





Monogamy of temporal correlations: Witnessing non-Markovianity beyond data processing

Matheus Capela ¹, Lucas C. Céleri ^{1,2}, Kavan Modi ³, and Rafael Chaves ^{4,5}

¹*Institute of Physics, Federal University of Goiás, POBOX 131, 74001-970, Goiânia, Brazil*

²*Department of Physical Chemistry, University of the Basque Country UPV/EHU, Apartado 644, E-48080 Bilbao, Spain*

³*School of Physics & Astronomy, Monash University, Clayton, Victoria 3800, Australia*

⁴*International Institute of Physics, Federal University of Rio Grande do Norte, 59070-405 Natal, Brazil*

⁵*School of Science and Technology, Federal University of Rio Grande do Norte, 59078-970 Natal, Brazil*



(Received 15 October 2019; accepted 4 March 2020; published 20 March 2020)

The modeling of natural phenomena via a Markov process—a process for which the future is independent of the past, given the present—is ubiquitous in many fields of science. Within this context, it is of foremost importance to develop ways to check from the available empirical data if the underlying mechanism is indeed Markovian. A paradigmatic example is given by data processing inequalities, the violation of which is an unambiguous proof of the non-Markovianity of the process. Here, our aim is twofold. First we show the existence of a monogamy-like type of constraint, beyond data processing, respected by Markov chains. Secondly, we show a connection between the quantification of causality and the violation of both data processing and monogamy inequalities. Apart from its foundational relevance in the study of stochastic processes we also consider the applicability of our results in a typical quantum information setup, showing it can be useful to witness the non-Markovianity arising in a sequence of quantum nonprojective measurements.

DOI: [10.1103/PhysRevResearch.2.013350](https://doi.org/10.1103/PhysRevResearch.2.013350)

I. INTRODUCTION

The importance of Markov processes can hardly be overstated. In short, a process is Markovian if, in order to predict its future state, the present state contains as much information as the full previous history. That is, a Markov process keeps only information about its immediate past. Applications of it range from computer science [1], causal inference [2,3], and statistics [4] to social sciences [5] and genetics [6]. Within physics, random walks [7] and the Brownian motion [8] are paradigmatic examples of Markov processes. In quantum information, they also play a key role [9], particularly in the understanding of open-system dynamics [10].

Mathematically, a stochastic discrete process $\{X_n, t_n \in T\}$ is Markovian if the probability that the random variable X_n takes a value x_n at time $t_n \in T$ is uniquely determined, and not affected by the possible values of X at previous times to t_{n-1} . That is,

$$p(x_n | x_{n-1}, \dots, x_1) = p(x_n | x_{n-1}), \quad \text{for all } t_n \in T. \quad (1)$$

Given a joint probability distribution $p(x_1, \dots, x_n)$, to check if it arises from a Markovian process we have to test all the n conditions expressed in (1). However, if the number n of variables or their cardinalities (the number of possible values they can assume) is large, it is practically impossible to gather enough statistical data to reconstruct $p(x_1, \dots, x_n)$ and thus check for its Markovianity. As an example, a source

producing strings with only 250 bits has $2^{250} \approx 10^{75}$ possible variations, close to the estimated number of atoms in the universe. For that reason, we often have to rely on marginal information, for instance, only the pairwise correlations between two variables. Because of this estimation limitation, we face a particular case of a marginal problem [11]: given some limited/marginal information, what can we conclude about the global object?

Formally, the problem is equivalent to quantifier elimination [12,13]: starting from a set of constraints—in this case the positivity and normalization of probabilities plus the Markov conditions (1)—we aim to eliminate from our description those variables to which we do not have empirical access (e.g., correlations involving more than two variables). The same problem arises in causal inference [14], an extremely demanding computational task that cannot be performed beyond very few variables [15]. A more treatable approach is to ask instead what are the constraints implied by Markovianity on the entropies of the variables of interest [11,16–19]. For instance, a simple and yet fundamental result implies that Markov processes fulfill data processing inequalities, basically stating that as we move along the Markov chain the correlation between the consecutive time steps (with respect to the initial one) cannot increase, but may remain constant. Data processing inequalities are a consequence of the Markov condition, imposing constraints to the correlations between the variables in different time steps, but are they the only consequence? This is the central question we explore in this article.

As we show, indeed data processing inequalities are the only consequence of the Markov condition for $n = 3$. However, for $n \geq 4$, Markovianity implies different kinds of monogamy constraints for the correlations in the Markov chain, violations of which can be seen as a device-independent test [20–22] of the non-Markovianity of the underlying

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

process. Interestingly, by employing an operational definition of divisibility [23], we show that such a condition is sufficient for a process satisfying all data processing and monogamy inequalities. Furthermore, we show how the violation of these constraints can also be connected with the quantification of causal influences [16,24,25] among the variables. Finally, we consider a quantum information related application of this framework, showing that nonprojective measurements can give rise to statistics compatible with data processing, but nonetheless can violate our monogamy inequalities, thus witnessing the non-Markovianity arising from such quantum measurements.

The paper is organized as follows. In Sec. II we review the basic toolbox to be used in this paper: the entropic approach to the marginal problem [11,16–18]. In Sec. III we employ this framework to show that another kind of constraint, beyond that given by data processing inequalities, arises for Markov chains with $n \geq 4$. The question regarding divisibility is addressed in Sec. IV. In Sec. V we show that the violation of both the data processing and the monogamy inequalities can be given a causal interpretation. Finally, in Sec. VI we conclude and discuss our findings. More technical details and proofs of our results can be found in the Appendix.

II. SHANNON ENTROPIES, ENTROPY CONES, AND MARGINAL PROBLEMS

The Shannon entropy of a random variable X is the fundamental building block in information theory [26]. It is defined

as

$$H(X) = - \sum_x p(x) \log p(x),$$

where the sum is taken over the support of X . If we are interested in the entropy of n variables X_1, \dots, X_n , it is useful to construct the entropy vector h associated with these variables as $h = [H(\emptyset), H(X_1), H(X_2), H(X_1, X_2), \dots, H(X_1, \dots, X_n)]$. That is, h is a vector with 2^n components given by all possible entropies among n variables [including the empty set \emptyset for which we define $H(\emptyset) = 0$]. Within this approach, a natural question is to understand which vectors in the real space \mathbb{R}^{2^n} define the entropy vectors. The first trivial constraint follows from the fact that entropies are positive quantities, that is, the entropy vector cannot have negative components. We are thus restricted to the positive-orthant of \mathbb{R}^{2^n} . The second constraint comes from the realization that entropies define an unbounded convex set, the so-called entropy cone [26]. We can think of it as a hyperdimensional and infinite ice cream cone, where the tip of cone lies in the origin of our coordinate system.

Annoyingly, however, the exact structure of the real space defining entropy vectors is still not precisely known, with a description being given by an outer approximation of the true entropy cone, the so-called Shannon cone [26]. The nice thing about the Shannon cone is the fact that it is defined in terms of finitely many linear inequalities of two types:

$$\begin{aligned} I(X_2 : X_3|X_1) \geq 0 &\Rightarrow H(X_1, X_2) + H(X_1, X_3) \geq H(X_1, X_2, X_3) + H(X_1), \\ H(X_1|X_2, X_3) \geq 0 &\Rightarrow H(X_1, X_2, X_3) \geq H(X_2, X_3). \end{aligned} \tag{2}$$

The first inequality is known as strong subadditivity (or submodularity), basically stating the positivity of the conditional mutual information. The second constraint is known as monotonicity, stating the positivity of conditional entropies or alternatively showing that the uncertainty of the whole is at least as large as the uncertainty of its parts. Given a collection of n variables there are $n + \binom{n}{2} 2^{n-2}$ nonredundant inequalities (that is, inequalities that cannot be obtained by combining the other inequalities) defining the Shannon cone. This minimum set of linear inequalities is known as the *elemental* set of Shannon type inequalities.

Given the Shannon cone (as represented by the elemental set), we can ask the following: what are the constraints following from the elemental set on the subspace where we eliminate some of the variables in the entropy vector [11]? There are practical and more fundamental reasons to do so. On the practical side, if we have N variables X_1, \dots, X_N , to fully describe the entropic vector h , one should estimate experimentally an exponential number of parameters, more precisely, a number $2^N - 1$ of them. In turn, if one is restricted to at most bipartite marginal information, there are N single body entropies $H(X_i)$ and $N(N - 1)/2$ two-body terms $H(X_i, X_j)$. That is, restricting to bipartite marginals offers an exponential advantage in terms of the number of

parameters to be estimated. The discussion above highlights the fact that checking the Markov condition (5) directly is not practically possible as we increase the number N of variables. This shows the practical relevance of looking at marginal information. For the reasons stated below, here we consider bipartite marginal information; however, the general approach we propose can also be readily applied beyond the bipartite information scenario. First, it is the simplest marginal scenario leading to nontrivial constraints. Second, from such single and bipartite entropies one can construct the mutual information, a measure of correlation ubiquitous in information theory. Not only, it provides a natural way to quantify correlations but also an intuitive way to understand the processing of information, for example, via the data processing inequality basically, stating that the local processing of information cannot increase correlations. Nicely, as we prove here, bipartite information contains much more relevance than simply the data processing.

As an example let us consider the following situation. Three people, seated in different rooms in such a way they cannot directly communicate (but are allowed to establish some preshared correlations), have to answer questions to a referee. Each of the answers of the three participants is being represented by the random variables X_1, X_2 , and X_3 .

However, at a given run of our experiment, a referee will only ask questions to two of them, that is, the referee does not have access to the entropy $H(X_1, X_2, X_3)$ corresponding to the event where the three participants would give their answers. This means that we have to eliminate this entropy from our description, implying that our object of interest is an entropy vector where we trace out one of its components. Formally, the problem is equivalent to a quantifier elimination problem [12]: we have a set of (linear) inequalities and we want to have the equivalent description of this set where some of the variables appearing in the inequalities have been eliminated from the problem.

Coming back to our problem. To obtain an inequality that does not depend on $H(X_1, X_2, X_3)$, we can simply sum the two Shannon type inequalities in (2) to obtain

$$H(X_1, X_2) + H(X_1, X_3) \geq H(X_2, X_3) + H(X_1), \quad (3)$$

or in terms of mutual information $I(X : Y) = H(X) + H(Y) - H(X, Y)$ (a measure of correlations between X and Y)

$$I(X_1 : X_2) + I(X_1 : X_3) - I(X_2 : X_3) \leq H(X_1). \quad (4)$$

As we can see, the simple assumption about the existence of a joint probability distribution describing the three variables already implies constraints about their pairwise correlations.

To illustrate the use of these marginal constraints, consider dichotomic answers (yes/no questions). Also, suppose that, after we run the experiment a sufficient number of times, the referee observes that all answers are unbiased [$H(X_1) = H(X_2) = H(X_3) = 1$]. Furthermore, answer X_1 is fully correlated with X_2 and X_3 [that is $I(X_1 : X_2) = I(X_1 : X_3) = 1$]; however, X_2 and X_3 are uncorrelated [$I(X_2 : X_3) = 0$]. If we plug in these values in (4) we see that this inequality is violated. What does this mean? Notice that (3) follows from the assumption that even if we cannot observe it, there is a well defined $H(X_1, X_2, X_3)$ joint entropy for all the answers. The violation of the inequality shows that this assumption is not valid: from the marginal observations we cannot construct a well-defined $H(X_1, X_2, X_3)$ (in such a way that all the elemental inequalities are respected). In fact, notice that the distribution violating the inequality is a bit odd. Since X_2 and X_3 are fully correlated with X_1 , by the transitivity of correlations we would expect that X_2 and X_3 are also fully correlated. We can understand this distribution as a violation of the rules of the game. At each run, parties 2 and 3 communicate through a secret channel: if in a given particular run one of them is excluded of the game (the referee does not ask one of them any question), then the other part (which was asked something) will use a strategy that correlates his answer with answer X_1 ; if both parties 2 and 3 are asked questions, they just give completely uncorrelated answers between them.

III. MARKOV PROCESSES BEYOND DATA PROCESSING INEQUALITIES

In the case of a Markov process we can follow a similar construction to the one delineated above. We have n variables respecting the usual elemental inequalities; however, in this case, we also have a set of constraints that follow from the Markov conditions (1). In terms of entropies, Eq. (1) can be

expressed as

$$H(X_n | X_{n-1}, \dots, X_1) = H(X_n | X_{n-1}), \quad \text{for all } n. \quad (5)$$

That is, in this case we are interested in the intersection of the Shannon cone with the hyperplanes defined by (5). We can proceed with the quantifier elimination and thus eliminate all terms but the marginal involving single and two-body terms.

Let us start with the simplest possible Markov chain with $n = 3$. In this case, the only Markov condition is given by $H(X_3 | X_1, X_2) = H(X_3 | X_2)$. Performing the quantifier elimination step we observe that the only nontrivial inequalities characterizing the marginal Shannon cone are given by

$$I(X_1 : X_2) \geq I(X_1 : X_3), \quad I(X_2 : X_3) \geq I(X_1 : X_3),$$

that is, we recover the usual data processing inequalities that we expect to hold in a Markov chain. By nontrivial, we mean inequalities that are not simple elemental inequalities (strong subadditivity and monotonicity). In this geometric perspective, data processing inequalities are nothing other than the facets of the Shannon cone intersected with the hyperplanes defining the entropic Markov conditions (5), marginalized to the subspace where the joint entropy between the three variables has been eliminated. For instance, the data processing inequality $I(X_1 : X_2) \geq I(X_1 : X_3)$ is a direct consequence of combining the strong subadditivity $I(X_1 : X_2 | X_3) \geq 0$ with the Markov condition $H(X_3 | X_2, X_1) = H(X_3 | X_2)$ (see Appendix A 1 for a simple proof).

We can now move to the case of a Markov chain with $n = 4$. In this case, we have two Markov conditions: $H(X_3 | X_1, X_2) = H(X_3 | X_2)$ and $H(X_4 | X_1, X_2, X_3) = H(X_4 | X_3)$. Performing the quantifier elimination, we observe that the only nontrivial (and nonredundant) inequalities are given by

$$\begin{aligned} I_{1,3} &\geq I_{1,4}, & I_{1,2} &\geq I_{1,3}, \\ I_{2,4} &\geq I_{1,4}, & I_{3,4} &\geq I_{2,4}, \\ I_{1,4} + I_{2,3} &\geq I_{1,3} + I_{2,4}, \end{aligned} \quad (6)$$

where we have used the short-hand notation $I_{i,j} \equiv I(X_i : X_j)$.

We highlight two aspects in the set of the inequalities above. First, we notice that there are other data processing inequalities that follow in this scenario (five more), for instance, $I_{1,2} \geq I_{1,4}$. However, all these other data processing inequalities are redundant in the sense that they follow from the set above combined with Shannon type inequalities (see Appendix A 3 for a proof); this set is the minimum nonredundant one [27]. Second, the most interesting feature comes from the fact that we have another kind of inequality emerging for $n = 4$ and that is not of the data processing type. It shows that in Markov processes not only the correlations should decrease as we move along the chain (as quantitatively expressed by the data processing inequality) but also that the pairwise correlations between the variables should respect a monogamy kind of constraint. Interestingly, as shown below there are non-Markovian processes respecting all data processing inequalities but nonetheless violating our derived inequality, thus showing the clear relevance of it.

A natural question is whether another kind of monogamy relation will appear as we increase the size of the Markov chain. As proven in Appendixes A 7, A 8, and A 9 this is indeed the case. For instance, for $n = 6, 8, 10$, the following

monogamy inequalities hold:

$$I_{1,6} + I_{2,5} + I_{3,4} \geq I_{1,4} + I_{2,6} + I_{3,5},$$

$$I_{1,8} + I_{2,7} + I_{3,6} + I_{4,5} \geq I_{1,5} + I_{2,8} + I_{3,7} + I_{4,6},$$

$$I_{1,10} + I_{2,9} + I_{3,8} + I_{4,7} + I_{5,6} \geq I_{1,6} + I_{2,10} + I_{3,9} + I_{4,8} + I_{5,7}.$$

In particular, notice that the sum of the distances between the nodes involved is the same in the left hand side (LHS) and right hand side (RHS) of the monogamy inequalities. For instance, the mutual information $I_{1,6}$ involves nodes X_1 and X_6 that have four nodes in between them. So, for $n = 6$ the sum of the distances in the LHS is $4 + 2 + 0 = 6$, while in the RHS we have $2 + 3 + 1 = 6$. A similar analysis shows that, for $n = 8$, the sum of distances in the LHS gives $6 + 4 + 2 + 0 = 12$ and the RHS gives $3 + 5 + 3 + 1 = 12$; for $n = 10$ the LHS gives $8 + 6 + 4 + 2 + 0 = 20$ and RHS gives $4 + 7 + 5 + 3 + 1 = 20$. Based on that and the clear pattern observed for $n = 4, 6, 8, 10$ we conjecture that the following monogamy inequality holds for arbitrary even n :

$$\sum_{i=1}^n I_{i,n-i+1} \geq I_{1,1+n/2} + \sum_{i=2}^{n/2} I_{i,n-i+2}. \quad (7)$$

This result is based on induction and an analytical general proof is missing.

Examples of monogamy violation while satisfying the data processing inequalities

Two bit process. We now given an explicit example of a classical non-Markovian process where the data processing inequalities hold, while the monogamy inequality does not. Consider that variable X_1 is binary ($x_1 = 0, 1$) and all the others can assume four values $x_2, x_3, x_4 = 0, 1, 2, 3$. The process is described by the following four steps.

- (1) Let $x_1 = 0, 1$ with probability $1/2$.
- (2) If $x_1 = 0$, then $x_2 = 0, 1$ with probability $1/2$. If $x_1 = 1$, then $x_2 = 2, 3$ with probability $1/2$.
- (3) If $x_2 = \{0, 1\}$, then $x_3 = 0, 1$ with probability $1/2$. If $x_2 = \{2, 3\}$, then $x_3 = 2, 3$ with probability $1/2$.
- (4) If $(x_1, x_2, x_3) = (0, 0, 0)$ or $(1, 2, 2)$, then $x_4 = 0$. If $(x_1, x_2, x_3) = (0, 1, 0)$ or $(1, 3, 2)$, then $x_4 = 1$. If $(x_1, x_2, x_3) = (0, 0, 1)$ or $(1, 2, 3)$, then $x_4 = 2$. If $(x_1, x_2, x_3) = (0, 1, 1)$ or $(1, 3, 3)$, then $x_4 = 3$.

This protocol generates a distribution $p(x_1, x_2, x_3, x_4)$ that with the same probability $1/8$ is equal to $(0,0,0,0), (0,1,0,1), (0,0,1,2), (0,1,1,3), (1,2,2,0), (1,3,2,1), (1,2,3,2)$, and $(1,3,3,3)$. Computing the associated entropies we get

$$H(X_1) = 1, \quad H(X_2) = H(X_3) = H(X_4) = 2,$$

$$H(X_1, X_2) = H(X_1, X_3) = 2,$$

$$H(X_1, X_4) = H(X_2, X_3) = H(X_2, X_4) = H(X_3, X_4) = 3.$$

We can check that this distribution respects all eight data processing inequalities [the four in (6) plus the four redundant ones] but nonetheless violates the monogamy constraint in (6). Notice that the process is Markovian among the three first nodes (X_2 is a function of X_1 alone and X_3 is a function of X_2 alone). However, X_4 has a direct dependence on the values of X_2 and X_3 (thus breaking the Markov condition).

One bit process. Consider the one bit process described by the following joint distribution:

$$p(X_1, X_2, X_3, X_4) = \frac{1}{100}(3, 9, 10, 9, 6, 9, 2, 1, 1, 5, 11, 4, 4, 8, 6, 12),$$

where the first entry in the vector corresponds to $p(0, 0, 0, 0)$ and the second to $p(0, 0, 0, 1)$ until the last entry corresponding to $p(1, 1, 1, 1)$. For this process all data processing inequalities are satisfied, while the monogamy is violated by -0.0027 . This example was found by random numerical search.

Non-Markovianity from nonprojective measurements. We now discuss another example where the data processing inequalities are satisfied but the monogamy inequality is violated. Here the non-Markovianity arises from a sequence of nonprojective quantum measurements. The generalized quantum measurements are defined by a collection of measurement operators M_x satisfying the completeness relation $\sum_x M_x^\dagger M_x = \mathbb{1}$, where each index x is associated with an experimental outcome and $\mathbb{1}$ is the identity operator on the system's Hilbert space. Here we consider the case of generalized measurements performed on a qubit system. The first collection of measurement operators, written in the computational basis, is defined by

$$A_1 = \begin{bmatrix} +0.4953 + i0.0687 & +0.0874 - i0.2751 \\ +0.2751 + i0.0874 & +0.1327 + i0.2564 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} +0.1327 + i0.2564 & 0.2751 + i0.0874 \\ +0.0874 - i0.2751 & +0.4953 + i0.0687 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} +0.1327 + i0.2564 & -0.2751 - i0.0874 \\ -0.0874 + i0.2751 & +0.4953 + i0.0687 \end{bmatrix},$$

$$A_4 = \begin{bmatrix} +0.4953 + i0.0687 & -0.0874 + i0.2751 \\ -0.2751 - i0.0874 & +0.1327 + i0.2564 \end{bmatrix}.$$

These measurement operators were chosen such that each $A_i^\dagger A_i$ ($i = 1, 2, 3, 4$) has the same eigenvalues, 0.3943 and 0.1057 , but the associated eigenvectors are different.

The second set of measurement operators considered here are defined by

$$B_1 = \sqrt{\frac{1+\alpha}{2}}|+\rangle\langle+| + \sqrt{\frac{1-\alpha}{2}}|-\rangle\langle-|,$$

$$B_2 = \sqrt{\frac{1-\alpha}{2}}|+\rangle\langle+| + \sqrt{\frac{1+\alpha}{2}}|-\rangle\langle-|,$$

where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ and the real parameter varies as $0 \leq \alpha \leq 1$. For $\alpha = 1$, the operators describe a projective measurement, while a weak one is described when $\alpha \ll 1$.

The protocol considered here is defined by four sequential measurements after the preparation of a qubit state ϱ : the first measurement is defined by $\{A_x\}_{x=1,\dots,4}$ and the second one by $\{B_y\}_{y=1,2}$. The third and fourth measurements are just a repetition of the first and second ones. Such sequential measurements are described by the operators $M_{i,j,k,l} = B_l A_k B_j A_i$ ($i, k = 1, 2, 3, 4$ and $j, l = 1, 2$), associated with the joint probability

$$p(i, j, k, l) = \text{Tr}[M_{i,j,k,l} \varrho M_{i,j,k,l}^\dagger]. \quad (8)$$

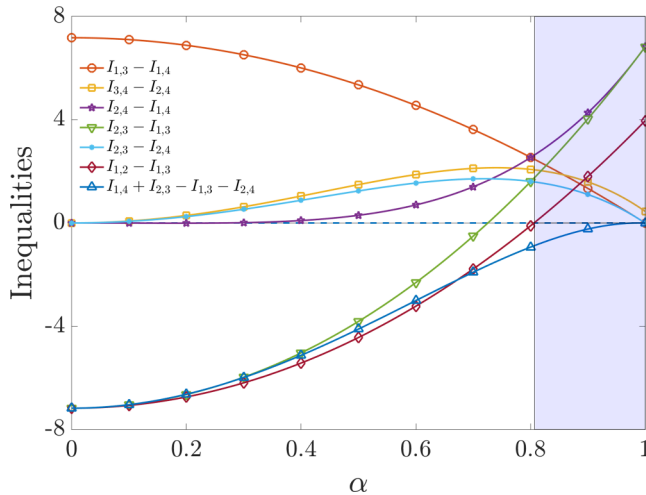


FIG. 1. Monogamy violation in series of nonprojective measurements. Data processing and monogamy inequalities discussed in the main text, given in Eqs. (6), as a function of the parameter α (defining the measurement operator). The shaded region shows the values of α for which all data processing inequalities are satisfied (are positive) while the monogamy inequality is violated (is negative). There are nine data processing inequalities for $n = 4$ but only six of them have to be considered here (see Appendix A 3 for more details).

If the initial state is $\varrho = |+\rangle\langle +|$ it is found that the monogamy inequality is violated in the region greater than $\alpha \simeq 0.8$, but none of the data processing inequalities are violated, as shown in Fig. 1.

This example is interesting for several reasons. First, while the underlying process is quantum, it leads to a classical process described by the distribution in Eq. (8). Secondly, the quantum process itself here is Markovian [28]. We can think of it simply as the identity channel or we can think of the measurements in the computational basis and the process as a unitary transformation between the measurements. In both cases the quantum process is Markovian, but it leads to a non-Markovian classical distribution. Therefore, the non-Markovianity must arise from the measurements themselves. In the classical domain, making coarse measurements can turn a Markov process into a non-Markov process; see Examples 4–6 in [29]. In this example the coarseness comes from the fact that the chosen quantum measurements are not sharp, i.e., rank-1 projections. Finally, the example illustrates a key difference between classical stochastic processes and quantum stochastic processes: in the former theory there is an assumption of noninvasiveness, while the latter requires invasive measurements to say anything about the process [30].

IV. INEQUALITIES FOR DIVISIBLE PROCESSES

Divisible processes form a special superset of Markov processes. Before we talk about divisible processes we first introduce the notion of stochastic matrices. Consider the process from $X \rightarrow Y$; the stochastic matrix $\Gamma_{Y:X}$ maps any initial distribution $p(X)$ to the corresponding final distribution $p(Y)$. The stochastic matrix can be acquired from the joint

probability distribution $p(X, Y)$ as

$$\Gamma_{Y:X} = \begin{pmatrix} \frac{p(X=0,Y=0)}{p(X=0)} & \frac{p(X=1,Y=0)}{p(X=1)} & \cdots \\ \frac{p(X=0,Y=1)}{p(X=0)} & \frac{p(X=1,Y=1)}{p(X=1)} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

Let us now consider the process $X \rightarrow Y$ with the stochastic matrix $\Gamma_{Y:X}$ and the process $X \rightarrow Z$ with the stochastic matrix $\Gamma_{Z:X}$. A process is called divisible if matrix $G_{Z:Y} := \Gamma_{Z:X} \Gamma_{Y:X}^{-1}$ is also a stochastic matrix [31]. This is known as divisibility by inversion. Usually in the literature when one talks about divisibility one refers to the above concept. However, there is also a stronger operational notion of divisibility [23], which we will exploit below. For divisibility by inversion, we have only used joint distributions $p(X, Y)$ and $p(X, Z)$. If we also have access to $p(Y, Z)$ then we can compute $\Gamma_{Z:Y}$ and check if $\Gamma_{Z:X} = \Gamma_{Z:Y} \Gamma_{Y:X}$. The latter condition is stronger than the former because $\Gamma_{Z:X}$ represents the actual process $Y \rightarrow Z$ (see [23] for details).

Importantly, it is known that there are non-Markovian processes that are divisible [23]. This is because divisibility only accounts for pairwise correlations and neglects higher-order correlations in time. Then a natural question is whether divisible processes satisfy the set of inequalities presented, for instance, in (6)? Recently, the equivalence between a nonentropic data processing inequality and divisibility was proved in Ref. [32]. It is important to stress that their data processing inequality is distinct from the ones considered here. Furthermore, Ref. [32] only considers divisibility by inversion, and the equivalence may not hold for when operational divisibility is considered. Here we show that the operational divisibility is sufficient for satisfying the data processing and the monogamy relations.

Suppose a process is operationally divisible; then for any map $\Gamma_{Z:X}$ can be written as $\Gamma_{Z:X} = \Gamma_{Z:Y} \Gamma_{Y:X}$ for some intermediate time step Y . Here, the RHS is a Markov process, meaning all pairwise correlations, i.e., the mutual information $\{I(Z : X), I(Y : X), \dots\}$, can be obtained from the underlying Markov process. Since the inequalities in Eq. (6) only require pairwise correlations, which in effect come from a Markov process, all inequities there will be satisfied because they are derived under the Markov assumption. It is worth stating that the same argument does not hold if the process is divisible by inversion.

Example relating divisibility and the inequalities

A divisible non-Markovian process. We give here an illustrative example of a non-Markovian process that is divisible and thus only carries higher-order correlations. Consider a one bit process with $x_1 = 0, 1$ with probability $1/2$. Let $x_j = y_j$ for $j = 2, 3$, where y_j are random bits. Finally, we let $x_4 = x_1 + y_2 + y_3$ (here we have modular addition). It's clear that the mutual information between any two marginals will be zero since they are all independent and random. In fact, the process is divisible and therefore all inequalities in (6) will be trivially satisfied. However, higher-order correlations, those containing correlations between three or more variables, cannot be obtained in the same way. Indeed this is exactly why a divisible process can be non-Markovian. For instance, the

mutual information such as $I(X_4 : X_1 X_2 X_3)$ or $I(X_4 : X_1 | X_2 X_3)$ will not vanish. The above example remains the same, in spirit, if we omit either step (2) or (3). That is, even for a three-step non-Markovian process divisibility may hold.

An indivisible non-Markovian process that satisfies all inequalities. It should be clear now that there are non-Markovian processes that have correlations that cannot be detected by the inequalities in (6). In fact, the converse to the above statement does not hold in general. That is, there are nondivisible (hence non-Markovian) processes that satisfy all of the inequalities in (6). As an example, consider the correlated process

$$p(X_1, X_2, X_3, X_4) = \frac{1}{100}(6, 9, 6, 6, 2, 1, 4, 10, 1, 9, 8, 8, 4, 10, 10, 6). \quad (9)$$

Here we find that the monogamy inequality and all data processing inequalities hold.

V. CAUSAL INTERPRETATION OF THE DATA PROCESSING AND MONOGAMY INEQUALITIES

The violation of the data processing or monogamy inequalities implies that the Markov constraint is not fulfilled by the process under investigation. It seems natural that the more we violate such constraints, the more non-Markovian the process should be. In order to formalize that quantitatively, we make use here of the causal Bayesian networks formalism [14].

A central concept is that of a *directed acyclic graph* (DAG) which has the variables X_i 's as vertices. The directed edges in the DAG represent relations of causal and effect, the reason why the graph should be acyclic; otherwise we would incur in paradoxical situations where the effect is its own cause. For the X_i 's to form a Bayesian network (with respect to the DAG), every variable should be expressed as a function of its graph-theoretical parents PA_i and an unobserved noise term N_i (such that the N_i 's are jointly independent). That is the case if and only if the distribution is of the form

$$p(x) = \prod_{i=1}^n p(x_i | pa_i).$$

We notice that this is equivalent to the so-called local Markov property stating that every X_i is conditionally independent of its nondescendants ND_i given its parents PA_i , that is, $p(x_i, nd_i | pa_i) = p(x_i | pa_i)p(nd_i | pa_i)$.

Within this context, a Markovian process is nothing other than a causal model where a given variable X_i has only X_{i-1} as a parent and X_{i+1} as a descendant (see Fig. 2). Thus a natural way to quantify non-Markovianity is to quantify how much causal dependence X_i has on X_{i-2}, \dots, X_1 . To that aim, we notice that a list of reasonable postulates any measure of causal strength should fulfill has been proposed in [24], in particular the axiom stating that

$$C_{X \rightarrow Y} \geq I(X : Y | PA_Y^X),$$

where $C_{X \rightarrow Y}$ is the causal strength of X into Y and PA_Y^X stands for the parents of variable Y other than X . We refer to [24] for a specific measure of causal influence fulfilling such axiom, quantified via the relative entropy between the

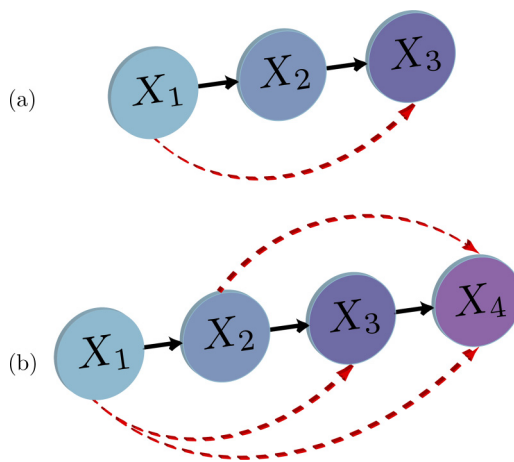


FIG. 2. DAGs illustrating a Markovian/non-Markovian chain. A DAG with (a) three and (b) four nodes. The arrows in solid black represent the causal relations respected by a Markov process. In dashed red are the causal arrows imposing a violation of the Markov condition.

observed probability distribution and another distribution obtained from a causal intervention. In general, if the Markovian condition is not fulfilled, $H(X_i | X_{i-1}, \dots, X_1) \neq H(X_i | X_{i-1})$, this means that the variables (X_1, \dots, X_{i-2}) have a direct causal influence over X_i (see Fig. 2) and that can be quantified as $C_{(X_1, \dots, X_{i-2}) \rightarrow X_i} \geq I(X_1, \dots, X_{i-2} : X_i | X_{i-1})$.

To draw a connection between causality, entropies, and Markovianity, we first notice that if a process is Markovian then

$$H(X_1, \dots, X_n) := \sum_{i=1}^{n-1} H(X_i, X_{i+1}) - \sum_{i=2}^{n-1} H(X_i), \quad (10)$$

as can be seen by a direct application of the chain rule plus the Markov conditions (5). A violation of (10) implies that at least one of the Markov conditions (5) is not fulfilled. In other terms, there is a causal influence over some X_i that is not simply given by its immediate past variable X_{i-1} . Nicely, it can be demonstrated (see Appendix A 4 for the proof) that the following inequality holds for the sum of all non-Markovian causal influences:

$$\sum_{i=3, \dots, n} C_{(X_1, \dots, X_{i-2}) \rightarrow X_i} \geq \sum_{i=1}^{n-1} H(X_i, X_{i+1}) - \sum_{i=2}^{n-1} H(X_i) - H(X_1, \dots, X_n),$$

where $n \geq 3$. The right-hand side is equal to zero if and only if the entropic Markov constraints (5) are fulfilled.

To further illustrate the connection between causality and Markovianity, consider the data processing gap, expressed as

$$DP_{23}^1 = -I_{1,2} + I_{1,3}, \quad (11)$$

which is always negative for Markov processes and may be positive for non-Markovian processes. A violation of it implies that $DP_{23}^1 > 0$ and thus $H(X_3 | X_2, X_1) \neq H(X_3 | X_2)$. Using the entropic framework described above we can prove

that (see Appendix A 5)

$$\mathcal{C}_{X_1 \rightarrow X_3} \geq \text{DP}_{23}^1,$$

thus showing that the violation of the data processing puts a lower bound on the causal effect the variable X_1 has over X_3 .

Similarly, considering the monogamy gap for $n = 4$, expressed as

$$M_4 = -I_{1,4} - I_{2,3} + I_{1,3} + I_{2,4},$$

is also negative for all Markov processes and can be positive for some non-Markovian processes. As proven in Appendix A 6, the monogamy inequality imposes a lower bound to the total non-Markovianity of the process, as quantified by the sum of causal influences

$$\mathcal{C}_{(X_1, X_2) \rightarrow X_4} + \mathcal{C}_{X_1 \rightarrow X_3} \geq M_4. \quad (12)$$

We could analytically prove (see Appendixes A 6, A 7, A 8, and A 9), up to $n = 10$, that the following inequality holds

$$\sum_{i=3, \dots, n} \mathcal{C}_{(X_1, \dots, X_{i-2}) \rightarrow X_i} \geq M_n,$$

where, similarly to M_4 defined above, M_n is the sum of mutual information terms appearing in the monogamy inequalities (7). Unfortunately, a general result for larger chains is still missing. Overall, this framework allows for an operational causal interpretation for the violation of data processing and monogamy inequalities, as a lower bound for the non-Markovian causal influences necessary to explain the observed correlations.

VI. DISCUSSION

The main question addressed in this article is the following: if a process is Markovian, what are the additional restrictions (beyond data processing inequality) imposed on the pairwise correlations between the variables in the Markov chain? This is a very important question since we do not have, in general, access to the full probability distribution describing the process. For instance, we cannot perform sequential time measurements in a continuous way, implying that a sequence of events in time must necessarily be discrete. Our contribution in the search for answering this question, based on an entropic approach, is threefold.

First, we know that Markov processes satisfy all data processing inequalities, which are restrictions on the possible correlations between distinct time steps. However, the converse to this statement is not true. Along these lines, we demonstrate that data processing constraints are not the only consequences of the Markov condition. We proved a class of relations, named monogamy inequalities, that can be violated by non-Markovian processes that satisfy all data processing inequalities. In this way, the violation of one of these monogamy inequalities can be employed as a device-independent test of the non-Markovianity of the underlying process. It is worthy to highlight that increasing the size of the Markov chain it is likely that different kinds of constraints, beyond data processing and monogamy, might appear. Also, in the derivation of the monogamy inequalities, only strong subadditivity constraints have been employed, meaning that

they also hold for von Neumann entropies, if the corresponding joint density matrix respecting the Markov condition can be defined [28,33,34]. Exploring these possibilities defines a clear venue for future research.

Our second contribution deals with the concepts of Markovianity and divisibility, sometimes seen as synonymous in the literature. By considering divisible processes, we showed that the definition of operational divisibility is sufficient to guarantee that the process will satisfy all data processing and monogamy inequalities, even if the process is non-Markovian (but operationally divisible). This result is very interesting because it points out the distinction between Markovianity and divisibility, in such a way that it can be experimentally investigated.

Finally, our third contribution is to build a connection between the violation of data processing and monogamy inequalities with causal influences. In short, the violation of these inequalities imposes nontrivial lower bounds on the causal influence from the past.

In summary, the ideas put forward in this article may have potential applications in several fields, like causal inference, statistical physics, and information theory.

ACKNOWLEDGMENTS

M.C., L.C.C., and R.C. thank the funding agency CNPq (Grants No. 307172/2017-1, No. 406574/2018-9, and No. 305740/2016-4 and INCT-IQ). M.C. acknowledges the warmly hospitality of International Institute of Physics. R.C. also acknowledges the Brazilian ministries MEC and MC-TIC, the Serrapilheira Institute (Grant No. Serra1708-15763), and the John Templeton Foundation via Grant Q-CAUSAL No. 61084. L.C.C. acknowledges FAPEG (PRONEX No. 201710267000503), CAPES (Finance Code 001) and support from Spanish MCIU/AEI/FEDER (Grant No. PGC2018-095113-B-I00), Basque Government IT986-16, the projects QMiCS (No. 820505) and OpenSuperQ (No. 820363) of the EU Flagship on Quantum Technologies and the EU FET Open Grant Quromorphic, and the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) quantum algorithm teams program, under field work proposal No. ERKJ333. K.M. is supported through ARC Grant No. FT160100073.

APPENDIX

1. Analytical derivation of the data processing inequality

Let us first see how the data processing inequalities can be proven. Simply use the Markov condition $H(X_3|X_1, X_2) = H(X_3|X_2)$ [rewritten as $H(X_1, X_2, X_3) = H(X_1, X_2) + H(X_2, X_3) - H(X_2)$] in the elemental inequality

$$H(X_1, X_3) + H(X_2, X_3) \geq H(X_1, X_2, X_3) + H(X_3)$$

to obtain

$$-H(X_1, X_2) + H(X_1, X_3) + H(X_2) - H(X_3) \geq 0,$$

which can be rewritten as

$$I_{1,2} \geq I_{1,3},$$

employing the same notation used in the main text.

2. Analytical derivation of the monogamy inequality for $n = 4$

Let us now prove the inequality beyond data processing. Add the following basic inequalities

$$\begin{aligned} &H(X_1, X_2, X_4) + H(X_1, X_3, X_4) \\ &\geq H(X_1, X_2, X_3, X_4) + H(X_1, X_4), \\ &H(X_1, X_2) + H(X_2, X_4) \geq H(X_1, X_2, X_4) + H(X_2), \\ &H(X_1, X_3) + H(X_3, X_4) \geq H(X_1, X_3, X_4) + H(X_3) \end{aligned}$$

to obtain

$$\begin{aligned} &H(X_1, X_2) + H(X_1, X_3) - H(X_1, X_4) + H(X_2, X_4) \\ &+ H(X_3, X_4) - H(X_2) - H(X_3) - H(X_1, X_2, X_3, X_4) \geq 0. \end{aligned}$$

Using the Markov conditions $H(X_3|X_1, X_2) = H(X_3|X_2)$ and $H(X_4|X_1, X_2, X_3) = H(X_4|X_3)$ we can write

$$\begin{aligned} &H(X_1, X_2, X_3, X_4) \\ &= H(X_1, X_2) + H(X_2, X_3) + H(X_3, X_4) - H(X_2) - H(X_3). \end{aligned}$$

Substituting that in the expression above we get

$$H(X_1, X_3) - H(X_1, X_4) + H(X_2, X_4) - H(X_2, X_3) \geq 0,$$

which can be rewritten as

$$-I_{1,3} + I_{1,4} - I_{2,4} + I_{2,3} \geq 0,$$

3. Redundant data processing inequalities

Data processing (DP) inequalities are of the form

$$I_{i,j} \geq I_{r,s}$$

such that $i \geq r$ and $j \leq s$, where we consider without loss of generality $i < j$ and $r < s$ as a matter of mathematical convention. For the case $n = 4$ we have nine DP inequalities:

$$\begin{aligned} &I_{1,2} \geq I_{1,3}, \\ &I_{1,2} \geq I_{1,4}, \\ &I_{1,3} \geq I_{1,4}, \\ &I_{2,3} \geq I_{1,3}, \\ &I_{2,3} \geq I_{2,4}, \\ &I_{2,3} \geq I_{1,4}, \\ &I_{3,4} \geq I_{1,4}, \\ &I_{3,4} \geq I_{2,4}, \\ &I_{2,4} \geq I_{1,4}. \end{aligned}$$

The nine DP inequalities above are implied by the following six, as can be seen by just adding two of

them:

$$\begin{aligned} &I_{1,2} \geq I_{1,3}, \\ &I_{1,3} \geq I_{1,4}, \\ &I_{2,3} \geq I_{1,3}, \\ &I_{2,3} \geq I_{2,4}, \\ &I_{3,4} \geq I_{2,4}, \\ &I_{2,4} \geq I_{1,4}. \end{aligned}$$

In turn, this set of six DP inequalities is implied by the four DP inequalities plus the monogamy inequality in (6). For instance, the DP inequality $I_{2,3} \geq I_{1,3}$ can be obtained by summing the DP inequality $I_{2,4} \geq I_{1,4}$ with the monogamy inequality. Similarly, the DP inequality $I_{2,3} \geq I_{2,4}$ can be obtained by summing the DP inequality $I_{1,3} \geq I_{1,4}$ with the monogamy inequality. Thus the nonredundant set of DP inequalities is given by those expressed in (6).

4. Causal influences and entropy

Consider a stochastic process $\{X_1, \dots, X_n\}$. One defines the joint Markov entropy by

$$H_{\text{Markov}}(X_1, \dots, X_n) := \sum_{i=1}^{n-1} H(X_i, X_{i+1}) - \sum_{i=2}^{n-1} H(X_i),$$

as the joint entropy of Markovian processes has the exact form above. The sum of the non-Markovian causal influences is lower bounded as follows:

$$\sum_{i=3}^n \mathcal{C}_{(X_1, \dots, X_{i-2}) \rightarrow X_i} \geq H_{\text{Markov}}(X_1, \dots, X_n) - H(X_1, \dots, X_n). \tag{A1}$$

The proof follows by mathematical induction on the number of random variables. Suppose that the following equality is valid for n :

$$\begin{aligned} &\sum_{i=1}^n I(X_1, \dots, X_{i-2} : X_i | X_{i-1}) \\ &= H_{\text{Markov}}(X_1, \dots, X_n) - H(X_1, \dots, X_n). \end{aligned} \tag{A2}$$

Then, it implies that the proposition is also valid for $n + 1$ as

$$\begin{aligned} &\sum_{i=1}^{n+1} I(X_1, \dots, X_{i-2} : X_i | X_{i-1}) \\ &= I(X_1, \dots, X_{n-1} : X_{n+1} | X_n) \\ &+ \sum_{i=1}^n I(X_1, \dots, X_{i-2} : X_i | X_{i-1}) \\ &= H(X_1, \dots, X_n) + H(X_n, X_{n+1}) \\ &- H(X_n) - H(X_1, \dots, X_{n+1}) \\ &+ H_{\text{Markov}}(X_1, \dots, X_n) - H(X_1, \dots, X_n) \\ &= H_{\text{Markov}}(X_1, \dots, X_{n+1}) - H(X_1, \dots, X_{n+1}). \end{aligned}$$

Now, since Eq. (A2) is true for $n = 3$,

$$I(X_1 : X_3 | X_2) = H_{\text{Markov}}(X_1, X_2, X_3) - H(X_1, X_2, X_3),$$

we conclude that the proof holds also for $n \geq 3$.

5. Proof of the causal interpretation of the data processing inequality

We add the term $I_{1,2} - I_{1,2} = 0$ to the strong subadditive constraint

$$H(X_1, X_3) + H(X_2, X_3) - H(X_1, X_2, X_3) - H(X_3) \geq 0$$

to obtain

$$H(X_1, X_3) + H(X_2, X_3) - H(X_1, X_2, X_3) - H(X_3) - H(X_1) - H(X_2) + H(X_1, X_2) \geq -I(X_1 : X_2),$$

which is equal to

$$H(X_1, X_2) + H(X_2, X_3) - H(X_1, X_2, X_3) - H(X_2) \geq -I(X_1 : X_2) + H(X_1) + H(X_3) - H(X_1, X_3),$$

thus leading to the causal bound

$$I(X_1 : X_3|X_2) \geq -I_{1,2} + I_{1,3}.$$

6. Proof of the causal interpretation of the monogamy inequality ($n = 4$)

The sum of the non-Markovian causal influences for the case $n = 4$ is lower bounded by

$$\begin{aligned} &I(X_1 : X_3|X_2) + I(X_1, X_2 : X_4|X_3) \\ &= H(X_1, X_2) + H(X_2, X_3) - H(X_1, X_2, X_3) - H(X_2) \\ &\quad + H(X_1, X_2, X_3) + H(X_3, X_4) \\ &\quad - H(X_1, X_2, X_3, X_4) - H(X_3) \\ &= H(X_1, X_2) + H(X_2, X_3) + H(X_3, X_4) \\ &\quad - H(X_2) - H(X_3) - H(X_1, X_2, X_3, X_4), \end{aligned}$$

in agreement with the previous general derivation for any number of random variables.

Adding the following strong subadditivity inequalities

$$\begin{aligned} I(X_2 : X_3|X_1, X_4) &= H(X_1, X_2, X_4) + H(X_1, X_3, X_4) \\ &\quad - H(X_1, X_2, X_3, X_4) - H(X_1, X_4) \geq 0, \\ I(X_1 : X_4|X_2) &= H(X_1, X_2) + H(X_2, X_4) - H(X_1, X_2, X_4) \\ &\quad - H(X_2) \geq 0, \\ I(X_1 : X_4|X_3) &= H(X_1, X_3) + H(X_3, X_4) - H(X_1, X_3, X_4) \\ &\quad - H(X_3) \geq 0, \end{aligned}$$

the valid inequality for an arbitrary process follows,

$$H(X_1, X_2) + H(X_2, X_3) + H(X_3, X_4) - H(X_2) - H(X_3) - H(X_1, X_2, X_3, X_4) \tag{A3}$$

$$+ H(X_2, X_4) + H(X_1, X_3) - H(X_2, X_3) - H(X_1, X_4) \geq 0. \tag{A4}$$

Writing it in terms of mutual information measures we have

$$I(X_1 : X_3|X_2) + I(X_1, X_2 : X_4|X_3) \geq -I_{1,4} - I_{2,3} + I_{1,3} + I_{2,4},$$

and the demonstration is completed.

7. Proof of the causal interpretation of the monogamy inequality ($n = 6$)

In the same way we did in Appendix A 6, for this case we have to play with the following strong subadditive inequalities

$$\begin{aligned} I(X_3 : X_4|X_1, X_2, X_5, X_6) &\geq 0, \\ I(X_1 : X_3|X_2, X_5, X_6) &\geq 0, \\ I(X_4 : X_6|X_1, X_2, X_5) &\geq 0, \\ I(X_3 : X_6|X_2, X_5) &\geq 0, \\ I(X_2 : X_4|X_1, X_5) &\geq 0, \\ I(X_2 : X_5|X_1, X_6) &\geq 0, \\ I(X_1 : X_6|X_5) &\geq 0, \\ I(X_1 : X_5|X_4) &\geq 0, \\ I(X_2 : X_5|X_3) &\geq 0, \\ I(X_1 : X_6|X_2) &\geq 0, \end{aligned}$$

in order to get the causal inequality

$$\sum_{i=3, \dots, 6} \mathcal{C}_{(X_1, \dots, X_{i-2}) \rightarrow X_i} \geq I_{1,4} + I_{2,6} + I_{3,5} - I_{1,6} - I_{2,5} - I_{3,4}.$$

8. Proof of the causal interpretation of the monogamy inequality ($n = 8$)

For this case, the strong subadditive inequalities are given by

$$\begin{aligned} I(X_4 : X_5|X_1, X_2, X_3, X_6, X_7, X_8) &\geq 0, \\ I(X_1 : X_4|X_2, X_3, X_6, X_7, X_8) &\geq 0, \\ I(X_5 : X_8|X_1, X_2, X_3, X_6, X_7) &\geq 0, \\ I(X_4 : X_8|X_2, X_3, X_6, X_7) &\geq 0, \\ I(X_3 : X_8|X_1, X_2, X_6, X_7) &\geq 0, \\ I(X_1 : X_3|X_2, X_5, X_6, X_7) &\geq 0, \\ I(X_2 : X_5|X_1, X_6, X_7) &\geq 0, \\ I(X_3 : X_5|X_2, X_6, X_7) &\geq 0, \\ I(X_2 : X_4|X_3, X_6, X_7) &\geq 0, \\ I(X_2 : X_7|X_1, X_6, X_8) &\geq 0, \\ I(X_1 : X_7|X_5, X_6) &\geq 0, \\ I(X_3 : X_6|X_2, X_7) &\geq 0, \\ I(X_4 : X_7|X_3, X_6) &\geq 0, \\ I(X_2 : X_6|X_1, X_8) &\geq 0, \\ I(X_1 : X_8|X_6, X_7) &\geq 0, \\ I(X_1 : X_8|X_2) &\geq 0, \\ I(X_2 : X_7|X_3) &\geq 0, \\ I(X_3 : X_6|X_4) &\geq 0, \\ I(X_1 : X_6|X_5) &\geq 0, \\ I(X_5 : X_7|X_6) &\geq 0, \\ I(X_6 : X_8|X_7) &\geq 0, \end{aligned}$$

thus leading us to the causal inequality

$$\sum_{i=3,\dots,8} \mathcal{C}_{(X_1,\dots,X_{i-2})\rightarrow X_i} \geq I_{1,5} + I_{2,8} + I_{3,7} + I_{4,6} - I_{1,8} - I_{2,7} - I_{3,6} - I_{4,5}.$$

9. Proof of the causal interpretation of the monogamy inequality ($n = 10$)

For this case, the strong subadditive inequalities are given by

$$\begin{aligned} I(X_5 : X_6 | X_1, X_2, X_3, X_4, X_7, X_8, X_9, X_{10}) &\geq 0, \\ I(X_1 : X_5 | X_2, X_3, X_4, X_7, X_8, X_9, X_{10}) &\geq 0, \\ I(X_6 : X_{10} | X_1, X_2, X_3, X_4, X_7, X_8, X_9) &\geq 0, \\ I(X_5 : X_{10} | X_2, X_3, X_4, X_7, X_8, X_9) &\geq 0, \\ I(X_1 : X_4 | X_2, X_3, X_6, X_7, X_8, X_9) &\geq 0, \\ I(X_4 : X_{10} | X_1, X_2, X_3, X_7, X_8, X_9) &\geq 0, \\ I(X_2 : X_5 | X_3, X_4, X_7, X_8, X_9) &\geq 0, \\ I(X_2 : X_6 | X_1, X_3, X_7, X_8, X_9) &\geq 0, \\ I(X_4 : X_6 | X_2, X_3, X_7, X_8, X_9) &\geq 0, \\ I(X_2 : X_7 | X_1, X_3, X_8, X_9, X_{10}) &\geq 0, \\ I(X_5 : X_9 | X_3, X_4, X_7, X_8) &\geq 0, \\ I(X_3 : X_6 | X_1, X_7, X_8, X_9) &\geq 0, \\ I(X_2 : X_4 | X_3, X_7, X_8, X_9) &\geq 0, \\ I(X_2 : X_8 | X_1, X_3, X_9, X_{10}) &\geq 0, \\ I(X_3 : X_8 | X_1, X_7, X_9, X_{10}) &\geq 0, \\ I(X_3 : X_5 | X_4, X_7, X_8) &\geq 0, \end{aligned}$$

$$\begin{aligned} I(X_6 : X_8 | X_1, X_7, X_9) &\geq 0, \\ I(X_4 : X_7 | X_3, X_8, X_9) &\geq 0, \\ I(X_1 : X_3 | X_2, X_9, X_{10}) &\geq 0, \\ I(X_3 : X_7 | X_1, X_9, X_{10}) &\geq 0, \\ I(X_1 : X_8 | X_7, X_9, X_{10}) &\geq 0, \\ I(X_5 : X_8 | X_4, X_7) &\geq 0, \\ I(X_6 : X_9 | X_1, X_7) &\geq 0, \\ I(X_4 : X_9 | X_3, X_8) &\geq 0, \\ I(X_2 : X_9 | X_1, X_{10}) &\geq 0, \\ I(X_3 : X_{10} | X_2, X_9) &\geq 0, \\ I(X_1 : X_9 | X_7, X_{10}) &\geq 0, \\ I(X_7 : X_9 | X_8, X_{10}) &\geq 0, \\ I(X_4 : X_7 | X_5) &\geq 0, \\ I(X_1 : X_7 | X_6) &\geq 0, \\ I(X_3 : X_8 | X_4) &\geq 0, \\ I(X_1 : X_{10} | X_2) &\geq 0, \\ I(X_2 : X_9 | X_3) &\geq 0, \\ I(X_1 : X_{10} | X_7) &\geq 0, \\ I(X_7 : X_{10} | X_8) &\geq 0, \\ I(X_8 : X_{10} | X_9) &\geq 0, \end{aligned}$$

thus leading us to the causal inequality

$$\begin{aligned} \sum_{i=3,\dots,10} \mathcal{C}_{(x_1,\dots,x_{i-2})\rightarrow x_i} &\geq I_{1,6} + I_{2,10} + I_{3,9} + I_{4,8} + I_{5,7} - I_{1,10} - I_{2,9} \\ &\quad - I_{3,8} - I_{4,7} - I_{5,6}. \end{aligned}$$

[1] G. Bolch, S. Greiner, H. De Meer, and K. S. Trivedi, *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications* (John Wiley & Sons, New York, 2006).

[2] D. M. Hausman and J. Woodward, Independence, invariance and the causal Markov condition, *British J. Philos. Sci.* **50**, 521 (1999).

[3] D. Janzing, R. Chaves, and B. Schölkopf, Algorithmic independence of initial condition and dynamical law in thermodynamics and causal inference, *New J. Phys.* **18**, 093052 (2016).

[4] W. R. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall/CRC, London, 1995).

[5] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann *et al.*, Computational social science, *Science* **323**, 721 (2009).

[6] O. Francois, S. Ancelet, and G. Guillot, Bayesian clustering using hidden Markov random fields in spatial population genetics, *Genetics* **174**, 805 (2006).

[7] F. Spitzer, *Principles of Random Walk* (Springer Science & Business Media, New York, 2013), Vol. 34.

[8] K. L. Chung, *Lectures from Markov Processes to Brownian Motion* (Springer Science & Business Media, New York, 2013), Vol. 249.

[9] O. Fawzi and R. Renner, Quantum conditional mutual information and approximate Markov chains, *Commun. Math. Phys.* **340**, 575 (2015).

[10] H.-P. Breuer, E.-M. Laine, J. Piilo, and B. Vacchini, Colloquium: Non-Markovian dynamics in open quantum systems, *Rev. Mod. Phys.* **88**, 021002 (2016).

[11] T. Fritz and R. Chaves, Entropic inequalities and marginal problems, *IEEE Trans. Inf. Theory* **59**, 803 (2012).

[12] D. Geiger and C. Meek, Quantifier elimination for statistical problems, in *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (Morgan Kaufmann Publishers Inc., New York, 1999), pp. 226–235.

[13] R. Chaves, Polynomial Bell Inequalities, *Phys. Rev. Lett.* **116**, 010402 (2016).

[14] J. Pearl, *Causality: Models, Reasoning and Inference* (Springer, New York, 2000), Vol. 29.

[15] L. D. Garcia, M. Stillman, and B. Sturmfels, Algebraic geometry of Bayesian networks, *J. Symbol. Comput.* **39**, 331 (2005).

- [16] R. Chaves, L. Luft, T. O. Maciel, D. Gross, D. Janzing, and B. Schölkopf, Inferring latent structures via information inequalities, in *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI 2014)* (AUAI Press, 2014), pp. 112–121.
- [17] M. Weilenmann and R. Colbeck, Analysing causal structures with entropy, *Proc. R. Soc. A* **473**, 20170483 (2017).
- [18] C. Budroni, N. Miklin, and R. Chaves, Indistinguishability of causal relations from limited marginals, *Phys. Rev. A* **94**, 042127 (2016).
- [19] R. Chaves, C. Majenz, and D. Gross, Information-theoretic implications of quantum causal structures, *Nat. Commun.* **6**, 5766 (2015).
- [20] 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**, 230501 (2007).
- [21] R. Gallego, N. Brunner, C. Hadley, and A. Acín, Device-Independent Tests of Classical and Quantum Dimensions, *Phys. Rev. Lett.* **105**, 230501 (2010).
- [22] R. Chaves, J. B. Brask, and N. Brunner, Device-Independent Tests of Entropy, *Phys. Rev. Lett.* **115**, 110501 (2015).
- [23] S. Milz, M. S. Kim, F. A. Pollock, and K. Modi, Completely Positive Divisibility does not Mean Markovianity, *Phys. Rev. Lett.* **123**, 040401 (2019).
- [24] D. Janzing, D. Balduzzi, M. Grosse-Wentrup, and B. Schölkopf, Quantifying causal influences, *Ann. Stat.* **41**, 2324 (2013).
- [25] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, Unifying Framework for Relaxations of the Causal Assumptions in Bell's Theorem, *Phys. Rev. Lett.* **114**, 140403 (2015).
- [26] R. W. Yeung, *Information Theory and Network Coding* (Springer Science & Business Media, New York, 2008).
- [27] Care has to be taken when searching for a process that satisfied the data processing inequalities, but not the monogamy inequality. This is because some of the remaining data processing inequalities are guaranteed to hold only if the monogamy inequality holds; see Appendix A 3 for details.
- [28] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Operational Markov Condition for Quantum Processes, *Phys. Rev. Lett.* **120**, 040405 (2018).
- [29] P. Taranto, S. Milz, F. A. Pollock, and K. Modi, Structure of quantum stochastic processes with finite Markov order, *Phys. Rev. A* **99**, 042108 (2019).
- [30] S. Milz, F. Sakuldee, F. A. Pollock, and K. Modi, Kolmogorov extension theorem for (quantum) causal modeling and general probabilistic theories, [arXiv:1712.02589](https://arxiv.org/abs/1712.02589).
- [31] Á. Rivas, S. F. Huelga, and M. B. Plenio, Quantum non-Markovianity: Characterization, quantification and detection, *Rep. Prog. Phys.* **77**, 094001 (2014).
- [32] F. Buscemi and N. Datta, Equivalence between divisibility and monotonic decrease of information in classical and quantum stochastic processes, *Phys. Rev. A* **93**, 012101 (2016).
- [33] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Non-Markovian quantum processes: Complete framework and efficient characterization, *Phys. Rev. A* **97**, 012127 (2018).
- [34] F. Costa and S. Shrapnel, Quantum causal modelling, *New J. Phys.* **18**, 063032 (2016).