Simulating extremal temporal correlations

Cornelia Spee1,2,∗, Costantino Budroni1,3 and Otfried Gühne2

1 Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Boltzmannasse 3, 1090 Vienna, Austria
2 Naturwissenschaftlich-Technische Fakultät, Universität Siegen, Walter-Flex-Straße 3, 57068 Siegen, Germany
3 Faculty of Physics, University of Vienna, Boltzmannasse 5, 1090 Vienna, Austria
∗ Author to whom any correspondence should be addressed.
E-mail: Cornelia.Spee@oeaw.ac.at

Keywords: quantum correlations, sequences of measurements, temporal correlations

Abstract

The correlations arising from sequential measurements on a single quantum system form a polytope. This is defined by the arrow-of-time (AoT) constraints, meaning that future choices of measurement settings cannot influence past outcomes. We discuss the resources needed to simulate the extreme points of the AoT polytope, where resources are quantified in terms of the minimal dimension, or ‘internal memory’ of the physical system. First, we analyze the equivalence classes of the extreme points under symmetries. Second, we characterize the minimal dimension necessary to obtain a given extreme point of the AoT polytope, including a lower scaling bound in the asymptotic limit of long sequences. Finally, we present a general method to derive dimension-sensitive temporal inequalities for longer sequences, based on inequalities for shorter ones, and investigate their robustness to imperfections.

1. Introduction

The study of spatial correlations, from Bell nonlocality [1, 2] to entanglement theory [3, 4], has had, on the one hand, a profound impact on the foundations of quantum mechanics. On the other hand, it stimulated plenty of applications in quantum information processing, such as quantum key distribution [5], randomness certification [6] and expansion [7], to mention a few. Moreover, correlations stronger than quantum ones, but still obeying the no-signaling constraints [8], have been extensively investigated both from a fundamental perspective and in relation with applications to quantum information processing.

Similarly, temporal correlations have been studied from the perspective of the difference between classical and quantum systems, mostly in the framework of Leggett–Garg inequalities [9, 10] and noncontextuality inequalities [11, 12] tested via sequential measurements [13, 14]. More recently, a notion of non-classical temporal correlations has been formulated also from a different perspective that does not require assumptions on the noninvasivity or compatibility of the measurements [15, 16]. Few quantum information processing tasks have been formulated directly in this framework, such as dimension witnesses [17–20], purity certification [21], and time-keeping devices [22]. Many other tasks, despite not being directly formulated in the language of temporal correlations, are closely related, since they naturally involve sequential operations. This is, for example, the case for prepare-and-measure scenarios [23, 24], quantum random access codes (QRACs) [25–30], classical simulations of quantum contextuality [31, 32], quantum simulation of classical stochastic processes [33], memory asymmetry between prediction and retrodiction [34], and optimal ticking clocks [35].

The temporal counterpart to the no-signaling constraints [8] are the arrow-of-time (AoT) constraints [36], stating the impossibility of signaling from the future to the past. These conditions define the AoT polytope [36]. It has been shown that the extreme points of this polytope are given by the deterministic assignments [19, 37, 38], where each output in a sequence is obtained as a deterministic function of the previous inputs and outputs. This implies that any such point is realizable by sequential measurements on a physical system, even for a classical theory, if the internal memory of the system is large enough to store the
information about previous inputs and outputs. This is in stark contrast to the spatial case, where different correlations correspond to different theories, often irrespectively of the system dimension. A difference in temporal correlations between classical and quantum theory is recovered if the system dimension is constrained, as noted already a long time ago in the context of QRACs [25–27], and similarly those theories can be distinguished from generalized probability theories [16]. This dimension dependence can be exploited to construct temporal inequalities that can certify a lower bound on the dimension of the system, i.e. they are dimension witnesses [16, 19, 20, 23, 24, 28, 39, 40].

In this paper, we investigate the minimal resources necessary to simulate a given extreme point of the AoT polytope. Here, the resource is quantified by the dimension (or memory) needed to reproduce the outcomes of a measurement sequence. First, we study the symmetries of the AoT polytope w.r.t. classical post-processing. Then, we determine the minimal dimension required for the realization of a given extreme point. As in references [16, 19, 20], no assumption on the concrete realization or quantum description of a measurement is made, it is only assumed that the same measurement may be carried out at different times. We then continue by providing a simple method to combine dimension-sensitive temporal inequalities for shorter sequences to obtain inequalities valid for longer sequences. Finally, we discuss the robustness of temporal inequalities if the measurements are not perfect, i.e. if they vary over time.

We remark that the simulation of extreme temporal correlations, for a given length and number of inputs and outputs, may or may not be sufficient to simulate all temporal correlations, depending on the assumption on the resources available. Consider the scenario in which the experimenter has two machines, able to generate the correlations $p_1$ and $p_2$, respectively. At the beginning of each measurement sequence she chooses with probability $\lambda$ the machine 1 and with probability $1 - \lambda$ the machine 2, and then she proceeds to measure the whole sequences with the initially chosen machine. In this way, she would obtain as correlation $p$ the convex mixture $p = \lambda p_1 + (1 - \lambda)p_2$. By straightforwardly extending this argument, one can see that the randomness available in the choice of the initial machine gives rise to all convex mixtures. In particular, by being able to simulate all extreme correlations of a given scenario (i.e., the extreme points of a given polytope), the experimenter can simulate all corresponding temporal correlations (i.e., the convex hull, corresponding to the full polytope). This is the scenario discussed in reference [16]. In contrast, if this initial randomness is not an available resource, the set of temporal correlations has a much more complicated structure [41].

The paper is organized as follows. In section 2, we introduce our notation and the considered scenario in detail and we review some important previous results. In section 3, we further explore the properties of the AoT polytope, in particular the symmetries of the AoT polytope under classical post-processing, i.e., possible relabeling of the inputs and outputs, as such transformations do not affect the quantum realization. In section 4, we investigate questions such as how much memory is required to realize a given extreme point within quantum theory, or stated differently, what is the minimal dimension that is necessary to obtain this correlation. In section 5, we provide a lower bound on the dimension needed to obtain an arbitrary extreme point and an estimate of the behavior in the asymptotic limit of arbitrary long measurement sequences, for any number of inputs and outputs. Then, in section 6 we show a general method to produce temporal inequalities for longer sequences by combining inequalities for shorter ones. Finally in section 7, we investigate how robust are our statements on temporal correlations in the case in which the assumption of repeated measurements is only approximately satisfied.

2. Notation and preliminaries

We consider the scenario of sequential measurements depicted in figure 1. A box receives a sequences of inputs, or measurement settings, $x_1, x_2, \ldots, x_L$ and produces a sequences of outputs, or measurement outcomes $a_1, a_2, \ldots, a_L$. The machines works by transforming probabilistically its internal state, e.g., a quantum state $\rho$, according to the measurement input and outcome and generating a measurement outcome according to the input and previous state. We are then interested in the correlations $p(a_1 a_2 \ldots a_L | x_1 x_2 \ldots x_L)$.

More concretely, an operation on a quantum system, associated with an input $x_i$, is represented by a quantum instrument, namely a collection of completely positive maps $\{I_{a | x} \}$, that sum up to a unital map, i.e., $\sum I_{a | x} (1) = 1$, where 1 denotes the identity operator, corresponding to the rule of preservation of probability in the Heisenberg picture, see, e.g., [42] for a textbook introduction. Each instrument $\{I_{a | x} \}$ defines a generalized measurement, i.e., a positive operator valued measure (POVM), through the formula $E_{a | x} := I_{a | x} (1)$. Correlations for a sequence of inputs $x_1, x_2$ and outputs $a_1, a_2$ are given by the formula

$$ p (a_1 a_2 | x_1 x_2) = \text{tr} \left[ \rho I_{a_1 | x_1} (E_{a_2 | x_2}) \right] = \text{tr} \left[ \rho I_{a_1 | x_1} \circ I_{a_2 | x_2} (1) \right],$$

(1)

where $\circ$ denotes the composition of maps, and analogous expressions hold for longer sequences. We assume
that the evolution of our box is time-independent, except for the external classical inputs that are provided at each time step. Practically, this assumption means two things. First, that the different correlations are generated by the transitions of the internal state of the machine; in quantum mechanical terms, this implies that for a given input \( x \) the machine applies the quantum instrument \( \{ I_{ax} \}_a \) independently of which time step \( t \) we are in. This is already implicit in equation (1), since we used only the symbol \( I_{ax} \) to denote the instruments, without any reference to the time step, e.g., we want to calculate the probability \( p(00|00) \) we apply the same mapping twice, i.e., \( p(00|00) = \text{tr}[\rho I_{00} \circ I_{00}(1)] \). Second, we assume that the inputs are provided at equally spaced time intervals and the free evolution of the system is always implemented by the same quantum channel, e.g., one can think about an evolution governed by a time-independent Hamiltonian. Hence, the time evolution can be reabsorbed wlog into the definition of the quantum instruments. Boxes satisfying these assumptions are called finite-state machines by generalizing a well-known classical notion [43] (see also reference [16] for more details on the quantum and generalized probability theory case). We equivalently say that the measurement operations are time-independent.

In this scenario with time-ordered measurements, any theory that respects causality must satisfy the so-called AoT conditions [36], namely, the future choice of inputs cannot modify the probabilities of past outcomes. For the simple case of a sequence of two measurements, the correlation \( p(a_1a_2|x_1x_2) \) must satisfy

\[
\sum_{a_1} p(a_1a_2|x_1x_2) = \sum_{a_2} p(a_1a_2|x_1x_2'), \quad \text{for all } a_1,x_1,x_2,x_2'.
\] (2)

This condition is analogous to the no-signaling conditions for spatial correlations [8], but it constrains only one direction, i.e., signaling from the future to the past. These linear constraints, together with positivity, \( p(a_1a_2|x_1x_2) \geq 0 \), and normalization, \( \sum_{a_1a_2} p(a_1a_2|x_1x_2) = 1 \), define a polytope called the AoT polytope [36], denoted in the general case as \( P_{L}^{0,S} \), where \( O \) denotes the number of outputs, \( S \) the number of inputs (or measurement settings) and \( L \) the length of the measurement sequence.

Such constraints are satisfied by classical and quantum mechanics and it has been proven that all extreme points are given by deterministic assignments, i.e., correlations which have the property that for any input one obtains a deterministic output [19, 37, 38]. Intuitively, this comes from the fact that the AoT constraints allow us to decompose the probability distribution as

\[
p(a_1a_2a_3|x_1x_2x_3) = p(a_1|x_1)p(a_2|x_1x_2)p(a_3|x_1a_2;x_2x_3);
\] (3)

the extreme points, then, are given by the products of deterministic functions, generating \( a_1, a_2 \) and \( a_3 \) respectively, from the previous inputs and outputs. It has been shown in [19] that the AoT polytope \( P_{L}^{0,S} \) has

\[
N_{L}^{0,S} = (O^S)^{\frac{L}{2}-1}
\] (4)

extreme points. The extreme points can, then, be reached if the machine has enough ‘internal memory’namely, a large enough set of perfectly distinguishable internal states [19], to remember previous inputs and outputs and generate deterministically the corresponding outputs.

For a sequence of length \( L \), an extreme point of the AoT polytope can be represented as a tree graph with \( \sum_{k=0}^{L-1} S^k = \frac{S^L - 1}{S - 1} \) nodes, as depicted in figure 2. The tree graph can be intuitively understood as follows. At
Each time-step the evolution of our system 'branches' depending on the received input, e.g., the history in which the system received 0 departs from the history in which it received 1, since the internal state of the machine will evolve differently. In this way, we can keep track of all possible sequences of inputs that are occurring in the next time-step after the measurements leading to them, coincide. (b) Example of two inequivalent futures. The 1-length futures of (1, 1) and (1, 2) are inequivalent as the respective subtrees do not coincide.

**Figure 2.** Assignment of tuples for an extreme point of $P^{1}_{1,2}$. In this example, for both measurements the outcome '0' is obtained in the first time step. Then, then evolution branches, depending whether $M_{1}$ or $M_{2}$ has been measured, giving, the branches with $\Gamma_{1,1}$ or $\Gamma_{2,2}$, respectively. After measuring $M_{1}$ in the first time step, one obtains in the second time step for both measurements deterministically the outcome '1' (corresponding to the tuple $\Gamma_{1,1}$) and after performing $M_{2}$ in the first time step one observes for measurement $M_{1}$ ($M_{2}$) in the second time step the outcome '0' (‘1’) respectively (corresponding to the tuple $\Gamma_{2,1}$). The 1-length future of (1, 1), i.e., $T_{1,1}$, is indicated by the dark blue triangle, and the history of the whole sequence, $T_{1,2}$, is represented by the triangle obtained by joining the dark blue and light blue regions.

**Figure 3.** We show two examples of, respectively, equivalent and inequivalent futures. In this case, since we want to compare the $(L - 1)$-th node with another one, in this case the first, we need to consider only the 1-length futures. In other words, in the comparison between tuples $\Gamma_{1}$ and $\Gamma_{L-1}$, we need to consider $T_{1,k} = T_{1,k}'$, with $r = \min\{L - 1, l\}$. (a) An example of two equivalent futures. The 1-length futures of (1, 1) and $(L - 1, 2^{L-1})$ are equivalent, since the nodes themselves, as well as the tuples occurring in the next time-step after the measurements leading to them, coincide. (b) Example of two inequivalent futures. The 1-length futures of (1, 1) and $(L - 1, 2^{L-2})$ are inequivalent as the respective subtrees do not coincide.
same sequence of future outcomes, namely,
\[ \rho_{r,l,k} = \rho_{r',l',k} \Rightarrow T_{l,k}^r = T_{l',k}^{r'}, \quad \text{for } r = \min\{L - l, L - l'\}. \] (5)

If two tuples \( \Gamma_{l,k} \) and \( \Gamma_{l',k}^{r'} \) satisfy \( T_{l,k}^r = T_{l',k}^{r'} \), for \( r = \min\{L - l, L - l'\} \), we say that they have equivalent futures, see figure 3. Notice that the notion of equivalent futures depends always on the maximal length \( L \) of the observed sequences, and that the comparison between subtrees makes sense only for equal lengths \( r \) chosen as above. For a given deterministic sequence, we call \( T_{1,1} \), i.e., the entire tree, the history of the sequence.

This observation provides a way of counting the minimal number of states necessary for reproducing an extreme point of the AoT polytope, since inequivalent futures must correspond to different states. In particular, in order to check whether \( T_{l,k}^r \neq T_{l',k}^{r'} \), one can simply check for shorter sequences, i.e., whether \( T_{l,k}^r \neq T_{l',k}^{r'} \), for \( s = 0, \ldots, r \).

3. Symmetries of the AoT polytope

It has been shown that, if no assumption on the dimension of the quantum system is made, any correlation in the polytope can be realized. The ability of realizing a correlation is independent of the chosen labeling of the outcomes and/or measurement settings as long as one performs exactly the same relabeling at every time step. This is due to the fact that any such relabeling can be implemented classically even after the measurement sequences have been performed, i.e. such relabelings correspond to some classical post-processing. Note that the condition that the same relabeling is applied to all time steps is necessary to be consistent with our assumption of time-independent measurements. In the following, we characterize the number of equivalence classes of extreme points under these symmetries for small numbers of settings and outcomes.

In particular, we define an outcome relabeling equivalence (ORE) class as an equivalence class of extreme points w.r.t. to the relation of being the same up to a relabeling of the outcomes. In particular, since relabeling is a classical post-processing, if one extreme point is obtainable by measurements on a physical system, the same is true for all elements in the class. Of course, also the measurement settings can be subject to relabeling and all extreme points that are equivalent up to relabeling of the measurement settings can be realized within the same physical implementation. Then, we define the relabeling equivalence (RE) classes of extreme points as the set of extreme points that are equal up to relabeling of outcomes and measurement settings.

3.1. General considerations

In the temporal scenario the only relevant symmetries are given by the relabeling of inputs and outputs of a given sequence. The corresponding symmetry groups are given by the symmetric groups \( S_O \) and \( S_S \), where as defined before and \( S \) are the number of outputs and inputs, respectively. The total group of symmetry is given by the direct product \( G := S_O \times S_S \).

To each element \( g \in G \), we associate a transformation \( S \) on the extreme points of the polytope \( P_L^{O,S} \). In this way, each extreme point \( v \) of \( P_L^{O,S} \) generates an orbit \( O_v \) defined as
\[ O_v = \{S_v|g \in G\} \] (6)

The action of a group on a set naturally induces an equivalence relations in terms of orbits given by \( v \sim w \Leftrightarrow O_v = O_w \). In this case, belonging to the same orbit means that the extreme point \( v \) can be obtained from the extreme point \( w \) via a relabeling of inputs and outputs, and vice versa.

The number of equivalence classes is then given by the number of different orbits. Hence, if one can evaluate the number of orbits one can deduce the number of RE classes for a given scenario. Below we show how to do so by identifying the elements that are invariant under a symmetry, their orbits and the cardinality of the orbits, which allows us to deduce the number of orbits, for the case of two outcomes and two and three settings. The same procedure can be applied to arbitrary number of outcomes and settings without extra conceptual difficulties; however, as the symmetric group grows, recall that there are \( n! \) permutations of \( n \) elements, the whole procedure becomes much longer and tedious.

3.2. Relabeling of outcomes and measurement settings

For the case \( O = 2 \) one can define for each ORE class as a representative an extreme point having the property that for any measurement the outcome in the first time step is ’0’. This allows us to count the number of ORE classes as given in the following lemma.

**Lemma 1.** The number of ORE classes of extreme points of \( P_L^{O,S} \) is given by \( N_{ORE}^{2,S,L} = (2^S)^{\frac{L}{r+s}} \).

\[ L \]
Figure 4. Illustration of the action of the group on an example. First we transform the extreme point to the representative of the ORE class, which has the property that in the first time step all outcomes are ‘0’ by relabeling the outcomes accordingly. Then we illustrate the action of the symmetry $S_{12}$ (which permutes the measurement settings) on this representative, where the tuples obtained by applying $S_{12}$ are denoted by $\tilde{\Gamma}$ and the operator $X$ permutes the elements of the tuples $\Gamma$ of the extreme point on which the transformation is performed. The other group element of $S_2$, i.e., the identity element $e$, leaves any extreme point invariant.

Proof. We consider one representative of each ORE class and show that the cardinality of the ORE class is $2^S$. The number of equivalence classes is the total number of extreme points divided by the cardinality of such classes. It is straightforward to see that one particular choice of a representative is given by demanding the outcome of each measurement setting at the first time step to be ‘0’. Note that, obviously, any relabeling of the outcomes must alter the outcome at the first time step. Note further that all subsequent outcomes specify the ORE class and any sequence of these outcomes is possible.

For each measurement setting there are 2 possible outcomes in the first time step and therefore there are $2^S$ possible relabelings of the outcomes at the first time step. This implies that the cardinality of an ORE class is given by $2^S$. Using that the number of extreme points of $P_{2,S}$ is given by $(2^S)^S$ (see reference [19]), the number of ORE classes is then

$$\frac{(2^S)^S}{2^S} = (2^S)^{\frac{S-1}{S}}, \quad (7)$$

which concludes the proof.

For the relabeling of the measurement settings, we first present a counting argument for $O = S = 2$. Then, we extend our investigation to the case of $O = 2$ and $S = 3$.

Lemma 2. The number of RE classes of extreme points of $P_{L,2}$ is given by $\frac{1}{2}(4^{2L-2} + 4^{(2L-1)-1})$.

Proof. As we know already from lemma 1 the number of equivalence classes under the relabeling of outcomes, it is sufficient to study the action of the permutation of the inputs on the representatives of the ORE classes. Since $S = 2$, we have the group $S_2 = \{e, S_{12}\}$, where $e$ is the identity element and $S_{12}$ exchange the first and second input (see also figure 4). In particular $S_2^2 = e$. For a given length $L$, the number of equivalence classes $N^{(L)}$ is given by

$$N^{(L)} = N_1^{(L)} + N_N^{(L)}, \quad (8)$$

where $N_1^{(L)}$ is the number of orbits consisting of only one vector, i.e., the vectors invariant under the action of the group (see figure 5), and $N_N^{(L)}$ the number of orbits consisting of two vectors, i.e., vectors not invariant under the action of the group. To count the number of invariant vectors, we apply the following
Figure 5. Equivalent tuples for an extreme point invariant under settings relabeling, in the $O = S = 2$ scenario. The operator $X$ permutes the elements of the tuple. E.g., the tuple $\Gamma_{2,2}$ corresponds to the outcomes for $M_1$ and $M_2$ in the second step, after $M_2$ has been measured. For the extreme point to be invariant under relabeling of outcomes, it must be that $\Gamma_{2,2} = X\Gamma_{2,1}$, i.e., the outcomes are, up to relabeling, those that would have been obtained had we measured $M_1$ in the first step.

Table 1. This table shows a representative for each of the 10 RE classes for $O = S = L = 2$ and the minimal dimension which allows to reach a member of the class (see also theorem 5). Note that for $O = S = L = 2$ the RE classes and their minimal dimension has been already identified in [37] and a corresponding table can be also found there (with a different choice of representatives).

| Extreme point [with $\Gamma_0 = (0, 0)$] | Minimal dimension |
|----------------------------------------|-------------------|
| $\Gamma_1 = (0, 0), \Gamma_2 = (0, 0)$ | 1                 |
| $\Gamma_1 = (0, 0), \Gamma_2 = (1, 1)$ | 2                 |
| $\Gamma_1 = (0, 0), \Gamma_2 = (0, 1)$ | 2                 |
| $\Gamma_1 = (0, 1), \Gamma_2 = (0, 1)$ | 2                 |
| $\Gamma_1 = (1, 1), \Gamma_2 = (1, 1)$ | 2                 |
| $\Gamma_1 = (0, 1), \Gamma_2 = (1, 0)$ | 3                 |
| $\Gamma_1 = (1, 0), \Gamma_2 = (0, 1)$ | 3                 |
| $\Gamma_1 = (0, 1), \Gamma_2 = (1, 1)$ | 3                 |
| $\Gamma_1 = (1, 1), \Gamma_2 = (0, 1)$ | 3                 |

argument. First, let us fix the outcome relabeling by choosing the first input as '0'. Given that the sequence is invariant under exchange of inputs until step $m-1$, the possible ways of completing it in an invariant way at the step $m$ are $4^{2m-2}$ out of $4^{2m-1}$ possible completions. In fact, for each fixed choice of inputs and outputs at the step $m-1$ there are 4 possible ways of completing the last two outcomes for the two settings. Out of the $2^{m-1}$ possible choices of inputs, we need to fix only half of them, because the other half is fixed by the input exchange symmetry $S_{12}$, see figure 5.

The number of possibilities to extend a vector at the step $m$, given that it is symmetric at the step $m-1$, is thus $4^{2m-2}$. We can then compute the number of invariant vectors up to length $L$ as

$$N_1^{(L)} := \prod_{i=2}^{L} 4^{2m-2} = 4^{4L-4}. \quad (9)$$

Moreover, we can compute the number of equivalence classes of non-invariant vectors as

$$N_N^{(L)} = \frac{1}{2} \left( N_{ORE}^{(L)} - N_1^{(L)} \right), \quad (10)$$

where $(N_{ORE}^{(L)} - N_1^{(L)})$ is the number of non-invariant vectors, and the factor $1/2$ comes from the fact that each orbit contains two elements.

Finally, we can write

$$N_N^{(L)} = N_1^{(L)} + N_N^{(L)} = \frac{1}{2} \left( N_{ORE}^{(L)} + N_1^{(L)} \right) = \frac{1}{2} \left( 4^L - 2 + 4^{L-1} \right). \quad (11)$$

□
For the most simple scenario, $O = S = L = 2$, this implies that there are ten RE classes. As already discussed in [19], six of these classes can be obtained with a qubit, whereas for four of these classes a qutrit is required. In Table 1, we provide a representative for each of these classes and indicate whether a qubit or a qutrit is necessary in order to realize a member of this class. In the following section, we then present a general theorem which allows us to deduce from a given extreme point (with arbitrary $O$, $S$ and $L$) the dimension that is necessary and sufficient to realize it.

After having gained an understanding of the case of two inputs, we generalize our approach for counting the RE classes to the case of three inputs.

**Lemma 3.** The number of RE classes of extreme points of $P_L^{2,3}$ is given by $2^{\frac{L+1}{2}} + 3^{\frac{L-1}{2}} - 1 + \frac{1}{4} \left[ \frac{3^{\frac{L-1}{2}}}{2} \right] + \frac{1}{2^{L+1}}$.

In order to prove the lemma, we start again from the equivalence classes of outcome relabeling and impose only the conditions for the relabeling of the inputs. In this case, we have the permutation group of three elements, $S_3$, representing the permutation of the inputs. The group $S_3$ consists of the following elements

$$e, S_{12}, S_{23}, S_{13} = S_{23}S_{12}S_{23} = S_{12}S_{23}S_{12},$$

$$\sigma_{123} = S_{12}S_{23} = S_{23}S_{13},$$

$$\sigma_{132} = S_{13}S_{23} = S_{23}S_{12}$$

We therefore write the total number of equivalence classes as

$$N(L) = N^{(L)}_1 + N^{(L)}_2 + N^{(L)}_4 + N^{(L)}_N, \quad (13)$$

where $N^{(L)}_1, N^{(L)}_2$ are defined as above, as the orbits of vectors that are invariant or non-invariant under any symmetry respectively, and $N^{(L)}_4, N^{(L)}_N$ are the orbits of vectors invariant under only one of the $S_j$ (vectors invariant only under $\sigma_{123}$ or $\sigma_{132}$) respectively. Note here that $S_0 = e$ and that $\sigma_{123}\sigma_{132} = \sigma_{132}\sigma_{123} = e$. In appendix A we count the number of invariant and non-invariant orbits and with this prove Lemma 3.

### 4. Minimal dimension for given extreme points

It is a well known result in quantum state discrimination that two states have orthogonal ranges, corresponding to a trace-distance of 1, if and only if they can be perfectly discriminated, i.e., with probability one are orthogonal.

Using this, we are able to determine the minimum dimension that is required for a quantum system to obtain a given extreme point of $P_2^{O,S}$.

**Theorem 5.** Given an extreme point $p$ of $P_2^{O,S}$, the minimal dimension $d$ needed to obtain it is given by the number of inequivalent tuples in the history of $p$, i.e., $T_{1,1} = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_{25}\}$. In particular, a system with dimension $d = S + 1$ can always reach all extreme points of $P_2^{O,S}$, independently of the number of outcomes, as this is the maximal number of tuples in $T_{1,1}$.

**Proof.** According to equation (5), to different futures correspond different states. In this particular case, namely, $L = 2$, we need to compare different $T_{ij}$, i.e., single tuples. By Observation 4, such states must have orthogonal ranges. These two conditions already provide the minimal number of orthogonal states necessary to reach a given extreme point of the AoT polytope. Intuitively, orthogonality is the only relevant property for obtaining different futures, hence a minimal realization requires only pure states. This is...
confirmed by the explicit construction below, which uses only pure states. Given the tuples
\( \Gamma_{1,1}, \Gamma_{1,2}, \ldots, \Gamma_{2,5} \), there may be repetitions, which in this simple case of \( L = 2 \) corresponds to having
equivalent futures. We can rewrite them as \( d \) tuples \( \{\Gamma_1, \ldots, \Gamma_d\} = \{\Gamma_{1,1}, \Gamma_{1,2}, \ldots, \Gamma_{2,5}\} \) with inequivalent futures, i.e., \( \Gamma_i \neq \Gamma_j \). We associate to each of them a vector \( |k\rangle \) from the ONB \( \{\{|k\rangle\}_{k=1}^{d}\} \). Without loss of
generality, we can assume that \( \Gamma_{1,1} = \Gamma_1 = (0, 0, 0) \), i.e., we fix all the measurement outcomes at the first
step to be zero. This simply means that we relabel the outcome of all measurements such that \( 0 \) is obtained for
all of them on the initial state. Then, we fix the initial state as \( \rho_{10} = |1\rangle\langle 1| \). The measurements are
constructed as follows
\[
E_{ax} = \sum_{j \in J_{ax}} |j\rangle\langle j|, \quad \text{with} \quad J_{ax} := \{ j \in \{1, \ldots, d\} \mid [\Gamma_j]_x = a \},
\] (14)
for \( a = 0, \ldots, O - 1, x = 1, \ldots, S \). Clearly, \( E_{ax} \geq 0 \) and \( \sum_a E_{ax} = 1 \), for all \( x \), so they are valid POVMs. In
particular, \( E_{0x} = |1\rangle\langle 1| + \sum_{j \in J_{0x}} |j\rangle\langle j| \) for all \( x \). The corresponding Kraus operators \( \{K^j_{0x}\}_{j \in J_{0x}} \)
providing the postmeasurement state, i.e., in the Schrödinger picture \( \rho \mapsto \sum_j K^j_{0x} \rho K^j_{0x}^\dagger \), are of the form
\( K^j_{0x} = |j\rangle\langle j| \) if \( \Gamma_i = \Gamma_{2,x} \) or a tuple with equivalent future and \( K^j_{0x} = |j\rangle\langle j| \) for \( j \in J_{0x} \) and \( j \neq 1 \). By
construction, there are at most \( S + 1 \) tuples, hence this number provides an upper bound on the minimal
dimension necessary to reach any extreme point of \( P_2^{1,1S} \). \( \square \)

We now discuss a specific example in \( P_2^{1,1S} \) to illustrate how theorem 5 can be applied in order to
determine the dimension which is necessary and sufficient to realize a given extreme point. For this we
consider the extreme point given by \( \Gamma_{1,1} = (0, 0, 0), \Gamma_{1,2} = (0, 0, 0), \Gamma_{2,2} = (1, 1, 1) \) and \( \Gamma_{2,3} = (0, 0, 1) \).
There are three inequivalent tuples in the history given by \( \Gamma_{1,1}, \Gamma_{1,2} \) and \( \Gamma_{2,3} \), as \( \Gamma_{1,1} = \Gamma_{2,1} \). Hence, we have
that any system which can realize this extreme point has at least dimension three. The Kraus operators of
the measurements \( x \in \{1, 2, 3\} \) for outcome 0, which allow us to obtain this extreme point from the initial
state \( \rho_{10} = |1\rangle\langle 1| \), can be chosen as \( K^j_{0x} = |x\rangle\langle 1| \) for all \( x \) and \( K^0_{0x} = 3/3 \) for \( x = 1, 2 \) (for \( x = 3 \) there
is only a single Kraus operator). Note that for the outcome 1 it is sufficient to know the effects (as the
post-measurement state does not need to be specified), which directly follow from the ones for outcome 0.
That is, we have \( \mathcal{E}_{1x} = |2\rangle\langle 2| \) for \( x = 1, 2 \) and \( \mathcal{E}_{13} = |2\rangle\langle 2| + |3\rangle\langle 3| \).

The same argument used to derive theorem 5 can be generalized to sequences of arbitrary length.

Theorem 6. The minimal dimension \( d \) required to reach an extreme point \( p \) of \( P_1^{1,1S} \), is given by the number of
inequivalent futures \( T_{k,l} \) in the history \( T_{1,1} \).

Proof. The proof generalizes straightforwardly from the case \( L = 2 \) above. Again, different futures must
correspond to different states, such states must be orthogonal, and can be chosen to be pure, providing a
minimal-dimension representation.

The explicit construction of the model can then be extended from the previous one. Let us assign the
state \( |l\rangle \) as initial state, i.e., to \( T_{1,1} \). Then compare \( T_{1,1} \) with \( T_{2,2} \), for \( k = 1, \ldots, S \), if they are equivalent,
assign the same state \( |l\rangle \) to \( T_{2,2} \), otherwise, assign a new orthogonal state \( |2\rangle \), \( |3\rangle \), . . . to the future \( T_{2,2} \).
Repeat the same operation for \( T_{3,3}, k' = 1, \ldots, S' \), assigning new state for any future inequivalent to \( T_{1,1} \) or
any \( T_{2,2} \). Repeat again until the end of the tree, i.e., \( T_{L,2} = \Gamma_{L,2} \), \( k = 1, \ldots, S^{L-1} \). To each node \( \Gamma_{L,2} \) of the tree
\( T_{L,1} \) is then assigned a pure state \( |l,k\rangle \in \{|j\rangle\}_{j=1}^{d} \), possibly with repetitions. As in theorem 5, POVMs
elements are constructed as projectors providing the correct outcomes for each state \( |j\rangle \), as in equation (14).
Similarly, Kraus operators \( \{K^j_{0x}\} \) associated with the POVM element \( E_{0x} = \sum_{j \in J_{0x}} |j\rangle\langle j| \) consist in
measure-and-prepare operations \( K^j_{0x} = |j\rangle\langle j| \), when the state \( |j\rangle \) emits the output \( a \) for the measurement \( x \)
and transition to the state \( |a\rangle \), all with probability one. For the last time-step, i.e., from \( L - 1 \) to \( L \), one can
use diagonal Kraus operators analogously to the construction in theorem 5. \( \square \)

Note that the above protocol does not involve any coherences, as all states and effects are diagonal in the
same basis and the state-update rule also involves transitions within the same basis, hence, it can be realized
with a classical system.

5. Lower bound on the dimension which is necessary to realize any extreme point

In the following, we provide a construction of extreme points for any \( L \) from which one can determine a
lower bound on the minimal dimension required for its realization. This lower bound is then automatically
also a lower bound on the dimension necessary to realize any extreme point. The main result is that the
minimal dimension scales, at least, exponentially in \( L \). Let us consider the polytope \( P_2^{1,1S} \). The main idea of
the proof can be briefly explained as follows. Consider a sequence of length \( L \) and take a time-step \( j < L \). If
the number of remaining time-steps $L - j$ is big enough, for the tuples $\{\Gamma_{j,\ell}\}$, we can choose their futures $\{T_{j,\ell}\}$ to be all different, hence, each $\Gamma_{j,\ell}$ will be associated to an orthogonal state and the number of such tuples will provide a lower bound on the minimal dimension necessary for their realization. Our argument consists in estimating the maximum $j$ such that this is possible.

For the history $T_{i,1}$ at time-step $j$, there exist $S_j^{-1}$ different subtrees $T_{j,\ell}$. If $j$ is properly chosen, namely, $L - j$ is large enough such that we can construct different futures $T_{j,\ell}$ for $k = 1, \ldots, S_j^{-1}$, then the realization of such an extreme point requires at least $d = S_j^{-1}$. First notice that the number of possible futures of length $x$ is given by $(O^x)^{S_j^{-1} + 1} = O^{S_j^{-1} + 2}$. Hence, $j$ must be selected in such a way that the remaining sequence allows us to assign different futures (which are of length $L - j$) to each node, namely as the largest integer such that

$$S_j^{-1} \leq O^{\frac{S_j^{-1} + 2 - x}{S_j^{-1} + 2}}.$$  

We can further simplify the expression using the identity $O^x = S^x O$

$$S_j^{-1} \leq S_j^{-1} + 2 - x \Rightarrow j - 1 \leq \frac{S_j^{-1} - S}{S - 1} \log_2 S = S_j^{-1} \left( \frac{S_j^{-1} - S}{S - 1} \log_2 O \right) = \frac{S_j^{-1} - S}{S - 1} \log_2 O.$$  

This equation can be solved in terms of the principal branch of Lambert function $W$, namely, the function implicitly defined as the solution to the equation $xe^x = k$, i.e., $xe^x = k \iff x = W(k)$. In this case, let us see how to solve it for the equation $x = aS^{-x} + b$

$$x = aS^{-x} + b \iff (x - b)S^{-b} = aS^{-b} \iff (x - b)e^{(x - b)\ln S} \ln S = aS^{-b} \ln S,$$

We have $ye^y = aS^{-b} \ln S$ for $y := (x - b) \ln S \iff y = W(aS^{-b} \ln S) \iff x = \frac{W(aS^{-b} \ln S)}{\ln S} + b$.

By substituting $x = j - 2$, $a = \left(\frac{S_j^{-1}}{S_j^{-1} + 2}\log_2 O\right)$, $b = -\frac{S}{S_j^{-1} + 2} \log_2 O$ one obtains the condition

$$j \leq W \left[ \left(\frac{S_j^{-1}}{S_j^{-1} + 2}\log_2 O\right) \frac{S_j^{S_j^{-1}} \log_2 O \ln S}{\ln S} \right] = \frac{S_j^{-1} - S}{S - 1} \log_2 O + 2,$$

which gives the maximal $j$ as

$$j = \left[ W \left[ \left(\frac{S_j^{-1}}{S_j^{-1} + 2}\log_2 O\right) \frac{S_j^{S_j^{-1}} \log_2 O \ln S}{\ln S} \right] = \frac{S_j^{-1} - S}{S - 1} \log_2 O + 2 \right],$$

where $\lfloor x \rfloor$ denotes the floor of $x$, i.e., the largest integer smaller than $x$.

To compute the asymptotic scaling, one can write

$$S_j^{-2} \leq O^{\frac{S_j^{-1} - S}{S - 1} \log_2 O},$$

Using that $\ln(x) - \ln[\ln(x)] + \frac{\ln\ln(x)}{\ln(x)} \leq W(x)$ for $x \geq e^{45}$ we can obtain a lower bound on the minimal dimension as follows. For $m \in \mathbb{R}$ such that

$$S_j^{m - 2} \leq O^{\frac{S_j^{-1} - S}{S - 1} \log_2 O} \left( \ln \left[ \frac{S_j^{\frac{S_j^{-1} - S}{S - 1} \log_2 O}}{\ln O} \right] \right)^{-1 - \frac{\ln \left[ \frac{S_j^{\frac{S_j^{-1} - S}{S - 1} \log_2 O}}{\ln O} \right]}{\ln \left[ \frac{S_j^{\frac{S_j^{-1} - S}{S - 1} \log_2 O}}{\ln O} \right]}},$$

the minimal dimension satisfies

$$d_{\text{min}} \geq S_j^{m - 2},$$

where the ‘$-2$’ term takes into account the fact that $m$ may not be an integer. For large $L$ such a lower bound scales as

$$AS_j^{-1}(\ln[BS_j^{1/4}])^{-1+\frac{1}{2}(\ln[BS_j^{1/4}])^{-1}} \approx A e^{\left[\frac{1}{2} - \gamma\right] \ln L},$$

for appropriately chosen constants $A, B, \alpha, \beta, \gamma, \delta$. This proves that the minimal dimension required to reach any extreme point scales at least exponentially (up to logarithmic corrections). In appendix C, we present a different construction of an extreme point which can provide an improved lower bound on the minimal dimension, however no closed formula on the scaling.
6. Combining temporal inequalities

In the following, we present a method for deriving new inequalities for temporal correlations for sequences of length $nL$ with $n \in \mathbb{N}^+$, based on the knowledge of inequalities for the shorter length $L$. It instructive to first describe the method for a simple example, based on the inequalities for the case $L = O = S = 2$ derived in reference [19]. The original inequalities were derived by computing the qubit bound for expressions of the form $\sum_{x,y} p(a_i,b_j|x,y)$ for a specific choice of the outputs $\{a_i, b_j\}_{x,y}$ where the algebraic bound $4$ is achieved by a extreme point of AoT polytope $P_{2,2}$, i.e., $p(a_i,b_j|x,y) = 1$ for all $x,y$. Up to symmetries, four expression were derived, namely

$$
B_1 := p(00|00) + p(00|11) + p(01|01) + p(01|10) \leq C_1,
B_2 := p(01|00) + p(01|11) + p(00|01) + p(00|10) \leq C_2,
B_3 := p(01|00) + p(00|11) + p(01|01) + p(01|10) \leq C_3,
B_4 := p(01|00) + p(01|11) + p(01|01) + p(00|10) \leq C_4.
$$

(24)

Each one corresponding to one of the extreme points of the AoT polytope $P_{2,2}$, which cannot be reached by qubit strategy, namely,

$$
e_1 : p(00|00) = p(00|11) = p(01|01) = p(01|10) = 1, \text{ and otherwise;}
$$

$$
e_2 : p(01|00) = p(01|11) = p(00|01) = p(00|10) = 1, \text{ and otherwise;}
$$

$$
e_3 : p(01|00) = p(00|11) = p(01|01) = p(01|10) = 1, \text{ and otherwise;}
$$

$$
e_4 : p(01|00) = p(01|11) = p(01|01) = p(00|10) = 1, \text{ and otherwise.}
$$

(25)

In general, to each extreme point $e_i$ with components labeled by $\vec{a} = (a_1, \ldots, a_l), \vec{x} = (x_1, \ldots, x_L)$, i.e., $|e_i|_{\vec{a},\vec{x}} = p(\vec{a}|\vec{x})$, we can associate a temporal inequality

$$
B_i = \sum_{\vec{a},\vec{x}} c_{\vec{a},\vec{x}} p(\vec{a}|\vec{x}) \leq C_i
$$

(26)

where $c_{\vec{a},\vec{x}} := |e_i|_{\vec{a},\vec{x}}$, $c_{\vec{a},\vec{x}} \in \{0,1\}$ since $e_i$ is a deterministic strategy, and $C_i$ is the bound for a given dimension $d$ of the quantum system, corresponding to the algebraic bound $\sum_{\vec{a},\vec{x}} c_{\vec{a},\vec{x}}$, if the extreme point $e_i$ can be reached in dimension $d$. Note that the same bounds also hold for any other element within the same RE class as $e_i$.

Given two deterministic strategies for length two, we can construct a strategy for length four simply by combining them in the following way:

$$
p(a_1, a_2, a_3, a_4|x_1,x_2,x_3,x_4) = p(a_1, a_2|x_1,x_2)p(a_3, a_4|a_1,a_2;x_1,x_2,x_3,x_4)
$$

(27)

where $p(a_1, a_2|x_1,x_2)$ and each $p(a_3, a_4|a_1,a_2;x_1,x_2,x_3,x_4)$, is a deterministic strategy associated with an extreme point $e_i \in \{e_1, e_2, e_3, e_4\}$, implying that

$$
p(\vec{a}_1|x_1), \vec{a}_2|x_1,x_2) \ p(\vec{a}_3|x_1,x_2,x_3), \vec{a}_4|x_1,x_2,x_3,x_4) \ |a_1,a_2;x_1,x_2,x_3,x_4) = 1,
$$

(28)

for properly chosen functions $\{\vec{a}_i\}$ of the inputs $\{x_i\}$, and 0 otherwise.

We denote the corresponding extreme point of $P_{2,2}$ as $e_k = (e_{k_1}, e_{k_2}, \ldots, e_{k_4})$ with $k = (k_1, k_2, \ldots, k_4)$, where $k_i$ labels the extreme point used for the first-time period, i.e., the two steps $t_1$ and $t_2$, and $k_i$ for $i = 1, \ldots, 4$ denote the possible extreme points for the second time period, i.e., steps $t_3$ and $t_4$, each belonging to one of the different branches of the evolution depending on the inputs $x_1, x_2$, as shown in equation (28). We, then, construct the associated inequality

$$
B_k = \sum_{\vec{a},\vec{x}} c_{\vec{a},\vec{x}} p(\vec{a}|\vec{x}) \leq C_k C_{k_3}
$$

(29)

where $C_k = \max_i C_{k_i}$, is the maximum taken for a given time period over all possible branches. The proof of the bound is straightforward.
Markovian time-evolutions can be still be absorbed in the definition of quantum instruments with a proper apparatus, e.g., a drift in the magnetic field orientation in the spin example. Notice, however, that notion of imperfect implementation, hence, does not deal with random fluctuations in the measurement still allowed, since such an uncertainty is already contained in the definition of quantum instrument. The representation of quantum instruments. This corresponds to take the dual important to remark that in the following, it is more convenient to use the Schrödinger picture for the assumption influences our results. Before proceeding further, it is helpful to remark what we mean by deviation from this measurement is implemented. Here, we discuss how a deviation from this example for this construction consider the extreme point \( C_{k} \) as an upper bound to the expression \( \sum_{x \in \mathcal{A}} p(\tilde{a}_{1}[x_{1}], \tilde{a}_{2}[x_{1}, x_{2}], \ldots) \sum_{x_{1}, x_{2}} p(\tilde{a}_{1}[x_{1}, x_{2}, x_{3}], \tilde{a}_{4}[x_{1}, x_{2}, x_{3}, x_{4}]) \sum_{x_{1}, x_{2}} p(\tilde{a}_{1}[x_{1}, x_{2}], \tilde{a}_{2}[x_{1}, x_{2}]; x_{1}, x_{2}, x_{3}, x_{4}) \) for any value \( x_{1}, x_{2}, \ldots, x_{n} \), and finally the bound \( C_{k} \).

It is obvious that the above result depends only on the way of choosing a deterministic strategy, i.e., an extreme point of the AoT polytope, as a product strategy as in equation (27), the way of constructing the corresponding expression \( B_{k} \), and the knowledge of the bounds \( \{ C_{k} \}_{i} \) for the single expressions \( \{ B_{k} \}_{i} \) for \( k = (k_{1}, \ldots, k_{n}) \). We can then generalize the result as follows.

**Theorem 7.** Given a collection of temporal inequalities involving \( O \) outcomes, \( S \) settings, and length \( L \),

\[
B_{i} = \sum_{x \in \mathcal{A}} c_{i}(x) p(\tilde{a}[x]) \leq C_{i}, \quad \text{for } i = 1, \ldots, N,
\]

associated to an extreme point \( e_{i} \) and valid for quantum systems of dimension \( d \), then, the following inequality for sequences of length \( nL \)

\[
B_{k} := \sum_{x} p(\tilde{a}_{1}[x_{1}], \tilde{a}_{2}[x_{1}, x_{2}], \ldots, \tilde{a}_{nL}[x_{1}, \ldots, x_{nL}]) \leq \prod_{j=1}^{n} C_{j},
\]

with \( k = (k_{1}, \ldots, k_{n}) \), \( c_{ij} \in \{ e_{1}, \ldots, e_{N} \} \) and \( C_{j} = \max_{i} C_{ij} \) also holds for quantum systems of the same dimension.

The proof of this theorem is analogous to the case of four time steps presented above. An explicit example for this construction consider the extreme point \( a_{k} = (e_{i}, \ldots, e_{i}) \) of length \( 2n \) with \( i \in \{ 1, 2, 3, 4 \} \) and \( n \in \mathbb{N}^{+} \) and the corresponding inequality \( B_{k} \). Then, according to theorem 7, it holds that for qubits \( B_{k} \leq (C_{i})^{n} \). Using theorem 6, it can be easily seen that with a three-dimensional system one can reach the algebraic maximum of \( B_{k} = 4^{n} \). It follows that the ratio of separation between a qubit and a qutrit is exponentially decreasing with the length of the sequence, i.e., \( (C_{i}/4)^{n} \).

### 7. Imperfect implementation of time-independent measurements

The results obtained so far assume that the measurements are time-independent, i.e. the same input indicates that also the same measurement is implemented. Here, we discuss how a deviation from this assumption influences our results. Before proceeding further, it is helpful to remark what we mean by imperfect implementation. What does it mean to ‘perform the same measurement twice’? Consider the basic example of an apparatus that measures the spin of a particle either along the X direction or the Z direction with probability 1/2 each. Clearly, in each round of the experiment when a sequence of two measurements is performed there is 50% chance that two different measurements are performed, i.e., \( X, Z \) or \( Z, X \). However, according to our definition of time-independent quantum instruments, this situation is still allowed, since such an uncertainty is already contained in the definition of quantum instrument. The notion of imperfect implementation, hence, does not deal with random fluctuations in the measurement apparatus, but rather with some time-dependent drift in the parameters describing the measurement apparatus, e.g., a drift in the magnetic field orientation in the spin example. Notice, however, that Markovian time-evolutions can be still be absorbed in the definition of quantum instruments with a proper choice of measurement times.

In the following, we quantify the effect of such imperfect implementations of quantum instruments on the observed correlations. Such deviations can be quantified in terms of the diamond norm [46]. It is also important to remark that in the following, it is more convenient to use the Schrödinger picture for the representation of quantum instruments. This corresponds to take the dual \( I^{+} \) of the instruments appearing in equation (1), acting now on states rather than observables. To avoid a heavy notation, however, we drop the superscript \(^{+}\) in the remaining part of this section.
If $\mathcal{I}_{a|xy}$ is the desired CP map for input $x$ and outcome $a$ and $\tilde{\mathcal{I}}_{a|xy}$ is the one that is instead implemented in the experiment then $\|\mathcal{I}_{a|xy} - \tilde{\mathcal{I}}_{a|xy}\|_\diamond \leq \epsilon$ for all $x$ and $a$, where the diamond norm of a CP map $\mathcal{I}$ is defined as $\|\mathcal{I}\|_\diamond := \max_{\rho_{AB}} ||\mathcal{I}_A \otimes \rho_{AB}||_\diamond$.

Note that from the definition of the diamond norm it straightforwardly follows that $\|\mathcal{I}_{a|xy} - \tilde{\mathcal{I}}_{a|xy}\|_\diamond \leq \|\mathcal{I}_{a|xy} - \tilde{\mathcal{I}}_{a|xy}\|_\diamond$ for all density matrices $\rho$. As we will see, this allows us to derive bounds on the influence of such a deviation on quantities that are linear in $p(ab \ldots |xy \ldots)$. In order to illustrate the basic idea we consider here first two time steps and then three time steps, however, it is straightforward to generalize the bound to an arbitrary number of time steps. In particular, we obtain that

$$p(ab|xy) = \text{tr}\{\tilde{\mathcal{I}}_{b|y}\{\tilde{\mathcal{I}}_{a|x}(\rho_{mn})\}\}$$

$$= \text{tr}\{(\tilde{\mathcal{I}}_{b|y} - \mathcal{I}_{b|y})\{\mathcal{I}_{a|x}(\rho_{mn})\}\} + \text{tr}\{\mathcal{I}_{b|y}\{\mathcal{I}_{a|x}(\rho_{mn})\}\} \leq \text{tr}\{\mathcal{I}_{b|y}\{\mathcal{I}_{a|x}(\rho_{mn})\}\} + \epsilon \text{ tr}\{\mathcal{I}_{a|x}(\rho_{mn})\} \tag{33}$$

and

$$p(abc|xyz) = \text{tr}\{\tilde{\mathcal{I}}_{c|z}\{\tilde{\mathcal{I}}_{b|y}\{\tilde{\mathcal{I}}_{a|x}(\rho_{mn})\}\}\}$$

$$= \text{tr}\{(\tilde{\mathcal{I}}_{c|z} - \mathcal{I}_{c|z})\{\tilde{\mathcal{I}}_{b|y}\{\tilde{\mathcal{I}}_{a|x}(\rho_{mn})\}\}\} + \text{tr}\{\mathcal{I}_{c|z}\{\mathcal{I}_{b|y}\{\mathcal{I}_{a|x}(\rho_{mn})\}\}\} \leq \text{tr}\{\mathcal{I}_{c|z}\{\mathcal{I}_{b|y}\{\mathcal{I}_{a|x}(\rho_{mn})\}\}\} + \epsilon \text{ tr}\{\mathcal{I}_{a|x}(\rho_{mn})\} \text{ tr}\{\mathcal{I}_{b|y}\{\mathcal{I}_{a|x}(\rho_{mn})\}\} + 2\epsilon \text{ tr}\{\mathcal{I}_{a|x}(\rho_{mn})\}. \tag{34}$$

Note that we used here multiple times that $\text{tr}\{\tilde{\mathcal{I}}_{a|x}(\rho) - \mathcal{I}_{a|x}(\rho)\} \leq \|\mathcal{I}_{a|x} - \tilde{\mathcal{I}}_{a|x}\|_\diamond \leq \epsilon$ with $\text{tr}(\rho) = 1$, that $\text{tr}\{\tilde{\mathcal{I}}_{b|y}\{\tilde{\mathcal{I}}_{a|x}(\rho_{mn})\}\} = \text{tr}\{\mathcal{I}_{a|x}(\rho_{mn})\} \text{ tr}\{\tilde{\mathcal{I}}_{b|y}\{\tilde{\mathcal{I}}_{a|x}(\rho_{mn})\}\}$, where $\tilde{\mathcal{I}}_{a|x}$ denotes the normalized post-measurement state for measurement $x$ and outcome $a$, and that $\text{tr}\{\mathcal{I}_{c|z}\{\tilde{\mathcal{I}}_{b|y}\{\tilde{\mathcal{I}}_{a|x}(\rho_{mn})\}\}\} \leq \|\mathcal{I}_{c|z} - \tilde{\mathcal{I}}_{c|z}\|_\diamond \text{ tr}\{\mathcal{I}_{a|x}(\rho_{mn})\} \leq \|\tilde{\mathcal{I}}_{b|y} - \mathcal{I}_{b|y}\|_\diamond \text{ tr}\{\mathcal{I}_{a|x}(\rho_{mn})\} \leq \epsilon \text{ tr}\{\mathcal{I}_{a|x}(\rho_{mn})\}$, which follows from the contractivity of the trace distance under completely positive trace-nonincreasing operations. From this argument it can be easily seen that the deviation of the probability $p(ab \ldots |xy \ldots)$ for $t$ time steps due to imperfect time-independent measurements can be bounded by $(t - 1)\epsilon \text{ tr}\{\mathcal{I}_{a|x}(\rho_{mn})\}$. Hence, temporal inequalities are still able to provide a reliable lower bound on the dimension.

As a final remark, it is interesting to notice the following. The above argument assumes certain quantum properties of the operations involved, hence, at least some partial characterization of the experimental devices. However, we simply noticed that instruments that are ‘close’ in the quantum mechanical sense (and arguably the diamond norm is the natural distance among them) give rise to probability distributions that are again ‘close’, with an error that scales linearly in the measurement length. Assumptions on such a distance, even if based on quantum mechanics, do not necessarily require a full characterization of the experimental devices. It would be interesting to estimate the diamond norm in a device independent way, then our result helps in the design of improved experimental tests of temporal quantum correlations that rely on minimal assumptions and do not require a complete characterization of the measurement devices.

8. Conclusion and outlook

In this work, we studied the resources required to realize AoT correlations within quantum mechanics. We first identified which extreme points of the AoT polytope can be obtained by using the same protocol followed by some classical post-processing of the input and output for the case of a small number of two-outcome measurements. Then we provided for an arbitrary given extreme point the dimension that is necessary and sufficient to realize it. In particular, we showed that this is given by the number of inequivalent futures in the history associated with a point and we gave an explicit protocol that allows one to obtain it. We observed that this protocol does not involve any coherences and hence can be also implemented with a classical system. Moreover, we derived a lower bound on the minimal dimension that is necessary to reach an arbitrary extreme point for a given number of settings $S$, outcomes $O$ and time steps $L$ and we showed that in the asymptotic limit of long sequences this scales as $e^{\alpha L}/L$ (with $\alpha$ being some constant that depends on $O$ and $S$).

In a previous work [19], extreme points of the AoT polytope have been used to construct dimension witnesses for sequences of short length. Here, we provided a general method to use these witnesses as building blocks for the construction of dimension witnesses for sequences of arbitrary length. Despite the fact that the bound on the so obtained temporal inequality is not necessarily tight, one finds inequalities
which show an exponential scaling with respect to the length of the sequence. Finally, we made quantitative statements on how the bounds on linear temporal inequalities are affected if the assumption that at any time step one is able to implement the same measurement is violated. We showed that small deviations from these assumptions still allow us to deduce lower bounds on the dimension.

There are several possible directions for future research. First, one can consider a general point in the correlation polytope and consider the resources needed for a simulation. This problem is challenging for two reasons: first, the quantum realizations for arbitrary points are difficult to find, and may require a larger dimension than the extremal points [41]. Second, for a general point in the polytope a deterministic protocol is not suitable, so more general concepts, such as hidden Markov models [47] or, more specifically, $\varepsilon$-transducers [48, 49] may be useful.

A second interesting problem comes from the observation that our simulation protocols were purely classical, in the sense that they can be implemented using quantum states diagonal in the computational basis. It would be interesting to develop a general theory of temporal correlations, for which the quantum mechanical simulation requires less resources than the classical one, due to effects like coherence [50]. This may open a further way to test quantum devices using temporal correlations.

Acknowledgments

We thank Jannik Hoffmann, Yuanyuan Mao, and Zhen-Peng Xu for useful discussions. This work has been supported by the Austrian Science Fund (FWF) through the projects J 4258-N27 (Erwin-Schrödinger Programm), Y879-N27 (START project), ZK 3 (Zukunftskolleg), and F7113-N48 (BeyondC), by the Austrian Academy of Sciences and by the ERC (Consolidator Grant 683107/TempoQ).

Appendix A. Proof of lemma 3

In the following we prove lemma 3, i.e. we show that the number of RE classes of extreme points of the polytope $P_{2,3}^{L}$ is given by

$$2^{\frac{L-1}{2}} + 2 \left[ 2^{\frac{L-3}{2}} + 2^{\frac{L-1}{2}} + 1 \right].$$  \hspace{1cm} (A1)

Proof. As mentioned in the main text, we consider the ORE classes and impose the conditions for the relabeling of the inputs. Hence, the relevant symmetry group has the following elements

\begin{align*}
e \\
S_{12} \\
S_{23} \\
S_{13} &= S_{23}S_{12}S_{23} = S_{12}S_{23}S_{12} \\
\sigma_{123} &= S_{12}S_{23} = S_{23}S_{13} \\
\sigma_{132} &= S_{13}S_{23} = S_{23}S_{12}.
\end{align*} \hspace{1cm} (A2)

The total number of equivalence classes can be written as

$$N^{(L)} = N^{(L)}_{I} + N^{(L)}_{S} + N^{(L)}_{\sigma} + N^{(L)}_{N},$$ \hspace{1cm} (A3)

where $N^{(L)}_{I}$, $N^{(L)}_{S}$ are the number of orbits of vectors that are invariant or non-invariant under the whole symmetry group respectively, and $N^{(L)}_{\sigma}$ ($N^{(L)}_{N}$) count the orbits of vectors invariant under only one of the $S_{ij}$ (vectors invariant only under $\sigma_{123}$ or $\sigma_{132}$) respectively. Notice that $S^{0}_{ij} = e$ and that $\sigma_{123}\sigma_{132} = \sigma_{132}\sigma_{123} = e$.

Let us start with the set of invariant vectors. For $S = 3$ the possible independent sequences of settings, i.e., sequences that are not generated one from the other by exchanging some settings, are given by:

\begin{align*}
l = 1 &: X; \quad l = 2 &: XX, XY^*; \quad l = 3 &: XXY; \quad \text{XXX, XYY, XYZ}
\end{align*} \hspace{1cm} (A4)

It is clear that, since $S = 3$, at each step the sequence of all identical measurements generate only two new sequences, i.e., XXX ... X and XXX ... XY, whereas all the other generate three new ones. We can then count

\begin{align*}
l = 1 &: X; \quad l = 2 &: XX, XY^*; \quad l = 3 &: XXY; \quad \text{XXX, XYY, XYZ}
\end{align*} \hspace{1cm} (A4)
the number of such sequences as \( Q_1 = 1 \) and \( Q_m = 3(Q_{m-1} - 1) + 2 = 3Q_{m-1} - 1 \), giving

\[
Q_m = \frac{1}{2}(1 + 3^{m-1}).
\]  

(A5)

At each step \( m \), then, we need to choose \( 2^{Q_m} \) possible values, i.e., two values for each extra measurement setting added, and we do not count the step \( m = 1 \), since this is fixed by the outcome relabeling symmetry. We then have

\[
N'^{(L)}_S = \prod_{i=2}^{L} 2Q_i = 2^{\sum_{i=2}^{L} i^{(1+\gamma^{-1})}} = 2^{\frac{L(L-1)+L}{\gamma}}.
\]  

(A6)

For calculating \( N^{(L)}_S \), we can first observe that if a vector \( v \) is invariant under the action of \( S_{12} \), i.e., \( S_{12}v = v \), then the corresponding orbit is given by \( O_i = \{v_{12}, v_{23}, v_{13}\} \), where \( v_{ij} \) is a vector invariant under the action of \( S_{ij} \). It is sufficient to look at the case of \( S_{12} \). If \( S_{12}v = v \), we define \( v_{12} := v \). Then, by action of the group we obtain \( v_{23} := S_{13}v \) and \( v_{13} := S_{23}v \). It can be straightforwardly verified that \( v_{23} \) is invariant under the action of \( S_3 \). In fact, \( S_{23}S_{23} = S_{23}S_{13}v = S_{23}S_{23}S_{13}S_{23}v = S_{13}S_{23}v = S_{12}S_{23}S_{12}v = S_{13}v = v_{23} \). A similar argument shows that \( v_{13} \) is invariant under \( S_3 \). Hence, each orbit of a vector invariant under exactly on \( S_2 \) generate the other two invariant vectors. We now need to count the number of representatives of such orbits.

\[
N^{(L)}_S = N'^{(L)}_S - N^{(L)}_I
\]  

(A7)

where \( N^{(L)}_S \) is the number of representative vectors invariant under exactly one \( S_{ij} \), but not fully invariant, \( N'^{(L)}_S \) is the number of vectors invariant under at least one of the \( S_{ij} \), but possibly fully invariant. If the vectors are invariant under two \( S_{ij} \), they are completely invariant, hence they are counted in \( N^{(L)}_I \). Note that the number of vectors invariant under exactly one \( S_{ij} \) is given by \( 3N'^{(L)}_S \), since each equivalence class contains three vectors.

For each sequence, say \( M_1 M_2 M_3 M_2 M_3 M_2 \), if we fix the outcomes of the sequences and then swap \( M_1 \leftrightarrow M_2 \), we obtain another sequence. This applies to all of them, except the sequence \( M_2 M_3 M_2 M_3 \) which remains invariant. This implies that, assuming to be invariant up to step \( m - 1 \), at step \( m \) we have \( (2^3)^\frac{m-1}{2} \) ways of extending these sequences from time step \( m - 1 \) to time step \( m \). For the case \( M_2 M_3 M_2 M_3 \) there are \( 2^2 \) possible assignments, giving in total \( (2^3)^\frac{m-1}{2} + 2^2 \). Again, we do not count the step \( m = 1 \) since it is fixed by outcome relabeling symmetry. We can then compute

\[
\tilde{N}^{(L)}_S = \prod_{i=2}^{L} \left(2^3\right)^{\frac{m-1}{2} + \frac{1}{2}} = 2^{\frac{L(L-1)}{2} + \frac{L}{2} - 1}
\]  

(A8)
implies that each orbit contains two invariant vectors. Then, with the same notation as in equation (A7) and the factor 1/2 coming from the fact that each orbit contains two invariant vectors. Then, \( \tilde{N}_\sigma^{(L)} \) can then be computed as follows. Given that a sequence of measurement outcomes is symmetric under cyclic permutation up to the step \( m - 1 \), there are \( 2^{3^{m-1}} \) way of completing it while still keeping it symmetric (see also figure 7).

Hence, we have that

\[
\tilde{N}_\sigma^{(L)} = \prod_{i=2}^{L} 2^{3^{m-1}} = 2^{\frac{m-3}{2}}. 
\]

Finally, we need to compute the number of orbits for vectors that are not invariant under any permutation. These can be obtained by removing all invariant ones from the total and divide by six, i.e., the number of vectors for each orbit, namely

\[
N_S^{(L)} = \frac{1}{6} \left[ N_{\text{ORE}}^{(L)} - \left( N_1^{(L)} + 3N_S^{(L)} + 2N_1^{(L)} \right) \right] = \frac{1}{6} \left[ N_{\text{ORE}}^{(L)} - 3N_S^{(L)} - \tilde{N}_\sigma^{(L)} + 3N_1^{(L)} \right] \tag{A11}
\]

Finally, we have

\[
N_S^{(L)} = \left[ N_1^{(L)} + \left( \tilde{N}_\sigma^{(L)} - N_1^{(L)} \right) + \frac{1}{2} \left( \tilde{N}_\sigma^{(L)} - N_1^{(L)} \right) + \frac{1}{6} \left( N_{\text{ORE}}^{(L)} - 3N_S^{(L)} - \tilde{N}_\sigma^{(L)} + 3N_1^{(L)} \right) \right]
\]

\[
= \left[ \frac{N_1^{(L)}}{6} + \frac{1}{2} N_S^{(L)} + \frac{2}{6} \tilde{N}_\sigma^{(L)} \right] = \frac{\frac{1}{2} + \frac{1}{2} \frac{3(3)^{m-3}}{2^{3^{m-1}} - 1}} {2^{3^{m-1}} + 2^{\frac{m-3}{2}} + 1}, \tag{A12}
\]

wich proves the lemma.
Appendix B. Proof of observation 4

For completeness we prove here observation 4 which is a well known result in quantum state discrimination (cf, e.g., reference [44] chapter 9) and which is used in the main text to identify the minimal dimension of a quantum system that is required to reach a given extreme point of $P_L^{O_A}$. We first show the following lemma which straightforwardly extends to observation 4.

**Lemma 8.** Let $E$ be an effect of a POVM, i.e. $E \geq 0$ and $E \leq 1$, and $\rho = \sum_{|i\rangle \in I} |i\rangle \langle i| + \sum_{|k\rangle \in K} c_k |k\rangle \langle k|$ be the spectral decomposition of a density matrix $\rho$. Then $\text{tr} \{ \rho E \} = 1$ if $E = \sum_{|i\rangle \in I} |i\rangle \langle i| + \sum_{|k\rangle \in K} c_k |k\rangle \langle k|$ with $I \cap K = \{ \}$ and $\{|i\rangle\}_{i \in I} \cup \{|k\rangle\}_{k \in K}$ being an ONB. The matrix $E_K = \sum_{|k\rangle \in K} c_k |k\rangle \langle k|$ is positive semidefinite and $E_K \leq 1$.

**Proof.** If: inserting $E = \sum_{|i\rangle \in I} |i\rangle \langle i| + \sum_{|k\rangle \in K} c_k |k\rangle \langle k|$ in $\text{tr} \{ \rho E \}$ and using that $\{|i\rangle\}_{i \in I} \cup \{|k\rangle\}_{k \in K}$ is an ONB as well as that $\sum_{|i\rangle \in I} p_i = 1$ readily proves the statement.

Only if: writing $E$ in the basis $\{|i\rangle\}_{i \in I} \cup \{|k\rangle\}_{k \in K}$, i.e. $E = \sum_{|i\rangle \in I} c_i |i\rangle \langle i|$, and inserting in $\text{tr} \{ \rho E \}$ it can be shown immediately from $E \leq 1$ that $\text{tr} \{ \rho E \} = 1$ and $\sum_{|i\rangle \in I} c_i p_i = 1$. As $0 \leq E \leq 1$ it holds that $0 \leq c_i \leq 1$. Moreover, using that $\sum_{|i\rangle \in I} p_i = 1$ and $p_i > 0$ it therefore follows that $c_i = 1$ for all $i \in I$. It can be easily seen that this condition and $E \leq 1$ can only be simultaneously fulfilled if $c_i = 0$ for $i \in I$ and $k \neq i$. More precisely, due to $E \leq 1$ it has to hold that $E = \sum_{|i\rangle \in I} c_i |i\rangle \langle i|$. Hence, $E$ is of the form $E = \sum_{|i\rangle \in I} |i\rangle \langle i| + \sum_{|k\rangle \in K} c_k |k\rangle \langle k|$. It follows immediately from $E \geq 0$ and $E \leq 1$ that $E = \sum_{|k\rangle \in K} c_k |k\rangle \langle k|$.

It follows that states giving, with probability one, different outcomes for the same sequence of measurements have ranges corresponding to orthogonal subspaces.

**Observation 4.** Let $E$ be an effect of a POVM and $\rho_i = \sum_{|i\rangle \in I} |i\rangle \langle i| + \sum_{|k\rangle \in K} c_k |k\rangle \langle k|$ with $p_i > 0$ and $\rho = \sum_{|k\rangle \in K} c_k |k\rangle \langle k|$ with $q_k > 0$ be the spectral decomposition of a density matrix $\rho$ and $\rho_2$ respectively. Then $\text{tr} \{ \rho_1 E \} = 1$ and $\text{tr} \{ \rho_2 E \} = 0$ only if $\langle i\rangle | \rho_1 \rangle = 0$ for all $i \in I$ and $l \in L$.

**Proof.** Using lemma 8 if follows from $\text{tr} \{ \rho_1 E \} = 1$ that $E = \sum_{|i\rangle \in I} |i\rangle \langle i| + \sum_{|k\rangle \in K} c_k |k\rangle \langle k|$ with $I \cap K = \{ \}$ and $\{|i\rangle\}_{i \in I} \cup \{|k\rangle\}_{k \in K}$ being an ONB. Denoting $\sum_{|k\rangle \in K} c_k |k\rangle \langle k|$ by $E_K$ we have that $\text{tr} \{ \rho_2 E \} = \sum_{|i\rangle \in I} \langle i\rangle | \rho_2 | i\rangle + \text{tr} \{ E_K \rho_2 \} = 0$. As $\rho_2 \geq 0$ and $E_K \geq 0$ we have that $\langle i\rangle | \rho_1 | i\rangle = 0 \quad \forall \ i \in I$ and it can be easily seen that this implies $\langle i\rangle | \rho_2 \rangle = 0$ for all $i \in I$ and $l \in L$.

Appendix C. A potentially improved lower bound on the dimension needed to realize any extreme point

In the main text we discussed a way to construct an extreme point that yields a lower bound on the necessary dimension to realize any extreme point. For any $O$ and $S$ this construction allowed to provide a closed formula for the scaling of the bound with respect to the length of the sequence. Here we discuss a different construction which gives a potentially better lower bound. In order to do so we consider the following extreme point. All tuples that are assigned to a time step $j < k$ correspond to $(0, 0, \ldots, 0)$, i.e. for the first $k - 1$ time steps one obtains outcome ‘0’ for all settings. In time step $t_k$ emerging subtrees $T_{k-1}$ have the property that in the root node all settings yield outcome ‘0’, however in the second time step at least one of the tuples is not of the form $(0, 0, \ldots, 0)$. Moreover, all of these subtrees are chosen to be different, see figure 8.

Note that therefore all possible futures assigned to a time step $i \leq k$ are not equivalent to each other. As discussed in the proof of theorem 6 the number of inequivalent futures corresponds to the necessary dimension. Hence, one obtains straightforwardly a lower bound on the dimension given by $\sum_{i=1}^{k+1} \binom{Q}{i}$. In order to obtain the best possible bound of this form it remains to identify the largest $k$ for which such a construction is possible. Recall that $s^{k-1}$ is the number of futures that can be assigned to time step $t_k$, $L - k$ is the length of these futures and $\binom{O'}{s^{k-1}}$ is the number of different futures of length $L - k$ for which the starting node is given by $(0, 0, \ldots, 0)$. The latter can be shown analogously to the proof of lemma 1 which can be straightforwardly extended to an arbitrary number of outcomes. Note, however, that for an arbitrary $O$ the condition that the first tuple corresponds to $(0, 0, \ldots, 0)$ does not uniquely identify one element of an ORE class. The number of futures of length $L - k$ for which all tuples in the second time step are equal to $(0, 0, \ldots, 0)$ is given by

$$
\binom{O'}{s^{k-1}} = \frac{(O')^{\frac{s^{k-1}}{s-1}}}{(O')^{s-1}}.
$$

(C1)
Figure 8. This figure illustrates the idea of the construction of the extreme point that allows to obtain a lower bound on the scaling. Here the subtrees \( e_i \) have the following properties: (a) in the first time step all measurements yield outcome ‘0’. (b) They are all chosen to be different. (c) At least one tuple assigned to the second time step does not correspond to \((0,0,\ldots,0)\).

This is due to the fact that the number of different possibilities to assign tuples in the second time step is \((O^3)^S\). With this it follows that one has to identify the largest natural number \(k\) such that

\[
S^{j-1} \leq (O^3)^{\frac{d-1+j-1}{j-1}} - (O^3)^{\frac{d-1+j-1}{j-1}} S^{j-1} \quad \text{(C2)}
\]

with \(j \in \mathbb{R}\). Given \(S\) and \(L\) one can obtain \(j\) for example graphically by determining the zero of \(\sum_{j=1}^k S^{j-1} \leq (O^3)^{\frac{d-1+j-1}{j-1}} - (O^3)^{\frac{d-1+j-1}{j-1}} S^{j-1}\) which is monotonically decreasing as a function of \(j\) and compute straightforwardly \(k\). The lower bound on the dimension is then given by \(\sum_{j=1}^k S^{j-1} \leq \frac{kS^{j-1}}{S^j}\). Compared to the construction presented in the main text it may be that the \(k\) obtained in the way presented here is smaller, however here all subtrees up to time step \(t_k\) are accounted for and not only the ones assigned to this time step. Hence, for certain scenarios the lower bound can be improved.

ORCID iDs

Cornelia Spee  https://orcid.org/0000-0002-2483-307X
Costantino Budroni  https://orcid.org/0000-0002-6562-7862
Otfried Gühne  https://orcid.org/0000-0002-6033-0867

References

[1] Bell A 1964 Physics 47 195
[2] Brunner N, Cavalcanti D, Pironio S, Scarani V and Wehner S 2014 Rev. Mod. Phys. 86 419
[3] Horodecki R, Horodecki P, Horodecki M and Horodecki K 2009 Rev. Mod. Phys. 81 865
[4] Gühne O and Tóth G 2009 Phys. Rep. 474 1
[5] Acín A, Brunner N, Gisin N, Massar S, Pironio S and Scarani V 2007 Phys. Rev. Lett. 98 230501
[6] Pironio S et al 2010 Nature 464 1021
[7] Colbeck R and Kent A 2011 J. Phys. A: Math. Theor. 44 095305
[8] Popescu S and Rohrlich D 1994 Found. Phys. 24 379
[9] Leggett A J and Garg A 1985 Phys. Rev. Lett. 54 857
[10] Emary C, Lambert N and Nori F 2014 Rep. Prog. Phys. 77 016001
[11] Klyachko A A, Can M A, Binicioglu S and Shumovsky A S 2008 Phys. Rev. Lett. 101 020403
[12] Cabello A 2008 Phys. Rev. Lett. 101 210401
[13] Kirchmair G, Zähringer F, Gerritsma R, Kleinmann M, Gühne O, Cabello A, Blatt R and Roos C F 2009 Nature 460 494
[14] Gühne O, Kleinmann M, Cabello A, Larsson J-A, Kirchmair G, Zähringer F, Gerritsma R and Roos C F 2010 Phys. Rev. A 81 022121
[15] Brzinski S, Kosowski A, Markiewicz M, Paterek T and Przywińska A 2015 Phys. Rev. Lett. 115 120404
[16] Budroni C, Fagundes G and Kleinmann M 2019 New J. Phys. 21 093018
[17] Budroni C and Emary C 2014 Phys. Rev. Lett. 113 050401
[18] Schild G and Emary C 2015 Phys. Rev. A 92 032101
[19] Hoffmann J, Spee C, Gühne O and Budroni C 2018 New J. Phys. 20 102001
[20] Spee C, Siebeneich H, Gloger T F, Kaufmann P, Johanning M, Kleinmann M, Wunderlich C and Gühne O 2020 New J. Phys. 22 023028
[21] Spee C 2020 Phys. Rev. A 102 012420
[22] Budroni C, Vitagliano G and Woods M P 2020 arXiv:2005.04241 [quant-ph]
[23] Gallego R, Brunner N, Hadley C and Acin A 2010 Phys. Rev. Lett. 105 230501
[24] Brunner N, Navascués M and Vértesi T 2013 Phys. Rev. Lett. 110 150501
[25] Wiesner S 1983 ACM Sigact News 15 78
[26] Ambainis A, Nayak A, Ta-Shma A and Vazirani U 1999 Proc. of the Thirty-First Annual ACM Symp. on Theory of Computing 376–83
[27] Ambainis A, Nayak A, Ta-Shma A and Vazirani U 2002 J. ACM 49 496
[28] Bowles J, Brunner N and Pawlowski M 2015 Phys. Rev. A 92 022351
[29] Aguilar E A et al 2018 Phys. Rev. Lett. 120 230503
[30] Miklin N, Borkala J J and Pawlowski M 2020 Phys. Rev. Research 2 033014
[31] Kleinmann M, Gühne O, Portillo J R, Larsson J -Å and Cabelló A 2011 New J. Phys. 13 113011
[32] Fagundes G and Kleinmann M 2017 J. Phys. A: Math. Theor. 50 325302
[33] Garner A J P, Liu Q, Thompson J, Vedral V and Gu M 2017 New J. Phys. 19 103009
[34] Thompson J, Garner A J P, Mahoney J R, Crutchfield J P, Vedral V and Gu M 2018 Phys. Rev. X 8 031013
[35] Woods M P, Silva R, Pütz G, Stupar S and Renner R 2018 arXiv:1806.00491 [quant-ph]
[36] Clemente L and Kofler J 2016 Phys. Rev. Lett. 116 150401
[37] Hoffmann J 2016 Temporal correlations in quantum theory MSc Thesis University of Siegen
[38] Abbott A A, Giarmatzi C, Costa F and Branciard C 2016 Phys. Rev. A 94 052131
[39] Wolf M M and Pérez-Garcia D 2009 Phys. Rev. Lett. 102 190504
[40] Gühne O, Budroni C, Cabelló A, Kleinmann M and Larsson J -Å 2014 Phys. Rev. A 89 062107
[41] Mao Y, Spee C, Xu Z -P and Gühne O 2020 arXiv:2005.13964 [quant-ph]
[42] Heinosaari T and Ziman M 2011 The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement (Cambridge: Cambridge University Press)
[43] Paz A 2003 Probabilistic Automata (New York: Wiley)
[44] Nielsen M A and Chuang I 2002 Quantum Computation and Quantum Information (Cambridge: Cambridge University Press)
[45] Hoorfar A and Hassani M 2008 J. Inequal. Pure and Appl. Math. 9 5
[46] Kitaev A Y 1997 Russ. Math. Surv. 52 1191
[47] Rabiner L and Juang B 1986 IEEE ASSP Mag. 3 4
[48] Barnett N and Crutchfield J P 2015 J. Stat. Phys. 161 404
[49] Cabelló A, Gu M, Gühne O and Xu Z -P 2018 Phys. Rev. Lett. 120 130401
[50] Gu M, Wiesner K, Riepe E and Vedral V 2012 Nat. Commun. 3 763