Non-Shannon inequalities in the entropy vector approach to causal structures

Mirjam Weilenmann[*] and Roger Colbeck[†]

Department of Mathematics, University of York, Heslington, York, YO10 5DD, UK.

(Dated: 6th October 2017)

A causal structure is a relationship between observed variables that in general restricts the possible correlations between them. This relationship can be mediated by unobserved systems, modelled by random variables in the classical case or joint quantum systems in the quantum case. One way to differentiate between the correlations realisable by two different causal structures is to use entropy vectors, i.e., vectors whose components correspond to the entropies of each subset of the observed variables. To date, the starting point for deriving entropic constraints within causal structures are the so-called Shannon inequalities (positivity of entropy, conditional entropy and conditional mutual information). In the present work we investigate what happens when non-Shannon entropic inequalities are included as well. We show that in general these lead to tighter outer approximations of the set of realisable entropy vectors and hence enable a sharper distinction of different causal structures. Since non-Shannon inequalities can only be applied amongst classical variables, it might be expected that their use enables an entropic distinction between classical and quantum causal structures. However, this remains an open question.

We also introduce techniques for deriving inner approximations to the allowed sets of entropy vectors for a given causal structure. These are useful for proving tightness of outer approximations or for finding interesting regions of entropy space. We illustrate these techniques in several scenarios, including the triangle causal structure.

I. INTRODUCTION

A common challenge in science is to make predictions based on incomplete information. Full details of the mechanism by which correlations between two or more variables come about is often not apparent and there may be several competing causal explanations. Experimentation with interventions is one way to decide between the candidate explanations [1]. However, in many situations such intervention is difficult (or unethical), for instance if certain involved systems are outside our control.

Considering a particular causal structure generally imposes restrictions on the set of correlations that can be produced. A well-known example of such a constraint is a Bell inequality [2]. That such relations can be violated using measurements on quantum states motivates the consideration of more general *quantum* causal structures. Correlations that can be generated in such structures but not in their classical analogue are the basis for several important cryptographic tasks [3], in particular for device-independent protocols for key distribution [4–7] or the generation of private randomness [8–11]. In a cryptographic scenario, an adversary is usually able to exert influence at particular points in the protocol, which can be conveniently encoded using a causal structure. Characterising the set of possible classical, quantum and post-quantum correlations within a specific causal structure provides a basis to understand further tasks and possible quantum and post-quantum advantages, which were initially studied in specific cases [12–16].

For a general causal structure with unobserved variables, deciding whether a given set of correlations can be generated is computationally difficult and only feasible for small examples [17, 18]. One way to get around this, is to use entropy to simplify the characterisation of the corresponding sets of correlations [19–30]. Rather than looking at the distributions themselves, we consider *entropy vectors* whose components are the joint entropies of each subset of the observed variables. This often[1] has the advantage that the set of entropies realisable in a given causal structure is convex, in contrast to the set of compatible distributions. In addition, the causal constraints can be represented by linear relations between entropies instead of polynomial constraints. It is also significant that entropic constraints on possible correlations in a causal structure are independent of the dimension of the involved random variables. Hence, the method enables the derivation of constraints that are valid for arbitrarily large alphabet sizes of all involved observed and unobserved systems. These properties make entropy vectors a convenient means to distinguish different causal structures in many situations.

In this paper we report the use of non-Shannon inequalities for distinguishing causal structures. After a short outline of the entropy vector approach and after introducing the necessary notation in Section II, we go on to show in Section III that non-Shannon inequalities play a central role for the distinction of causal structures. This is illustrated with the triangle causal structure (Section III A), one of the simplest causal structures in which

---

[*] msw518@york.ac.uk
[†] roger.colbeck@york.ac.uk

---

[1] For classical causal structures, the set is always convex, but for quantum causal structures it is not known whether this is always the case.

there is a separation between classical and quantum at the level of correlations. For this example, we present numerous new entropic constraints, which involve several infinite families of valid inequalities, that together form the tightest entropic characterisation of the classical triangle causal structure known to date. This also leads us to disprove a claim that previously known entropic approximations to this causal structure were tight [22, 26]. Whether our new inequalities are sufficient to separate classical and quantum versions of causal structures is left as an open problem.

In Section III B, we analyse a number of other causal structures, taking into account non-Shannon inequalities for their entropic characterisation. These inequalities are relevant for distinguishing different classical causal structures as well as for settling the question of whether there is a classical-quantum separation in the entropy vector approach.

We further analyse the role of non-Shannon inequalities for the entropic characterisation of the causal structure relevant in the context of information causality [31] in Section III C, where the combination of non-Shannon inequalities with post-selection allows us to derive numerous new entropy inequalities.

In Section IV, we provide the first inner approximations to the entropy cones of causal structures. These are useful for certifying that particular entropy vectors are realisable in a causal structure as well as for showing that for certain entropy cones we have fully characterised their boundary with linear constraints, e.g. for some of the causal structures analysed in Section III B. In scenarios where this is not the case, we find inner approximations that share extremal rays with our outer approximations, which allows us to identify regions of the true entropy cone as well as those regions where identifying the cone's actual boundary requires further analysis.

For comparison with the classical case, we also briefly consider non-Shannon inequalities in the context of quantum and hybrid causal structures in Section V, which is illustrated with the example of the triangle causal structure, before concluding in Section VI.

II. ENTROPIC CONES AND THE ENTROPY VECTOR APPROACH TO CAUSAL STRUCTURES

In this section, we briefly outline the entropy vector approach and introduce the required notation. An elaborate introduction to the topic can for instance be found in the review [32].

A. Entropic cones

For a set of n jointly distributed random variables $\Omega = \{X_1, X_2, \ldots, X_n\}$ taking values in the alphabet $\mathcal{X}_\Omega = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n$ we denote the set of all possible joint

probability distributions as $\mathcal{P}_n$. For a set of variables with joint distribution $P_\Omega \in \mathcal{P}_n$ its *Shannon entropy* [33] is

$$H(\Omega) \coloneqq -\sum_{x \in \mathcal{X}_\Omega} P_\Omega(x) \log_2 (P_\Omega(x)).$$

The Shannon entropy of Ω and of all its subsets can be expressed in an *entropy vector*,

$$\mathbf{H}(P) \coloneqq (H(X_1), \ H(X_2), \ \ldots, \ H(X_n), H(X_1 X_2),$$
$$H(X_1 X_3), \ \ldots, \ H(X_1 X_2 \ldots X_n)) \in \mathbb{R}^{2^n - 1}.$$

The closure of the set of all possible entropy vectors, Γ_n^*, is a convex cone, denoted as $\overline{\Gamma_n^*}$ [34].[2] While for $n \leq 3$, the *entropy cone* $\overline{\Gamma_n^*}$ is polyhedral [3] [35], an infinite number of linear inequalities are required to characterise $\overline{\Gamma_n^*}$ for $n \geq 4$ [36]. Hence, considering approximations to $\overline{\Gamma_n^*}$ is common practice.

1. Approximations to $\overline{\Gamma_n^*}$

Before specifying approximations to $\overline{\Gamma_n^*}$, we define a few quantities, that are relevant in the following. The *conditional entropy* of two disjoint subsets X_S, $X_T \subseteq \Omega$ is defined as

$$H(X_S \mid X_T) \coloneqq H(X_S \cup X_T) - H(X_T)$$

and for three mutually disjoint subsets X_S, X_T, $X_U \subseteq \Omega$ the *conditional mutual information* of X_S and X_T conditioned on X_U is

$$I(X_S : X_T \mid X_U) \coloneqq H(X_S \mid X_U) - H(X_S \mid X_T \cup X_U).$$

Note that the entropy of the empty set is $H(\varnothing) = 0$, so that $H(X_S) = H(X_S \mid \varnothing)$, for example. Two other entropic quantities we will make use of in this article are the *interaction information* [37] of three mutually disjoint subsets X_S, X_T, $X_U \subseteq \Omega$,

$$I(X_S : X_T : X_U) \coloneqq I(X_S : X_T) - I(X_S : X_T \mid X_U),$$

and the *Ingleton quantity* of four mutually disjoint subsets X_S, X_T, X_U, $X_V \subseteq \Omega$,

$$I_{\mathrm{ING}}(X_S, X_T; X_U, X_V) \coloneqq I(X_S : X_T | X_U)$$
$$+ I(X_S : X_T | X_V) + I(X_U : X_V) - I(X_S : X_T). \quad (1)$$

For any entropy vector of a joint distribution of the random variables Ω the following *Shannon inequalities* hold:

---

[2] The closure is taken because there isn't in general a good reason to put an upper bound on the alphabet sizes and it is known that $\overline{\Gamma_n^*} \neq \Gamma_n^*$ for $n \geq 3$ [34].

[3] In fact $\overline{\Gamma_3^*}$ equals the corresponding Shannon cone, Γ_3, introduced below.

- For any $X_S \subseteq \Omega$, $H(X_S) \geq 0$.

- For any disjoint X_S, $X_T \subseteq \Omega$, $H(X_S \mid X_T) \geq 0$.

- For any disjoint X_S, X_T, $X_U \subseteq \Omega$, $I(X_S : X_T \mid X_U) \geq 0$.

They are known to constrain a convex polyhedral cone, the *Shannon cone*, Γ_n [38]. Because the Shannon inequalities hold for any entropy vector we have $\overline{\Gamma_n^*} \subseteq \Gamma_n$.

The first entropy inequality that is not of Shannon type was found in [34] and is presented in the following.

Proposition 1 (Zhang & Yeung). *For any four discrete random variables X_1, X_2, X_3 and X_4 the following inequality holds:*

$$I(X_1 : X_2|X_3) + I(X_1 : X_2|X_4) + I(X_3 : X_4) - I(X_1 : X_2)$$
$$+I(X_1 : X_3|X_2) + I(X_2 : X_3|X_1) + I(X_1 : X_2|X_3) \geq 0.$$

In the following the lhs of this inequality is abbreviated as $\Diamond_{X_1 X_2 X_3 X_4}$.

The first account of infinite families of inequalities was given in [36].

Proposition 2 (Matúš). *Let X_1, X_2, X_3 and X_4 be random variables and let $s \in \mathbb{N}$. Then the following inequalities hold:*

$$s[I(X_1 : X_2|X_3)+I(X_1 : X_2|X_4)+I(X_3 : X_4)-I(X_1 : X_2)]$$
$$+I(X_1 : X_3|X_2)+\frac{s(s+1)}{2}[I(X_2 : X_3|X_1)+I(X_1 : X_2|X_3)]\geq 0, \tag{2}$$

$$s[I(X_1 : X_2|X_3)+I(X_1 : X_2|X_4)+I(X_3 : X_4)-I(X_1 : X_2)]$$
$$+ s[I(X_2 : X_3|X_1) + I(X_1 : X_2|X_3)] + I(X_1 : X_3|X_2)$$
$$+ \frac{s(s-1)}{2}[I(X_2 : X_4|X_1) + I(X_1 : X_2|X_4)] \geq 0. \tag{3}$$

For $s = 1$ both inequalities are equivalent to $\Diamond_{X_1 X_2 X_3 X_4} \geq 0$. For the current state of the art on non-Shannon inequalities we refer to [39]. To our knowledge, all known non-Shannon entropy inequalities in four variables that are not (known to be) rendered redundant by tighter ones can be written as the sum of the Ingleton quantity and (conditional) mutual information terms [35, 36, 39–42].

Complementary to outer approximations (such as the Shannon cone, Γ_n) it is also interesting to consider inner approximations, Γ_n^I, to the n-variable entropy cone $\overline{\Gamma_n^*}$. Such approximations can be defined in terms of so-called linear rank inequalities. In the case of $n = 4$ the Shannon inequalities and the *Ingleton inequality*, i.e.,

$$I_{\mathrm{ING}}(X_1, X_2; X_3, X_4) \geq 0 \tag{4}$$

(and its permutations), define such an inner approximation Γ_4^I [43]. Γ_5^I is defined by the Shannon inequalities, all instances of the Ingleton inequality and 24 additional classes of inequalities [44]. For 6 or more variables, a complete list of all linear rank inequalities is not known, and it is even unknown whether such a list of inequalities would be finite. A list of over a billion inequalities (counting permutations) has been found [45].[4]

B. The entropy vector approach to causal structures

A *causal structure*, C, is a set of variables arranged in a directed acyclic graph (DAG). The *parents*, $X^{\downarrow_1}$, of a variable X in a DAG are the variables from which an arrow is directly pointing at X, and the *descendants*, $X^\uparrow$ of X are all variables that may be reached from X along a directed path within the DAG. We use C^C and C^Q to denote the classical and the quantum version of a causal structure respectively.

1. Classical causal structures

The graph of a classical causal structure, C^C, with random variables X_1, X_2, ..., X_n, encodes the independence relations of X_1, X_2, ..., X_n in the sense that the distribution $P_{X_1 X_2 ... X_n}$ is said to be *compatible* with C^C if it can be decomposed as

$$P_{X_1 X_2 ... X_n} = \prod_i P_{X_i | X_i^{\downarrow_1}}.$$

This interpretation of classical causal structures follows the theory of Bayesian networks [1]. The set of all compatible distributions is in the following denoted $\mathcal{P}(C^C)$. The compatibility requirement is equivalent to the condition that for each variable X_i,

$$I(X_i : X_i^{\updownarrow} \mid X_i^{\downarrow_1}) = 0. \tag{5}$$

$X_i^{\updownarrow}$ denotes the *non-descendants* of X_i, i.e., all variables in the causal structure except for the variable itself and its descendants. [5]

The entropic description of causal structures has first been considered in [19, 21]. The n equalities (5) restrict the n-variable entropy cone $\overline{\Gamma_n^*}$ to the cone of all entropy vectors compatible with C^C, denoted $\overline{\Gamma^*}(C^C)$. An outer approximation to $\overline{\Gamma^*}(C^C)$ is constructed by supplementing Γ_n with the same n equalities, which leads to the cone $\Gamma(C^C)$.

———

[4] Note that we can always obtain some inner approximations with other methods, e.g., by constructing a set of achievable entropy vectors and taking their convex hull.

[5] In particular, all other (conditional) independence relations of variables in the causal structure are implied by these n equalities.

When k out of the n variables of C^{C} are observed, we take these to be the first k variables, $X_1, X_2, \ldots, X_k$, without loss of generality. For $k < n$ we are then interested in deriving constraints for the observed variables only. For a compatible distribution, $P_{\mathrm{X_1 X_2 \cdots X_n}} \in \mathcal{P}\left(C^{\mathrm{C}}\right)$, this is achieved by marginalising over the unobserved variables $X_{k+1}, X_{k+2}, \ldots, X_n$ which yields a distribution in the set of all compatible marginal distributions $P_{\mathrm{X_1 X_2 \cdots X_k}} \in \mathcal{P}_{\mathcal{M}}\left(C^{\mathrm{C}}\right)$. Entropically, marginalisation corresponds to a projection of the entropy cone to the corresponding k-variable marginal cone $\overline{\Gamma^*_{\mathcal{M}}}\left(C^{\mathrm{C}}\right) \subsetneqq \mathbb{R}^{2^k-1}$, which would be obtained by dropping all components involving any of the $n - k$ unobserved variables from each vector in $\overline{\Gamma^*}\left(C^{\mathrm{C}}\right)$. The outer approximation $\Gamma\left(C^{\mathrm{C}}\right)$ can be analogously projected to an approximation, $\Gamma_{\mathcal{M}}\left(C^{\mathrm{C}}\right)$, of $\overline{\Gamma^*_{\mathcal{M}}}\left(C^{\mathrm{C}}\right)$. Computationally, $\Gamma_{\mathcal{M}}\left(C^{\mathrm{C}}\right)$ is usually found by considering $\Gamma\left(C^{\mathrm{C}}\right)$ characterised by means of bounding hyperplanes and applying a Fourier-Motzkin elimination algorithm to the system of linear inequalities [46].

2. Quantum causal structures

A quantum causal structure C^{Q} differs from its classical analogue in that the unobserved nodes correspond to quantum systems. Here, we only consider causal structures with two generations of nodes, where the nodes of the first generation are unobserved quantum systems whereas the nodes of the second generation represent observed (classical) variables. Note that this also allows for the description of causal structures with observed input nodes, as is illustrated in Figure 1.

For such causal structures, each edge has an associated Hilbert space, which can be labelled by the parent and child, e.g., for a DAG with an edge $X \to Y$, there is an associated $\mathcal{H}_{X_Y}$. Each unobserved node is labelled by a quantum state, a density operator on the tensor product of the Hilbert spaces associated with the edges originating at that node. For each observed node there is an associated POVM that acts on the tensor product of the Hilbert spaces associated with the edges that meet at that node. The distributions of the observed variables that are *compatible* with a causal structure C^{Q}, $P \in \mathcal{P}_{\mathcal{M}}\left(C^{\mathrm{Q}}\right)$, are those resulting from performing the specified POVMs on the relevant systems via the Born rule.

A technique to analyse these sets entropically was proposed by Chaves et al. [26] and is outlined in the following, where the idea of considering entropy cones of multi-party quantum states goes back to Pippenger [47]. The set of compatible observed distributions $P \in \mathcal{P}_{\mathcal{M}}\left(C^{\mathrm{Q}}\right)$ can be mapped to a set of compatible entropy vectors, the closure of which is denoted $\overline{\Gamma^*_{\mathcal{M}}}\left(C^{\mathrm{Q}}\right)$. To approximate this set, a system is assigned to each observed variable as well as to each outgoing edge of each unobserved node. As opposed to the classical case, where we can always
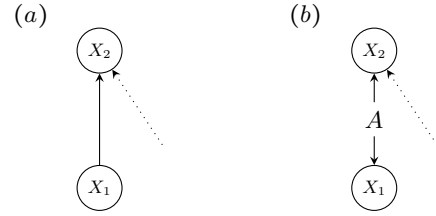

Figure 1. For a quantum causal structure with an observed input node, X_1—meaning a parentless node from which there is only one arrow to another observed node, X_2—there always exists another (quantum) causal structure that allows for exactly the same correlations and where the observed input is replaced by a shared quantum parent of X_1 and X_2. To simulate any correlations in (a) within scenario (b) we can use a quantum system that sends perfectly correlated classical states to both nodes X_1 and X_2, distributed as X_1. On the other hand, any correlations obtained in scenario (b) can be created in scenario (a) by having a random variable X_1 sent to node X_2, where the relevant quantum states (the reduced states that would be present in (b) conditioned on the value of X_1) are locally generated. Note that these considerations are not restricted to quantum causal structures but apply also in the classical case (or even if considering states from a generalised probabilistic theory).

define a joint distribution over all variables in a causal structure C^{C}, there is in general no joint quantum state over all systems in C^{Q}. In particular, the systems corresponding to the edges that meet at an observed node do not coexist with the outcome at that node and hence there is no joint quantum state from which a joint entropy could be derived. The approach is therefore based on a notion of coexistence: two systems are said to *coexist* if neither is a quantum ancestor of the other in C^{Q}, and a set of systems that pairwise coexist form a *coexisting set*. For each coexisting set, $X_S \subseteq \Omega$, the von Neumann entropy $H(X_S) := -\mathrm{tr}\left(\rho_{X_S} \log_2 \rho_{X_S}\right)$ of their joint state ρ_{X_S} is defined; all of these von Neumann entropies are considered as components of an entropy vector.

For each coexisting set the entropies of all its subsets as well as all conditional mutual informations of its systems are positive [48]. The conditional entropy may not be positive in general, but for three mutually disjoint subsets of a coexisting set, X_T, X_U, $X_V \subseteq X_S$, $H(X_T) + H(X_U) \leq H(X_T \cup X_V) + H(X_U \cup X_V)$ holds instead. These three types of inequality hold for the components of any entropy vector. For the von Neumann entropy of a multi-party quantum state no additional entropy inequalities are known. It has been suggested, however, that any classical 'balanced entropy inequality' [49] (which includes all known non-Shannon inequalities) may also hold for multi-party quantum states [50]. It is worth remarking that the lack of a joint state for all nodes within a quantum causal structure would restrict the applicability of such inequalities in the causal context

if they were to hold.[6]

In many circumstances the conditional entropy of certain sets of systems is known to be positive, e.g. if all systems in a coexisting set are classical. Such constraints on the entropy vectors are also added (cf. [32] for further details). The causal restrictions encoded in the graph are accounted for by the condition that two subsets of a coexisting set are independent (and hence have zero mutual information between them) if they have no shared ancestors.[7] To relate the entropies of systems in different coexisting sets, data processing inequalities (DPIs) are used: Let $\rho_{X_S X_T} \in \mathcal{S}(\mathcal{H}_{X_S} \otimes \mathcal{H}_{X_T})$ and $\mathcal{E}$ be a completely positive trace preserving (CPTP) map on $\mathcal{S}(\mathcal{H}_{X_T})$ leading to a state $\rho'_{X_S X_T}$. Then

$$I(X_S : X_T)_{\rho'_{X_S X_T}} \le I(X_S : X_T)_{\rho_{X_S X_T}}. \tag{6}$$

Results on the redundancy of certain DPI have been presented in [32]. All constraints on the possible entropy vectors taken together define a polyhedral cone, which we denote $\Gamma(C^Q)$. Its projection to the observed variables, $\Gamma_{\mathcal{M}}(C^Q)$, is an outer approximation to $\overline{\Gamma^*_{\mathcal{M}}}(C^Q)$, that can be computed from $\Gamma(C^Q)$ with a Fourier-Motzkin elimination algorithm [46].

III. IMPROVING CURRENT ENTROPIC CHARACTERISATIONS WITH NON-SHANNON INEQUALITIES

In this section we show how non-Shannon inequalities allow us to improve the previous outer approximations to the entropy cones of classical causal structures. We give an improved entropic description of the triangle causal structure of Figure 2(e) (Section III A), discuss the application of non-Shannon inequalities to further causal structures (Section III B) and demonstrate that non-Shannon inequalities are also applicable in combination with post-selection using information causality as an example (Section III C).

The computational procedure that we use in order to derive these new inequalities is roughly outlined in the following. (1) We take the Shannon inequalities for the joint distribution of all variables in a causal structure C^C, (2) we add a set of valid non-Shannon inequalities to these, (3) we add all conditional independence equalities that are implied by C^C, (4) we eliminate all entropies of unobserved variables from the full set of inequalities (by means of a Fourier-Motzkin elimination algorithm [46]),
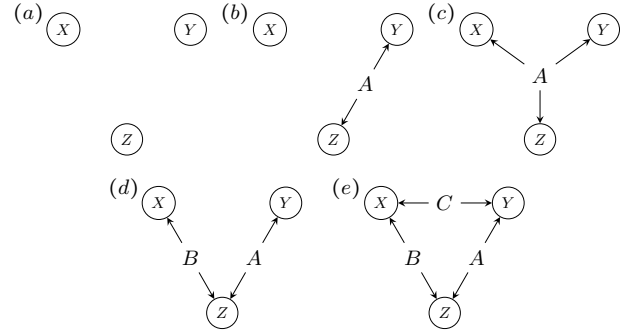
————

[6] In the triangle causal structure, for instance, even if the known non-Shannon inequalities held for arbitrary quantum states, they would not allow us to derive any new entropy inequalities for the quantum version of this causal structure.

[7] Since a node and its (quantum) ancestors never coexist, conditional inependences don't have to be taken into account in two-generation causal structures.

Figure 2. Assuming no ancestral relations between any of the three observed variables X, Y and Z (i.e., no member of $\{X, Y, Z\}$ is an ancestor of any other), the above are the only possible causal structures (up to relabelling). A, B and C correspond to unobserved variables.

which leads to constraints on the entropies of the observed variables only.

Note that the same procedure, but missing out step (2) corresponds to the computation of $\Gamma_{\mathcal{M}}(C^C)$ as in [19, 21, 23] and outlined in Section II B 1. Thus, the inclusion of (2) is responsible for the new constraints. In addition to deriving entropy inequalities computationally, we also provide analytic derivations of (infinite families of) new inequalities.

A. Improved outer approximation to the entropy cone of the classical triangle scenario

The triangle causal structure, called C_3, is one of the simplest examples with interesting features [22, 24, 26, 51, 52]. It may represent a situation where three parties make observations, X, Y and Z respectively, on systems, A, B and C, that are shared between two parties each. This may for instance be realised in a communication protocol where three parties aim to obtain (correlated) data without ever having interacted as a group but after having previously shared systems pairwise.

C_3 is one of only five distinct causal structures involving three observed random variables that exhibit no ancestral relations among the observed variables (cf. Figure 2). All except for the causal structures (c) and (e) may be distinguished by looking at independences among the observed variables, X, Y and Z, listed in Table I. However, while the causal structure of Figure 2(c) does not impose any restrictions on the compatible P_{XYZ}, the distributions that are compatible with the triangle causal structure obey additional constraints [25].[8] This illustrates that causal structure encodes more than the observed independences.

————

[8] For instance, perfectly correlated bits X, Y and Z, i.e., those

Causal Structure	Compatible Distributions	Observed Independence
(a)	$P_{XYZ} = P_X P_Y P_Z$	$I(X:YZ) = 0$ $I(Y:XZ) = 0$ $I(Z:XY) = 0$
(b)	$P_{XYZ} = \sum_A P_X P_{Y\mid A} P_{Z\mid A} P_A$	$I(X:YZ) = 0$
(c)	$P_{XYZ} = \sum_A P_{X\mid A} P_{Y\mid A} P_{Z\mid A} P_A$	None
(d)	$P_{XYZ} = \sum_{A,B} P_{X\mid B} P_{Y\mid A} P_{Z\mid AB} P_A P_B$	$I(X:Y) = 0$
(e)	$P_{XYZ} = \sum_{A,B,C} P_{X\mid AC} P_{Y\mid AC} P_{Z\mid AB} P_A P_B P_C$	None

Table I. Distributions compatible with the three-variable causal structures displayed in Figure 2.
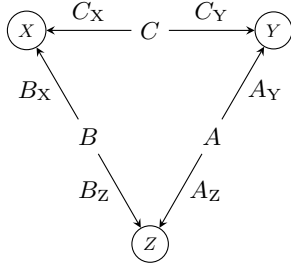

Figure 3. Triangle causal structure C_3. Three observed random variables X, Y and Z have pairwise common causes. In the classical case these common causes are random variables, A, B and C, while in the quantum case these are replaced by quantum systems, (A_Y, A_Z), (B_X, B_Z) and (C_X, C_Y).

Furthermore, C_3 is unique among these five causal structures, it being the only one that features quantum correlations that are not classically reproducible, i.e., $P_\mathcal{M}\left(C_3^C\right) \subsetneq P_\mathcal{M}\left(C_3^Q\right)$, this uniqueness was proven

in Ref. [52] (see Section V A for further details regarding the quantum scenario). [9]

In the following, we derive new and improved outer approximations to $\overline{\Gamma_\mathcal{M}^*}(C_3^C)$ by using non-Shannon entropy inequalities. These show that the Shannon approximation to $\overline{\Gamma_\mathcal{M}^*}(C_3^C)$ is not tight, i.e., that $\overline{\Gamma_\mathcal{M}^*}(C_3^C) \subsetneq \Gamma_\mathcal{M}(C_3^C)$. We remark that our findings contradict the considerations of [22, 26], which together argue that in the marginal scenario there is no separation between the Shannon cone and the classical entropy cone, i.e., they argue that $\Gamma_\mathcal{M}(C_3^C) = \overline{\Gamma_\mathcal{M}^*}(C_3^C)$, which would imply that non-Shannon inequalities are irrelevant.[10] For further discussion of the discrepancy with [22, 26], see Appendix A.

The set of all observed distributions compatible with C_3^C is[11]

$$\mathcal{P}_\mathcal{M}\left(C_3^C\right) = \left\{ P_{XYZ} \in \mathcal{P}_3 \mid \right.$$

$$\left. P_{XYZ} = \sum_{A,B,C} P_A P_B P_C P_{X\mid BC} P_{Y\mid AC} P_{Z\mid AB} \right\}.$$

The compatible entropy vectors are,

$$\Gamma_\mathcal{M}^*\left(C_3^C\right) = \left\{ v \in \mathbb{R}_{\geq 0}^7 \mid \exists P \in \mathcal{P}_\mathcal{M}\left(C_3^C\right) \text{ s.t. } v = \mathbf{H}(P) \right\},$$

and $\overline{\Gamma_\mathcal{M}^*}\left(C_3^C\right)$ is a convex cone (cf. [32]).

The Shannon outer approximation,

$$\Gamma_\mathcal{M}\left(C_3^C\right) = \left\{ w \in \Gamma_3 \mid M_\mathcal{M}\left(C_3^C\right) \cdot w \geq 0 \right\},$$

was explicitly computed by Chaves et al. [22, 26],[12] where $M_\mathcal{M}\left(C_3^C\right)$ is the coefficient matrix of the following three equivalence classes of inequalities (where permutations of X, Y and Z lead to a total of 7 inequalities):[13]

---

with joint distribution

$$P_{XYZ}(x,y,z) = \begin{cases} \frac{1}{2} & x = y = z \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

are not achievable in this causal structure. This is not only true classically, but also in any generalised probabilistic theory [24, 25].

[9] In structures (a), (b) and (c) all joint distributions are allowed for the variables that share a common cause in the classical case. Hence, quantum systems do not enable any stronger correlations. This can also be seen as for any quantum state ρ_A shared at A and measured later the correlations can be classically reproduced if A sends out the same classical output statistics to the parties directly. In structure (d) no non-classical quantum correlations exist either [52]. This is also fairly intuitive: the quantum measurements performed at X and Y could be equivalently performed at the sources B and A respectively, such that these sources distribute cq-states of the form $\sum_x P_X(x) |x\rangle\langle x| \otimes \rho_{B_Z}^x$ and $\sum_y P_Y(y) |y\rangle\langle y| \otimes \rho_{A_Z}^y$ instead. The same correlations can be achieved classically by taking random variables $B = X$ and

$A = Y$ (these being distributed according to P_X and P_Y). Since $\rho_{B_Z}^x$ and $\rho_{A_Z}^y$ are functions of X and Y, the statistics formed by measuring such states can be computed classically via a probabilistic function (this function could be made deterministic by taking $B = (X, W)$, where W is distributed appropriately).

[10] The details of this are in the Supplementary Information of [26].

[11] Note that this set is not convex, which can be seen by considering the perfect correlations P_{XYZ} of (7) (which are not in $\mathcal{P}_\mathcal{M}\left(C_3^C\right)$) as a convex combination of the distribution where X, Y and Z are always 0 and the distribution where X, Y and Z are always 1 (both of which are).

[12] The outer approximation obtained from all six variable Shannon inequalities and the conditional independence equalities (5), which are in this case $I(A : BCX) = 0$, $I(X : AYZ \mid BC) = 0$ and appropriate permutations.

[13] Recall that an explicit linear description of their entropy cone is generally only available for causal structures with up to three nodes. In particular, such a description is not available for $\overline{\Gamma^*}\left(C_3^C\right)$, which involves six nodes. Hence, it is impossible to directly compute $\overline{\Gamma_\mathcal{M}^*}\left(C_3^C\right)$ with a variable elimination algorithm.

$$-H(X) - H(Y) - H(Z) + H(XY) + H(XZ) \geq 0,$$
$$-5H(X) - 5H(Y) - 5H(Z) + 4H(XY) + 4H(XZ) + 4H(YZ) - 2H(XYZ) \geq 0, \qquad (8)$$
$$-3H(X) - 3H(Y) - 3H(Z) + 2H(XY) + 2H(XZ) + 3H(YZ) - H(XYZ) \geq 0.$$

We now show that tighter outer approximations of the set of achievable entropy vectors in the marginal scenario of the triangle, $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^{\mathrm{C}}\right)$, can be derived by using non-Shannon type inequalities. However, there are infinitely many such linear entropy inequalities. To restrict the number of inequalities to be considered, the following reasoning can be applied. As mentioned in Section II A, all known non-Shannon entropy inequalities for four variables can be written as the sum of the *Ingleton quantity* (1) and (conditional) mutual information terms. Since the latter are always positive, any non-Shannon inequality is irrelevant (i.e., implied by existing ones) if the causal restrictions imply that the Ingleton term is non-negative. This significantly reduces the choices of variable sets for which the known additional inequalities may be relevant.

Example 1. Consider Proposition 1 with $(X_1, X_2, X_3, X_4) = (A, B, C, X)$. The corresponding inequality is

$$I(A:B|C) + I(A:B|X) + I(C:X) - I(A:B)$$
$$+ I(A:C|B) + I(B:C|A) + I(A:B|C) \geq 0.$$

Whenever a causal structure C^{C} implies $I(A:B) = 0$, i.e., independence of A and B, the above inequality is implied by the Shannon inequalities and the independence constraint $I(A:B) = 0$. Hence it cannot improve our outer approximation.

The following proposition restricts the permutations of each non-Shannon inequality that may be relevant for the derivation of our improved approximations to $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^{\mathrm{C}}\right)$.

Proposition 3. *Consider an entropy inequality on four variables that enforces the non-negativity of a positive linear combination of the Ingleton quantity (1) and (conditional) mutual information terms. This inequality is implied by the Shannon inequalities and the conditional independences of C_3^{C} (i.e., $I(A:XBC) = 0$, $I(X:YZA|BC) = 0$ and appropriate permutations) for all choices of four out of the six involved random variables, except*

$$(X_1, X_2)(X_3, X_4) = (X, Y)(Z, C)$$
$$= (X, Z)(Y, B)$$
$$= (Y, Z)(X, A),$$

up to exchange of X_1 and X_2 or exchange of X_3 and X_4.

All known irredundant non-Shannon inequalities satisfy the conditions of this proposition. Note also that the application of non-Shannon inequalities to subsets of four out of the six random variables in C_3^{C} does not encompass all possible applications of these inequalities. Specifically, each inequality can also be applied to sets of five or to all six random variables, where the joint distribution of some sets of two or three random variables are interpreted as those of one of the four random variables in the non-Shannon inequality. We have not looked into such configurations.

Proof. For four random variables X_1, X_2, X_3 and X_4, the Ingleton inequality

$$I(X_1:X_2|X_3) + I(X_1:X_2|X_4) + I(X_3:X_4)$$
$$- I(X_1:X_2) \geq 0 \quad (9)$$

can be equivalently rewritten in four more ways with the following equalities:

$$I(X_1:X_2|X_3) - I(X_1:X_2) = I(X_1:X_3|X_2) - I(X_1:X_3)$$
$$= I(X_2:X_3|X_1) - I(X_2:X_3),$$
$$I(X_1:X_2|X_4) - I(X_1:X_2) = I(X_1X_4|X_2) - I(X_1:X_4)$$
$$(10)$$
$$= I(X_2:X_4|X_1) - I(X_2:X_4).$$

For the inequality (9) not to be implied by the Shannon inequalities and the conditional independences we need X_1, X_2, X_3 and X_4 to be such that

$$I(X_1:X_2) > 0,$$
$$I(X_1:X_3) > 0,$$
$$I(X_1:X_4) > 0, \qquad (11)$$
$$I(X_2:X_3) > 0,$$
$$I(X_2:X_4) > 0,$$

hold simultaneously. If the conditional independences of C_3^{C} imply that one of these mutual informations is zero then the Ingleton inequality can be expressed as a positive linear combination of (conditional) mutual information terms in one of its five equivalent forms and the corresponding non-Shannon inequality is redundant.

For the five constraints (11) to hold simultaneously, X_1 and X_2 have to be correlated with one another as well as with two further variables. This excludes the independent sources A, B and C as candidates for X_1 and X_2; therefore X_1, $X_2 \in \{X, Y, Z\}$. Furthermore, the variables X_3 and X_4 have to be correlated with both, X_1 and X_2. This excludes the two variables in $\{A, B, C\}$ that do not lie between X_1 and X_2 in C_3^{C}. Hence, for each choice of X_1 and X_2, the variables X_3 and X_4 have to be chosen as the remaining element of $\{X, Y, Z\}$ and the variable positioned opposite it in C_3^{C}.

In summary, $(X_1, X_2)(X_3, X_4)$ can only be $(X, Y)(Z, C)$, $(X, Z)(Y, B)$ and $(Y, Z)(X, A)$ up to permutations of the variables within a tuple. $\quad\square$

If we were to take one non-Shannon inequality into account and applied it to any subset of four out of the total of six random variables in the causal structure, this would leave us with 360 permutations of the inequality (if the inequality is not invariant under the permutation of any of the four involved variables). Proposition 3 reduces this to only 12 (potentially) irredundant permutations.

For each non-Shannon inequality, these 12 permutations are candidates for improving the outer approximation to $\overline{\Gamma^*_{\mathcal{M}}}\left(C_3^{\mathrm{C}}\right)$. We remark here that for most known non-Shannon inequalities, several of these 12 permutations can be shown to be redundant[14]. Despite accounting for this reduction in the permutations of each inequality, the number of different inequalities to be considered is infinite, and, the outer approximation to $\overline{\Gamma^*_{\mathcal{M}}}\left(C_3^{\mathrm{C}}\right)$ could (potentially) be tightened further by including additional inequalities.

In principle, the more inequalities that are added, the better the approximation to $\overline{\Gamma^*_{\mathcal{M}}}\left(C_3^{\mathrm{C}}\right)$. However, adding too many inequalities at a time renders the task of marginalising infeasible; the main reason is that the Fourier-Motzkin elimination algorithm is often too slow. Applied to a system of n_0 inequalities it can yield up to $\left(\frac{n_0}{2}\right)^2$ inequalities in the first elimination step. Iterating the procedure for n steps produces up to $4 \cdot \left(\frac{n_0}{4}\right)^{2^n}$ inequalities. This doubly exponential growth of the algorithm in the worst case is the main reason for its inefficiency. The elimination algorithm can be adapted by implementing a few rules to remove some of the many redundant inequalities produced in each step. These rules are collectively known as Černikov rules [53, 54] and comprehensively explained in [55]. It is known, however, that the number of necessary inequalities can still grow exponentially [56]. That said, the worst case scaling may not be exhibited in our case. In fact, the inequalities defining $\Gamma\left(C_3^{\mathrm{C}}\right)$ contain few variables each and thus lead to far fewer than the maximal number of inequalities. However, computational resources still limit us to adding a relatively small number of different supplementary inequalities to the standard Shannon cone at a time.

We have used the previously outlined technique to compute tighter outer approximations to $\overline{\Gamma^*_{\mathcal{M}}}\left(C_3^{\mathrm{C}}\right)$, by including a manageable number of non-Shannon inequalities at a time:

Case 1: We include the inequality from Proposition 1 as well as all six inequalities from [41] applied to all subsets of four out of the six variables of C_3^{C}. This leads to 45 classes of inequalities, of which 41 are not part of the outer approximation $\Gamma_{\mathcal{M}}\left(C_3^{\mathrm{C}}\right)$. All classes (together

with the number of members in each class) are provided as Supplementary Information.

Case 2: We include the inequalities of the form given in (2) and (3) for $s = 1, 2, 3$ and for all subsets of four out of the six variables in C_3^{C}. In this case, we find 114 classes of inequalities, of which 110 are not part of the outer approximation $\Gamma_{\mathcal{M}}\left(C_3^{\mathrm{C}}\right)$. All classes (together with the number of members in each class) are provided as Supplementary Information.

We have compared our new approximations to the Shannon outer approximation by sampling uniformly over the surface of the positive sector of the unit hypersphere around $\mathbf{0}$ in $\mathbb{R}^7$ [57][15]. A measure for the hyperdimensional solid angle included by these approximations is given in terms of the fraction of points within the respective cones, α. We have sampled 3.2×10^9 points each, which led to the following estimates for α:

Shannon Cone: $\alpha_S = (3.308 \pm 0.010) \times 10^{-5}$.

Case 1: $\alpha_1 = (3.090 \pm 0.010) \times 10^{-5}$.

Case 2: $\alpha_2 = (3.072 \pm 0.010) \times 10^{-5}$.

This shows that the difference between the three approximations it relatively small: the hyperdimensional solid angle encompassed by the cones of the Case 1 and Case 2 approximations are both roughly 93% of that of the Shannon cone. An explicit entropy vector that lies in the Shannon approximation, but not in either of the new outer approximations to $\overline{\Gamma^*_{\mathcal{M}}}\left(C_3^{\mathrm{C}}\right)$ is

$$(H(X), H(Y), H(Z), H(XY), H(XZ), H(YZ), H(XYZ))$$
$$= (11, 14, 14, 20, 20, 23, 28).$$

We also derive some valid families of inequalities.

Proposition 4. *All entropy vectors* $v \in \Gamma^*_{\mathcal{M}}\left(C_3^{\mathrm{C}}\right)$ *obey*

$$\left(-\frac{1}{2}s^2 - \frac{3}{2}s\right)(H(X) + H(Z)) - (s+1)H(Y)$$
$$+ \left(\frac{1}{2}s^2 + \frac{3}{2}s + 1\right)(H(XY) + H(YZ))$$
$$+ \left(s^2 + 2s\right)H(XZ) - \left(s^2 + 2s + 1\right)H(XYZ) \geq 0, \quad (12)$$

$$\left(-\frac{1}{2}s^2 - \frac{3}{2}s - 2\right)(H(X) + H(Y) + H(Z) - H(XY))$$
$$+ \left(\frac{1}{2}s^2 + \frac{3}{2}s + 1\right)H(XZ) + (s+2)H(YZ)$$
$$- (s+1)H(XYZ) \geq 0, \quad (13)$$

$$\left(-\frac{1}{2}s^2 - \frac{3}{2}s - 2\right)(H(X) + H(Z) - H(XY))$$
$$- (2s+2)H(Y) + \left(s^2 + 2\right)H(XZ)$$
$$+ \left(\frac{1}{2}s^2 + \frac{3}{2}s + 1\right)H(YZ) - \left(s^2 + 1\right)H(XYZ) \geq 0, \quad (14)$$

---

[14] For instance, if the non-Shannon inequality in question is invariant under the permutation of some of its variables then some of the 12 permutations are equivalent, or, if the marginalisation of different permutations of the same inequality (that are not equal) imply the same inequalities for the marginal scenario then some of these inequalities may be redundant for our purposes.

[15] I.e., from the set $\{\mathbf{v} \in \mathbb{R}^7 : v_i \geq 0, \sum_{i=1}^7 v_i^2 = 1\}$.

for all $s \in \mathbb{N}$. The same holds for all permutations of X, Y and Z.

The proof of this proposition can be found in Appendix B.

Further families of inequalities can be derived by separately considering different inequalities from a family, e.g. the same permutation of (2) for each $s \in \mathbb{N}$, and combining them with the same Shannon inequalities to obtain new constraints on the marginal scenario by means of the Fourier-Motzkin elimination algorithm. Tighter inequalities are often obtained by combining several permutations of an inequality (2).

Combining instances of (2) for several $s \in \mathbb{N}$ leads to an even larger number of new inequalities, which render many of the families derived with the previously explained method redundant. For the few orders s up to which we were able to run our calculations, the families (12) and (13) from Proposition 4 were the only two for which none of the inequalities were implied by others. Similar considerations can be applied to (3) (from which (14) is derived) and to further families of inequalities [39].

One might imagine that adding genuine five and six variable inequalities to $\Gamma\left(C_3^{\mathrm{C}}\right)$ leads to further entropy inequalities for C_3^{C}. It turns out that applying the five and six variable inequalities from [35, 40] to five and six variables of the triangle causal structure respectively does not lead to a tighter outer approximation to C_3^{C} than the inequality from Proposition 1. This can be shown by expanding the inequalities into a linear combination of mutual information terms and applying a similar reasoning to that in the proof of Proposition 3. As they are not particularly instructive, the technical details of these arguments are omitted here. The same is not known to hold for the inequality derived in [58].

Conjecture 5. *Infinitely many linear inequalities are needed to characterise $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^{\mathrm{C}}\right)$.*

Our main evidence for this is that the families of inequalities (2), used by Matúš to prove that the analogue of this conjecture holds for $\overline{\Gamma_4^*}$, lead to infinite families of inequalities for C_3^{C} after marginalising (cf. Proposition 4). The curve constructed by Matúš in Ref. [36] to prove his statement for $\overline{\Gamma_4^*}$ can be adapted to our scenario, which can be used to show that the inequalities (12) are independent. However, we were not able to show that this curve can be realised with entropy vectors that are compatible with the triangle causal structure, and hence we cannot exclude the possibility that the marginal cone $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^{\mathrm{C}}\right)$ is polyhedral.

The infinite families of inequalities (cf. Proposition 4) that we obtained from Matúš's original family of inequalities may indicate that this region of entropy space retains a non-polyhedral segment after the causal constraints are included and the set is projected to the marginal scenario. However, we cannot rule out the possibility that no non-polyhedral boundary regions survive the mapping to entropy vectors for C_3^{C}. If this were the case then (most
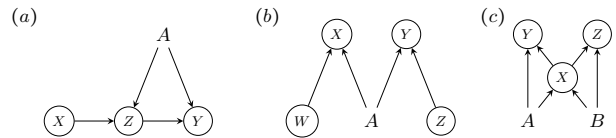

Figure 4. Three causal structures, C, for which the outer approximation, $\Gamma_{\mathcal{M}}(C^{\mathrm{C}})$ tightly approximates the classical entropy cone $\overline{\Gamma_{\mathcal{M}}^*}(C^{\mathrm{C}})$, which also coincides with $\overline{\Gamma_{\mathcal{M}}^*}(C^{\mathrm{Q}})$. The observed variables are labelled W, X, Y and Z, the unobserved nodes are called A and B.

of) our infinite set of inequalities would be rendered redundant by another inequality.

B. Application of non-Shannon inequalities to various causal structures

The concept of a generalised DAG was introduced in [24], the idea being to have a framework in which classical, quantum and even more general systems can be shared by unobserved nodes. For the details, we refer to the original paper. The part that is of interest here is that the authors of Ref. [24] list 21 generalised DAGs with up to six nodes for which there may be a separation between the correlations realisable classically and quantum mechanically, i.e., between $\mathcal{P}_{\mathcal{M}}(C^{\mathrm{C}})$ and $\mathcal{P}_{\mathcal{M}}(C^{\mathrm{Q}})$ [24, 28].[16] We analyse these from an entropic perspective, looking for a causal structure C in which there is a separation between $\overline{\Gamma_{\mathcal{M}}^*}\left(C^{\mathrm{C}}\right)$ and $\overline{\Gamma_{\mathcal{M}}^*}\left(C^{\mathrm{Q}}\right)$. Among these structures there are three that have fewer than six nodes, displayed in Figure 4. For these three, we find that the vertices of the corresponding Shannon cone, $\Gamma_{\mathcal{M}}(C^{\mathrm{C}})$, are achievable with entropy vectors of classical probability distributions compatible with the causal structure, from which it follows that this cone is equal to the entropy cone $\overline{\Gamma_{\mathcal{M}}^*}(C^{\mathrm{C}})$. (This can also be shown by computing an inner approximation to the corresponding entropy cones and showing that the inner and outer approximations coincide, e.g. by employing linear rank inequalities as outlined in Section IV.) Our results also imply that the consideration of non-Shannon inequalities cannot lead to any further constraints in these three causal structures. In the following, we furthermore show that there is no entropic separation between classical and quantum versions of these causal structures.

Proposition 6. *Let C be any of the causal structures shown in Figure 4. Then $\overline{\Gamma_{\mathcal{M}}^*}(C^{\mathrm{C}}) = \overline{\Gamma_{\mathcal{M}}^*}(C^{\mathrm{Q}})$.*

Remark 7. Note that there are causal structures involving up to five variables that reduce to those shown

---

[16] Note that there are further causal structures with 5 and 6 nodes that have this property, which can all be reduced to these 21 examples with rules specified in [24].

in Figure 4 under the reduction rules from [24]. Our proof does not rule out that these exhibit a classical to quantum separation.

Further details, including the proof of Proposition 6 are given in Appendix C.

The 18 remaining example causal structures involve six variables. For all of them we have found that several instances of the non-Shannon inequality from Proposition 1 lead to tighter entropic constraints for the classical marginal scenarios than those listed in [24].

For the causal structures with four observed variables, instances of this inequality are relevant even without considering the unobserved nodes. These instances thus hold whether or not the unobserved nodes are classical or quantum. Hence, they allow us to tighten the outer approximations to the sets of achievable entropy vectors in both cases, in contrast to non-Shannon inequalities that are applied involving unobserved variables (for which the quantum analogue is not known to hold).[17]

The above considerations have not enabled us to show a separation between the achievable entropy vectors in the classical and quantum cases, hence we are left with the following open problem.

Open Problem 8. Find a causal structure C with a set of observed nodes $\mathcal{M}$ in which the sets $\overline{\Gamma_{\mathcal{M}}^*}(C^{\mathrm{C}})$ and $\overline{\Gamma_{\mathcal{M}}^*}(C^{\mathrm{Q}})$ are provably different, or show that this can never occur.

C. Application of non-Shannon inequalities with post-selection

In the discussion so far we have not considered a related technique that allows for post-selection on particular outcomes of certain variables. The idea of doing this first appeared in [13] based on results by Fine [59, 60] and was later generalised [19–21, 26–28]. We refer to [32] for an explanation of this technique.

Here we illustrate that non-Shannon inequalities can be used in combination with post-selection by discussing a specific example relevant for information causality [31]. Information causality is an information theoretic principle obeyed by classical and quantum physics but not by general probabilistic theories in which there are correlations that violate Tsirelson's bound [61], e.g. generalized no signalling theory [62], which allows PR-Boxes as a resource [63, 64]. The principle is stated in terms of the optimal performance of two parties in a game, which we
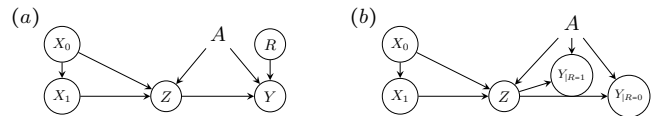

Figure 5. (a) Causal structure underlying the Information Causality game, IC. Alice holds a database, here made up of two bits X_0 and X_1. These need not be independent, which is expressed by a potential causal influence from X_0 to X_1. She is then allowed to send a message Z to Bob, who, depending on which bit R a referee asks for, takes a guess Y of either X_0 or X_1. Alice and Bob may have shared some resources (represented by A) before performing the protocol, either some classical randomness, a quantum system, or a resource from a more general non-signalling theory, which Alice may use in order to choose her message and Bob may use to make his guess. (b) The effective causal structure of the Information Causality game after post-selecting on binary R, labelled $\mathrm{IC_R}$. This causal structure shares some of its marginal distributions with conditional distributions of IC, i.e., if we use P for the distribution in $\mathrm{IC_R}$ and Q for that in IC then $P_{X_0 X_1 Z Y_{|R=r}} = Q_{X_0 X_1 Z Y | R=r}$ for $r = 0, 1$.

describe below, and is quantified in terms of an entropic quantity.

Alice holds two pieces of information[18], X_0 and X_1, she can send classical information Z to Bob, who is later given a message R indicating whether he should guess X_0 or X_1. Bob's guess is denoted Y. Alice and Bob are able to use a pre-shared resource (depicted as A) to help them. The relevant causal structure of the game is displayed in Figure 5(a) and it is often analysed after post-selecting on the value of R, which can be analysed using the causal structure of Figure 5(b) (note that in the quantum case the variables $Y_{|R=0}$ and $Y_{|R=1}$ do not coexist, so it doesn't make sense to consider $\mathrm{IC_R^Q}$; instead a restricted set of entropies needs to be considered – see later). A theory is said to obey information causality if for all pre-shared resources allowed by the theory, $I(X_0 : Y_{|R=0}) + I(X_1 : Y_{|R=1}) \le H(Z) + I(X_0 : X_1)$.

A stronger set of entropic constraints for this causal structure were found in [26], including the relation

$$I(X_0 : ZY_{|R=0}) + I(X_1 : ZY_{|R=1}) + I(X_0 : X_1 \mid ZY_{|R=1})$$
$$\le H(Z) + I(X_0 : X_1), \qquad (15)$$

which holds for both classical and quantum shared resources.[19]

---

[17] Note that this reasoning is not restricted to distinguishing classical and quantum, but, it may also apply to the comparison of different causal structures with the same set of observed variables. While non-Shannon inequalities derived from latent variables may lead to a separation between the two causal structures, non-Shannon inequalities valid only for the observed variables may not.

[18] In general the game is formulated for more, but we restrict to two here for simplicity.

[19] Because the existence of a joint distribution of $Y_{|R=0}$ and $Y_{|R=1}$ with appropriate marginals is not clear in the quantum case, the two variables have to be interpreted as alternatives and are part of different coexisting sets. Therefore, the analysis of $\mathrm{IC_R^C}$ does not carry over to the quantum case, but a separate analysis is required there.

We show that using non-Shannon inequalities leads to a tighter outer approximation of the information causality scenario in the case of a classical shared resource. Considering just the inequality from Proposition 1 (and permutations) has led us to derive a total of 265 classes of entropy inequalities, including the 52 classes that were obtained without non-Shannon constraints in [26] (a list of all 265 classes together with the number of representatives of each class is available as Supplementary Information). Moreover, we expect further non-Shannon inequalities to lead to numerous additional constraints potentially rendering our inequalities redundant. In principle, infinite families of inequalities, similar to those found in Proposition 4 for the triangle scenario could also be derived here.

In the quantum case, we can only apply the non-Shannon inequalities to the two coexisting sets of exclusively classical variables $\{X_0, X_1, Z, Y_{|R=0}\}$ and $\{X_0, X_1, Z, Y_{|R=1}\}$, which means that we can impose a set of 24 additional constraints (including permutations) just by adding all permutations of the inequality from Proposition 1 to the outer approximation that is obtained without these (no further variable elimination is required).

It is worth pointing out that although our results imply that previous entropic characterisations of IC_R were not tight and even though our new inequalities improve on the entropic characterisation of IC_R, the inequality (15) is not rendered redundant by our new inequalities.

IV. INNER APPROXIMATIONS TO THE ENTROPY CONES OF CAUSAL STRUCTURES

To complement the outer approximations, it is sometimes useful to consider inner approximations to the entropy cones of causal structures. This is particularly useful when one can show that inner and outer approximations coincide, as they then identify the actual boundary of the entropy cone. Examples for this are the three causal structures of Figure 4, also discussed in the previous section. Hence, inner and outer approximation together serve as a relatively simple means to identify the boundary of certain entropy cones. Such findings also immediately imply that non-Shannon inequalities are irrelevant for improving on the outer approximation to the entropy cone for the causal structure in question.

Furthermore, we can often find inner approximations that share extremal rays with the outer approximations derived from the Shannon and independence constraints (even when the two do not coincide). They hence allow us to identify the regions of entropy space where our approximations are tight and those regions where there is a gap between inner and outer approximation.[20] Such a

gap can be explored, e.g. by using non-Shannon inequalities, as was explained in the previous section.

Inner approximations also serve as a tool to decide whether entropy vectors are suitable for certifying the unattainability of particular distributions that are suspected not to be achievable within the causal structure at hand. If such a distribution leads to an entropy vector within an inner approximation to the entropy cone in question, this means either that the distribution is in fact achievable within the causal structure or that the causal structure allows for another distribution with the same entropy vector (or an arbitrarily good approximation of such). Hence, to determine whether the distribution in question is achievable, switching to a more fine-grained method is necessary [28, 65].

In the following we show how inner approximations can be found in different scenarios.

A. Techniques to find inner approximations for causal structures with up to five observed variables

For a causal structure, C, that involves a total of four or five variables, inner approximations to its entropy cone can be derived from Γ_4^I or Γ_5^I respectively (as defined in Section II A) combined with the conditional independence constraints of C^C, which together constrain a cone $\Gamma^I(C^C)$. An inner approximation to the corresponding marginal scenarios, $\Gamma_{\mathcal{M}}^I(C^C)$, is then obtained from $\Gamma^I(C^C)$ with a Fourier-Motzkin elimination, like for outer approximations. It is guaranteed that $\Gamma_{\mathcal{M}}^I(C^C)$ is an inner approximation to $\overline{\Gamma_{\mathcal{M}}^*}(C^C)$, as it is a projection of an inner approximation $\Gamma^I(C^C) \subseteq \overline{\Gamma^*}(C^C)$. Hence, inner approximations can be straightforwardly computed for such causal structures. Examples where this applies are the three causal structures of Figure 4.

Example 2 (Inner approximation to the instrumental scenario.)**.** For the classical instrumental scenario, C_I of Figure 4(a), we can compute an inner approximation by adding the conditional independence constraints $I(A : X) = 0$ and $I(X : Y | AZ) = 0$ to the Ingleton cone Γ_4^I, as prescribed above. We can, however, also directly prove that $\overline{\Gamma_{\mathcal{M}}^*}(C_I^C) = \Gamma_{\mathcal{M}}(C_I^C)$ by showing that all permutations of the Ingleton inequality are implied by Shannon and conditional independence constraints and, hence, inner and outer approximations coincide for C_I^C. Since $I(A : X) = 0$, $I_{ING}(A, X; Y, Z) \geq 0$ is immediately implied by Shannon and independence constraints. Furthermore, the rewritings of I_{ING} according to (10) imply that $I(A : X) = 0$ (together with the Shannon inequalities) implies all permutations of the Ingleton inequality

---

[20] Such a comparison of inner and outer approximations can be performed for the entropy cone of a causal structure including

its unobserved variables, i.e., before marginalisation, as well as for the respective approximations to its marginal cone, which we are mainly interested in here.

except for $I_{\text{ING}}(Y, Z; A, X) \geq 0$. Now, we can rewrite

$$
\begin{aligned}
& I_{\text{ING}}(Y, Z; A, X) \\
& = I(Y:Z|A) + I(Y:X|Z) + I(X:A|Y) - I(X:Y|A) \\
& = I(Y:X|Z) + I(X:A|Y) + I(Y:Z|AX) - I(X:Y|AZ),
\end{aligned}
$$

the positivity of which is implied by the Shannon inequalities and the independence constraint $I(X:Y|AZ) = 0$.

Implementing all relevant linear rank inequalities of four and five variables (which includes their permutations and the application of the Ingleton inequality to each four variable subset as well as grouping several variables to one) [44] and then performing a variable elimination may be impractical and computationally challenging for certain causal structures. Furthermore, for causal structures that involve more than five nodes not all possible linear rank inequalities are known and their number may even be infinite [45]. It is therefore useful to derive inner approximations by other methods. For a causal structure, C, one may consider different methods to derive inner approximations, $\Gamma_{\mathcal{M}}^{I}(C^{\text{C}})$, for instance the following.

- Construct (random) entropy vectors from distributions compatible with C^{C} and take their convex hull.

- Take the vertices of $\Gamma_{\mathcal{M}}(C^{\text{C}})$ that are reproducible with distributions compatible with the causal structure, their convex hull is an inner approximation.

- Take the outer approximation to the classical causal structure, $\Gamma(C^{\text{C}})$, as a starting point and add a manageable number of linear rank inequalities to derive further constraints. These inequalities may be employed either before or after marginalising, which leads to different cones.[21] The convex hull of the reproducible rays is an inner approximation.

For the three examples of Figure 4 it is rather straightforward to recover all extremal rays of the outer approximation to the marginal scenario, $\Gamma_{\mathcal{M}}(C^{\text{C}})$ (cf. also Appendix C), i.e., the second method above is effective.

Overall, we found that whenever the extremal rays are not all straightforwardly recovered, the third method is effective. This is our preferred technique because by starting out with extremal rays of the Shannon cone we obtain approximations that in some regions are already tight (as opposed to the first method), and, at the same time adding linear rank inequalities helps us identify those extremal rays that are likely to be reproducible

---

[21] If for instance all linear rank inequalities in up to k observed variables are added after marginalisation, the resulting cone corresponds to the intersection $\Gamma_{\mathcal{M}}(C^{\text{C}}) \cap \Gamma_k^{I}$, where k is the number of observed variables.

with distributions in C^{C} (this may help us avoid dropping reproducible rays in some situations). The entropy cones obtained in this way are not necessarily inner approximations, and, if they are, they have to be proven as such, for example by explicitly constructing distributions that reproduce entropy vectors on each of the extremal rays (as with the second method above). However, in all our examples this method allowed us to recover a cone of which all extremal rays were easily seen to be reproducible after adding only few linear rank inequalities to $\Gamma(C^{\text{C}})$. (If this is not the case one may still have to drop several irreproducible rays from the resulting cones to obtain an inner approximation.) The method is illustrated in the example below.

We also remark here that in order to improve on inner approximations obtained with the second or third method above, the first method is applicable.

Example 3. Consider the classical causal structure of Figure 5(a) and remove the node R to give a 5-variable causal structure, $\hat{\text{IC}}^{\text{C}}$. We can in principle consider all linear rank inequalities of five random variables combined with all Shannon inequalities and the conditional independence constraints, which would give us an inner approximation, $\Gamma_{\mathcal{M}}^{I}(\hat{\text{IC}}^{\text{C}})$, to the entropy cone, $\overline{\Gamma_{\mathcal{M}}^{*}}(\hat{\text{IC}}^{\text{C}})$. This procedure would involve a (impractically) large number of inequalities.

Instead, we can consider the outer approximation in terms of Shannon inequalities and conditional independence constraints, $\Gamma_{\mathcal{M}}(\hat{\text{IC}}^{\text{C}})$, and intersect this cone with the Ingleton cone for the four observed variables, Γ_4^{I}, i.e., we add all permutations of the Ingleton inequality for the four observed variables to $\Gamma_{\mathcal{M}}(\hat{\text{IC}}^{\text{C}})$. This is easily obtained but does not result in any restrictions beyond those of the Shannon outer approximation, which is characterised by 52 extremal rays.

Adding the Ingleton inequality for all subsets of four out of the five random variables to $\Gamma(\hat{\text{IC}}^{\text{C}})$ before performing the variable elimination, only 46 extremal rays are recovered. These are straightforward to reproduce with entropy vectors in $\hat{\text{IC}}^{\text{C}}$.[22] A detailed exposition of this is presented in Appendix D.

This method can also be applied to causal structures with more than five variables. For the first few causal structures from [24] we have recovered inner approximations by adding the Ingleton inequality to $\Gamma_{\mathcal{M}}(C^{\text{C}})$, i.e.,

---

[22] We can show that the 6 extremal rays of the Shannon cone that are not part of this inner approximation are not achievable in $\hat{\text{IC}}^{\text{C}}$, as they violate the entropy inequalities we obtain when taking non-Shannon inequalities into account in the computation of the outer approximations to $\overline{\Gamma_{\mathcal{M}}^{*}}(\hat{\text{IC}}^{\text{C}})$.

by taking the intersection $\Gamma_{\mathcal{M}}(C^{\mathrm{C}}) \cap \Gamma_k^I$ (the extremal rays as well as distributions recovering entropy vectors on each of them are available as Supplementary Information).

In the following we give a detailed analysis of the inner approximation to the triangle causal structure and compare this to the outer approximations presented in previous sections.

B.   Example: Inner approximation to $\overline{\Gamma_{\mathcal{M}}^*}(C_3^{\mathrm{C}})$

Here, we derive an inner approximation to the entropy cone compatible with C_3^{C}. An inner approximation to $\overline{\Gamma_6^*}$ in terms of linear rank inequalities is not available (see also Section II A). Nonetheless, we are able to derive an inner approximation to $\overline{\Gamma_{\mathcal{M}}^*}(C_3^{\mathrm{C}})$ by relying on Ingleton's inequality. In the following, we apply (4) to any subset of four out of the six random variables of C_3^{C} and take all their permutations into account. We concisely write these inequalities in a matrix M_I and consider the cone

$$\Gamma^I(C_3^{\mathrm{C}}) := \left\{ v \in \Gamma_6 \mid M_{CI}(C_3^{\mathrm{C}}) \cdot v = 0, \ M_I \cdot v \geq 0 \right\}.$$

When marginalising this cone we obtain

$$\Gamma_{\mathcal{M}}^I(C_3^{\mathrm{C}}) := \left\{ w \in \Gamma_3 \mid M_{I,\mathcal{M}}(C_3^{\mathrm{C}}) \cdot w \geq 0 \right\},$$

where $M_{I,\mathcal{M}}(C_3^{\mathrm{C}})$ contains only one inequality,[23]

$$-I(X:Y:Z) \geq 0. \tag{16}$$

This relation can also be analytically derived from the Ingleton inequality and the conditional independence constraints of C_3^{C}.[24]

Proposition 9. $\Gamma_{\mathcal{M}}^I(C_3^{\mathrm{C}})$ *is an inner approximation to the marginal entropy cone of the triangle causal structure,*

$$\Gamma_{\mathcal{M}}^I(C_3^{\mathrm{C}}) \subsetneq \overline{\Gamma_{\mathcal{M}}^*}(C_3^{\mathrm{C}}).$$

The proof of Proposition 9 is deferred to Appendix E. $\Gamma_{\mathcal{M}}^I(C_3^{\mathrm{C}})$ provides a certificate for vectors to be realisable as entropy vectors in C_3^{C}: if a vector $v \in \mathbb{R}^7$ obeys all Shannon constraints as well as (16), then it lies in $\overline{\Gamma_{\mathcal{M}}^*}(C_3^{\mathrm{C}})$. Compared to the different outer approximations to $\overline{\Gamma_{\mathcal{M}}^*}(C_3^{\mathrm{C}})$ analysed in Section III the hyperdimensional solid angle for this inner approximation is considerably smaller. Sampling over the unit hypersphere

around $\mathbf{0}$ in $\mathbb{R}^7$ as before (meaning 3.2×10^9 samples), we obtain $\alpha_I = (2.147 \pm 0.008) \times 10^{-5}$.

It is worth emphasising that not all correlations whose entropy vectors lie in $\Gamma_{\mathcal{M}}^I(C_3^{\mathrm{C}})$ can be realised in C_3^{C}. Instead, if $\mathbf{H}(P) \in \Gamma_{\mathcal{M}}^I$ then there exists $P' \in \mathcal{P}_{\mathcal{M}}(C_3^{\mathrm{C}})$ such that $\mathbf{H}(P') = \mathbf{H}(P)$. The correlations of Figure 6, realised in the quantum version of the triangle causal structure, C_3^{Q}, which will be considered in detail in Section V A, are one such example. These are not in $\mathcal{P}_{\mathcal{M}}(C_3^{\mathrm{C}})$, but their entropy vector nevertheless satisfies (16). Our argument implies that there must be another distribution realisable in C_3^{C} with the same entropy vector.[25]

V.   NON-SHANNON INEQUALITIES IN QUANTUM AND HYBRID CAUSAL STRUCTURES

In this section, we compare classical and quantum versions of a given causal structure (the distinction reflecting the nature of the unobserved nodes). We also consider hybrid scenarios—causal structures where some of the unobserved systems are restricted to be classical whereas others may be quantum—which turn out to be insightful for understanding the gap between classical and quantum causal structures. We analyse whether non-Shannon inequalities lead to improved entropic characterisations for such scenarios. Our considerations are illustrated with the example of the triangle causal structure.

A.   Quantum triangle scenario

It was first shown in Ref. [52], that there are joint distributions among the three observed variables X, Y and Z in C_3^{Q} that cannot be reproduced in C_3^{C}, based on the CHSH scenario (see Figure 6 and Appendix F for the details). Hence C_3^{Q} might also lead to a larger set of compatible entropy vectors than C_3^{C}.

Entropically, C_3^{Q} can be analysed with the technique outlined in Section II B 2. An outer approximation, $\Gamma_{\mathcal{M}}(C_3^{\mathrm{Q}})$, to the set of achievable entropy vectors, $\overline{\Gamma_{\mathcal{M}}^*}(C_3^{\mathrm{Q}})$, was constructed in [26]. It led to the Shannon inequalities for the jointly distributed X, Y and Z and the additional inequality

$$I(X:Y) + I(X:Z) \leq H(X), \tag{17}$$

as well as its permutations in X, Y and Z [26].

---

[23] Inequality (16) renders the three Shannon inequalities of the form $I(X:Y|Z) \geq 0$ redundant. $\Gamma_{\mathcal{M}}^I(C_3^{\mathrm{C}})$ is thus fully characterised by the six remaining three variable Shannon inequalities (constraining Γ_3) and (16).

[24] The proof proceeds as follows. There are only three instances of the Ingleton inequality that are not implied by the conditional independences and the Shannon inequalities (cf. also Proposition 3). The independence constraint $I(X:Y|C) = 0$ and its permutations $I(X:Z|B) = 0$ and $I(Y:Z|A) = 0$ lead to (16) in all three cases.

[25] Another such example is given by the distribution

$$P_{XYZ}(x,y,z) = \begin{cases} \frac{1}{3} & (x,y,z) = (1,0,0), (0,1,0), (0,0,1) \\ 0 & \text{otherwise}, \end{cases}$$

which is not compatible with C_3^{C}, as shown in [65].

It is natural to ask whether tighter approximations to $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^Q\right)$ can be realised by a similar procedure to the one that led to tighter approximations in the classical case. Unfortunately, we don't know of any similar inequalities for the von Neumann entropy of multi-party quantum states. Furthermore, even if the known non-Shannon inequalities were to hold for von Neumann entropy we would not be able to use them to add constraints to C_3^Q due to the lack of large enough sets of coexisting, interdependent variables.[26]

Open Problem 10. Do the closures of the sets of compatible entropy vectors coincide in the classical and the quantum triangle scenario, i.e., does $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^C\right) = \overline{\Gamma_{\mathcal{M}}^*}\left(C_3^Q\right)$ hold?

Note that if this were to be answered in the affirmative, it would point towards deficiencies of the current entropic techniques for approximating $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^Q\right)$, which are not able to recover any additional inequalities similar to the non-Shannon inequalities found in the classical case.

One way to solve this problem would be to find an entropy vector compatible with C_3^Q that lies outside one of our outer approximations to $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^C\right)$. Random searches where the sources A, B and C distribute up to four qubits each did not yield violations. However, the evidence from these random searches against a separation of the classical and the quantum sets is relatively weak. For one, our classical outer approximations might be so loose that they contain $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^Q\right)$. To counter this, we have attempted to randomly search for vectors that lie in $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^Q\right)$ but not in the classical inner approximation $\Gamma_{\mathcal{M}}^I\left(C_3^C\right)$. In spite of the fact that we know such vectors exist, we were unable to randomly find any. This shows the weakness of random searching, and also that the region we are looking for (if it even exists) is small with respect to our sampling measure.[27]

A natural candidate for an entropy vector that might violate some of our classical inequalities is the one corresponding to the CHSH correlations that were shown not

to be reproducible in C_3^C in Ref. [52] (detailed in Figure 6 where $Z = (A', B')$ and in Appendix F). However, the corresponding entropy vector lies inside $\Gamma_{\mathcal{M}}^I\left(C_3^C\right)$ so is classically reproducible. This particular distribution is also achievable in the causal structure P_4 (a causal structure equivalent to the one in Figure 4(b)). Any distribution compatible with P_4 may be realised in C_3 by choosing one of the variables, e.g. Z, to have two outputs, one depending only on the input from node A and the other one depending on the input from B. Distributions realisable in P_4^Q or P_4^C are thus always realisable in C_3^Q or C_3^C respectively. According to the results of [66], all entropy vectors realised with distributions in P_4^Q are also classically achievable, i.e., realisable in P_4^C (at least asymptotically). Hence, no distribution in P_4^Q can violate any of the classical entropy inequalities valid for C_3^C. A way that might still allow us to use our knowledge about quantum correlations that are not classically reproducible in the Bell scenario to violate our entropic constraints to $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^C\right)$, is by processing the inputs to all three nodes X, Y and Z, so as to get around the results from [66].[28]

In the following, we generalise the distribution that was utilised in Ref. [52] to show that there is a separation between the achievable distributions in C_3^C and C_3^Q to a scenario where there is local processing at each output node. This also allows us reduce the required dimension of the output at Z for which one can provably detect a difference between classical and quantum distributions from two bits to one bit.

Proposition 11. *There are non-classical quantum correlations in C_3 in the case where X and Y output two bits each while Z outputs only one.*

A proof of Proposition 11 can be found in Appendix F. It is interesting in so far as the example in [52] relies on a Bell inequality violation. Given this, one might have expected that all information about the measurement choices in the Bell setup, $\tilde{A}$ and $\tilde{B}$, has to be exposed at the observed node Z. Proposition 11 shows that this is not the case.

Nonetheless, we find that the entropy vector used to prove this proposition does not violate our classical inequalities. We have also taken Z to be determined by different functions in A and B and have additionally considered local processing of X and Y. However, even after such post-processing, for instance by applying all possible functions from two bits to one, we have not been able to detect any violations of the classical entropic bounds.

Note that vectors outside $\Gamma_{\mathcal{M}}^I\left(C_3^C\right)$ can be constructed with appropriate post-processing of the (quantum) distribution. A possible way to achieve this is applying AND or OR functions appropriately. One may for instance

---

[26] Note that in causal structures that involve four or more observed variables non-Shannon inequalities can be applied to these. However, non-Shannon inequalities cannot be applied to unobserved quantum systems (see also Section III C).

[27] This is not a statement about the geometric extent of this region (for instance in terms of a hyperdimensional solid angle as was previously considered for inner and outer approximations to $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^C\right)$). Instead, since we are sampling quantum states here, and since these are not in a one-to-one correspondence with the entropy vectors, it is a statement about the fraction of states and measurements that may produce entropy vectors outside $\Gamma_{\mathcal{M}}^I\left(C_3^C\right)$ (in low dimensions) according to our sampling distribution. This must be a very small proportion of states and measurements (we didn't sample any). Note also that if there is a gap between $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^C\right)$ and $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^Q\right)$ it is smaller than that between $\Gamma_{\mathcal{M}}^I\left(C_3^C\right)$ and $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^Q\right)$. Hence, constructing a vector in the first gap by sampling quantum states is even more difficult than for the second.

[28] Two distributions that share the same entropy vector can be very different and hence may be separated by local processing.
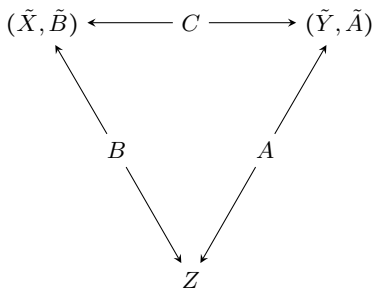
Figure 6. Scenario involving unobserved quantum systems, leading to a distribution which is not reproducible with classical A, B and C [52]. The observed variables $X = (\tilde{X}, \tilde{B})$ and $Y = (\tilde{Y}, \tilde{A})$ are chosen such that $P_{\tilde{X}\tilde{Y}|AB}$ maximally violates the CHSH inequality [12]. $Z = (A', B')$ is such that $B' = \tilde{B} = B$ and $A' = \tilde{A} = A$. In essence the reason that this cannot be realised in the causal structure C_3^{C} is the CHSH violation. Note though that it is also important that information about A is present in both Y and Z (and analogously for B), otherwise the correlations could be mocked up. In Proposition 11, we prove that a strategy where $Z = \mathrm{AND}(A', B')$ also leads to correlations that cannot be classically realised (see Appendix F for further details).

consider the quantum scenario detailed above, and take $X = \mathrm{AND}(\tilde{X}, \tilde{B})$, $Y = \mathrm{AND}(\tilde{Y}, \tilde{A})$ and $Z = \mathrm{OR}(A', B')$. This renders the interaction information of the entropy vector of the joint distribution of X, Y and Z positive, so the vector is not in $\Gamma_{\mathcal{M}}^{I}(C_3^{\mathrm{C}})$.

We have similarly tried to violate our entropy inequalities by relying on games other than the CHSH scenario, for which we know that there is distinctive quantum behaviour (i.e., a separation at the level of correlations); these include input states and measurements known to lead to violations of the chained Bell inequalities [13] or the Mermin-Peres magic square game [67, 68], all with post-processing at (X, Y and) Z.

We have further considered scenarios where all three parties measure entangled states and use the measurement outputs as inputs for further measurements. We have also attempted to incorporate functions known to lead to a positive interaction information in the classical case, as well as functions from two to one bits in general, into these scenarios. None of these attempts has led to a violation of the classical inequalities so far. In a number of scenarios we have also considered shared PR-boxes instead of entangled states, again without detecting any violations of the inequalities. In most cases the corresponding entropy vectors have a negative interaction information, and hence lie in $\Gamma_{\mathcal{M}}^{I}(C_3^{\mathrm{C}})$, so can be realised with a classical distribution as well, like in the case of the correlations mentioned at the end of Section IV B.

B. Hybrid triangle scenarios

In a *hybrid causal structure* some of the unobserved nodes are allowed to be quantum, whereas others are restricted to be classical. One motivation for this is that sharing entanglement over large distances is challenging due to noise, so two distant observations might be assumed to have a classical cause while nearby ones could have quantum causes. In the case of the causal structure C_3, there are two such hybrid scenarios: either one or two of the three unobserved variables can be chosen to be classical, whereas the others are quantum. We call these two causal structures C_3^{CQQ} and C_3^{CCQ} respectively. In the following, we will approximate the sets of compatible entropy vectors for both scenarios. We show that in hybrid scenarios of the triangle causal structure non-Shannon inequalities are relevant.

1. C_3^{CQQ} scenario

In this scenario one of the unobserved variables is classical (we take this to be A). The techniques introduced in Sections II B 1 and II B 2 allow us to compute approximations of the set of allowed entropy vectors. We find

$$\Gamma_{\mathcal{M}}(C_3^{\mathrm{CQQ}}) = \Gamma_{\mathcal{M}}(C_3^{\mathrm{Q}}),$$

i.e., the outer approximation to $\overline{\Gamma_{\mathcal{M}}^{*}}(C_3^{\mathrm{CQQ}})$ obtained without taking non-Shannon inequalities into account coincides with the outer approximation to $\overline{\Gamma_{\mathcal{M}}^{*}}(C_3^{\mathrm{Q}})$. However, unlike in the fully quantum case C_3^{Q}, non-Shannon constraints can be included for C_3^{CQQ}, for instance the inequality from Proposition 1 with variable choices

$$\diamondsuit_{YZAX} \geq 0, \quad \diamondsuit_{YZXA} \geq 0.$$

This results in a tighter approximation to $\overline{\Gamma_{\mathcal{M}}^{*}}(C_3^{\mathrm{CQQ}})$, which comprises the Shannon inequalities for three variables, the constraint (17) and[29]

$$-3H(X) - 3H(Y) - 3H(Z) + 2H(XY) \\ + 2H(XZ) + 3H(YZ) - H(XYZ) \geq 0,$$
$$-2H(X) - 2H(Y) - 2H(Z) + 3H(XY) \\ + 3H(XZ) + 3H(YZ) - 4H(XYZ) \geq 0.$$

Further non-Shannon constraints could also be exploited to improve these approximations. Hence, some of the extremal rays of the Shannon outer approximation $\Gamma_{\mathcal{M}}(C_3^{\mathrm{Q}})$ are provably not achievable if A, B and C do

---

[29] The second of these inequalities can be easily derived from $\diamondsuit_{YZXA} \geq 0$ and the conditional independences, analogously to Proposition 4. To derive the first inequality, on the other hand, several inequalities have to be combined.

not all share entangled states. Note that this does not imply that the sets of achievable entropy vectors in C_3^{Q} and C_3^{CQQ} differ. However, the difference in their outer approximations may prove useful for analysing whether there is a difference between the two.

If one were to prove by other means that there is no such difference the inequalities for C_3^{CQQ} would give us a way to better approximate the set of achievable entropy vectors of C_3^{Q}.

2. C_3^{CCQ} scenario

In this scenario we take A and B to be classical. This scenario can be understood as a Bell scenario, where the measurement choices of the two parties are unobserved and processed to one single observed output, Z.[30] The distributions from Section V A, that are provably not reproducible in C_3^{C} can be generated in this causal structure. Its entropic analysis thus restricts the violations of our classical inequalities we may hope to achieve with such distributions. To approximate the set of compatible entropy vectors of this scenario, $\overline{\Gamma_{\mathcal{M}}^*}\left(C_3^{\mathrm{CCQ}}\right)$, we proceed analogously to the C_3^{Q} and C_3^{CQQ} scenarios before. How-

ever, the result differs and leads to a tighter cone, even without considering non-Shannon inequalities,

$$\Gamma_{\mathcal{M}}\left(C_3^{\mathrm{CCQ}}\right) \subsetneq \Gamma_{\mathcal{M}}\left(C_3^{\mathrm{CQQ}}\right).$$

It is given by the three variable Shannon inequalities and the following additional inequalities:

$$\begin{aligned}
-H(X)-H(Y)-H(Z)+H(XY)+H(XZ) &\geq 0, \\
-3H(X)-3H(Y)-3H(Z)+2H(XY) & \\
+3H(XZ)+2H(YZ)-H(XYZ) &\geq 0,
\end{aligned} \tag{18}$$

up to permutations of X, Y and Z in the first inequality and of Y and Z in the second. Note that these five inequalities are a subset of the seven inequalities (8) delimiting $\Gamma_{\mathcal{M}}\left(C_3^{\mathrm{C}}\right)$ and that

$$\Gamma_{\mathcal{M}}\left(C_3^{\mathrm{C}}\right) \subsetneq \Gamma_{\mathcal{M}}\left(C_3^{\mathrm{CCQ}}\right).$$

The inequalities

$$\diamond_{XZBY} \geq 0, \quad \diamond_{YZAX} \geq 0, \quad \diamond_{YZXA} \geq 0,$$

lead to the additional inequalities

$$\begin{aligned}
-2H(X)-2H(Y)-2H(Z)+3H(XY)+3H(XZ)+3H(YZ)-4H(XYZ) &\geq 0, \\
-6H(X)-6H(Y)-6H(Z)+5H(XY)+5H(XZ)+5H(YZ)-3H(XYZ) &\geq 0, \\
-4H(X)-4H(Y)-4H(Z)+3H(XY)+4H(XZ)+3H(YZ)-2H(XYZ) &\geq 0,
\end{aligned} \tag{19}$$

(including permutations of X and Y in the last inequality). They render the second inequality (and its permutations) in (18) redundant, while the first remains (for all of its permutations). Note that the first inequality of (19) is also present in $\Gamma_{\mathcal{M}}\left(C_3^{\mathrm{CQQ}}\right)$. As in the previous example, further constraints could likely be derived by considering additional non-Shannon inequalities.

VI. CONCLUSIONS

We have shown that non-Shannon inequalities tighten the entropic approximations of the classical entropy cones in many causal structures including the triangle scenario and the causal structure relevant for information causality. Our newly derived inequalities improve on the entropic distinction of these from other (classical) causal

structures, which is of interest for inferring (classical) causal relations. They also constitute a set of restrictions on the classical entropy cones that we cannot derive in the quantum case, which may point towards differences between the sets of achievable entropy vectors in classical and quantum case.

Since it is known from the Bell scenario that quantum correlations can be detected when considering the entropies of the variables in a post-selected causal structure [13], our analysis of the information causality scenario is the one that is most likely to be useful for this purpose. In this context, non-Shannon inequalities may also be important with regard to the discussion of whether entropic techniques may even be sufficient for certifying classical reproducibility in certain scenarios, a question that has previously been explored for the CHSH scenario in Ref. [20].

While the entropy vector approach is known to be a useful means for distinguishing different classical causal structures, its ability to differentiate between classical and quantum versions of the same causal structure is known to be limited [66]. The present work has unveiled further limitations of the approach: for all causal structures classified in [24] we found either that the sets of

---

[30] Note that even though the sets of achievable entropy vectors in classical and quantum case coincide in the Bell scenario (cf. Section III B and [66]) this may not be the case here as very different distributions may lead to the same entropy vector in the classical and quantum case, which may be separated by local processing.

achievable entropy vectors in classical and quantum case coincide (for the causal structures of Figure 4), or that non-Shannon inequalities play a role in their characterisation leaving us unable to make such a statement.

One of the reasons why it is difficult to make such a statement when non-Shannon inequalities play a role is our relatively poor understanding of the structure of entropy space. Even in the absence of a causal structure we lack a tight characterisation of the set of allowed entropy vectors for four random variables. In the quantum case, it is an important open problem whether any further general constraints on the von Neumann entropy exist. This partly explains our inability to show whether there is some causal structure in which the described entropy vector approach can be useful for distinguishing classical and quantum.

Behind all this is the question, of whether there is a novel technique that allows for an efficient and accurate way to distinguish classical and quantum versions of the same causal structure. Such a technique needs to simplify the description of the set of allowed distributions but remain complex enough to retain the distinctive features of classical, quantum and post-quantum probability distributions. Identifying such a quantity would provide further insight into the meaning of cause in quantum mechanics.

[1] Pearl, J. *Causality: Models, Reasoning, and Inference* (Cambridge University Press, 2009), 2nd edn.

[2] Bell, J. S. On the Einstein-Podolsky-Rosen paradox. *Physics* **1**, 195–200 (1964).

[3] Ekert, A. K. Quantum cryptography based on Bell's theorem. *Physical review letters* **67**, 661–663 (1991).

[4] Mayers, D. & Yao, A. Quantum cryptography with imperfect apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS-98)*, 503–509 (IEEE Computer Society, Los Alamitos, CA, USA, 1998).

[5] Barrett, J., Hardy, L. & Kent, A. No signaling and quantum key distribution. *Physical review letters* **95**, 010503 (2005).

[6] Acín, A., Gisin, N. & Masanes, L. From Bell's theorem to secure quantum key distribution. *Physical review letters* **97**, 120405 (2006).

[7] Vazirani, U. & Vidick, T. Fully device-independent quantum key distribution. *Physical review letters* **113**, 140501 (2014).

[8] Colbeck, R. *Quantum and Relativistic Protocols For Secure Multi-Party Computation*. Ph.D. thesis, University of Cambridge (2007). Also available as arXiv:0911.3814.

[9] Pironio, S. *et al.* Random numbers certified by Bell's theorem. *Nature* **464**, 1021–1024 (2010).

[10] Colbeck, R. & Kent, A. Private randomness expansion with untrusted devices. *Journal of Physics A: Mathematical and Theoretical* **44**, 095305 (2011).

[11] Miller, C. A. & Shi, Y. Universal security for randomness expansion from the spot-checking protocol. *SIAM Journal on Computing* **46**, 1304–1335 (2017).

[12] Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Physical Review Letters* **23**, 880–884 (1969).

[13] Braunstein, S. L. & Caves, C. M. Information-theoretic Bell inequalities. *Physical Review Letters* **61**, 662–665 (1988).

[14] Greenberger, D. M., Horne, M. & Zeilinger, A. Going beyond Bell's theorem. In Kafatos, M. (ed.) *Bell's Theorem, Quantum Mechanics and Conceptions of the Universe*, 69–72 (Kluwer Academic, Dordrecht, The Netherlands, 1989).

[15] Cerf, N. J. & Adami, C. Entropic Bell inequalities. *Physical Review A* **55**, 3371–3374 (1997).

[16] Collins, D., Gisin, N., Linden, N., Massar, S. & Popescu, S. Bell inequalities for arbitrarily high-dimensional systems. *Physical review letters* **88**, 040404 (2002).

[17] Garcia, L. D., Stillman, M. & Sturmfels, B. Algebraic geometry of Bayesian networks. *Journal of Symbolic Computation* **39**, 331–355 (2005).

[18] Lee, C. M. & Spekkens, R. W. Causal Inference via Algebraic Geometry: Feasibility Tests for Functional Causal Structures with Two Binary Observed Variables. *Journal of Causal Inference* **5** (2017).

[19] Chaves, R. & Fritz, T. Entropic approach to local realism and noncontextuality. *Physical Review A* **85**, 032113 (2012).

[20] Chaves, R. Entropic inequalities as a necessary and sufficient condition to noncontextuality and locality. *Physical Review A* **87**, 022102 (2013).

[21] Fritz, T. & Chaves, R. Entropic inequalities and marginal problems. *IEEE Transactions on Information Theory* **59**, 803–817 (2013).

[22] Chaves, R., Luft, L. & Gross, D. Causal structures from entropic information: geometry and novel scenarios. *New Journal of Physics* **16**, 043001 (2014).

[23] Chaves, R. *et al.* Inferring latent structures via information inequalities. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, 112–121 (AUAI Press, Corvallis, Oregon, 2014).

[24] Henson, J., Lal, R. & Pusey, M. F. Theory-independent limits on correlations from generalized Bayesian networks. *New Journal of Physics* **16**, 113043 (2014).

[25] Steudel, B. & Ay, N. Information-theoretic inference of common ancestors. *Entropy* **17**, 2304–2327 (2015).

[26] Chaves, R., Majenz, C. & Gross, D. Information-theoretic implications of quantum causal structures. *Nature communications* **6**, 5766 (2015).

[27] Chaves, R. & Budroni, C. Entropic nonsignaling correlations. *Phys. Rev. Lett.* **116**, 240501 (2016).

[28] Pienaar, J. Which causal structures might support a quantum–classical gap? *New Journal of Physics* **19**, 043021 (2017).

[29] Kela, A., von Prillwitz, K., Aberg, J., Chaves, R. & Gross, D. Semidefinite tests for latent causal structures. e-print `arXiv:1701.00652` (2017).

[30] Miklin, N., Abbott, A. A., Branciard, C., Chaves, R. & Budroni, C. The entropic approach to causal correlations. e-print `arXiv:1706.10270` (2017).

[31] Pawłowski, M. *et al.* Information causality as a physical principle. *Nature* **461**, 1101–1104 (2009).

[32] Weilenmann, M. & Colbeck, R. Analysing causal structures with entropy. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **473** (2017).

[33] Shannon, C. E. A mathematical theory of communication. *Bell System Technical Journal* **27**, 379–423 (1948).

[34] Zhang, Z. & Yeung, R. W. A non-Shannon-type conditional inequality of information quantities. *IEEE Transactions on Information Theory* **43**, 1982–1986 (1997).

[35] Zhang, Z. & Yeung, R. On characterization of entropy function via information inequalities. *IEEE Transactions on Information Theory* **44**, 1440–1452 (1998).

[36] Matus, F. Infinitely many information inequalities. In *IEEE International Symposium on Information Theory*, 41–44 (IEEE, 2007).

[37] McGill, W. J. Multivariate information transmission. *Psychometrika* **19**, 97–116 (1954).

[38] Yeung, R. A framework for linear information inequalities. *IEEE Transactions on Information Theory* **43**, 1924–1934 (1997).

[39] Dougherty, R., Freiling, C. & Zeger, K. Non-Shannon information inequalities in four random variables. e-print `arXiv:1104.3602` (2011).

[40] Makarychev, K., Makarychev, Y., Romashchenko, A. & Vereshchagin, N. A new class of non-Shannon-type inequalities for entropies. *Communications in Information and Systems* **2**, 147–166 (2002).

[41] Dougherty, R., Freiling, C. & Zeger, K. Six new non-Shannon information inequalities. In *IEEE International Symposium on Information Theory*, 233–236 (IEEE, 2006).

[42] Weidong Xu, Jia Wang & Jun Sun. A projection method for derivation of non-Shannon-type information inequalities. In *IEEE International Symposium on Information Theory*, 2116–2120 (IEEE, 2008).

[43] Ingleton, A. W. Representation of matroids. In Welsh, D. J. A. (ed.) *Combinatorial Mathematics and its applications*, 149–167 (Academic Press, 1971).

[44] Dougherty, R., Freiling, C. & Zeger, K. Linear rank inequalities on five or more variables. e-print `arXiv:0910.0284` (2009).

[45] Dougherty, R. Computations of linear rank inequalities on six variables. In *IEEE International Symposium on Information Theory*, 2819–2823 (IEEE, 2014).

[46] Williams, H. P. Fourier's method of linear programming and its dual. *The American Mathematical Monthly* **93**, 681–695 (1986).

[47] Pippenger, N. The inequalities of quantum information theory. *IEEE Transactions on Information Theory* **49**, 773–789 (2003).

[48] Lieb, E. H. & Ruskai, M. B. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics* **14**, 1938 (1973).

[49] Chan, T. H. Balanced information inequalities. *IEEE Trans. Inf. Theor.* **49**, 3261–3267 (2003).

[50] Cadney, J., Linden, N. & Winter, A. Infinitely many constrained inequalities for the von Neumann entropy. *IEEE Transactions on Information Theory* **58**, 3657–3663 (2012).

[51] Branciard, C., Rosset, D., Gisin, N. & Pironio, S. Bilocal versus nonbilocal correlations in entanglement-swapping experiments. *Physical Review A* **85**, 032119 (2012).

[52] Fritz, T. Beyond Bell's theorem: correlation scenarios. *New Journal of Physics* **14**, 103001 (2012).

[53] Chernikov, S. N. Contraction of systems of linear inequalities. *Soviet mathematics Doklady* **1** (1960).

[54] Chernikov, S. N. The convolution of finite systems of linear inequalities. *USSR Computational Mathematics and Mathematical Physics* **5**, 1–24 (1965).

[55] Bastrakov, S. I. & Zolotykh, N. Y. Fast method for verifying Chernikov rules in Fourier-Motzkin elimination. *Computational Mathematics and Mathematical Physics* **55**, 160–167 (2015).

[56] Monniaux, D. Quantifier elimination by lazy model enumeration. In Touili, T., Cook, B. & Jackson, P. (eds.) *Computer Aided Verification: 22nd International Conference, CAV 2010, Edinburgh, UK*, 585–599 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010).

[57] Muller, M. E. A note on a method for generating points uniformly on n-dimensional spheres. *Communications of the ACM* **2**, 19–20 (1959).

[58] Zhang, Z. On a new non-Shannon type information inequality. *Communications in Information and Systems* **3**, 47–60 (2003).

[59] Fine, A. Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics* **23**, 1306 (1982).

[60] Fine, A. Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters* **48**, 291–295 (1982).

[61] Cirel'son, B. Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics* **4**, 93–100 (1980).

[62] Barrett, J. Information processing in generalized probabilistic theories. *Physical Review A* **75**, 032304 (2007).

[63] Tsirelson, B. Some results and problems on quantum Bell-type inequalities. *Hadronic Journal Supplement* **8**, 329–345 (1993).

[64] Popescu, S. & Rohrlich, D. Quantum nonlocality as an axiom. *Foundations of Physics* **24**, 379–385 (1994).

[65] Wolfe, E., Spekkens, R. W. & Fritz, T. The inflation technique for causal inference with latent variables. e-print `arXiv:1609.00672` (2016).

[66] Weilenmann, M. & Colbeck, R. Inability of the entropy vector method to certify nonclassicality in linelike causal structures. *Physical Review A* **94**, 042112 (2016).

[67] Mermin, N. D. Simple unified form for the major no-hidden-variables theorems. *Physical Review Letters* **65**, 3373–3376 (1990).

[68] Peres, A. Incompatible results of quantum measurements. *Physics Letters A* **151**, 107–108 (1990).

Appendix A: Discussion of discrepancy with [22, 26]

Our new approximations to $\overline{\Gamma^*_{\mathcal{M}}}\left(C_3^C\right)$ presented in Section III A contradict the claim in [22, 26] that $\overline{\Gamma^*_{\mathcal{M}}}\left(C_3^C\right) = \Gamma_{\mathcal{M}}(C_3^C)$. This appendix reviews these results and explains the discrepancy.

In [22, 26], the inequalities defining the set $\Gamma_{\mathcal{M}}\left(C_3^C\right)$ as well as its vertex description were calculated. Furthermore, probability distributions $P \in \mathcal{P}_{\mathcal{M}}$ that achieve the rays of $\Gamma_{\mathcal{M}}\left(C_3^C\right)$ are presented in the Supplementary Information of [26]. However, it is not shown there, that the corresponding distributions, P, lie in $\mathcal{P}_{\mathcal{M}}\left(C_3^C\right)$, and hence that the corresponding entropy vectors are achievable in C_3^C.

Our results imply that $\Gamma_{\mathcal{M}}\left(C_3^C\right) \neq \overline{\Gamma^*_{\mathcal{M}}}\left(C_3^C\right)$ and that three of the extremal rays of $\Gamma_{\mathcal{M}}\left(C_3^C\right)$ cannot lie within $\overline{\Gamma^*_{\mathcal{M}}}\left(C_3^C\right)$, specifically the ray containing the vector $v = (2, 3, 3, 4, 4, 5, 6)$ and its permutations. In the Supplementary Information of [26], v is shown to be achieved with the probability distribution

$$P_{\mathrm{XYZ}}(x,y,z) = \begin{cases} \frac{1}{64}, & x \oplus y + x \oplus z + y \oplus z = 0 \\ 0 & \text{otherwise,} \end{cases} \quad \text{(A1)}$$

where $x \in \{1, \ldots, 4\}$, $y, z \in \{1, \ldots, 8\}$ and $\oplus$ denotes addition modulo 2. This means that $P_{\mathrm{XYZ}}(x,y,z) = \frac{1}{64}$ if and only if either x, y and z are all odd or they are all even. This distribution can be mapped to the perfect correlations of (7) by locally mapping all odd outcomes to 1 and all even outcomes to 0 at X, Y and Z. Since perfect correlations are known not to be achievable in $\mathcal{P}_{\mathcal{M}}\left(C_3^C\right)$, the distribution (A1) is not compatible with the triangle causal structure.

This resolves the apparent contradiction of our results with those from [22, 26]. What was shown there is that

$$\Gamma_{\mathcal{M}}\left(C_3^C\right) \subseteq \overline{\Gamma^*_3},$$

i.e., that all vectors $v \in \Gamma_{\mathcal{M}}\left(C_3^C\right)$ can be written as the entropy of a valid probability distribution, or arbitrarily well approximated by such (but not necessarily one that is achievable in the triangle causal structure).

Appendix B: Infinite families of inequalities

Infinite families of inequalities may be derived to tighten the entropic approximation to $\overline{\Gamma^*_{\mathcal{M}}}\left(C_3^C\right)$. Here we give the proof for the three examples provided in Proposition 4. However, there are numerous other examples that can be derived in a similar way.

The families (12) and (13) are derived from (2) by combining the inequalities for one s-value at a time with Shannon and conditional independence constraints. These are the only families derived from (2) in this way for which none of the resulting inequalities are rendered redundant by those found in the calculations for Case 2 in Section III A.

Proof of Proposition 4. We tackle the three inequalities separately.

(12): The instance of inequality (2) with $(X_1, X_2, X_3, X_4) = (X, Y, Z, C)$ can be rewritten as

$$\left(-\frac{1}{2}s^2 - \frac{3}{2}s\right)H(X) - (s+1)H(Y)$$
$$-\left(\frac{1}{2}s^2 + \frac{1}{2}s\right)H(Z) + sH(CX)$$
$$+sH(CY) - sH(CZ) + \left(\frac{1}{2}s^2 + \frac{3}{2}s + 1\right)H(XY)$$
$$+\left(s^2 + 2s\right)H(XZ) + \left(\frac{1}{2}s^2 + \frac{3}{2}s + 1\right)H(YZ)$$
$$-sH(CXY) - \left(s^2 + 2s + 1\right)H(XYZ) \geq 0.$$

Applying $I(X : Y|C) = 0$ and $I(Z : C) = 0$, all terms containing the variable C cancel and we recover (12).

(13): Inequality (2), with variable choices $(X_1, X_2, X_3, X_4) = (Y, X, C, Z)$ and using the independences $I(X : Y|C) = 0$ and $I(Z : C) = 0$, can be rewritten as

$$\left(\frac{1}{2}s^2 + \frac{1}{2}s + 1\right)H(C) - (s+1)H(X) - \left(\frac{1}{2}s^2 + \frac{3}{2}s\right)H(Y)$$
$$-\left(\frac{1}{2}s^2 + \frac{1}{2}s\right)H(CX) - H(CY) + \left(\frac{1}{2}s^2 + \frac{3}{2}s + 1\right)H(XY)$$
$$-sH(Z) + sH(XZ) + sH(YZ) - sH(XYZ) \geq 0. \quad \text{(B1)}$$

We also marginalise $\Gamma\left(C_3^C\right)$ to obtain constraints on the vectors

$$(H(C), H(X), H(Y), H(Z), H(CX), H(CY),$$
$$H(XY), H(XZ), H(YZ), H(XYZ))$$

that arise from Shannon and independence inequalities. Two of the constraints this elimination yields are the following inequalities,

$$-2H(C) - 2H(X) - 2H(Y) - 3H(Z) + H(CX) + H(CY)$$
$$+H(XY) + 2H(XZ) + 2H(YZ) - H(XYZ) \geq 0, \quad \text{(B2)}$$

$$-H(C) - H(X) - H(Z) + H(CX) + H(XZ) \geq 0. \quad \text{(B3)}$$

We now use (B2) to remove $H(CY)$ from (B1), which yields

$$\left(\frac{1}{2}s^2 + \frac{1}{2}s - 1\right)H(C) - \left(\frac{1}{2}s^2 + \frac{3}{2}s + 2\right)H(Y)$$
$$-(s+3)H(X) - (s+3)H(Z) + \left(-\frac{1}{2}s^2 - \frac{1}{2}s + 1\right)H(CX)$$
$$+\left(\frac{1}{2}s^2 + \frac{3}{2}s + 2\right)H(XY) + (s+2)H(XZ)$$
$$+(s+2)H(YZ) + (-s+1)H(XYZ) \geq 0. \quad \text{(B4)}$$

With (B3), $H(CX)$ and $H(C)$ are eliminated from (B4), which concludes the proof for this case.

(14): In a similar manner as for the family (13), we consider inequality 3 with variable choices

$(X_1, X_2, X_3, X_4) = (X, Y, C, Z)$ and the independences $I(X : Y|C) = 0$ and $I(Z : C) = 0$ to obtain

$$
\begin{aligned}
&(s+1)\,H(C) - \left(\frac{1}{2}s^2 + \frac{3}{2}s\right)H(X) - (s+1)\,H(Y) \\
&- \left(\frac{1}{2}s^2 + \frac{1}{2}s\right)H(Z) - H(CX) - sH(CY) \\
&+ \left(\frac{1}{2}s^2 + \frac{3}{2}s + 1\right)H(XY) + s^2 H(XZ) \\
&+ \left(\frac{1}{2}s^2 + \frac{1}{2}s\right)H(YZ) - s^2 H(XYZ) \geq 0 \; .
\end{aligned} \tag{B5}
$$

We also consider two inequalities that are obtained from marginalising $\Gamma\!\left(C_3^{\mathrm{C}}\right)$ to vectors

$$
\begin{aligned}
(H(C), H(X), H(Y), H(Z), H(CX), H(CY), \\
H(XY), H(XZ), H(YZ), H(XYZ)) \; ,
\end{aligned}
$$

namely (B2) as well as

$$
-H(C) - H(Y) - H(Z) + H(CY) + H(YZ) \geq 0 \; . \tag{B6}
$$

Inequality (B2) allows us to eliminate $H(CX)$ from (B5) and (B6) allows us to eliminate $H(C)$ and $H(CY)$, which concludes the proof. $\qquad\square$

Appendix C: Entropy cones for causal structures with up to five variables

For most causal structures with up to five nodes, the sets of compatible distributions generated with classical and quantum resources are identical, and hence, so are their entropic cones [24]. Ref. [24] reports one causal structure with four nodes (Figure 4(a)) and 96 causal structures with five, for which this equivalence does not hold. These were reduced to the three causal structures shown in Figure 4, using reduction criteria. In the following, we show that, for the three causal structures in question, the classical and quantum entropy cones coincide. Note that this does not imply that the same holds true for the remaining 94 causal structures. An example where we have not been able to establish this is the causal structure IĈ.

Proof of Proposition 6. We begin by showing that in the causal structures shown in Figure 4, their classical entropy cones coincide with the corresponding Shannon approximations. For the instrumental scenario of Figure 4(a) this is shown in Example 2, for the Bell scenario of Figure 4(b) this was previously shown in [66]. In the following, we hence consider Figure 4(c).

The Shannon inequalities and independence constraints lead to an outer approximation that is the conic hull of the following vectors, denoted here as lists of their components, ordered as

$$
(H(X), H(Y), H(Z), H(XY), H(XZ), H(YZ), H(XYZ)).
$$

(1)	1122222
(2)	1212222
(3)	1112222
(4)	1222222
(5)	1112212
(6)	1111111
(7)	1011111
(8)	1101111
(9)	1001101
(10)	0101011
(11)	0010111

The following strategies confirm that all of the extremal rays are achievable within the causal structure and, hence, that we have found the associated entropy cone. Note that $\oplus$ denotes addition modulo 2.

- The entropy vectors (1) and (2) are recovered by choosing A and B to be uniform bits and $X = A \oplus B$, $Y = B$, $Z = (A, X)$, or $X = A \oplus B$, $Y = (B, X)$, $Z = A$ respectively.

- (3) is recovered by letting A and B be uniform bits and $X = A \oplus B$, $Y = B$, $Z = A$.

- The entropy vector (4) is recovered by letting A and B be uniform bits and $X = A \oplus B$, $Y = (B, X)$, $Z = (A, X)$.

- Let A and B be uniform bits and let $X = A \oplus B$, $Y = B$, $Z = A \oplus X$ to recover (5).

- X is a uniform bit and $Y = X = Z$ to recover (6).

- To recover vectors (7) and (8), A or B are taken to be a uniform bit, and $X = A = Z$ or $X = B = Y$ respectively. The remaining variable is deterministic.

- Entropy vectors (9)-(11) are obtained by choosing either X, Y, or Z respectively to be uniform bits and the other two variables to take a value deterministically.

We next show that in all three examples the Shannon outer approximation also coincides with the set of compatible entropy vectors in the quantum case $\overline{\Gamma^*_{\mathcal{M}}}\!\left(C^{\mathrm{Q}}\right)$. For this, we rely on the facet description of the respective cones and show that each of the inequalities also holds in the quantum case.

1. For the instrumental scenario of Figure 4(a) the only inequality in addition to the Shannon inequalities for three observed variables is $I(X : ZY) \leq H(Z)$ [24].

This holds in the quantum case because

$$I(X : ZY) \leq I(X : ZA_Y)$$
$$\leq H(X) + H(Z) + H(A_Y) - H(XZA_Y)$$
$$\leq H(Z) + H(XA_Y) - H(XZA_Y)$$
$$\leq H(Z),$$

where the first inequality is a DPI, then we use submodularity, the independence of X and A_Y and monotonicity for the cq-state ρ_{XZA_Y}.

2. For the Bell scenario of Figure 4(b) the only constraints (in addition to the four variable Shannon inequalities) are the independencies $I(W : YZ) = 0$ and $I(Z : WX) = 0$, which hold in the quantum case.

3. For the causal structure of Figure 4(c) the only additional inequality is $I(Y : Z|X) \leq H(X)$ [24]. This holds in the quantum case because

$$I(Y : Z|X)$$
$$\leq I(Y : B_Z|X)$$
$$\leq I(A_Y : B_Z|X)$$
$$\leq H(A_Y X) + H(B_Z X) - H(X) - H(A_Y B_Z)$$
$$= H(A_Y X) + H(B_Z X) - H(X) - H(A_Y) - H(B_Z)$$
$$\leq H(X),$$

where the first two inequalities are DPIs and the third holds by monotonicity. The equality holds because A_Y and B_Z are independent and the last inequality follows from two submodularity constraints. $\square$

Appendix D: Inner approximation to $\overline{\Gamma^*_{\mathcal{M}}}\left(\hat{\mathrm{IC}}^{\mathrm{C}}\right)$

Computing an outer approximation in terms of Shannon and independence (in)equalities as well as including all permutations of the Ingleton inequality for four of the five variables, yields a cone with 46 extremal rays. In the following, we list an entropy vector on each such extremal ray, with components

$$(H(X_0), H(X_1), H(Z), H(Y), H(X_0X_1), H(X_0Z),$$
$$H(X_0Y), H(X_1Z), H(X_1Y), H(ZY), H(X_0X_1Z),$$
$$H(X_0X_1Y), H(X_0ZY), H(X_1ZY), H(X_0X_1ZY)),$$

where rays that are obtained from others by permuting X_0 and X_1 are omitted.

(1)	111111111111111
(2)	111011111111111
(3)	101111111111111
(4)	110011111011111
(5)	101011110111111
(6)	001101111111111
(7)	100011100011101
(8)	001001010110111
(9)	000100101101111
(10)	111122222122222
(11)	111112222222222
(12)	111112121221222
(13)	101112111221222
(14)	101112211222222
(15)	111022121122222
(16)	011211222222222
(17)	111212222222222
(18)	112122222222222
(19)	112123232333333
(20)	111222323233333
(21)	212123232332333
(22)	112123222332333
(23)	112123232332333
(24)	111122222233333
(25)	112223332333333
(26)	112223323333333
(27)	112123231332333
(28)	111122222232333
(29)	112223333333333
(30)	112223232332333
(31)	112323333333333
(32)	112223333443444
(33)	212134332343444
(34)	112323434444444.

We have identified probability distributions compatible with $\hat{\mathrm{IC}}^{\mathrm{C}}$ that reproduce vertices on each of the rays. Hence the convex hull of these rays is an inner approximation to $\overline{\Gamma^*_{\mathcal{M}}}\left(\hat{\mathrm{IC}}^{\mathrm{C}}\right)$. It is characterised by 23 classes of inequalities, giving a total number of 35 inequalities when including permutations. In the following we list distributions recovering one vector on each extremal ray (again not listing strategies for the rays that are obtained from others by permuting X_0 and X_1). For this purpose, let C_1, C_2, C_3, C_4, C_5 and C_6 be random bits and let $\oplus$

denote addition mod 2.

- (1) Let $Y = Z = X_1 = X_0 = C_1$.

- (2) Let $Z = X_1 = X_0 = C_1$ and $Y = 1$; (3) let $Y = Z = X_0 = C_1$ and $X_1 = 1$.

- (4) Let $X_1 = X_0 = C_1$ and $Y = Z = 1$; (5) let $Z = X_0 = C_1$, $Y = 1$ and $X_1 = 1$; (6) let $Y = Z = C_1$ and $X_1 = X_0 = 1$.

- (7) Let $X_0 = C_1$ and $Y = Z = X_1 = 1$. (8)-(9) are permutations of this.

- (10) Let $X_0 = C_1$, $X_1 = C_2$ and $Y = Z = X_0 \oplus X_1$; (11) Let $X_1 = X_0 = C_1$, let $Y = A = C_2$ and $Z = X_1 \oplus A$.

- (12) Let $X_1 = X_0 = C_1$, $A = C_2$, $Z = X_1 \oplus A$ and $Y = Z \oplus A$.

- (13) Let $X_0 = C_1$ and $X_1 = 1$, let $A = C_2$, $Z = X_0 \oplus A$ and $Y = Z \oplus A$.

- (14) Let $X_0 = C_1$ and $X_1 = 1$, let $Y = A = C_2$ and $Z = X_0 \oplus A$; (15) let $X_0 = C_1$ and $X_1 = C_2$, let $Z = X_0 \oplus X_1$ and $Y = 1$.

- (16) Let $X_0 = 1$ and $X_1 = C_1$, let $A = C_2$, $Z = X_1 \oplus A$ and $Y = (A, Z \oplus A)$.

- (17) Let $X_0 = X_1 = C_1$, let $A = C_2$, $Z = X_1 \oplus A$ and $Y = (A, Z \oplus A)$.

- (18) Let $X_0 = C_1$ and $X_1 = C_2$, let $Z = (Z_1, Z_2) = (X_0, X_1)$ and $Y = Z_1 \oplus Z_2$.

- (19) Let $X_0 = C_1$, $X_1 = C_2$ and $A = C_3$, let $Z = (X_0 \oplus A, X_1 \oplus A)$ and $Y = A$; (20) let $X_0 = C_1$, $X_1 = C_2$ and $A = C_3$, let $Z = X_0 \oplus X_1 \oplus A$ and $Y = (A, Z \oplus A)$.

- (21) Let $X_0 = (C_1, C_2)$, $X_1 = C_1$ and $A = C_3$, let $Z = (Z_1, Z_2) = (C_2 \oplus A, X_1 \oplus A)$ and $Y = Z_1 \oplus A$.

- (22) Let $X_0 = C_1$, $X_1 = C_2$ and $A = C_3$, let $Z = (Z_1, Z_2) = (X_0 \oplus A, X_1)$ and $Y = Z_1 \oplus Z_2 \oplus A$.

- (23) Let $X_0 = (C_1, C_2)$, $X_1 = (C_3, C_4)$ and $A = (C_5, C_6)$, let $Z = (Z_1, Z_2, Z_3, Z_4) = (C_1 \oplus C_5, C_2 \oplus C_6, C_3 \oplus C_6, C_4 \oplus C_5 \oplus C_6)$ and $Y = (Z_1 \oplus Z_3 \oplus C_5 \oplus C_6, Z_2 \oplus Z_4 \oplus C_5)$.[31]

- (24) Let $X_0 = C_1$, $X_1 = C_2$ and $A = C_3$, let $Z = X_0 \oplus X_1 \oplus A$ and $Y = A$.

- (25) Let $X_0 = C_1$, $X_1 = C_2$ and $A = C_3$, let $Z = (Z_1, Z_2) = (X_0 \oplus A, X_1 \oplus A)$ and $Y = (A, Z_2 \oplus A)$; (26) let $X_0 = C_1$, $X_1 = C_2$ and $A = C_3$, let $Z = (Z_1, Z_2) = (X_0 \oplus A, X_1)$ and $Y = (Z_1, Z_2 \oplus A)$.

- (27) Let $X_0 = C_1$, $X_1 = C_2$ and $A = C_3$, let $Z = (Z_1, Z_2) = (X_0 \oplus A, X_1 \oplus A)$ and $Y = Z_2 \oplus A$.

- (28) Let $X_0 = C_1$, $X_1 = C_2$ and $A = C_3$, let $Z = X_0 \oplus X_1 \oplus A$ and $Y = Z \oplus A$.

- (29) Let $X_0 = C_1$, $X_1 = C_2$ and $A = C_3$, let $Z = (Z_1, Z_2) = (X_0 \oplus A, X_1 \oplus A)$ and $Y = (Z_1 \oplus Z_2 \oplus A, A)$; (30) Let $X_0 = C_1$, $X_1 = C_2$ and $A = C_3$, let $Z = (Z_1, Z_2) = (X_0 \oplus A, X_1 \oplus A)$ and $Y = (Z_1 \oplus A, Z_2 \oplus A)$; (31) let $X_0 = C_1$, $X_1 = C_2$ and $A = C_3$, let $Z = (Z_1, Z_2) = (X_0 \oplus A, X_1 \oplus A)$ and $Y = (Z_1 \oplus A, Z_2 \oplus A, A)$.

- (32) Let $X_0 = C_1$, $X_1 = C_2$ and $A = (C_3, C_4)$, let $Z = (Z_1, Z_2) = (X_0 \oplus C_3, X_1 \oplus C_4)$ and $Y = (C_3, Z_1 \oplus Z_2 \oplus C_3 \oplus C_4)$; (33) let $X_0 = (C_1, C_2)$, $X_1 = C_3$ and $A = C_4$, let $Z = (Z_1, Z_2) = (C_1 \oplus X_1, C_2 \oplus A)$ and $Y = Z_1 \oplus Z_2 \oplus A$.

- (34) Let $X_0 = C_1$, $X_1 = C_2$ and $A = (C_3, C_4)$, let $Z = (Z_1, Z_2) = (X_0 \oplus C_3, X_1 \oplus C_4)$ and $Y = (C_3, C_4, Z_1 \oplus Z_2 \oplus C_3 \oplus C_4)$.

The searches for these distributions were performed by hand; they could, however, also be straightforwardly automated.

The Shannon outer approximation to $\overline{\Gamma^*_{\mathcal{M}}}\left(\hat{\text{IC}}^{\text{C}}\right)$ shares the 46 extremal rays of the inner approximation (given above) but has six additional ones, where in the following we list one vector on each ray, omitting rays obtained through permutation of X_0 and X_1 as above,

$$
\begin{array}{ll}
(35) & 222233334344444 \\
(36) & 222233343344444 \\
(37) & 223234343554555 \\
(38) & 223244343454555.
\end{array}
$$

We can show that these vectors are all outside $\overline{\Gamma^*_{\mathcal{M}}}\left(\hat{\text{IC}}^{\text{C}}\right)$, by resorting to non-Shannon inequalities. The Shannon outer approximation is characterised by 19 classes of inequalities, or a total of 29 inequalities including permutations.

Appendix E: Proof of Proposition 9

In the following we prove Proposition 9. First, we have computed the 7 extremal rays of $\Gamma^{\text{I}}_{\mathcal{M}}\left(C_3^{\text{C}}\right)$. We list one

---

[31] Note that the entropy values here are double the ones given in the description of the extremal ray.

vector on each such extremal ray in the following[32]:

$$
\begin{array}{ll}
(1) & 1112222 \\
(2) & 0010111 \\
(3) & 0101011 \\
(4) & 1001101 \\
(5) & 0111111 \\
(6) & 1011111 \\
(7) & 1101111.
\end{array}
$$

These rays can also be analytically shown to be the extremal rays of $\Gamma^I_{\mathcal{M}}\left(C^C_3\right)$.[33] For each extremal ray we show how to generate a probability distribution $P \in \mathcal{P}_{\mathcal{M}}\left(C^C_3\right)$ whose entropy vector $v \in \overline{\Gamma^*_{\mathcal{M}}}\left(C^C_3\right)$ lies on the ray. To do so, let A, B and C be uniform random bits.

- (1): Take $X = C$, $Z = A$ and $Y = A \oplus C$ where $\oplus$ stands for addition modulus 2.

- (2): Take $Z = A$ and let $X = 1$ and $Y = 1$ deterministic. (3) and (4) are permutations of this.

- (5): Choose $Y = A = Z$ and let $X = 1$ deterministic. (6) and (7) are permutations of this.

In this way, all extremal rays of $\Gamma^I_{\mathcal{M}}\left(C^C_3\right)$ are achieved by vectors in $\overline{\Gamma^*_{\mathcal{M}}}\left(C^C_3\right)$ and, by convexity of $\overline{\Gamma^*_{\mathcal{M}}}\left(C^C_3\right)$, we have $\Gamma^I_{\mathcal{M}}\left(C^C_3\right) \subseteq \overline{\Gamma^*_{\mathcal{M}}}\left(C^C_3\right)$.

To show that the inclusion is strict, let A, B and C be uniform random bits. Let $X = \mathrm{AND}(B,C)$, $Y = \mathrm{AND}(A,C)$ and $Z = \mathrm{OR}(A,B)$. The marginal distribution $P_{\mathrm{XYZ}} \in \mathcal{P}_{\mathcal{M}}\left(C^C_3\right)$ leads to an entropy vector $v_1 = (0.81,\ 0.81,\ 0.81,\ 1.55,\ 1.5,\ 1.5,\ 2.16)$ and an interaction information of $I(X:Y:Z) \approx 0.04 > 0$ (where all numeric values are rounded to two decimal places). Hence, $v_1 \in \overline{\Gamma^*_{\mathcal{M}}}\left(C^C_3\right)$ but $v_1 \notin \Gamma^I_{\mathcal{M}}\left(C^C_3\right)$ and therefore $\Gamma^I_{\mathcal{M}}\left(C^C_3\right) \subsetneq \overline{\Gamma^*_{\mathcal{M}}}\left(C^C_3\right)$. $\qquad\square$

Appendix F: Proof of Proposition 11

In this section, we prove Proposition 11 based on reasoning from [52]. There, the idea is that in C^Q_3 one can take X and Y to correspond to two bits, which we call $(\tilde{X}, \tilde{B})$ and $(\tilde{Y}, \tilde{A})$ respectively. The quantum state corresponding to node C is a maximally entangled state $\Psi_C = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$, the first half of which is the subsystem to C_X and the second half is C_Y. A and B can be taken to be uniform classical bits. We introduce

$\Pi_\theta = |\theta\rangle\langle\theta|$, where $|\theta\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + \sin\left(\frac{\theta}{2}\right)|1\rangle$, and the four POVMs

$$
\begin{array}{ll}
E_0 = \{\Pi_0, \Pi_\pi\} & E_1 = \{\Pi_{\pi/2}, \Pi_{3\pi/2}\} \\
F_0 = \{\Pi_{\pi/4}, \Pi_{5\pi/4}\} & F_1 = \{\Pi_{3\pi/4}, \Pi_{7\pi/4}\}.
\end{array}
$$

Consider a measurement on the C_X subsystem with POVM E_B (i.e., if $B = 0$ then E_0 is measured and otherwise E_1), and likewise a measurement on C_Y with POVM F_A. Let us denote the corresponding outcomes $\tilde{X}$ and $\tilde{Y}$. With this choice $P_{\tilde{X}\tilde{Y}|AB}$ violates the CHSH inequality [12]. The observed variables are then $X = (\tilde{X}, \tilde{B})$, $Y = (\tilde{Y}, \tilde{A})$ and $Z = (A', B')$, with the correlations set up such that $B' = \tilde{B} = B$ and $A' = \tilde{A} = A$. In essence the reason that this cannot be realised in the causal structure C^C_3 is the CHSH violation. Note though that it is also important that information about A is present in both Y and Z (and analogously for B). If for example, we consider the same scenario but with $Y = \tilde{Y}$ then we could mock-up the correlations classically. This can be done by removing A, replacing B with (B_1, B_2) and taking B_1, B_2 and C to each be a uniform random bit. We can then take $Y = C$, $Z = (B_1, B_2)$ and $X = (f(C, B_1, B_2), B_1)$, where f is chosen appropriately. Since f can depend on all of the other observed variables it can generate any correlations between them[34]. In the causal structure C_3, taking $\tilde{A} = A'$ ensures that these are shared through A and hence information about them cannot be used to generate X.

Our Proposition 11 requires a restriction of Z to one bit of information, which we prove to be possible in the following.

Proof of Proposition 11. First, since all classical distributions can be realised using quantum systems, $\mathcal{P}_{\mathcal{M}}\left(C^C_3\right) \subseteq \mathcal{P}_{\mathcal{M}}\left(C^Q_3\right)$. We now show that this inclusion is strict in the case where two nodes output two bits and one node outputs only one. To do so, we consider the setup in Figure 6 of Section V A, taking $Z = \mathrm{AND}(A, B)$, $\tilde{A} = A$, $\tilde{B} = B$ and $P_{\tilde{X}\tilde{Y}|\tilde{A}\tilde{B}}$ to violate the CHSH inequality. This yields an observed distribution of the form

$$
P_{\tilde{A}\tilde{B}\tilde{X}\tilde{Y}Z} = \frac{1}{4} P_{\tilde{X}\tilde{Y}|\tilde{A}\tilde{B}} \delta_{Z,\mathrm{AND}(\tilde{A},\tilde{B})}. \tag{F1}
$$

In a slight abuse of notation we refer to this as P_{XYZ}. We start with the following lemma in which we ignore $\tilde{X}$ and $\tilde{Y}$.

Lemma 12. *If $P_{\mathrm{XYZ}} \in \mathcal{P}_{\mathcal{M}}\left(C^C_3\right)$, then $P_{\tilde{A}|AC} = P_{\tilde{A}|A}$ and $P_{\tilde{B}|BC} = P_{\tilde{B}|B}$.*

Proof. Due to the causal constraints we can write

$$
P_{Z\tilde{A}\tilde{B}} = \sum_{abc} P_{Z|ab} P_{\tilde{A}|ac} P_{\tilde{B}|bc} P_A(a) P_B(b) P_C(c).
$$

Because $Z = \text{AND}(\tilde{A}, \tilde{B})$, we can derive the following two conditions:

1. Using $P_{Z\tilde{A}\tilde{B}}(0,1,1) = 0$, it follows that for each triple (a,b,c) either $P_{Z|ab}(0) = 0$ or $P_{\tilde{A}|ac}(1) = 0$ or $P_{\tilde{B}|bc}(1) = 0$.

2. Using $P_{Z\tilde{A}\tilde{B}}(1,0,0) = P_{Z\tilde{A}\tilde{B}}(1,0,1) = P_{Z\tilde{A}\tilde{B}}(1,1,0) = 0$, it follows that for each triple (a,b,c) either $P_{Z|ab}(1) = 0$ or $P_{\tilde{A}|ac}(1) = P_{\tilde{B}|bc}(1) = 1$.

We first argue that Z is a deterministic function of A and B, i.e., $P_{Z|ab} \in \{0,1\}$ for all pairs (a,b). From condition 2 we know that either $P_{Z|ab}(1) = 0$ (and thus $P_{Z|ab}(0) = 1$) deterministically, or that $P_{\tilde{A}|ac}(1) = P_{\tilde{B}|bc}(1) = 1$. But in the latter case condition 1 implies that $P_{Z|ab}(0) = 0$ (and thus $P_{Z|ab}(1) = 1$).

Now let us consider the two cases (a) $P_{Z|ab}(1) = 1$; and (b) $P_{Z|ab}(1) = 0$ separately.

(a) Let (a,b) be such that $P_{Z|ab}(1) = 1$. According to condition 2, $P_{\tilde{A}|ac}(1) = P_{\tilde{B}|bc}(1) = 1$ for all c, and thus we have $P_{\tilde{A}|ac} = P_{\tilde{A}|a}$ as well as $P_{\tilde{B}|bc} = P_{\tilde{B}|b}$.

(b) Let (a,b) be such that $P_{Z|ab}(1) = 0$. Then $P_{Z|ab}(0) = 1$ and thus by condition 1 for all c either $P_{\tilde{A}|ac}(1) = 0$ or $P_{\tilde{B}|bc}(1) = 0$. We further divide into two cases: either (i) (a,b) are such that $P_{Z|ab'}(1) = 0$ for all b' and $P_{Z|a'b}(1) = 0$ for all a'; or (ii) they are not.

(ii) Suppose $\exists b'$ such that $P_{Z|ab'}(1) = 1$. In this case $P_{\tilde{A}|ac}(1) = 1$ for all c due to condition 2 and thus from condition 1 we have $P_{\tilde{B}|bc}(1) = 0$. Thus for such pairs (a,b), the relations $P_{\tilde{A}|ac} = P_{\tilde{A}|a}$ as well as $P_{\tilde{B}|bc} = P_{\tilde{B}|b}$ hold. Symmetric considerations can be made in the case where $\exists a'$ such that $P_{Z|a'b}(1) = 1$ instead.

(i) It cannot be the case that all pairs (a,b) have $P_{Z|ab'}(1) = 0$ for all b' and $P_{Z|a'b}(1) = 0$ for all a' (otherwise $P_Z(1) = 0$). Hence there exists (a'',b'') for which $P_{Z|a''b''}(1) = 1$. By condition 2, this implies that $P_{\tilde{A}|a''c}(1) = P_{\tilde{B}|b''c}(1) = 1$ for all c. Thus, as $P_{Z|ab''}(1) = 0$ and $P_{\tilde{B}|b''c}(1) = 1$, it follows from condition 1 that $P_{\tilde{A}|ac}(1) = 0$ for any c; $P_{\tilde{B}|bc}(1) = 0$ follows analogously, which concludes the proof. $\qquad\square$

To prove the proposition, we will suppose that $P_{XYZ} \in \mathcal{P}_\mathcal{M}(C_3^C)$ and derive a contradiction.

First note that the previous lemma together with the form of C_3^C implies

$$P_{\tilde{A}|C} = P_{\tilde{A}}, \text{ and } P_{\tilde{B}|C} = P_{\tilde{B}}. \qquad (\text{F2})$$

Furthermore, from $P_{XYZ} \in \mathcal{P}_\mathcal{M}(C_3^C)$ we have

$$P_{\tilde{A}\tilde{B}\tilde{X}\tilde{Y}} = \sum_c P_C(c) P_{\tilde{A}|c} P_{\tilde{B}|c} P_{\tilde{X}\tilde{Y}|\tilde{A}\tilde{B}c}$$

which, using (F2), and the form of C_3^C can be rewritten

$$P_{\tilde{A}\tilde{B}\tilde{X}\tilde{Y}} = \sum_c P_C(c) P_{\tilde{A}} P_{\tilde{B}} P_{\tilde{X}|\tilde{B}c} P_{\tilde{Y}|\tilde{A}c}.$$

However, that $P_{\tilde{X}\tilde{Y}|\tilde{A}\tilde{B}}$ violates a Bell inequality means that this equation cannot hold, establishing a contradiction. $\qquad\square$