^{dc} is a direct-cause process (6.4), if and only if the operator $\operatorname{tr}_{C_I B_O} W^{\text{dc}}$ is separable with respect to the bipartition $(A_I, A_O B_I)$, we effectively require S to be an *entanglement witness* [HHHH09, CS14] of the reduced process for the bipartition $(A_I, A_O B_I)$. The full characterization of the set of entanglement witnesses is known to be computationally hard [Gur03]. Instead, we will use the *positive partial transpose* [Per96, HHH96] criterion as a relaxation to define an efficiently computable measure of nonclassicality.

Enforcing that S is positive on common-cause process matrices in terms of semidefinite constraints is straightforward: since the condition for W^{cc} (6.5) to be a common-cause process is already a semidefinite constraint, the "dual" constraint for S to be positive on all common-cause process matrices is semidefinite as well.

The operators in the set $\mathcal{S}_{\text{SDP}}$ (explicitly constructed in Appendix D.4) are defined as those that obey the condition of having a positive parital transpose and being positive on all common-cause process matrices. Every $S \in \mathcal{S}_{\text{SDP}}$ has positive trace with any W^{conv}. Conversely, $\operatorname{tr}[S\,W] < 0$ certifies that the process W is a genuinely nonclassical causal structure—the operators S can therefore be used as *nonclassicality of causality witnesses*[4].

It is crucial to realize that for every given genuinely quantum W, one can *efficiently optimize*—the optimization is a semidefinite program [NN87]—over the set of nonclassicality witnesses to find the one that has minimal trace with W:

$$\begin{aligned} &\min \operatorname{tr}[S\,W] \\ \text{s.t.} \quad &S \in \mathcal{S}_{\text{SDP}}, \quad \mathbb{1}/d_O - S \in \mathcal{W}^*, \end{aligned} \tag{6.9}$$

where $\mathcal{W}^*$ is the dual cone of $\mathcal{W}$, given in Appendix D.3. The *normalization condition* $\mathbb{1}/d_O - S \in \mathcal{W}^*$ is necessary for the optimization to reach a finite minimum and confers an operational meaning to $\mathcal{C}(W) := -\operatorname{tr}[S_{\text{opt}}\,W]$: it is the amount of "worst-case noise" the process can tolerate before its quantum features stop being detectable by witnesses in $\mathcal{S}_{\text{SDP}}$ (in analogy to the "generalized robustness of entanglement" [Ste03]). Because of its ability to certify the quantum nonclassicality of causal structures, we will refer to $\mathcal{C}(\cdot)$ as the "nonclassicality of causality". Note that $\mathcal{C}(\cdot)$ satisfies the natural properties of *convexity* and *monotonicity under local operations* (see Appendix D.5).

To experimentally verify the properties of a process like W^{coherent}, one can use the semidefinite program (6.9) to compute the optimal nonclassicality of causality witness S_{opt} for W^{coherent}. The nonclassicality of causality $\mathcal{C}(W^{\text{coherent}})$ can be measured by decomposing S_{opt} in a convenient basis of local operations. In general, this is as demanding as performing a full "causal tomography" [ABC+15, RAV+15, CS16].

[4]The "causal witnesses" introduced in Ref. [ABC+15] are conceptually different, since they examine whether a process can be decomposed as a convex mixture of *causally ordered* processes. All of the processes we study here have a fixed causal order $A \prec B \prec C$.

6.6 Causal inference under experimental constraints

There are two reasons to consider witnesses that are subject to certain *additional restrictions*. First, there might be various technical limitations arising from the experimental setup [RAV+15, PMA+15], which make full tomography impractical. Second, in analogy to the classical case, it is of *conceptual* interest to investigate the power of *quantum causal inference mechanisms* working on *limited data*. In particular, one might want to investigate differences between quantum and classical causal inference algorithms under such constraints.

As an application of this method, we will examine witnesses for the process W^{coherent}. In the following, we will consider qubit input and output spaces, i.e., $\dim A_I = \dim A_O = \dim B_I = \dim C_I^{(0,1,2)} = 2$ for simplicity and computational speed. The optimal witness for W_{coherent}, obtained from the optimization (6.9) using YALMIP [Lof04] with the solver MOSEK [mos15], leads to a nonclassicality of causality of $\mathcal{C}(W^{\text{coherent}}) = -\operatorname{tr}[S_{\text{opt}} W_{\text{coherent}}] \approx 0.2278$.

An intriguing feature of quantum causal models is that direct-cause correlations (Fig. 6.1 (a)) and common-cause correlations (Fig. 6.1 (b)) can be distinguished through a restricted class of informationally symmetric operations [LS13], sometimes called "observations" [FJV13, RAV+15] that are non-demolition measurements (we refer the reader to Appendix D.8 for certain issues with this definition). We can constrain a witness S^{ndmeas} to consist of linear combinations of such non-demolition measurements through an additional condition to the semidefinite program (6.9), given in Appendix D.6.

Surprisingly, purely "observational" witnesses are sufficient not only to distinguish common-cause from direct-cause processes, but *also* to distinguish a classical mixture of direct-cause and common-cause processes from a genuine quantum superposition, since $-\operatorname{tr}[S_{\text{opt}}^{\text{ndmeas}} W_{\text{coherent}}] \approx 0.0732$.

Since measurements and repreparations and even non-demolition measurements are often challenging to implement [GLP98], it can also be useful to consider a nonclassicality of causality witness S^{unitary} which can be decomposed into *unitary operations* for A and B, and arbitrary measurements for C. The requirement can also easily be translated in a semidefinite constraint, given in Appendix D.6. One finds that $-\operatorname{tr}[S_{\text{opt}}^{\text{unitary}} W_{\text{coherent}}] \approx 0.1686$. A summary of the different constraints mentioned in this section can be found in Appendix D.6.

6.7 Conclusions

We presented a three-event quantum causal model compatible with the causal order $A \prec B \prec C$ which is a quantum controlled *coherent superposition between common-cause and direct-cause models*, not a classical mixture thereof.

The experimental implementation we proposed is of conceptual interest, since it relies both on general relativity and the quantum superpositions principle, two elements we expect to feature in a full theory unifying quantum theory and general relativity. Interestingly, both the mass of the object and the separation between the two amplitudes can be arbitrarily small, as long as Bob has access to a sufficiently precise clock to define the instant of his event B.

In order to experimentally certify a genuinely quantum causal structure, we introduced and characterized *nonclassicality of causality witnesses* and provided a semidefinite program to efficiently compute them. Experimental and conceptual constraints are readily included in the framework.

The potential of quantum causal structures as a quantum information resource was recently demonstrated in terms of query complexity [ACB14] and communication complexity [FAB15, AFAB16], but is still poorly understood. It would be interesting to understand which advantages could be obtained from the coherent superpositions of and common- and direct-cause processes.

Remark — In the final stages of completing this manuscript, a related work by MacLean et al. [MRSR17] appeared independently. The difference in the definitions of direct-cause processes between the two papers and its implications are discussed in Appendix D.7.

Acknowledgements

We thank Mateus Araújo, Fabio Costa, Flaminia Giacomini, Nikola Paunković, Jacques Pienaar and Katia Ried for useful discussions. We acknowledge support from the Austrian Science Fund (FWF) through the Special Research Programme FoQuS, the Doctoral Programme CoQuS, the project I-2526-N27 and the research platform TURIS. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

Chapter 7

Conclusion

In standard quantum information theory, quantum operations are embedded in a fixed, global causal order, which means they can be represented as being "plugged into" a quantum circuit. The process matrix formalism allows for the description of more general causal structures, including ways of composing quantum operations that cannot be represented as plugging the operations into a circuit. Certifying that a process is neither causally ordered nor a classical mixture thereof (not "causally separable") requires the parties to implement a suitable set of quantum operations and to examine their experimental outcomes, and potentially, additional assumptions regarding how the operations were actually implemented.

In this thesis, we explored questions related to three broad classes of certification of the causally nonseparable nature of processes. First, *device-independent* certification, where the conditional probability distribution resulting from the quantum operations alone (by violating a "causal inequality") is sufficient to prove that the underlying process is not causally separable, without any additional assumption. Second, *semi-device-independent* certification, where in addition to the probability distribution, an assumption on the dimension of the quantum systems operated is required to prove the causal nonseparability. Third, *device-dependent* certification, which requires the full knowledge of the applied quantum operations to prove nonseparability (by implementing a "causal witness").

In Chapter 2 we examined the conditions required for a process to produce (device-independent) violations of causal inequalities and showed that operations on classical bits and "classical process matrices" can produce maximal violations of causal inequalities for three or more parties. This shows that the two-party scenario—where genuinely quantum operations are required to violate causal inequalities—is an exception. In Chapter 3, we investigated the relationship between the ability to violate causal inequalities and the physical implementability of process matrices. By giving examples of causally nonseparable processes which cannot violate any causal inequalities and yet seem to lack any physical implementation, we provided evidence against the conjecture that all processes that do not violate causal inequalities are physically implementable.

In Chapter 4 we presented a task for the semi-device-independent certification of the causal nonseparability of the "quantum switch", a process known to be physically implementable. The certification relies on a simple three-party task, which the quantum switch can perform while using less communication than any causally separable process, by putting the direction of communication between two of the parties in coherent superposition. In Chapter 5, we extended this result by showing that the reduction in communication scales *exponentially* in the length of the inputs given to the parties, even when allowing for a bounded probability of failing. The result indicates that, beyond being of foundational interest, nonseparable processes can provide a significant quantum information processing advantage compared to causally ordered ones.

In Chapter 6 we applied the concept of device-dependent certification in a slightly different scenario, with an underlying fixed causal structure with three events A, B and C, where A is before B and C and B is before C. In this scenario, one can distinguish "common-cause" causal structures (where A and B share a common cause) from "direct-cause" causal structures (where A directly influences B). We showed that a quantum superposition of both falls into neither category by developing a framework of witnesses of "causal nonclassicality". We then described a thought experiment implementing such a process through a spatial superposition of a mass and general relativistic time dilation.

Many aspects of quantum causal structures are still poorly understood. To conclude, let me briefly mention three areas which seem particularly promising for future research. The first one is the study of multipartite quantum causal structures. It is plausible that novel types of causal nonseparability, along with corresponding foundational interpretations and applications, will appear when increasing the number of parties. In this regard, a framework device-independent certification of genuine multipartite causal nonseparability was recently developed [AWCB17], while the corresponding device-dependent and semi-device dependent definitions still have to be formulated. With these frameworks in place, it is intriguing to find out whether causality is a phenomenon that emerges when "coarse-graining" a typical process with a large number of parties into a process with fewer parties (by tracing out a subset of parties) and in what cases the causal nonseparability is unaffected by such a coarse-graining.

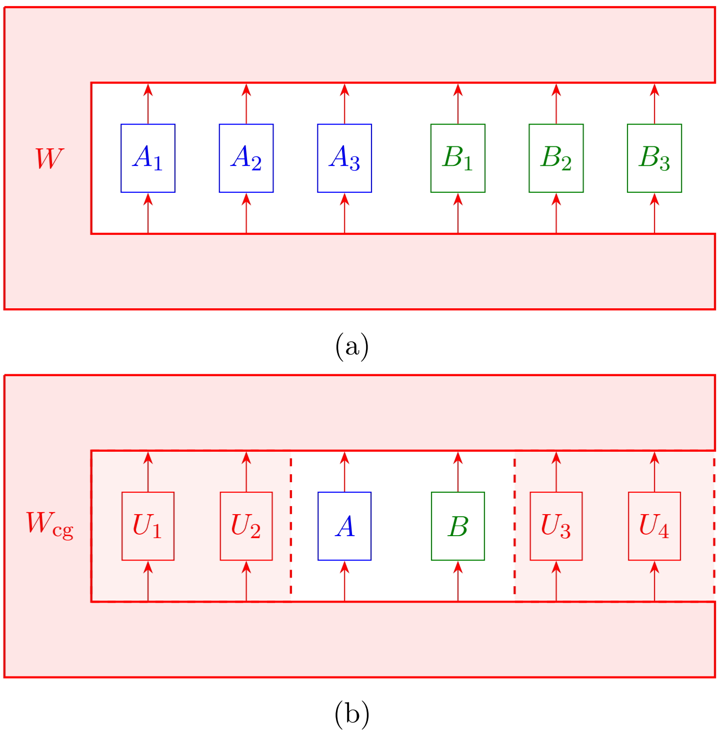

Figure 7.1: "Coarse-graining" of a six-party multipartite causally nonseparable process W to a bipartite process W_{cg}. The parties A_1, A_2 and B_2, B_3 implement unitaries and are traced out. Under which conditions can causal nonseparability survive in the remaining process W_{cg}?

Finding novel information-theoretic applications of process matrices, in particular of those that have a physical implementation, is another infant field of research. In this regard, recent work to characterise the computational power of process matrices [AGB17] and a surprising result showing that a quantum superposition of two fully depolarising channels allows for the transmission of information [ESC18] have made further progress in this direction. The main challenge would be to combine these different results to build a non-causal algorithm significantly outperforming any known quantum causal algorithm, for instance in runtime or memory.

From the foundational point of view, the most interesting open question concerns the physical implementability of process matrices. In this area, some progress was recently made by devising a suitable "purity" criterion for process matrices and proposing that all physical processes are those that can be purified [AFNB17]. This postulate is based on the assumption that reversibility is a basic physical principle, which non-purifiable processes would violate. Another way of approaching the question of physical implementability is to consider the *dynamics* of causal structures through transformations between process matrices. One reasonable conjecture is that only processes that can be reached through *reversible transformations* are physically meaningful, which allows to rule out a number of nontrivial processes [CRGB18].

Ultimately, it might be necessary to first qualify what is meant by physical implementability. The class of processes that can be realized in standard quantum experiments is of interest, because it is related to potential quantum information applications. On the other hand, the class of processes that might provide an effective description of quantum gravity phenomena is—a priori—broader. A recent paper provides a clearer picture of how a process and its experimental implementation are linked [Ore18], which is an important step for defining processes with a standard experimental implementation. Conversely, it would be interesting to try pinpointing the characteristics of process matrices that do not have such a standard quantum realization. If it were possible to conceive of a quantum gravity thought experiment represented by such an exotic process matrix, this would provide a signature for genuine post-quantum behaviour.

Appendix A

Causally nonseparable processes admitting a causal model

A.1 Analytic proofs for the nonseparabilities in Chapter 3

First, we prove that the class of processes, defined by

$$\begin{aligned} W^{A\prec B} :=& \mathbb{1}^\circ + \frac{1}{12}(\mathbb{1}ZZ\mathbb{1} + \mathbb{1}XX\mathbb{1} + \mathbb{1}YY\mathbb{1}), \\ W^{B\prec A} :=& \mathbb{1}^\circ + \frac{1}{4}(Z\mathbb{1}XZ), \\ W :=& qW^{A\prec B} + (1-q+\epsilon)W^{B\prec A} - \epsilon\mathbb{1}^\circ, \end{aligned} \tag{A.1}$$

has causal separability $R_r(W) = \epsilon$. To do so, we define a causal witness S_W for it:

$$S_W = \mathbb{1}^\circ - \frac{1}{4}(\mathbb{1}ZZ\mathbb{1} + \mathbb{1}XX\mathbb{1} + \mathbb{1}YY\mathbb{1}) - \frac{1}{4}(Z\mathbb{1}XZ). \tag{A.2}$$

We first verify that S_W is a causal witness: $\mathrm{tr}_{A_O} S_W \geq 0$ and $\mathrm{tr}_{B_O} S_W \geq 0$, which is a sufficient condition [ABC+15] for S to have positive trace with any causally separable process. Therefore, $\mathrm{tr}[W_{\mathrm{sep}} S_W] \geq 0$ and S_W is indeed a causal witness.

We compute $\mathrm{tr}[S_W(W + \lambda\mathbb{1}^\circ)] = -\epsilon + \lambda$, which is negative for $\lambda < \epsilon$ and implies that $R_r(W) \geq \epsilon$. From Eq. (A.1), it is clear that $W + \lambda\mathbb{1}^\circ$ is causally separable for $\lambda \geq \epsilon$, so $R_r(W) \leq \epsilon$. This establishes that $R_r(W) = \epsilon$.

Using the same approach, we can show that the process

$$W_{\mathrm{mix}}(\alpha) := \alpha W_{\mathrm{opt}} + (1-\alpha)W_{\mathrm{OCB}}, \tag{A.3}$$

where W_{opt} is defined in (3.15) and W_{OCB} in (3.16), has nonseparability $R_r(W_{\mathrm{mix}}(\alpha)) = \alpha R_r(W_{\mathrm{opt}}) + (1-\alpha)R_r(W_{\mathrm{OCB}}) = 1 + \alpha(\frac{4}{\sqrt{3}} - 3)$.

The convexity of random robustness [ABC+15] implies that the random robustness of a convex combination is smaller than the convex combination of the random robustnesses, so $R_r(W_{\mathrm{mix}}) \leq 1+\alpha(\frac{4}{\sqrt{3}}-3)$. Using the same witness S_W as before, we compute $\mathrm{tr}[S_W(W_{\mathrm{mix}}+\lambda\mathbb{1}^\circ)] = -1 - \alpha(\frac{4}{\sqrt{3}} - 3) + \lambda$ which is strictly negative when $\lambda > 1 + \alpha(\frac{4}{\sqrt{3}} - 3)$ and implies that $R_r(W_{\mathrm{mix}}) \geq 1 + \alpha(\frac{4}{\sqrt{3}} - 3)$. We conclude that $R_r(W_{\mathrm{mix}}) = 1 + \alpha(\frac{4}{\sqrt{3}} - 3)$.

The same proof (with the same witness S_W given in Eq. (A.2)) can also be used to show that the causal nonseparability of $W_{\mathrm{mix}}^{\mathrm{T}_B}$ is again the convex combination $R_r(W_{\mathrm{mix}}^{\mathrm{T}_B}) = \alpha R_r(W_{\mathrm{opt}}^{\mathrm{T}_B}) + (1-\alpha)R_r(W_{\mathrm{OCB}})$.

A.2 The dimension of the set causally separable processes

Here we show that the set of causally separable processes $\mathcal{W}_{\text{sep}}$ has the same dimension as the set of valid processes $\mathcal{W}$, which establishes that the set of causally separable processes has nonzero measure in the set of valid processes.

We will use the Hilbert-Schmidt decomposition of operators. An arbitrary process W can be decomposed as $W = \sum_{ijkl=0}^{3} \alpha_{ijkl} \sigma_i^{A_I} \otimes \sigma_j^{A_O} \otimes \sigma_k^{B_I} \otimes \sigma_l^{B_O}$.

The condition of normalization of probabilities, i.e., $\text{tr}[W \cdot M_A^{\text{CPTP}} \otimes M_B^{\text{CPTP}}] = 1$ for all CJ-representations of completely positive *trace-preserving maps* M_A^{CPTP} and M_B^{CPTP}, implies that some terms of the Hilbert-Schmidt decomposition, corresponding to "causal loops" are excluded. In particular, $\alpha_{0j00} = \alpha_{000l} = \alpha_{0j0l} = \alpha_{0jkl} = \alpha_{ij0l} = \alpha_{ijkl} = 0$ for $i, j, k, l \geq 1$ (see the Supplementary Material of Ref. [OCB12]).

Counting all the "allowed terms" in the Hilbert-Schmidt decomposition, we find that the dimension d_W of $\mathcal{W}$ is:

$$d_W = (1 + d_{A_I}^2(d_{A_O}^2 - 1))(d_{B_I}^2 - 1) + (d_{A_I}^2 - 1)d_{B_I}^2 d_{B_O}^2.$$

For causally ordered processes in $W^{A \prec B} \in \mathcal{W}^{A \prec B}$ compatible with the causal order $A \prec B$, some additional terms, which allow for signaling from Bob to Alice, are excluded in the Hilbert-Schmidt decomposition, reducing the dimension to

$$d_{W^{A \prec B}} = d_{A_I}^2(1 + (d_{B_I}^2 - 1)d_{A_O}^2) - 1.$$

This means that the set of causally ordered processes $\mathcal{W}^{A \prec B}$ has measure zero within the set of all process matrices.

Separable processes are convex combinations of $W^{A \prec B}$ and $W^{B \prec A}$. This means that all the terms allowed in the Hilbert-Schmidt decomposition of a valid process matrix are also allowed in the decomposition of separable processes. Therefore $\mathcal{W}$ and $\mathcal{W}_{\text{sep}}$ share the same basis and $d_W = d_{W_{\text{sep}}}$.

A.3 Generating uniformly distributed processes

We consider the space of process matrices $\mathcal{W}$ as being embedded in $\mathbb{R}^{d_W}$. We wish to obtain a *uniform sample* of $\mathcal{W}$ according to the d_W-dimensional volume (Lebesgue measure), which also corresponds to the measure generated by the Hilbert-Schmidt metric. We use an adaptation of the "hit-and-run" Markov chain sampler [Smi80, Smi84] for this task. The iteration works as follows:

Algorithm 1

1. Select a starting point W_0.

2. Choose a traceless matrix Q_{i+1} from a set of d_W orthogonal traceless matrices and generate a random sign variable $s = \pm 1$.

3. Find μ such that $W_i + \mu(\mathbb{1}^\circ + sQ_{i+1})$ is on the boundary of the set of valid processes.

4. Generate a random real scalar $\theta \in [0, \mu]$. Take $W_{i+1} = W_i + \theta(\mathbb{1}^\circ + sQ_{i+1})$ and go to step 2.

The set of directions is simply the Hilbert-Schmidt basis of allowed terms; it has dimension d_W (see Appendix A.2). For bipartite processes with $d_{A_I} = d_{B_I} = d_{A_O} = d_{B_O} = 2$, there are $d_W = 87$ possible directions to choose from.

Finding the intersection with the boundary of the set of positive processes in step 3 turns out to be a semidefinite program

$$\begin{aligned} &\max \mu \\ \text{s.t.} \quad &W_i + \mu(\mathbb{1}^\circ + sQ_{i+1}) \geq 0. \end{aligned} \tag{A.4}$$

However, the SDP which computes μ at each step of the Markov chain is a bottleneck of the algorithm. Instead, we can skip it and generate $\theta \in [0,1]$, rejecting and retrying if the resulting process is not positive:

Algorithm 2

1. Select a starting point W_0.
2. Choose a traceless matrix Q_{i+1} from a set of d_W orthogonal traceless matrices and generate a random sign variable $s = \pm 1$.
3. Generate a random real scalar $\theta \in [0,1]$. Take $W_{i+1} = W_i + \theta(\mathbb{1}^\circ + sQ_{i+1})$.
4. If $W_{i+1} \geq 0$, go to step 2, otherwise repeat step 3.

The matrices $\mathbb{1}^\circ + Q_i$ are chosen to be slightly *outside* the set $\mathcal{W}$ by having slightly negative eigenvalues. Therefore, there is always a finite probability of rejection at step 4, which guarantees that the algorithm samples uniformly all the way up to the boundary.

The resulting sample $\{W_i\}_{i=0}^{\infty}$ is uniform when two conditions hold [Smi84]. First, from every W_i, W' the probability to have $W_{i+d_W} = W'$ is nonzero, which is indeed true: In d_W steps, one can reach any W' starting from any W_i. Second, the uniform distribution is a stationary distribution of the Markov chain. This is also the case: for any W_i, W' the probability to reach W' starting from W_i in d_W steps is the same as the probability to reach W_i starting from W' in d_W steps.

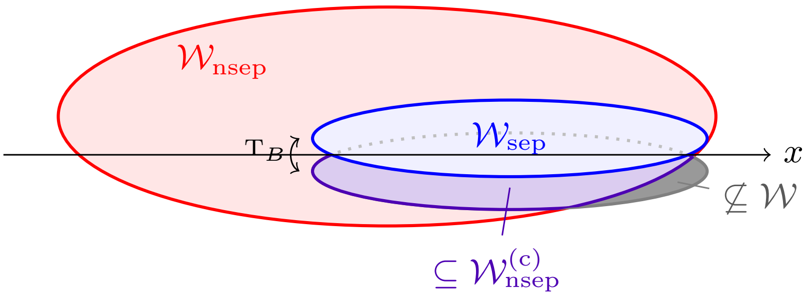

Figure A.1: Schematic two-dimensional cut of $\mathcal{W}$ in R^{d_W}, the partial transpose ^{T_B} here corresponds to a reflection along the horizontal axis x. The partial transpose of the set of causally separable processes $\mathcal{W}_{\text{sep}}$ consists of three parts: (i) causally separable matrices ($\subseteq \mathcal{W}_{\text{sep}}$), (ii) non-valid processes ($\not\subseteq \mathcal{W}$) and (iii) valid, causally nonseparable matrices with a causal model ($\subseteq \mathcal{W}_{\text{nsep}}^{(c)}$).

An upper bound on the *convergence* of the hit-and-run algorithm for convex sets (which is the case for the set $\mathcal{W}$) is known—in particular, the mixing time scales as $\tilde{O}(d_w^3) := O(d_w^3 \operatorname{polylog} d_w)$, which matches the best known mixing times for other algorithms [Lov99]. For $d_w = 87$ (which is the case for $d_{A_I} = d_{B_I} = d_{A_O} = d_{B_O} = 2$), we would need around $7 \cdot 10^7$ samples to achieve the same statistical significance as from a one-dimensional hit-and-run with 100 samples, which we deem sufficient for our purposes.

To sample uniformly distributed *causally separable* processes, we use the *rejection method*: after sampling $7 \cdot 10^7$ process matrices (with a warm-up period of 10^6 discarded steps), we randomly select 1000 causally separable processes (rejecting the $\approx 92.5\%$ of nonseparable ones) in the sample.

The map ${}^{\mathrm{T}_B}$ preserves the Lebesgue measure, since it corresponds to reflections in $\mathbb{R}^{d_W}$. Therefore, if upon applying ${}^{\mathrm{T}_B}$ to random causally separable matrices, there is a finite probability to obtain a *valid, causally nonseparable process*, this means that the set of causally nonseparable processes with a separable partial transpose is full dimensional (see Fig. A.1).

Appendix B

Proof of the causally separable quantum bound on the equality game

Here we prove the validity of the quantum bound $p^{\mathcal{Q}}_{\text{succ.}} \leq \frac{5}{6}$. We start with Eq. (4.9):

$$p^{\mathcal{Q}}_{\text{succ.}} \leq \frac{1}{9} \max_{\rho_x, \{\mathcal{B}_y\}, \{C_m\}} \left(\sum_x \operatorname{tr}\{C_1 \mathcal{B}_x(\rho_x)\} + \sum_{xy, x\neq y} \operatorname{tr}\{C_0 \mathcal{B}_y(\rho_x)\} \right). \tag{B.1}$$

We now use the fact that the POVM preceded by a CPTP map is still a POVM (the elements of which we will call B^0_x and B^1_x) which can be thought of as being applied by Bob and Charlie together. This allows us to drop the optimization over $\{C_m\}$:

$$p^{\mathcal{Q}}_{\text{succ.}} \leq \frac{1}{9} \max_{\rho_x, \{\mathcal{B}^{0,1}_y\}} \left(\sum_x \operatorname{tr}\{B^1_x \rho_x\} + \sum_{x,y,x\neq y} \operatorname{tr}\{B^0_y \rho_x\} \right). \tag{B.2}$$

Since $\operatorname{tr}\{B^0_x \rho\} = 1 - \operatorname{tr}\{B^1_x \rho\}, \forall x$ (the probabilities sum to one), we can rewrite (B.2) as

$$9 p^{\mathcal{Q}}_{\text{succ.}} \leq 6 + \max_{\rho_x, \{B^1_y\}} \sum_y \operatorname{tr}\left\{ B^1_y \left(\rho_y - \sum_{x, x\neq y} \rho_x \right) \right\}. \tag{B.3}$$

We notice that each optimization over B^1_y is independent; similarly to the one for optimal state distinguishability [NC00], we find that the optimal POVM elements B^1_y are projectors on the positive eigenvalue subspace of $\rho_y - \sum_{x,x\neq y} \rho_x$. Using the Bloch vector decomposition $\rho_x = (\mathbb{1} + \boldsymbol{\sigma} \cdot \mathbf{a}_x)/2$, this leads to the result:

$$p^{\mathcal{Q}}_{\text{succ.}} \leq \frac{1}{2} + \frac{1}{18} \max_{\|\mathbf{a}_x\|_2 \leq 1, \mathbf{a}_x \in \mathbb{R}^3} (\|\mathbf{a}_0 - \mathbf{a}_1 - \mathbf{a}_2\|_2 + \|\mathbf{a}_1 - \mathbf{a}_0 - \mathbf{a}_2\|_2 + \|\mathbf{a}_2 - \mathbf{a}_0 - \mathbf{a}_1\|_2). \tag{B.4}$$

Choosing $\mathbf{a}_0 = (1, 0, 0)^{\mathrm{T}}$, parametrizing $\mathbf{a}_1, \mathbf{a}_2$ using spherical coordinates, and optimizing (B.4), the analytical maximum turns out to be

$$p^{\mathcal{Q}}_{\text{succ.}} \leq \frac{5}{6},$$

with the optimal preparation and measurement strategies given in Eq. (4.11). Charlie measures in $|x\pm\rangle$-basis; the channels of Bob are explicitly given by

$$\begin{aligned}
\mathcal{B}_0(\rho) &= \Pi_0^+ \rho \Pi_0^+ + \Pi_0^- \rho \Pi_0^- , \\
\mathcal{B}_1(\rho) &= U_1 \Pi_1^+ \rho \Pi_1^+ U_1^\dagger + U_1 \Pi_1^- \rho \Pi_1^- U_1^\dagger, \\
\mathcal{B}_2(\rho) &= U_2 \Pi_2^+ \rho \Pi_2^+ U_2^\dagger + U_2 \Pi_2^- \rho \Pi_2^- U_2^\dagger,
\end{aligned} \tag{B.5}$$

where $\Pi_0^+ = |a_0\rangle\langle a_0|$, $\Pi_1^+ = |a_1\rangle\langle a_1|$, $\Pi_2^+ = |a_2\rangle\langle a_2|$ and the corresponding $\Pi_{0,1,2}^- = \mathbb{1} - \Pi_{0,1,2}^+$. The unitaries $U_{1,2}$ correspond to a basis transformation such that $U_{1,2} |a_{1,2}\rangle = |a_0\rangle$.

Appendix C

VC-dimension bounds on the bounded error one-way quantum communication complexity

In this section we show that if the protocol allows for some error probability, bounded by ϵ for all inputs, the one-way communication complexity of EE_n still scales as $\Omega(2^n)$. As in Fig. 5.1, we assume that Alice and Bob share unlimited prior entanglement, and that Alice sends a quantum state to Bob. We note that under the promise that Bob's input function is the zero function $g = 0$, the Exchange Evaluation game reduces to a random access code [ANTSV99], for which optimal bounds on the bounded error communication complexity are known [Nay99]. However, it is more straightforward to apply a bound that uses the concept of VC-dimension [VC15].

Definition *VC-dimension.* Let $f : X \times Y \to \{0,1\}$. A subset $S \subseteq Y$ is shattered, if $\forall R \subseteq S, \exists x \in X$ such that

$$f(x,y) = \begin{cases} 1, & \text{if } \mathbf{y} \in R. \\ 0, & \text{if } \mathbf{y} \in S \setminus R. \end{cases} \tag{C.1}$$

The VC-dimension VC(f) is the size of the largest shattered subset of Y.

Given a function $f(x,y)$, we denote by $Q^1_\epsilon(f)$ the one-way (from Alice to Bob) bounded error quantum communication, where ϵ is the allowed worst-case error, and arbitrary prior shared entanglement is available. We make use of a theorem by Klauck (Theorem 3 of [Kla00]) that relates the bounded error quantum communication complexity of a function to its VC-dimension.

Theorem 3 *For all functions* $f : X \times Y \to \{0,1\}$, $Q^1_\epsilon(f) \geq \frac{1}{2}(1 - H(\epsilon))\,VC(f)$, *where* $H(\epsilon)$ *is the binary entropy* $H(\epsilon) = \epsilon \log(\epsilon) + (1-\epsilon)\log(1-\epsilon)$

Let us bound the VC-dimension of $EE_n : X \times Y \to \{0,1\}$, where $X = Y = \mathbb{Z}_2^n \times F_n$, by showing that $S = \{(\mathbf{y}, g) | g = 0, \mathbf{y} \neq \mathbf{0}\} \subset Y$ is shattered. This is clear, since for any $R \subseteq S$, there exists the indicator function

$$f_R(\mathbf{y}) = \begin{cases} 1, & \text{if } (\mathbf{y}, 0) \in R. \\ 0, & \text{otherwise,} \end{cases} \tag{C.2}$$

so that S is shattered.

Therefore $\mathrm{VC}(EE_n) \geq |S| = 2^{n-1}$, and Theorem 3 implies that the one-way quantum communication complexity $Q^1_\epsilon(EE_n) \geq (1 - H(\epsilon))2^{n-2}$. This establishes that the number of communicated qubits scales exponentially with n even in the bounded error case, so that the exponential separation between the quantum switch and one-way quantum communication, established in Chapter 5 for the deterministic case, continues to hold.

Appendix D

Quantum superposition of causal structures

D.1 Choi-Jamiołkowski isomorphism

The Choi-Jamiołkowski (CJ) representation of a CP map $\mathcal{M}_A : A_I \to A_O$ is

$$M_A := \left[(\mathcal{I} \otimes \mathcal{M}_A)(|I\rangle\rangle\langle\langle I|)\right]^{\mathrm{T}} \in A_I \otimes A_O, \tag{D.1}$$

where $\mathcal{I}$ is the identity map, $|I\rangle\rangle := \sum_{j=1}^{d_{\mathcal{H}_I}} |jj\rangle \in \mathcal{H}_I \otimes \mathcal{H}_I$ is a non-normalized maximally entangled state and $^{\mathrm{T}}$ denotes matrix transposition in the computational basis.

The inverse transformation is then defined as:

$$\mathcal{M}_A(\rho) = \mathrm{tr}_I[(\rho \otimes \mathbb{1})M_A]^{\mathrm{T}}. \tag{D.2}$$

For operations which have a single Kraus operator ($\mathcal{M}_A(\rho) = A\rho A^\dagger$), one also define a "pure CJ-isomorphism" [Roy91, BDMS00], which maps the operation to a *vector*[1]:

$$|A^*\rangle\rangle := (\mathbb{1} \otimes A^*)\,|I\rangle\rangle \in \mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O} \tag{D.3}$$

The usual CJ-representation of such an operation is simply the *projector onto the CJ-vector*: $M_A = |A^*\rangle\rangle\langle\langle A^*|$.

D.2 Causally ordered and common-cause process matrices

We first introduce a shorthand that we will use throughout the following appendices:

$${}_X W := \frac{\mathbb{1}^X}{d_X} \otimes \mathrm{tr}_X\, W, \tag{D.4}$$

where d_X is the dimension of the Hilbert space X.

In this paper, we consider three parties, where the C's output space C_O can be disregarded. The process matrix $W \in A_I \otimes A_O \otimes B_I \otimes B_O \otimes C_I$, which encodes the quantum causal model, is defined on the dual space to the tensor products of the maps. Since both the "common-cause" and the "direct-cause" scenarios are compatible with the causal order $A \prec B \prec C$, we can also represent the process matrix W as a circuit. (see Fig. D.1).

For instance, the coherent superposition of common cause and direct cause, defined in (6.7), would consist of $|\psi\rangle = |\phi^+\rangle \otimes (|0\rangle + |1\rangle)/\sqrt{2}$, W_1 and W_2 being control-SWAPs (where the control is the last qubit, initially in the state $(|0\rangle + |1\rangle)/\sqrt{2}$).

[1] Note that there are differing conventions, where the conjugation is omitted.

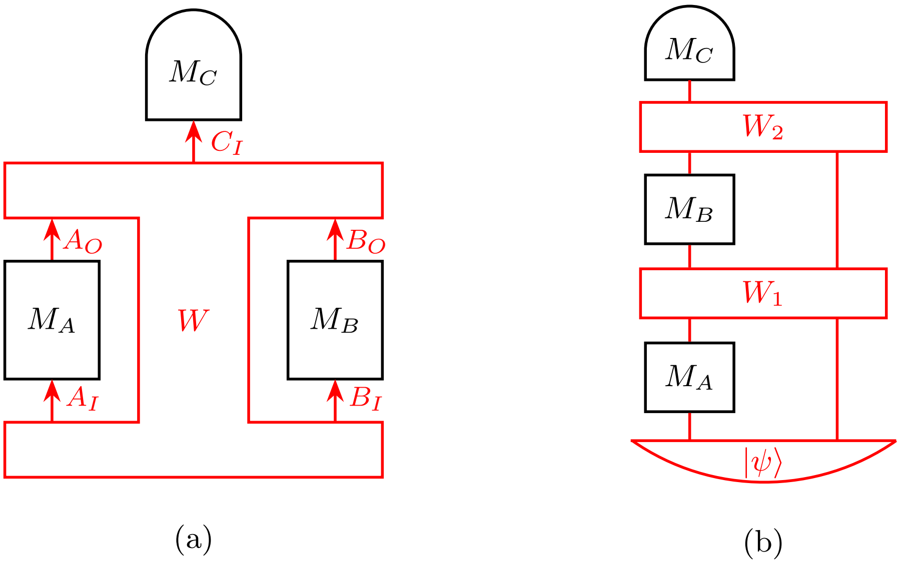

Figure D.1: (a) General three-party process matrix $W \in A_I \otimes A_O \otimes B_I \otimes B_O \otimes C_I$. (b) Since, in our scenarios, W is compatible with the causal order $A \prec B \prec C$, we can also represent W as a "causal network", which can be implemented as a quantum circuit ($|\psi\rangle$ is a state, W_1 and W_2 CPTP maps; lines can represent quantum systems of different dimensions.)

We now define the projection $\mathcal{L}_{A\prec B\prec C}(\cdot)$ onto the linear subspace of process matrices compatible with the causal order $A \prec B \prec C$, which can be derived from the conditions given in Ref. [ABC+15]:

$$\mathcal{L}_{A\prec B\prec C}(W) := W -_{C_I} W +_{B_O C_I} W -_{B_I B_O C_I} W +_{A_O B_I B_O C_I} W. \tag{D.5}$$

$W^{A\prec B\prec C}$ is compatible with the causal order $A \prec B \prec C$ if and only if $W^{A\prec B\prec C} = \mathcal{L}_{A\prec B\prec C}(W^{A\prec B\prec C})$ holds.

The projection onto the subspace of common-cause process matrices $\mathcal{L}_{\mathrm{cc}}(\cdot)$ is given by composing the projection $\mathcal{L}_{A\prec B\prec C}$ with the projection onto processes which have no channel from A_O to B_I:

$$\mathcal{L}_{\mathrm{cc}}(W) := \mathcal{L}_{A\prec B\prec C}(W) -_{C_I} \mathcal{L}_{A\prec B\prec C}(W) +_{C_I A_O} \mathcal{L}_{A\prec B\prec C}(W). \tag{D.6}$$

D.3 Dual cones

Given the definition (6.2) of the cone $\mathcal{W}$, we can characterize the *dual cone* $\mathcal{W}^*$ of all operators whose product with operators in $\mathcal{W}$ has positive trace. Remember that $\mathcal{W}$ is the *intersection* of the cone of positive operators $\mathcal{P}$ with a linear subspace defined by the conditions for causal order: $\mathcal{W} := \mathcal{P} \cap \mathcal{L}_{A\prec B\prec C}$.

The dual of the linear subspace $\mathcal{L}^*_{A\prec B\prec C}$ is its orthogonal complement [NN87, ABC+15]

$$\mathcal{L}^*_{A\prec B\prec C} = \mathcal{L}^\perp_{A\prec B\prec C}, \tag{D.7}$$

i.e., the space of operators with a support that is orthogonal to the original subspace.

Additionally, the dual of the intersection of two closed convex cones containing the origin is the convex union of their duals [NN87, ABC+15], so that

$$\mathcal{W}^* = (\mathcal{P} \cap \mathcal{L}_{A\prec B\prec C})^* = \mathrm{conv}(\mathcal{P}^* \cup \mathcal{L}^\perp_{A\prec B\prec C}). \tag{D.8}$$

Since the cone of positive operators is self-adjoint ($\mathcal{P}^* = \mathcal{P}$), we can combine (D.7) and (D.8) into $\mathcal{W}^* = \text{conv}(\mathcal{P} \cup \mathcal{L}^{\perp}_{A \prec B \prec C})$. Explicitly, this means that any operator $Q \in \mathcal{W}^*$ can be decomposed as

$$\begin{gathered} Q = Q_1 + Q_2 \\ \text{s.t. } Q_1 \geq 0, \quad \mathcal{L}_{A \prec B \prec C}(Q_2) = 0. \end{gathered} \tag{D.9}$$

D.4 Nonclassicality of causality witnesses

We will now explicitly construct the set of nonclassicality of causality witnesses $\mathcal{S}_{\text{SDP}}$.

The semidefinite relaxation of the direct-cause constraint (6.4) in terms of positive partial transposition is (using the shorthand introduced in (D.4)):

$$({}_{C_I B_O} W^{\text{dc}})^{\text{T}_{A_I}} \geq 0. \tag{D.10}$$

The dual cone (D.11) to the cone of relaxed direct-cause processes defined by the intersection of $\mathcal{W}$ with the cone defined in (D.10) and the dual cone (D.12) to the cone of common-cause processes defined by the intersection of $\mathcal{W}$ with the linear subspace (6.5) can be constructed in the same way as in Appendix D.3.

The set of witnesses positive on all positive partial transpose operators is a *subset* of entanglement witnesses. Every witness belonging to this set satisfies[2]:

$$\begin{gathered} S^{\text{dc}} = {}_{C_I B_O} (S_1^{\text{T}_{A_I}}) + S_2 + S_3 \\ \text{s.t. } S_1, S_2 \geq 0, \quad \mathcal{L}_{A \prec B \prec C}(S_3) = 0. \end{gathered} \tag{D.11}$$

If $\text{tr}[S^{\text{dc}}\, W] < 0$, this implies that W is not a direct-cause process as defined in Eq. (6.4). Note that since we are only considering a subset of entanglement witnesses, *the converse does not hold.*

We can now turn to the requirement that S is positive on common-cause processes. Since condition (6.5) (corresponding to (D.6) together with positivity) defines a convex cone, we can use the techniques of Appendix D.3 to construct the dual cone, of which the witness will be an element. This leads us to write S as

$$\begin{gathered} S^{\text{cc}} = S_4 + S_5 \\ \text{s.t. } S_4 \geq 0, \quad \mathcal{L}_{\text{cc}}(S_5) = 0, \end{gathered} \tag{D.12}$$

where the projection onto the common-cause subspace $\mathcal{L}_{\text{cc}}$ is defined in Appendix D.2. W is *not* a common-cause process as defined in (6.4) if and only if there exists an S^{cc} such that $\text{tr}[S^{\text{cc}}\, W] < 0$.

Now, combining both conditions, we can construct a set of operators positive on all mixtures of direct-cause and common-cause processes *only in terms of semidefinite constraints.* To test whether an arbitrary W process is of this type, we can run the following semidefinite program (SDP) [NN87]:

$$\begin{aligned} &\min \text{tr}[S\, W] \\ \text{s.t.} \quad & S = {}_{C_I B_O} (S_1^{\text{T}_{A_I}}) + S_2 + S_3 = S_4 + S_5, \\ & S_1 \geq 0, \quad S_2 \geq 0, \quad S_4 \geq 0, \\ & \mathcal{L}_{A \prec B \prec C}(S_3) = \mathcal{L}_{\text{cc}}(S_5) = 0, \\ & \mathbb{1}/d_O - S \in \mathcal{W}^*. \end{aligned} \tag{D.13}$$

[2] We included the term S_2 and S_3 although they do not make the witnesses "more powerful" to detect entanglement. S_2 will become relevant when combining the conditions on direct-cause and common-cause processes in Eq. (D.13); S_3 is included because it could appear in restricted types of witnesses [ABC+15].

The last condition, where $\mathcal{W}^*$ is the cone dual to $\mathcal{W}$ (see Appendix D.3), imposes a normalization on S. It gives the nonclassicality of causality $\mathcal{C}(W) = -\operatorname{tr}[S_{\text{opt}}W]$ the operational meaning of "generalized robustness", quantifying resistance of the nonclassicality detectable by $\mathcal{S}_{\text{SDP}}$ to *worst possible noise* [Ste03, ABC+15]. This becomes more intuitive from the dual SDP, given by

$$\begin{aligned} &\min \operatorname{tr}[\Omega/d_O] \\ \text{s.t.} \quad & W + \Omega = W^{\text{cc}} + W^{\text{dc}}, \\ &({}_{C_I B_O}W^{\text{dc}})^{\mathrm{T}_{A_I}} \geq 0, \quad W^{\text{dc}} \in \mathcal{W}, \\ &{}_{C_I}W^{\text{cc}} = {}_{C_I A_O} W^{\text{cc}}, \quad W^{\text{cc}} \in \mathcal{W}. \end{aligned} \tag{D.14}$$

The process $\Omega \cdot d_O/\operatorname{tr}[\Omega]$ can be interpreted as worst-case noise with respect to the optimal witness S_{opt}, resulting from the SDP (D.13).

D.5 Convexity and monotonicity

Here we prove that the *nonclassicality of causality* defined as $\mathcal{C}(W) := -\operatorname{tr}[S_{\text{opt}}W]$, which results from the SDP (D.13), satisfies the natural properties of *convexity* and *monotonicity*, following analogous proofs of Ref. [ABC+15].

Convexity means that $\mathcal{C}(\sum_i p_i W_i) \leq \sum_i p_i \mathcal{C}(W_i)$, for any $p_i \geq 0, \sum_i p_i = 1$. Take S_{W_i} to be the optimal witness for W_i. Any other witness, in particular the optimal witness S_W for $W := \sum_i p_i W_i$ will be less robust to noise with respect to W_i:

$$\operatorname{tr}[S_{W_i}\, W_i] \leq \operatorname{tr}[S_W\, W_i]. \tag{D.15}$$

Averaging over i we have

$$-\operatorname{tr}\left[S_W \sum_i p_i W_i\right] \leq -\sum_i p_i \operatorname{tr}[S_{W_i} W_i], \tag{D.16}$$

which is exactly the statement of convexity for $\mathcal{C}$.

Monotonicity under local operation means that $\mathcal{C}(W) \geq \mathcal{C}(\$(W))$, where $\$(\cdot)$ is the composition of W with local operations.

We wish to show that $-\operatorname{tr}\big[S_{\$(W)}\$(W)\big] \leq -\operatorname{tr}[S_W W]$. By duality, this is equivalent to

$$-\operatorname{tr}\big[\$^*\big(S_{\$(W)}\big)W\big] \leq -\operatorname{tr}[S_W\, W], \tag{D.17}$$

where $\$^*(\cdot)$ is the map dual to $\$(\cdot)$. Eq. (D.17) is satisfied if $\$^*\big(S_{\$(W)}\big)$ is a witness, i.e., is positive on all mixtures of direct-cause and common-cause operators ($\operatorname{tr}\big[\$^*\big(S_{\$(W)}\big)W^{\text{mix}}\big] \geq 0$), and is normalized appropriately ($1/d_O - \$^*\big(S_{\$(W)}\big) \in \mathcal{W}^*$).

The first condition can be seen to hold by applying duality and using the fact that local operations map any mixture of direct-cause and common-cause processes to a mixture of direct-cause and common-cause processes. The second condition is equivalent to

$$\operatorname{tr}\big[\mathbb{1}/d_O - \$^*\big(S_{\$(W)}\big)\Omega\big] \geq 0 \tag{D.18}$$

for every process matrix Ω. We apply duality and linearity of the trace to find that

$$\operatorname{tr}\big[S_{\$(W)}\$(\Omega)\big] \leq \operatorname{tr}[\Omega]/d_O. \tag{D.19}$$

This relation holds because $\$(\cdot)$ maps normalized ordered process matrices to normalized ordered process matrices and $\mathbb{1}/d_O - S_{\$(W)} \in \mathcal{W}^*$ is the normalization condition for the SDP (D.13).

The condition of *discrimination* (or *faithfulness*), which would mean that $\mathcal{C}(W) \geq 0$ *if and only if* the process matrix is not a mixture of direct-cause and common-cause processes (6.6), is *not satisfied*. Since we relied on a relaxation of the direct-cause condition by using the positive partial transpose criterion, there are processes which are not a mixture satisfying (6.6) but for which the nonclassicality of causality is zero.

Therefore, the nonclassicality of causality is not a *faithful measure* of the nonclassicality of the causal structure. This is reasonable, since finding such a measure would be equivalent to finding a fully general *entanglement criterion*—a problem known to be computationally hard [Gur03].

D.6 Experimental constraints on witnesses

In this appendix, we give the explicit form of the experimental constraints mentioned in the main text. When using a constrained class of witnesses, the value $-\operatorname{tr}[S_{\text{opt}}^{\text{restricted}} W_{\text{coherent}}]$. can be interpreted as the *amount of noise* tolerated before the *constrained set of witnesses* becomes incapable of detecting the nonclassicality of causality of W_{coherent}.

A simple example of a restriction simplifying the experimental implementation consists in disregarding the space $C_I^{(1,2)}$, i.e., to have $S =_{C_I^{(1,2)}} S$ as an additional constraint. The nonclassicality of causality is *unaffected* by this restriction, which shows that the input spaces $C_I^{(1,2)}$ do not carry any additional information about the nonclassicality of causality.

The constraint for the witness to consist only of non-demolition measurements is:

$$S^{\text{ndmeas}} = \sum_{ijl} \alpha_{ijl} (\mathbb{1} + \sigma_i^{A_I}) \otimes (\mathbb{1} + \sigma_i^{A_O}) \otimes (\mathbb{1} + \sigma_j^{B_I}) \otimes (\mathbb{1} + \sigma_j^{B_O}) \otimes E_l^{C_I}, \tag{D.20}$$

where σ_k $(k = 1, 2, 3)$ are the qubit Pauli matrices and E_l, $l = 1, \ldots, 8$ is an arbitrary basis of projectors on C_I's three qubits.

The constraint for the witness to only consist of unitary operations[3] for A and B is:

$$S^{\text{unitary}} = \sum_{ijl} \beta_{ijl} \, |U_i^*\rangle\rangle\langle\langle U_i^*|^{A_I A_O} \otimes |U_j^*\rangle\rangle\langle\langle U_j^*|^{B_I B_O} \otimes E_l^{C_I}, \tag{D.21}$$

where $i, j = 1, \ldots, 10$ indexes a basis[4] of the CJ-vectors (see Appendix D.1) of unitaries.

Table D.1: Constrained nonclassicality of causality for different types of constraints on S, in descending order.

Constraint on the witness S	$-\operatorname{tr}[S\, W^{\text{coherent}}]$
No constraint	0.2278
Discarding $C_I^{(1,2)}$	0.2278
Unitary operations A, B	0.1686
ND measurement A, B	0.0732

D.7 Definition of direct-cause processes and relationship to the definitions of Ref. [MRSR17]

Since Ref. [MRSR17] considers two party case, we can merge B and C to make our scenario comparable to the one of Ref. [MRSR17]. More precisely, B_I and C_I are relabeled as B_I'

[3] Note that according to definition of Ref. [RAV+15], unitary witnesses should also be considered as "observations" although operationally they are standardly understood as interventions.

[4] This is because there are ten linearly independent projectors on CJ-vectors for unitaries acting on qubits [ABC+15].

and B_O is disregarded, eliminating the necessity to trace over B_O and C_I. The condition for direct-cause processes (6.4) then becomes

$$W^{\text{dc}} = \sum_i p_i \rho_i^{A_I} \otimes \tilde{W}_i^{A_O B_I'}, \tag{D.22}$$

which implies that the states given to A and the channel connecting A and B can be *classically correlated*.

In the terminology of DAGs this convex mixture would correspond to tracing over a (hidden) classical[5] common cause between A and B. An alternative, more restricted definition would exclude such classical correlations, i.e.,

$$W^{\text{dc}} = \rho^{A_I} \otimes \tilde{W}^{A_O B_I'}. \tag{D.23}$$

It is used in Ref. [MRSR17]. To make the difference apparent, consider the convex mixture of two direct-cause processes between A and B (here, $\dim A_I = \dim A_O = \dim B_I' = 2$):

$$W^{\text{mem}} = \frac{1}{4} |0\rangle \langle 0|^{A_I} (\mathbb{1}^{A_O B_I'} + \sigma_z^{A_O} \sigma_z^{B_I'}) + \frac{1}{4} |1\rangle \langle 1|^{A_I} (\mathbb{1}^{A_O B_I'} - \sigma_z^{A_O} \sigma_z^{B_I'}), \tag{D.24}$$

where the tensor products between the Hilbert spaces are implicit. W^{mem} classically correlates the channel between A_O and B_I' (a classical channel with or a without bit flip) to the state in A_I, as shown in Fig. D.2. It is of the type (D.22) but *not* of the type (D.23).

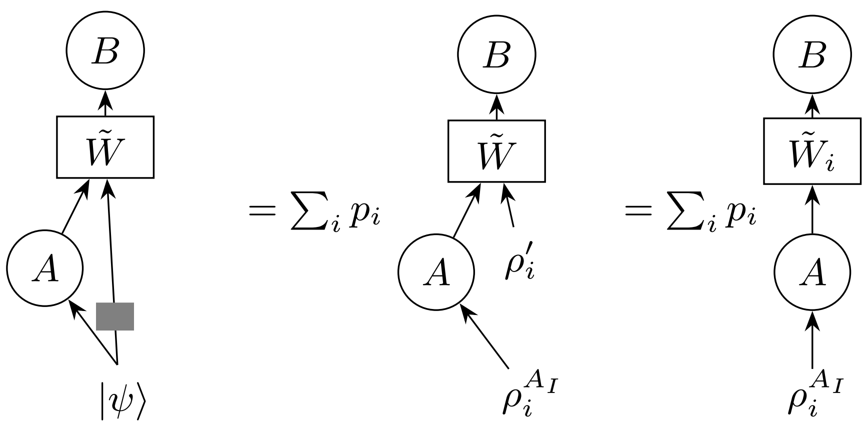

Figure D.2: Quantum causal models respecting the extended "direct-cause" condition (D.22) can be thought of as a general channel with *classical* memory (left), or equivalently as a convex combination of direct-cause processes with no memory (right). $\tilde{W}$ and $\tilde{W}_i$ are general quantum channels, $|\psi\rangle$ an arbitrary quantum state and the gray square represents a fully dephazing channel (in an arbitrary basis).

In Ref. [MRSR17], (D.24) is not considered to be a direct-cause process, nor a convex mixture (called "probabilistic mixture") of direct-cause and common-cause processes. It is instead termed a "physical mixture" of common-cause and direct-cause processes.

We instead use the broader definition (D.22) because we ultimately intend to study convex combinations of common-cause and direct-cause processes (6.6), which means we should also allow for convex combinations of direct-cause processes. The restricted definition (D.23) for direct-cause processes would lead to consider a convex combination of a

[5]Strictly speaking, it just needs not to produce any entanglement between A_I and (A_O, B_I), see Fig. D.2.

direct-cause and a common-cause process to be a "probabilistic mixture", but *not* a convex combination of *two cause-effect processes*.

Finally note that the class of processes, which, when post-selected on CP maps being implemented at B_I', result in an entangled conditional process on $A_I A_O$, is defined to be "coherent mixtures" in Ref. [MRSR17]. All of these "coherent mixtures" are nonclassical in our terminology (the processes that can be decomposed as (6.6) never result in an entangled conditional process on $A_I A_O$). It is not clear whether the converse is true.

D.8 Issues in defining a quantum "observational scheme"

Ried et al. [RAV$^+$15] define the "observational scheme" (as opposed to the "interventionist scheme") on a quantum causal structure as composed of operations satisfying the "informational symmetry principle". We examine the subtleties and issues involved in this definition, in particular regarding the dependence on the initially assigned state.

Ref. [RAV$^+$15] assumes that before the observation, one assigns the (epistemic) state ρ_{A_I} to the system coming into A's laboratory. A quantum operation (described by the Choi-Jamiołkowski representation of the quantum instrument [DL70] $\{M_A^i\}$, where i labels the outcome) is applied. This updates the information about the outgoing state $\rho_{A_O}^{(i)}$ *but also* (through retrodiction) about the incoming state $\rho_{A_I}^{(i)}$. These states are found by applying the update rules [LS13]:

$$\rho_{A_O}^{(i)} = \frac{\mathrm{tr}_{A_I}[M_A^i \cdot \rho_{A_I} \otimes \mathbb{1}_{A_O}]^{\mathrm{T}}}{\mathrm{tr}[M_A^i \cdot \rho_{A_I} \otimes \mathbb{1}_{A_O}]}, \tag{D.25}$$

$$\rho_{A_I}^{(i)} = \frac{\mathrm{tr}_{A_O}\left[(\sqrt{\rho_{A_I}} \otimes \mathbb{1}_{A_O}) M_A^i (\sqrt{\rho_{A_I}} \otimes \mathbb{1}_{A_O})\right]}{\mathrm{tr}\left[(\sqrt{\rho_{A_I}} \otimes \mathbb{1}_{A_O}) M_A^i (\sqrt{\rho_{A_I}} \otimes \mathbb{1}_{A_O})\right]}. \tag{D.26}$$

The informational symmetry principle holds if and only if after the operation, the states assigned to the incoming and outgoing systems are the same:

$$\rho_{A_I}^{(i)} = \rho_{A_O}^{(i)}. \tag{D.27}$$

For Ried et al., an instrument for which this informational symmetry holds is *defined* to be an "observation" [RAV$^+$15]. In this sense, there can obviously be "non-passive" observations such as non-demolition measurements. Any non-demolition measurement in a basis in which the initially assigned state ρ_{A_I} is *diagonal* will be an observation in this sense. This matches the intuition that a *classical* measurement only *reveals* information and does not disturb the system.

If one wishes to implement measurements in *arbitrary bases*, the *only* initially assigned state which results in informational symmetry is the maximally mixed state $\rho_{A_I} = \mathbb{1}/d$ [RAV$^+$15]. This shows how problematic the definition of observational scheme is, since it not only crucially depends on an initial (epistemic) assignment ρ_{A_I} but also because there is *only one* such assignment which allows all measurements to be "observations"—which tolerates no amount and no type of noise. In this sense, as soon as the experimenter *changes her beliefs* about the incoming state *in any way*, she will be intervening on the system, not merely observing it.

Leaving aside these interpretative difficulties, it is interesting to realize that operations which are *unitary* also turn out to be "observations" if the initially assigned state is $\rho_{A_I} = \mathbb{1}/d$: for a unitary operation, $\rho_{A_I}^{(i)} = \rho_{A_O}^{(i)} = \rho_{A_I} = \mathbb{1}/d$. The unitary provides exactly the same information about input and output states, namely *none*.

Finally, note that both the framework of Ref. [RAV$^+$15] and the one we developed rely on the assumption that quantum theory is valid and the correct operations were implemented—the analysis is *device-dependent*. This means that any "quantum advantage" in inference will not be based on *mere correlations* in the sense of a conditional probability distribution of outputs given inputs. This makes the comparison with the power of classical causal models somewhat problematic.

Bibliography

[ABC+15] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, and Č. Brukner. "Witnessing causal nonseparability". *New J. Phys.*, **17**(10):102001, 2015.

[ABG+07] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani. "Device-Independent Security of Quantum Cryptography against Collective Attacks". *Phys. Rev. Lett.*, **98**(23):230501, 2007.

[ACB14] M. Araújo, F. Costa, and Č. Brukner. "Computational advantage from quantum-controlled ordering of gates". *Phys. Rev. Lett.*, **113**(25):250402, 2014.

[AFAB16] 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**(10):100502, 2016.

[AFCB14] M. Araújo, A. Feix, F. Costa, and Č. Brukner. "Quantum circuits cannot control unknown operations". *New J. Phys.*, **16**(9):093026, 2014.

[AFNB17] M. Araújo, A. Feix, M. Navascués, and Č. Brukner. "A purification postulate for quantum mechanics with indefinite causal order". *Quantum*, **1**:10, 2017.

[AGB17] M. Araújo, P. A. Guérin, and A. Baumeler. "Quantum computation with indefinite causal structures". *Phys. Rev. A*, **96**:052315, Nov 2017.

[ALMO08] A. Ambainis, D. Leung, L. Mancinska, and M. Ozols. "Quantum random access codes with shared randomness". `arXiv:0810.2937 [quant-ph]`, 2008.

[ANTSV99] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. "Dense quantum coding and a lower bound for 1-way quantum automata". STOC '99, pp. 376–383. ACM, 1999.

[ANTSV02] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. "Dense quantum coding and quantum finite automata". *J. ACM*, **49**(4):496–511, 2002.

[Ari33] Aristotle. *Metaphysics.* William Heinemann Ltd, London, 1933.

[Asp82] A. Aspect. "Experimental test of Bell's inequalities using time-varying analyzers". *Phys. Rev. Lett.*, **49**(25):1804–1807, 1982.

[Asp99] A. Aspect. "Bell's inequality test: more ideal than ever". *Nature*, **398**:189–190, 1999.

[AW09] S. Aaronson and J. Watrous. "Closed timelike curves make quantum and classical computing equivalent". *Proc. R. Soc. Lond. A*, **465**(2102):631–647, 2009.

[AWCB17] A. A. Abbott, J. Wechs, F. Costa, and C. Branciard. "Genuinely multipartite noncausality". *Quantum*, **1**:39, 2017.

[BAF+16] C. Branciard, M. Araújo, A. Feix, F. Costa, and Č. Brukner. "The simplest causal inequalities and their violation". *New J. Phys.*, **18**(1):013008, 2016.

[BB16] V. Baumann and i. c. v. Brukner. "Appearance of causality in process matrices when performing fixed-basis measurements for two parties". *Phys. Rev. A*, **93**:062324, Jun 2016.

[BBLG13] T. J. Barnea, J.-D. Bancal, Y.-C. Liang, and N. Gisin. "Tripartite quantum state violating the hidden-influence constraints". *Phys. Rev. A*, **88**(2), 2013.

[BCD+10] A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti. "Optimal quantum learning of a unitary transformation". *Phys. Rev. A*, **81**(3):032324, 2010.

[BCDP11] A. Bisio, G. Chiribella, G. D'Ariano, and P. Perinotti. "Quantum Networks: General Theory and Applications". *Acta Physica Slovaca.*, **61**(3), January 2011.

[BCMd10] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf. "Nonlocality and communication complexity". *Rev. Mod. Phys.*, **82**(1):665–698, 2010.

[BCP+13] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. "Bell nonlocality". *arXiv:1303.2849*. `arXiv:1303.2849`, 2013.

[BCS12] H. Buhrman, M. Christandl, and C. Schaffner. "Complete insecurity of quantum protocols for classical two-party computation". *Phys. Rev. Lett.*, **109**(16), 2012.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson. "Quantum vs. classical communication and computation". STOC '98, pp. 63–68. ACM, 1998.

[BCWd01] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. "Quantum fingerprinting". *Phys. Rev. Lett.*, **87**(16):167902, 2001.

[BDMS00] S. L. Braunstein, G. M. D'Ariano, G. J. Milburn, and M. F. Sacchi. "Universal teleportation with a twist". *Phys. Rev. Lett.*, **84**(15):3486–3489, 2000.

[BDPC11] A. Bisio, G. M. D'Ariano, P. Perinotti, and G. Chiribella. "Minimal Computational-Space Implementation of Multiround Quantum Protocols". *Phys. Rev. A*, **83**(2), February 2011.

[BDPS14] A. Bisio, G. M. D'Ariano, P. Perinotti, and M. Sedlák. "Optimal processing of reversible quantum channels". *Phys. Lett. A*, **378**(26–27):1797–1808, 2014.

[Bel64] J. S. Bell. "On the Einstein Podolsky Rosen paradox". *Physics*, **1**(3):195–200, 1964.

[Bel81] J. S. Bell. "Bertlmann's socks and the nature of reality". In *Journal de physique. Colloque*, vol. 42, pp. 41–62, 1981.

[BFW14] Ä. Baumeler, A. Feix, and S. Wolf. "Maximal incompatibility of locally classical behavior and global causal order in multiparty scenarios". *Phys. Rev. A*, **90**(4):042106, 2014.

[BG11] J. Barrett and N. Gisin. "How much measurement independence is needed to demonstrate nonlocality?". *Phys. Rev. Lett.*, **106**(10):100406, 2011.

[BLMS87] L. Bombelli, J. Lee, D. Meyer, and R. D. Sorkin. "Space-time as a causal set". *Phys. Rev. Lett.*, **59**(5):521–524, 1987.

[Bra05] G. Brassard. "Is information the key?". *Nat. Phys.*, **1**(1):2–4, 2005.

[Bro09] H. R. Brown. "The behaviour of rods and clocks in general relativity, and the meaning of the metric field". `arXiv:0911.4440 [gr-qc, quant-ph]`, 2009.

[Bru] Č. Brukner. Private communication.

[Bru14] Č. Brukner. "Quantum causality". *Nat. Phys.*, **10**(4):259–263, 2014.

[Bru15] Č. Brukner. "Bounding quantum correlations with indefinite causal order". *New J. Phys.*, **17**(8):083034, 2015.

[BW92] C. H. Bennett and S. J. Wiesner. "Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states". *Phys. Rev. Lett.*, **69**(20):2881–2884, 1992.

[BW13] Ä. Baumeler and S. Wolf. "Perfect signaling among three parties violating predefined causal order". *arXiv:1312.5916 [quant-ph]*. `arXiv:1312.5916 [quant-ph]`, 2013.

[BW14] Ä. Baumeler and S. Wolf. "Perfect signaling among three parties violating predefined causal order". In *2014 IEEE International Symposium on Information Theory*, pp. 526–530, 2014.

[BYJK04] Z. Bar-Yossef, T. S. Jayram, and I. Kerenidis. "Exponential separation of quantum and classical one-way communication complexity". STOC '04, pp. 128–137. ACM, 2004.

[BZ03] Č. Brukner and A. Zeilinger. "Information and fundamental elements of the structure of quantum theory". In *Time, Quantum and Information*, pp. 323–354. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[CBH03] R. Clifton, J. Bub, and H. Halvorson. "Characterizing quantum theory in terms of information-theoretic constraints". *Found Phys*, **33**(11):1561–1591, 2003.

[CDFP12] T. Colnaghi, G. M. D'Ariano, S. Facchini, and P. Perinotti. "Quantum computation with programmable connections between gates". *Phys. Lett. A*, **376**(45):2940–2943, 2012.

[CDP08a] G. Chiribella, G. M. D'Ariano, and P. Perinotti. "Quantum circuit architecture". *Phys. Rev. Lett.*, **101**(6):060401, 2008.

[CDP08b] G. Chiribella, G. M. D'Ariano, and P. Perinotti. "Transforming quantum operations: Quantum supermaps". *EPL*, **83**(3):30004, 2008.

[CDP09] G. Chiribella, G. M. D'Ariano, and P. Perinotti. "Theoretical framework for quantum networks". *Phys. Rev. A*, **80**(2):022339, 2009.

[CDP11] G. Chiribella, G. M. D'Ariano, and P. Perinotti. "Informational derivation of quantum theory". *Phys. Rev. A*, **84**(1):012311, 2011.

[CDPV13] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron. "Quantum computations without definite causal structure". *Phys. Rev. A*, **88**(2):022318, 2013.

[CFS02] C. M. Caves, C. A. Fuchs, and R. Schack. "Quantum probabilities as Bayesian probabilities". *Phys. Rev. A*, **65**(2):022305, 2002.

[Chi12] G. Chiribella. "Perfect discrimination of no-signalling channels via quantum superposition of causal structures". *Phys. Rev. A*, **86**(4):040301, 2012.

[Cho75] M.-D. Choi. "Completely positive linear maps on complex matrices". *Linear Algebra Appl.*, **10**(3):285–290, 1975.

[CHSH69] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. "Proposed experiment to test local hidden-variable theories". *Phys. Rev. Lett.*, **23**(15):880–884, 1969.

[CIJT] G. Chiribella, R. Ioniciou, T. Jennewein, and D. Terno. Private communication.

[CL14] 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**(42):424018, 2014.

[CMG15] R. Chaves, C. Majenz, and D. Gross. "Information–theoretic implications of quantum causal structures". *Nat. Commun.*, **6**:5766, 2015.

[Cos] F. Costa. Private communication.

[Cos13] F. Costa. *Local and Causal Structures in Quantum Theory.* Phd, University of Vienna, Vienna, 2013.

[CR11] R. Colbeck and R. Renner. "No extension of quantum theory can have improved predictive power". *Nat. Commun.*, **2**:411, 2011.

[CRGB17] E. Castro Ruiz, F. Giacomini, and a. Brukner. "Entanglement of quantum clocks through gravity". *PNAS*, **114**(12):E2303–E2309, 2017.

[CRGB18] E. Castro-Ruiz, F. Giacomini, and i. c. v. Brukner. "Dynamics of Quantum Causal Structures". *Phys. Rev. X*, **8**:011047, Mar 2018.

[CS14] D. Chruściński and G. Sarbicki. "Entanglement witnesses: construction, analysis and classification". *J. Phys. A: Math. Theor.*, **47**(48):483001, 2014.

[CS16] F. Costa and S. Shrapnel. "Quantum causal modelling". *New J. Phys.*, **18**(6):063032, 2016.

[CvNT13] R. Cleve, W. van Dam, M. Nielsen, and A. Tapp. "Quantum entanglement and the communication complexity of the inner product function". *Theoretical Computer Science*, **486**:11–19, 2013.

[D'A11] G. M. D'Ariano. "Physics as quantum information processing". *AIP Conference Proceedings*, **1363**(1):63–68, 2011.

[Deu91] D. Deutsch. "Quantum mechanics near closed timelike lines". *Phys. Rev. D*, **44**(10):3197–3217, 1991.

[DJ92] D. Deutsch and R. Jozsa. "Rapid solution of problems by quantum computation". *Proc. R. Soc. Lond. A*, **439**(1907):553–558, 1992.

[DL70] E. B. Davies and J. T. Lewis. "An operational approach to quantum probability". *Comm. Math. Phys.*, **17**(3):239–260, 1970.

[EGA+13] S. Eibenberger, S. Gerlich, M. Arndt, M. Mayor, and J. Tüxen. "Matter–wave interference of particles selected from a molecular library with masses exceeding 10000 amu". *Phys. Chem. Chem. Phys.*, **15**(35):14696–14700, 2013.

[EPR35] A. Einstein, B. Podolsky, and N. Rosen. "Can quantum-mechanical description of physical reality be considered complete?". *Phys. Rev.*, **47**(10):777–780, 1935.

[ESC18] D. Ebler, S. Salek, and G. Chiribella. "Enhanced Communication with the Assistance of Indefinite Causal Order". *Phys. Rev. Lett.*, **120**:120502, Mar 2018.

[FAB15] A. Feix, M. Araújo, and Č. Brukner. "Quantum superposition of the order of parties as a communication resource". *Phys. Rev. A*, **92**(5):052326, 2015.

[FAB16] A. Feix, M. Araújo, and Č. Brukner. "Causally nonseparable processes admitting a causal model". *New J. Phys.*, **18**:083040, 2016.

[FJV13] J. Fitzsimons, J. Jones, and V. Vedral. "Quantum correlations which imply causation". `arXiv:1302.2731 [gr-qc, quant-ph]`, 2013.

[Fri16] T. Fritz. "Beyond Bell's theorem II: scenarios with arbitrary causal structure". *Commun. Math. Phys.*, **341**(2):391–434, 2016.

[GHSZ90] D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger. "Bell's theorem without inequalities". *American Journal of Physics*, **58**(12):1131–1143, 1990.

[GHZ89] D. M. Greenberger, M. A. Horne, and A. Zeilinger. "Going beyond Bell's theorem". In *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, pp. 69–72. Springer Netherlands, Dordrecht, 1989.

[Gia] C. Giarmatzi. Private communication.

[Gis90] N. Gisin. "Weinberg's non-linear quantum mechanics and supraluminal communications". *Phys. Lett. A*, **143**(1-2):1–2, 1990.

[GLP98] P. Grangier, J. A. Levenson, and J.-P. Poizat. "Quantum non-demolition measurements in optics". *Nature*, **396**(6711):537–542, 1998.

[GMR+13] M. Giustina, A. Mech, S. Ramelow, B. Wittmann, J. Kofler, J. Beyer, A. Lita, B. Calkins, T. Gerrits, S. W. Nam, R. Ursin, and A. Zeilinger. "Bell violation using entangled photons without the fair-sampling assumption". *Nature*, **497**:227–230, 2013.

[Göd49] K. Gödel. "An example of a new type of cosmological solutions of Einstein's field equations of gravitation". *Rev. Mod. Phys.*, **21**(3):447–450, 1949.

[GR13] G. Ghirardi and R. Romano. "About possible extensions of quantum theory". *Foundations of Physics*, **43**(7):881–894, 2013.

[Gur03] L. Gurvits. "Classical deterministic complexity of Edmonds' problem and quantum entanglement". STOC '03, pp. 10–19. ACM, 2003.

[GW07] G. Gutoski and J. Watrous. "Toward a general theory of quantum games". STOC '07, pp. 565–574. ACM, 2007.

[Hal11] M. J. W. Hall. "Relaxed Bell inequalities and Kochen-Specker theorems". *Phys. Rev. A*, **84**(2), 2011.

[Har05] L. Hardy. "Probability theories with dynamic causal structure: a new framework for Quantum Gravity". `arXiv:gr-qc/0509120`, 2005.

[Har07] L. Hardy. "Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure". *J. Phys. A: Math. Theor.*, **40**(12):3081, 2007.

[Har09] L. Hardy. *Quantum gravity computers: on the theory of computation with indefinite causal structure*. Springer Netherlands, Dordrecht, 2009.

[HHH96] M. Horodecki, P. Horodecki, and R. Horodecki. "Separability of mixed states: necessary and sufficient conditions". *Phys. Lett. A*, **223**(1–2):1–8, 1996.

[HHHH09] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. "Quantum entanglement". *Rev. Mod. Phys.*, **81**(2):865–942, 2009.

[HIN+06] M. Hayashi, K. Iwama, H. Nishimura, R. Raymond, and S. Yamashita. "(4,1)-Quantum random access coding does not exist—one qubit is not enough to recover one of four bits". *New J. Phys.*, **8**(8):129, 2006.

[HJS+96] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters. "Classical information capacity of a quantum channel". *Phys. Rev. A*, **54**(3):1869–1876, 1996.

[HLP14] J. Henson, R. Lal, and M. F. Pusey. "Theory-independent limits on correlations from generalized Bayesian networks". *New J. Phys.*, **16**(11):113043, 2014.

[Hum75] D. Hume. "An Enquiry concerning Human Understanding". In L. A. Selby-Bigge and P. H. Nidditch, editors, *Enquiries concerning Human Understanding and concerning the Principles of Morals*. Clarendon Press, Oxford, 1975.

[IRW11] P. Illari, F. Russo, and J. Williamson, editors. *Causality in the Sciences*. Oxford University Press, 2011.

[Jam72] A. Jamiołkowski. "Linear transformations which preserve trace and positive semidefiniteness of operators". *Rep. Math. Phys.*, **3**(4):275–278, 1972.

[Kla00] H. Klauck. "On Quantum and Probabilistic Communication: Las Vegas and One-Way Protocols". STOC '00, pp. 644–651. ACM, 2000.

[KN06] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, 2006.

[LLWJ13] W. Liu, C. Liu, H. Wang, and T. Jia. "Quantum private comparison: a review". *IETE Technical Review*, **30**(5):439, 2013.

[Lo97] H.-K. Lo. "Insecurity of quantum secure computations". *Phys. Rev. A*, **56**(2):1154–1162, 1997.

[Lof04] J. Lofberg. "YALMIP : A Toolbox for Modeling and Optimization in MATLAB". In *Proceedings of the CACSD conference*, Taipei, Taiwan, 2004.

[Lov99] L. Lovász. "Hit-and-run mixes fast". *Math. Program.*, **86**(3):443–461, 1999.

[LS13] 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**(5):052130, 2013.

[LVB11] Y.-C. Liang, T. Vértesi, and N. Brunner. "Semi-device-independent bounds on entanglement". *Phys. Rev. A*, **83**(2):022108, 2011.

[MM13] M. P. Müller and L. Masanes. "Three-dimensionality of space and the quantum bit: an information-theoretic approach". *New J. Phys.*, **15**(5):053040, 2013.

[Mor14] T. Morimae. "Acausal measurement-based quantum computing". *Phys. Rev. A*, **90**(1):010101, 2014.

[mos15] "The MOSEK optimization toolbox for MATLAB 7.0". 2015.

[MRSR17] J.-P. W. MacLean, K. Ried, R. W. Spekkens, and K. J. Resch. "Quantum-coherent mixtures of causal relations". *Nat. Commun.*, **8**:15149, 2017.

[Nay99] A. Nayak. "Optimal lower bounds for quantum automata and random access codes". In *40th annual symposium on foundations of computer science, 1999*, pp. 369–376, 1999.

[NC00] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge ; New York, 2000.

[NCH+15] T. L. Nicholson, S. L. Campbell, R. B. Hutson, G. E. Marti, B. J. Bloom, R. L. McNally, W. Zhang, M. D. Barrett, M. S. Safronova, G. F. Strouse, W. L. Tew, and J. Ye. "Systematic evaluation of an atomic clock at $2 \cdot 10^{-18}$ total uncertainty". *Nat. Commun.*, **6**:6896, 2015.

[NN87] Y. Nesterov and A. Nemirovskii. *Interior Point Polynomial Algorithms in Convex Programming*. Studies in Applied Mathematics. Society for Industrial and Applied Mathematics, 1987.

[NS02] A. Nayak and J. Salzman. "On communication over an entanglement-assisted quantum channel". STOC '02, pp. 698–704. ACM, 2002.

[OCB12] O. Oreshkov, F. Costa, and Č. Brukner. "Quantum correlations with no causal order". *Nat. Commun.*, **3**:1092, 2012.

[OG16] O. Oreshkov and C. Giarmatzi. "Causal and causally separable processes". *New J. Phys.*, **18**(9):093020, 2016.

[Ore18] O. Oreshkov. "On the whereabouts of the local operations in physical realizations of quantum processes with indefinite causal order". `arXiv:1801.07594 [quant-ph]`, 2018.

[Par03] Parmenides. *Paramenides of Elea: a verse translation with interpretative essays and commentary to the text.* Praeger, Westport, Conn., 2003.

[PB11] M. Pawłowski and N. Brunner. "Semi-device-independent security of one-way quantum key distribution". *Phys. Rev. A*, **84**(1):010302, 2011.

[PB15] J. Pienaar and Č. Brukner. "A graph-separation theorem for quantum causal models". *New J. Phys.*, **17**(7):073020, 2015.

[Pea00] J. Pearl. *Causality: models, reasoning, and inference.* Cambridge University Press, Cambridge, 2000.

[Per96] A. Peres. "Separability criterion for density matrices". *Phys. Rev. Lett.*, **77**(8):1413–1415, 1996.

[PMA+15] L. M. Procopio, A. Moqanaki, M. Araújo, F. Costa, I. Alonso 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.

[PPK+09] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski. "Information causality as a physical principle". *Nature*, **461**(7267):1101–1104, 2009.

[PRB+14] G. Pütz, D. Rosset, T. J. Barnea, Y.-C. Liang, and N. Gisin. "Arbitrarily small amount of measurement independence is sufficient to manifest quantum nonlocality". *Phys. Rev. Lett.*, **113**(19), 2014.

[PW13] C. Pfister and S. Wehner. "An information-theoretic principle implies that any discrete physical theory is classical". *Nat. Commun.*, **4**:1851, 2013.

[RAV+15] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, and K. J. Resch. "A quantum advantage for inferring causal structure". *Nat. Phys.*, **11**(5):414–420, 2015.

[Raz99] R. Raz. "Exponential separation of quantum and classical communication complexity". STOC '99, pp. 358–367. ACM, 1999.

[RBK11] O. Regev and B. Bo'az Klartag. "Quantum ene-way communication can be exponentially stronger than classical communication". STOC '11, pp. 31–40. ACM Press, 2011.

[Rei56] H. Reichenbach. *The direction of time.* University of California Press, Berkeley, 1956.

[RKM+01] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland. "Experimental violation of a Bell's inequality with efficient detection". *Nature*, **409**(6822):791–794, 2001.

[Roy91] A. Royer. "Wigner function in Liouville space: a canonical formalism". *Phys. Rev. A*, **43**(1):44–56, 1991.

[RRF+17] 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". *Science Advances*, **3**(3). `http://advances.sciencemag.org/content/3/3/e1602589.full.pdf`, 2017.

[SBv+08] D. Salart, A. Baas, J. A. W. van Houwelingen, N. Gisin, and H. Zbinden. "Spacelike separation in a Bell test assuming gravitationally induced collapses". *Phys. Rev. Lett.*, **100**(22), 2008.

[SGS93] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction and Search*. Springer, Berlin, 1993.

[Smi80] R. L. Smith. "Monte Carlo procedures for generating random feasible solutions to mathematical programs". In *A Bulletin of the ORSA/TIMS Joint National Meeting, Washington, DC*, vol. 101, 1980.

[Smi84] R. L. Smith. "Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions". *Oper. Res.*, **32**(6):1296–1308, 1984.

[SS97] A. Suarez and V. Scarani. "Does entanglement depend on the timing of the impacts at the beam-splitters?". *Phys. Lett. A*, **232**(1-2):9–14, 1997.

[Ste03] M. Steiner. "Generalized robustness of entanglement". *Phys. Rev. A*, **67**(5):054305, 2003.

[Tsi80] B. S. Tsirelson. "Quantum generalizations of Bell's inequality". *Lett. Math. Phys.*, **4**(2):93–100, 1980.

[VC15] V. N. Vapnik and A. Y. Chervonenkis. "On the Uniform Convergence of Relative Frequencies of Events to their Probabilities". In *Measures of Complexity*, pp. 11–30. Springer International Publishing, Cham, 2015.

[Wer89] R. F. Werner. "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model". *Phys. Rev. A*, **40**(8):4277–4281, 1989.

[Whe90] J. A. Wheeler. "Information, physics, quantum: The search for links". In *Complexity, Entropy, and the Physics of Information*, vol. 2, pp. 200–19. Redwood city, 1990.

[Wie83] S. Wiesner. "Conjugate coding". *SIGACT News*, **15**(1):78–88, 1983.

[Wis14] H. M. Wiseman. "The two Bell's theorems of John Bell". *J. Phys. A: Math. Theor.*, **47**(42):424001, 2014.

[WJS+98] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger. "Violation of Bell's Inequality under Strict Einstein Locality Conditions". *Phys. Rev. Lett.*, **81**(23):5039–5043, 1998.

[Woo84] W. K. Wootters. "'Time' replaced by quantum correlations". *Int. J. Theor. Phys.*, **23**(8):701–711, 1984.

[WS15] 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**(3):033002, 2015.

[Yao79] A. C.-C. Yao. "Some complexity questions related to distributive computing". STOC '79, pp. 209–213. ACM, 1979.

[Yao82] A. C. Yao. "Protocols for secure computations". In *23rd Annual Symposium on Foundations of Computer Science*, pp. 160–164, 1982.

[Yao93] A. C.-C. Yao. "Quantum circuit complexity". In *Proceedings of the IEEE 34th Annual Foundations of Computer Science*, pp. 352–361, 1993.

[ZCPB17] M. Zych, F. Costa, I. Pikovski, and Č. Brukner. "Bell's Theorem for Temporal Order". arXiv:1708.00248 [quant-ph], 2017.

[Zyc15] M. Zych. *Quantum systems under gravitational time dilation*. PhD thesis, University of Vienna, Vienna, 2015.