Quantum Inflation: A General Approach to Quantum Causal Compatibility

Elie Wolfe,[1] Alejandro Pozas-Kerstjens,[2, 3] Matan Grinberg,[1, 4]
Denis Rosset,[1] Antonio Acín,[5, 6] and Miguel Navascués[7]

[1]*Perimeter Institute for Theoretical Physics, Waterloo, N2L 2Y5 Ontario, Canada*
[2]*Departamento de Análisis Matemático, Universidad Complutense de Madrid, 28040 Madrid, Spain*
[3]*ICFO-Institut de Ciències Fotòniques, The Barcelona Institute of
Science and Technology, 08860 Castelldefels (Barcelona), Spain*
[4]*Princeton University, Princeton, New Jersey 08544, USA*
[5]*ICFO-Institut de Cienciès Fotòniques, The Barcelona Institute of
Science and Technology, 08860 Castelldefels (Barcelona), Spain*
[6]*ICREA, Passeig Lluis Companys 23, 08010 Barcelona, Spain*
[7]*Institute for Quantum Optics and Quantum Information (IQOQI), Boltzmanngasse 3 1090 Vienna, Austria*

Causality is a seminal concept in science: Any research discipline, from sociology and medicine to physics and chemistry, aims at understanding the causes that could explain the correlations observed among some measured variables. While several methods exist to characterize classical causal models, no general construction is known for the quantum case. In this work, we present quantum inflation, a systematic technique to falsify if a given quantum causal model is compatible with some observed correlations. We demonstrate the power of the technique by reproducing known results and solving open problems for some paradigmatic examples of causal networks. Our results may find applications in many fields: from the characterization of correlations in quantum networks to the study of quantum effects in thermodynamic and biological processes.

I. INTRODUCTION

It can be argued that one of the main challenges in any scientific discipline is to identify which causes are behind the correlations observed among some measured variables, encapsulated by their joint probability distribution. Understanding this problem is crucial in many situations, such as, for example, the development of medical treatments, taking data-based social policy decisions, the design of new materials or the theoretical modeling of experiments. More precise characterizations of causal correlations enable better decision among competing explanations for given statistics, a task known as **causal discovery**. Advances in causal understanding also enable quantification of causal effects from purely observational data, thus extracting counterfactual conclusions even in instances where randomized or controlled trials are not feasible, a task known as **causal inference** [1–3].

Bayesian causal networks, in the form of directed acyclic graphs (DAGs), provide the tools to formalize such problems. These graphs, examples of which are shown in Fig. 1, encode the causal relations between the various variables in the problem, which could be either observed or non-observed. The latter, also known as latent, are required in many relevant situations in order to explain correlations among the observed. The fundamental task addressed in this work underlies both causal discovery and causal inference, and is known as the **causal compatibility** problem. It consists of deciding whether a given joint probability distribution over some observed variables can be explained by a given candidate Bayesian causal network. Equivalently, the objective of causal compatibility can be viewed as

characterizing the set of distributions compatible with a given Bayesian network.

In all cases, the measured variables in a causal network are, by definition, classical. However, causal networks may be classical or quantum depending on whether correlations are established by means of classical or quantum information. Because of its importance and broad range of applications, there is a vast literature devoted to understanding the problem of causal compatibility for classical causal networks; see for instance Ref. [1]. On the contrary, very little is known for the quantum case despite the fact that nature is ultimately quantum and quantum effects are expected to be crucial for the understanding of many relevant phenomena in many scientific disciplines. Moreover the two problems are known to be different, as one of the consequences of Bell's theorem [4, 5] is that quantum causal networks can explain correlations for which the analogous classical network fails [6–10]. Our work addresses these issues and provides a systematic construction to tackle the problem of causal compatibility for quantum causal networks.

As mentioned, several results already exist in the classical case. Whenever the network does not contain any latent variable the solution is rather simple and it suffices to check whether all the conditional independences associated to the network topology are satisfied [1]. The problem, however, becomes much more difficult as soon as the network also includes latent variables, as their presence generally implies nontrivial inequalities on the observed probabilities. A general method to tackle the causal compatibility problem, known as the **inflation technique**, is obtained in [11]. It consists of a hierarchy of conditions, organized according to their computational cost, that are necessary for a Bayesian network to be

able to explain the observed correlations. Moreover, the hierarchy is asymptotically sufficient, in the sense that the candidate Bayesian network is compatible if, and only if, all conditions in the hierarchy are satisfied [12].

When moving to quantum causal scenarios, the problem of causal compatibility presents several new features. In the classical case, the cardinality of the latent variables can be upper bounded [13] and, therefore, the problem is decidable. In the quantum case, however, a similar upper bound cannot exist because the problem of quantum causal compatibility is undecidable, as implied by recent results on quantum correlations [14, 15]. Yet, this fact does not preclude the existence of a method similar to inflation to tackle the question. Unfortunately, the inflation technique cannot be straightforwardly adapted to the quantum case because it relies on information broadcasting, a primitive that is not plausible with quantum information [16, 17]. Other causal analysis techniques which are fundamentally quantum have been proposed. Notable among these is the quantum entropy vector approach in Ref. [18], which is applicable to all causal structures but uses only those constraints on entropies imposed by the causal structure, the scalar extension in Ref. [19], which imposes stronger constraints but cannot be applied to causal structures in which all observed nodes are causally connected, such as the triangle scenario in Fig. 1(b) or the covariance matrix approach in Ref. [20], which provides low-degree polynomial (and, thus, rough) approximations of the sets of quantum correlations in networks.

The main result of our work is the construction of **quantum inflation**, a systematic technique to study causal compatibility in any quantum Bayesian network. It can be seen as a quantum analog of the classical inflation technique which avoids the latter's reliance on information broadcast. In Sec. II we introduce the graphical notation to be used throughout this work and define the causal compatibility problem. We next explain quantum inflation by means of a simple example in Sec. III, deferring its detailed construction for arbitrary two-layer DAGs to Sec. IV. Section V generalizes the technique to apply to *every* possible multilayer causal structure involving unobserved quantum systems, utilizing the two-layer construction as its elementary constituent. Section VI then considers a slight modification to quantum inflation to obtain a new tool for assessing causal compatibility with classical models. The next three sections illustrate the power of quantum inflation through different applications. First, in Sec. VII, we focus on the study of quantum correlations and use quantum inflation to characterize correlations achievable in various tripartite quantum causal networks, including the derivation of quantum causal incompatibility witnesses. Second, we show in Sec. VIII how quantum inflation can be used for cryptography purposes in quantum networks, in particular, to bound the eavesdropper's predictability on the observed measurement outcomes. Third, in Sec. IX, we consider a standard application in causality and show how to use quantum inflation to bound the strength of causal effects (this is, to perform do-conditional estimation) in the presence of quantum common causes. Finally, we conclude in Sec. X.

We expect quantum inflation to find an application in many different contexts. The most immediate one is probably the characterization of correlations in quantum networks and its use for quantum information protocols. However, tools to characterize classical causal connections are commonly used in many scientific disciplines; see, for instance, Refs. [21, 22] for examples in the context of biomolecular investigations or genomics. Progress in our understanding of these and other processes at smaller scales will eventually have to face quantum phenomena and understand how causal explanations are affected by them. In fact, some recent works advocate, admittedly not always in fully convincing ways, that quantum effects may play a role in biological processes [23, 24]. We expect quantum inflation to be a fundamental method to address all these scientific challenges that are to come.
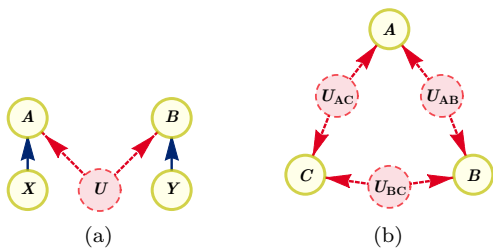

FIG. 1. DAG representation of different causal scenarios. The red, dashed circles are latent nodes, and the yellow, single-lined circles denote observed variables. (a) The Bell scenario is one of the simplest causal structures exhibiting a classical-quantum gap, that is, where there exist distributions that can be realized upon taking the latent nodes to represent quantum states (subsequently denoted as $U=\rho$), but not when the latent nodes are taken to represent distributions over classical hidden variables (subsequently denoted as $U=\Lambda$). (b) The triangle scenario, with three observed and three latent nodes, also presents a classical-quantum gap.

II. DEVICE-INDEPENDENT QUANTUM CAUSAL STRUCTURES

In this work we consider physical scenarios where an experimenter can observe a number of random variables. Some of such observed variables can be influenced by other observed variables and also by unknown quantum processes. Our goal is to determine whether the postulated causal relations between the observed variables and such hidden processes are compatible with the observed statistics.

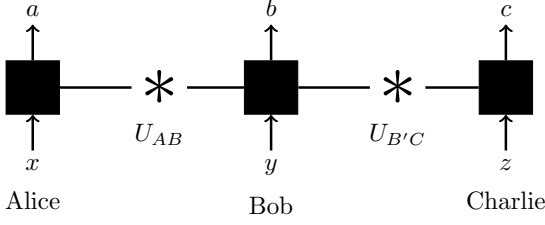Think, for instance, of the scenario depicted in Fig. 2.

FIG. 2. The quantum tripartite line in the device-independent framework.

The quantum states U_{AB} and $U_{B'C}$ are, respectively, distributed to the parties Alice and Bob, and Bob and Charlie. Alice (respectively, Charlie) probes her (respectively, his) subsystem with a measurement labeled by X (respectively, Z), obtaining the result A (respectively, C). Bob probes both of his parts of states U_{AB} and $U_{B'C}$ via a collective measurement labeled Y, obtaining the outcome B. If Alice, Bob, and Charlie repeat this experiment several times, then they can estimate the probabilities $\{P_{\mathrm{ABCXYZ}}(a,b,c,x,y,z)\}$, that, from now on, we just refer to as $P(a,b,c,x,y,z)$. We further assume that, for each party, the choice of measurement setting is done independently of the other two parties.

This scenario is known as the *tripartite-line causal scenario* [25, 26], and it is meant to model the action of a quantum repeater in a device-independent way, where we assume no knowledge of the state preparations or measurement devices involved or, for that matter, the Hilbert spaces $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_{B'}$, and $\mathcal{H}_C$ where those objects act.

For any Hilbert space $\mathcal{H}$, let $\mathcal{B}(\mathcal{H})$ denote the set of linear bounded operators mapping $\mathcal{H}$ to itself. The distribution P_{ABCXYZ} is realizable in the tripartite-line causal scenario if and only if there exist Hilbert spaces $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_{B'}$, and $\mathcal{H}_C$, quantum states

$$
\begin{aligned}
U_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B), & \quad U_{AB} \succeq 0, \quad \mathrm{Tr}[U_{AB}] = 1, \\
U_{B'C} \in \mathcal{B}(\mathcal{H}_{B'} \otimes \mathcal{H}_C), & \quad U_{B'C} \succeq 0, \quad \mathrm{Tr}[U_{B'C}] = 1,
\end{aligned} \tag{1}
$$

and positive operator valued measures (POVMs)

$$
E_{a|x} \in \mathcal{B}(\mathcal{H}_A), \qquad E_{a|x} \succeq 0, \quad \sum_a E_{a|x} = \mathbb{1}_A,
$$

$$
F_{b|y} \in \mathcal{B}(\mathcal{H}_B \otimes \mathcal{H}_{B'}), \quad F_{b|y} \succeq 0, \quad \sum_b F_{b|y} = \mathbb{1}_{BB'}, \tag{2}
$$

$$
G_{c|z} \in \mathcal{B}(\mathcal{H}_C), \qquad G_{c|z} \succeq 0, \quad \sum_c G_{c|z} = \mathbb{1}_C,
$$

such that

$$
\begin{aligned}
P_{\mathrm{ABCXYZ}}(a,b,c,x,y,z) = & \\
P_{\mathrm{X}}(x) P_{\mathrm{Y}}(y) P_{\mathrm{Z}}(z) \times & \\
\mathrm{Tr}\big[(U_{AB} \otimes U_{B'C})(E_{a|x} \otimes F_{b|y} \otimes G_{c|z})\big]. &
\end{aligned} \tag{3}
$$

For simplicity, rather than sketching the corresponding quantum experiment, we prefer to represent the relations between the variables X, Y, Z, A, B, C, U_{AB},

and $U_{B'C}$, by means of what we hereby call *device-independent quantum causal structures*, or **quantum causal structures**, for short. This structure fully determines the form of the observed statistics, as in Eq. (3).

A quantum causal structure is a Directed Acyclic Graph (**DAG**) where each node represents a classical or a quantum system. All quantum nodes are assumed to be latent: This assumption means that we can extract information from them only through their classical successors. Conversely, all classical variables are regarded as directly observable. The outgoing edges of a quantum node represent different subsystems or sectors of the corresponding quantum state, whereas the outgoing edges of a classical node denote copies of the node variable. The interpretation of the DAG is that each node is generated by applying an unknown deterministic quantum operation to all of its incoming edges. A *root node* (a node without incoming edges) can be regarded as the result of applying a transformation to a system with no information content: The corresponding random variable or quantum system is, therefore, assumed uncorrelated with any other root nodes.

The quantum causal structure corresponding to the tripartite-line causal scenario (Fig. 2) is depicted in Fig. 3(a). Since the variables X, Y, and Z and the quantum states U_{AB} and $U_{B'C}$ are assumed independent, they are represented by root nodes. The random variable A is the result of applying a quantum operation from the random variable X and one part of state U_{AB} to A's sample space. Since the outcome of the map is a classical variable, any such map corresponds to a quantum measurement on $\mathcal{H}_A$, labeled by X, i.e., to the POVMs (2).

When we specify the quantum operations mapping incoming edges to nodes in a quantum causal structure $\mathcal{G}$, we speak of a *quantum realization* of $\mathcal{G}$. A distribution P_{obs} over the observed nodes $\boldsymbol{V}_O$ of $\mathcal{G}$ is said to be
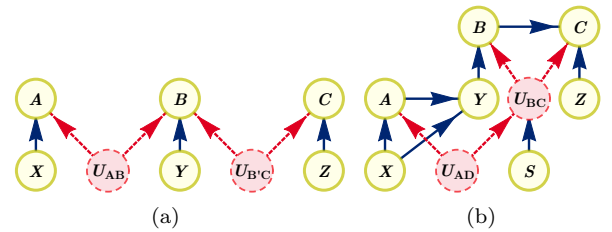

FIG. 3. DAG representation of two example causal scenarios with quantum latent nodes. (a) The tripartite-line causal scenario, where two causally independent parties A and C each share some quantum entanglement with a central party B, later denoted $\mathcal{Q}^{\curlyvee}$ for brevity. (b) Arbitrary quantum-latent causal structures contain directed edges beyond the traditional network connections of latent-to-terminal edges and root-to-terminal edges. A method for analyzing correlations in general structures with quantum latent nodes is given in Sec. V.

compatible with $\mathcal{G}$ if there exists a quantum realization that generates the distribution $P_{\mathrm{obs}}(\boldsymbol{V}_O)$.

Coming back to our example, P_{ABCXYZ} is compatible with the quantum causal structure in Fig. 3(a) if and only if P admits a decomposition of the form (3), for some Hilbert spaces $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_{B'}$, and $\mathcal{H}_C$ and operators satisfying Eqs. (1) and (2).

As another example, consider the DAG in Fig. 3(b), with quantum nodes U_{AD} and U_{BC}. A distribution $\{P_{\mathrm{ABCXYZS}}\}$ is compatible with this causal structure if and only if there exists a state U_{AD}, quantum channels $\Omega_s \in \mathcal{B}(\mathcal{H}_D) \to \mathcal{B}(\mathcal{H}_B \otimes \mathcal{H}_C)$, and POVMs $E_{a|x} \in \mathcal{B}(\mathcal{H}_A)$, $F_{b|y} \in \mathcal{B}(\mathcal{H}_{B'})$, and $G_{c|bz} \in \mathcal{B}(\mathcal{H}_C)$ such that

$$
\begin{aligned}
P_{\mathrm{ABCXYZS}}&(a,b,c,x,y,z,s) = \\
& P_{\mathrm{X}}(x) P_{\mathrm{Y|AX}}(y|a,x) P_{\mathrm{Z}}(z) P_{\mathrm{S}}(s) \times \\
& \mathrm{Tr}\big[(\mathbb{1}_A \otimes \Omega_s)(U_{AD})(E_{a|x} \otimes F_{b|y} \otimes G_{c|bz})\big].
\end{aligned} \tag{4}
$$

In general terms, the quantum causal structures we consider are arbitrary combinations of the following elements:

- *preparations*, controlled or not, of multipartite quantum states of arbitrary dimension (as an extension to the formalism, we also add sources of global shared randomness in Sec. VII C 1),

- *transformations* of (multipartite) quantum systems, described by quantum channels which may be classically controlled, and

- *measurements*, i.e., operations which generate classical random variables from (a set of) quantum states.

Without loss of generality we can assume that all measurements are projective and that all transformations are unitary, since we can dilate all such operations and then subsume all the local ancilla states into the causal structure's preparations.

When we depict such a process as a directed acyclic graph, the vertices in the DAG represent the outcomes of these process elements, not the process elements themselves. That is, DAGs are distinct from circuit diagrams. For instance, in Fig. 3(b) the node B indicates the post-measurement random variable arising from a measurement on some subsystem of U_{BC} classically controlled by Y. The node U_{BC} there indicates the (bipartite) posttransformation state arising from a transformation (classically controlled by S) applied to a fraction of the subsystems of U_{AD}.

Note that combinations of these elements can describe quite complex operations. For instance, an operation with both classical and quantum inputs and outputs is captured in this formalism by a controlled transformation followed by a projective measurement on a subspace of the resulting state.

Given a quantum causal structure $\mathcal{G}$, the purpose of this work is to tackle the following problem.

Causal Compatibility Problem
Input: a device-independent causal structure $\mathcal{G}$, and a probability distribution $P_{\mathrm{obs}}(\boldsymbol{V}_O)$ over its observed nodes $\boldsymbol{V}_O$.
Output: if $P_{\mathrm{obs}}(\boldsymbol{V}_O)$ is compatible with $\mathcal{G}$, output COMPATIBLE. Otherwise, output INCOMPATIBLE.

The main challenge when addressing the causal compatibility problem comes from the fact that the Hilbert spaces $\mathcal{H}_Q^{(i)}$ required to reproduce the observed statistics $P_{\mathrm{obs}}(\boldsymbol{V}_O)$ are not known *a priori*, and they might well be infinite dimensional. In fact, for the causal structure depicted in Fig. 1(a) (the quantum correlations scenario), the causal compatibility problem is known to be undecidable [14]. Moreover, even the weaker problem of deciding if $P_{\mathrm{obs}}(\boldsymbol{V}_O)$ can be roughly approximated by a compatible distribution cannot be solved, in general, by a Turing machine [15].

The quantum inflation technique, introduced in the next sections, is a tool to prove that certain distributions $P_{\mathrm{obs}}(\boldsymbol{V}_O)$ are incompatible with the considered causal structure $\mathcal{G}$. That is, for some instances, the inflation technique allows us to solve the causal compatibility problem in the negative. Even though the no-go results of Refs. [14, 15] imply that this tool, as any other tool designed to tackle the quantum causal compatibility problem, cannot be complete,[1] in the next sections we illustrate its applicability by proving the incompatibility of large sets of observed distributions for different quantum causal structures.

III. QUANTUM INFLATION BY EXAMPLE

For the sake of clarity, it is convenient to start the presentation of quantum inflation through an example. Consider the quantum causal network depicted in Fig. 4(a), whereby three random variables A, B, and C (taking the values a, b, and c, respectively) are generated by conducting bipartite measurements over the ends of three bipartite quantum states ρ_{AB}, ρ_{BC}, and ρ_{AC} (from now on, quantum latent variables are denoted with ρ, instead of U). We are handed the distribution $P_{\mathrm{obs}}(a,b,c)$ of observed variables and asked if it is compatible with this model. How to proceed?

―――――

[1] Namely, that for many quantum-latent causal structures $\mathcal{G}$, there will exist incompatible distributions $P_{\mathrm{obs}}(\boldsymbol{V}_O)$ whose incompatibility cannot be detected through the quantum inflation technique. It is important to note that the undecidability proofs of Refs. [14, 15] imply a limitation not specific to quantum inflation, however: Any further method for quantum causal compatibility must also be incomplete.

Suppose that there existed indeed bipartite states ρ_{AB}, ρ_{BC}, and ρ_{AC} of systems $A''B'$, $B''C'$, and $A'C''$ and commuting measurement operators E_a, F_b, and G_c, acting on systems $A'A''$, $B'B''$, and $C'C''$, respectively, which are able to reproduce the correlations $P_{\mathrm{obs}}(a,b,c)$. Now imagine how the scenario would change if n independent copies ρ_{AB}^i, ρ_{BC}^i, and ρ_{AC}^i, $i=1,...,n$ of each of the original states are distributed instead, as depicted in Fig. 4(b). Call ρ the overall quantum state before any measurement is carried out. For any $i,j=1,...,n$ we can, in principle, implement measurement $\{E_a\}_a$ on the ith copy of ρ_{AC} and the jth copy of ρ_{AB}: We denote by $\{E_a^{i,j}\}_a$ the corresponding measurement operators. Similarly, call $\{F_b^{i,j}\}_b$, (respectively, $\{G_c^{i,j}\}_c$) the measurement $\{F_b\}_b$ (respectively, $\{G_c\}_c$) over the states ρ_{AB}^i and ρ_{BC}^j (respectively, ρ_{BC}^i and ρ_{AC}^j).

classical

The newly defined operators and their averages, $\langle X \rangle_\rho := \mathrm{Tr}[\rho X]$, satisfy nontrivial relations. For example, for $H=E,F,G$ and $i \neq k, j \neq l$ the operators $H^{i,j}$ and $H^{k,l}$ act on different Hilbert spaces and, hence, commute, $\left[H^{i,j}, H^{k,l} \right] = 0$. Similarly, expressions such as $\left\langle E_a^{1,1} E_{a'}^{1,2} F_b^{2,2} \right\rangle$ and $\left\langle E_a^{1,2} E_{a'}^{1,1} F_b^{1,2} \right\rangle_\rho$ can be shown identical, since one can arrive at the second one from the first one just by exchanging ρ_{AB}^1 with ρ_{AB}^2. More generally, for any function $Q(\{E_a^{i,j}, F_b^{k,l}, G_c^{m,n}\})$ of the measurement operators and any three permutations π,
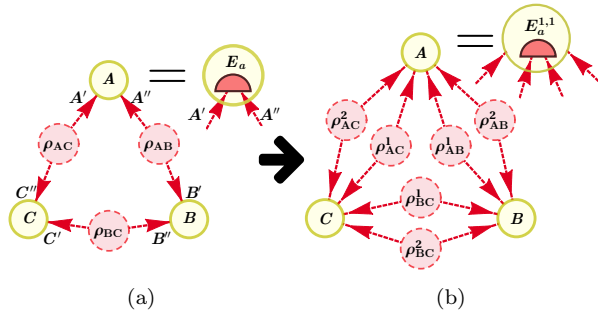

(a)   (b)

FIG. 4. Quantum inflation in the triangle scenario. (a) In the original scenario, by probing systems A' and A'' with the quantum measurement $\{E_a\}_a$, a value a for the random variable A is generated. The values b and c for the random variables B and C are produced similarly. (b) In quantum inflation, we distribute n (in the case shown, $n=2$) independent copies of the same states to the parties, which now use the original measurement operators on different pairs of copies of the states they receive. For instance, the measurement operators $\{E_a^{1,1}\}_a$ act on the states corresponding to copies ρ_{AB}^1 and ρ_{AC}^1, and the measurements with other superindices are defined in an analogous way.

π', and π'' of the indices $1,...,n$, one should have

$$\left\langle Q(\{E_a^{i,j}, F_b^{k,l}, G_c^{m,n}\}) \right\rangle_\rho$$
$$= \left\langle Q(\{E_a^{\pi(i),\pi'(j)}, F_b^{\pi'(k),\pi''(l)}, G_c^{\pi''(m),\pi(n)}\}) \right\rangle_\rho. \quad (5)$$

Finally, note that, if we conduct the measurements $\{E^{i,i}, F^{i,i}, G^{i,i}\}_{i=1}^n$ at the same time (something we can do, as they all commute with each other), then the measurement outcomes $a^1,...,a^n,b^1,...,b^n,c^1,...,c^n$ are distributed according to

$$\left\langle \prod_{i=1}^n E_{a^i}^{i,i} F_{b^i}^{i,i} G_{c^i}^{i,i} \right\rangle_\rho = \prod_{i=1}^n P_{\mathrm{obs}}(a^i, b^i, c^i). \quad (6)$$

That is, if the original distribution $P_{\mathrm{obs}}(a,b,c)$ is compatible with the network in Fig. 4(a), then there should exist a Hilbert space $\mathcal{H}$, a state $\rho : \mathcal{H} \to \mathcal{H}$ and operators $\{E_a^{i,j}\}_{i,j,a}$, $\{F_b^{k,l}\}_{k,l,b}$, and $\{G_c^{m,n}\}_{m,n,c}$ satisfying the above relations. If such is the case, we say that $P_{\mathrm{obs}}(a,b,c)$ admits an **n^{th}-order quantum inflation**. By increasing the index of n, we arrive at a hierarchy of conditions, each of which must be satisfied by any compatible distribution $P_{\mathrm{obs}}(a,b,c)$.

At first glance, disproving the existence of a quantum inflation looks as difficult as the original feasibility problem. However, the former task can be tackled via noncommutative polynomial optimization (NPO) theory [27]. Originally developed to characterize quantum correlations in Bell scenarios through the Navascues-Pironio-Acin (NPA) hierarchy [28, 29], the general goal of NPO theory is to optimize the expectation value of a polynomial over operators subject to a number of polynomial operator and statistical constraints. This optimization is achieved by means of a hierarchy of semidefinite programming (SDP) tests [30]; see also Appendix A. In our particular case, we are dealing with a feasibility problem. The polynomial operator constraints we wish to enforce on $E_a^{i,j}$, $F_b^{k,l}$, and $G_c^{m,n}$ are that they define complete families of projectors, which commute when acting on different quantum systems. The statistical constraints are given by Eqs. (5) and (6). If for some n we are able to certify, via NPO theory, that $P_{\mathrm{obs}}(a,b,c)$ does not admit an nth-order quantum inflation, then we would prove that $P_{\mathrm{obs}}(a,b,c)$ does not admit a realization in the quantum network in Fig. 4(a). An application of this method for this precise scenario is given in Sec. VII A 1.

The method just described can be easily adapted to bound the statistics of any network in which the observed variables are defined by measurements on the quantum latent variables, the triangle scenario described in this section being only an example of such networks. To test the incompatibility of a distribution P_{obs}, one would

- consider a modified network with n copies of each of the latent variables;

- extend the original measurement operators to act on all possible copies of each system;

- work out how operator averages relate to P_{obs} and to each other; and

- use NPO theory to disprove the existence of a state and operators satisfying the inferred constraints.

In Sec. V we further show how to extend the notion of quantum inflation to prove infeasibility in general quantum causal structures, where there might be causal connections among observed variables, as well as from observed to latent variables.

IV. DETAILED DESCRIPTION

To illustrate the details of the construction, we first consider a subset of causal scenarios in which single measurements are applied to different quantum states. They correspond to two-layer DAGs in which arrows coming from a first layer, consisting in both observed and latent variables, go to a second layer of observed variables. Each of the variables in the second layer is regarded as an *outcome variable*, since it is the result of conducting a measurement on a quantum state. The tuple of values of all the classical observed parents of such a variable can be understood as the *measurement setting* used to produce this outcome; this case is, for instance, variables X, Y, and Z in the tripartite-line scenario in Fig. 3(a).

The essential premise of quantum inflation is to ask what would happen if multiple copies of the original (unspecified) quantum states are simultaneously available to each party. In this gedanken experiment the parties use copies of their original measurement apparatus to perform n simultaneous measurements on the n copies of the original quantum states now available to them. There are different ways in which a party can align her measurements to act on the states now available; thus, we must explicitly specify *upon which unique set* of Hilbert spaces a given measurement operator acts nontrivially. Let us therefore denote measurement operators by

$$\hat{O}_{i|m}^{\boldsymbol{s}|k} \equiv \hat{O}\big(\text{outcome variable}=k,\ \text{spaces}=\boldsymbol{s},$$
$$\text{setting}=m,\ \text{outcome}=i\big),$$

where the four indices specify
1. k, the index or name of the *outcome variable* in the original causal graph,
2. $\boldsymbol{s}$, the Hilbert spaces the given operator acts on,
3. m, the measurement *setting* being used, and
4. i, the *outcome* associated with the operator.
In the example in Fig. 4, we have

$$E_a^{i,j} = E_a^{\{A_i', A_j''\}} = \hat{O}_{a|\emptyset}^{\{A_i', A_j''\}|A},$$
$$F_b^{i,j} = F_b^{\{B_i', B_j''\}} = \hat{O}_{b|\emptyset}^{\{B_i', B_j''\}|B}, \qquad (7)$$
$$G_c^{i,j} = G_c^{\{C_i', C_j''\}} = \hat{O}_{c|\emptyset}^{\{C_i', C_j''\}|C}.$$

Now, using an $n{=}2$ quantum inflation as in Fig. 4(b), one finds that $\boldsymbol{s}$ for outcome variable $k{=}A$ may be sampled from precisely four possibilities, each value being a different *tuple*:

$$\boldsymbol{s} \in \Big\{ \{A_1', A_1''\}, \{A_1', A_2''\}, \{A_2', A_1''\}, \{A_2', A_2''\} \Big\},$$

where A_i' (respectively, A_i'') denotes the factor A of the Hilbert space where ρ_{AC}^i (respectively, ρ_{AB}^i) acts.

These operators are regarded as the noncommuting variables of an NPO problem where the polynomial constraints are derived according to rules pertaining to the operators' projective nature as well as a number of commutation rules. The statistical constraints are then enforced by imposing symmetry under permutations of the state indices, and also equating certain expectation values with products of observed probabilities.

Projection rules

For fixed $\boldsymbol{s},k,m$, the noncommuting variables $\{\hat{O}_{i|m}^{\boldsymbol{s}|k}\}_i$ must correspond to a complete set of measurement operators. Since we do not restrict the dimensionality of the Hilbert space where they act, we can take them to be a complete set of projectors. That is, they must obey the relations

$$\hat{O}_{i|m}^{\boldsymbol{s}|k} = (\hat{O}_{i|m}^{\boldsymbol{s}|k})^\dagger, \hat{O}_{i|m}^{\boldsymbol{s}|k} \hat{O}_{i'|m}^{\boldsymbol{s}|k} = \delta_{ii'} \hat{O}_{i|m}^{\boldsymbol{s}|k}, \forall \boldsymbol{s},k,i,i',m \quad (8a)$$
$$\sum_i \hat{O}_{i|m}^{\boldsymbol{s}|k} = \mathbb{1}, \text{ for all } \boldsymbol{s},k,m. \quad (8b)$$

These relations imply, in turn, that each of the noncommuting variables is a bounded operator. Hence, by Ref. [27], the hierarchy of SDP programs provided by NPO is complete; i.e., if said distribution does not admit an nth-order inflation, then one of the NPO SDP relaxations will detect its infeasibility.

Commutation rules

Operators acting on different Hilbert spaces must commute. More formally,

$$\Big[\hat{O}_{i_1|m_1}^{\boldsymbol{s}_1|k_1}, \hat{O}_{i_2|m_2}^{\boldsymbol{s}_2|k_2}\Big] = 0 \qquad \text{if } \boldsymbol{s}_1 \cap \boldsymbol{s}_2 = \emptyset. \quad (9)$$

As shown in Sec. VI, it is possible to construct an alternative SDP for constraining the correlations of *classical* causal structures by imposing that all measurement operators commute.

Symmetry under permutations of the indices

Because of the way inflated networks are constructed from the original network, all averages of products of the

noncommuting variables must be invariant under any permutation π of the source indices. Call ρ the overall quantum state of the inflated network (since we do not cap the Hilbert space dimension, we can assume that all state preparations in the original network are pure). Then we have that

$$\left\langle \hat{O}_{i_1|m_1}^{\boldsymbol{s}_1|k_1} \cdot \hat{O}_{i_2|m_2}^{\boldsymbol{s}_2|k_2} \cdot \ldots \cdot \hat{O}_{i_n|m_n}^{\boldsymbol{s}_n|k_n} \right\rangle_\rho \\ = \left\langle \hat{O}_{i_1|m_1}^{\pi(\boldsymbol{s}_1)|k_1} \cdot \hat{O}_{i_2|m_2}^{\pi(\boldsymbol{s}_2)|k_2} \cdot \ldots \cdot \hat{O}_{i_n|m_n}^{\pi(\boldsymbol{s}_n)|k_n} \right\rangle_\rho. \quad (10)$$

An example of such statistical constraints imposed in the triangle scenario is given in Eq. (5). Another example, for an inflation level $n=3$, is the following:

$$\left\langle E_0^{\{A_1', A_1''\}} E_1^{\{A_2', A_2''\}} F_0^{\{B_1', B_3''\}} G_0^{\{C_1', C_1''\}} \right\rangle_\rho$$

$$\text{apply } \rho_{AB}^1 \leftrightarrow \rho_{AB}^2$$

$$= \left\langle E_0^{\{A_1', A_2''\}} E_1^{\{A_2', A_1''\}} F_0^{\{B_2', B_3''\}} G_0^{\{C_1', C_1''\}} \right\rangle_\rho \quad (11a)$$

$$\text{apply } \rho_{AB}^1 \leftrightarrow \rho_{AB}^3$$

$$= \left\langle E_0^{\{A_1', A_2''\}} E_1^{\{A_2', A_3''\}} F_0^{\{B_2', B_3''\}} G_0^{\{C_1', C_1''\}} \right\rangle_\rho \quad (11b)$$

$$\text{apply } \rho_{BC}^1 \leftrightarrow \rho_{BC}^3$$

$$= \left\langle E_0^{\{A_1', A_2''\}} E_1^{\{A_2', A_3''\}} F_0^{\{B_2', B_1''\}} G_0^{\{C_3', C_1''\}} \right\rangle_\rho, \quad (11c)$$

where, for readability, we identify $\hat{O}_{a|\emptyset}^{\boldsymbol{s}|A}$ ($\hat{O}_{b|\emptyset}^{\boldsymbol{s}|B}$) $[\hat{O}_{c|\emptyset}^{\boldsymbol{s}|C}]$ with $E_a^{\boldsymbol{s}}$ ($F_b^{\boldsymbol{s}}$) $[G_c^{\boldsymbol{s}}]$ per Eq. (7).

Consistency with identifiable monomials

Finally, the crucial ingredient that connects an inflated network to the observed correlations is that, as described by Eq. (6) in Sec. III, certain expectation values pertaining to the inflated network can be related to products of the probabilities of P_{obs}. We refer to those products of probabilities as the **identifiable monomials**. Let n be the order of the considered inflation. We first single out the jth copy of each state, for $j=1,...,n$. For each random variable k, we write $\vec{j}_k$ for the set of Hilbert spaces on which k acts with the copy labels of all Hilbert spaces set to j (e.g., $\vec{1}_A = \{A_1', A_1''\}$ and $\vec{2}_A = \{A_2', A_2''\}$ in Fig. 4); we write $i_{j,k}$ for the measurement outcome of k and $m_{j,k}$ for its measurement setting, so we end up describing the operator $\hat{O}_{i_{j,k}|m_{j,k}}^{\vec{j}_k|k}$. For fixed j, the product of these operators over all random variables k has an expectation value that reproduces the observed probability $P_{\text{obs}}(\bigcap_k(i_{j,k}|m_{j,k},k))$. Now, if we furthermore take the product of j over all n copies, the expectation value of the resulting operator reproduces a degree-n monomial comprised of observable probabilities. For-

mally, for any set of indices $\{i_{j,k},m_{j,k}\}_{j,k}$, we have that

$$\left\langle \prod_{j=1}^n \prod_k \hat{O}_{i_{j,k}|m_{j,k}}^{\vec{j}_k|k} \right\rangle_\rho = \prod_{j=1}^n P_{\text{obs}}\left(\bigcap_k (i_{j,k}|m_{j,k},k) \right), \quad (12)$$

where $(i|m,k)$ denotes the event of probing k with setting m and obtaining the result i, and where $\vec{j}_k$ indicates that this operator acts nontrivially on only the Hilbert spaces associated with copy index $\boldsymbol{\sigma}=\{jj...\}$ of those original sources which are parents of party k. See Eq. (6) for an example of this consistency rule applied to the triangle scenario.

The constraints (8), (9), (10) and (12) are satisfied by the overall state and operators involved in an nth-order quantum inflation of a particular causal process with observed distribution P_{obs}. All have a form suitable to assess their physical realizability through NPO via a feasibility problem.

The same machinery can easily be adapted to optimization problems beyond feasibility. Crucially, the left-hand side of Eq. (12), under the constraints (8)-(10), can be interpreted as a relaxation of the convex hull of n independent copies of a distribution P_{obs} compatible with the considered causal structure. Therefore, quantum inflation can be exploited not only for assessing whether an observed probability distribution can be generated in a particular quantum causal scenario, but also to optimize Bell-like polynomial expressions $B(P_{\text{obs}})$ over all quantum P_{obs} admitting a given causal explanation, so long as the polynomial $B(P_{\text{obs}})$ makes reference only to identifiable monomials. The key idea is to express any Bell-like polynomial B as a linear function over the expectation values $\left\langle \prod_j O_j \right\rangle$, via the correspondence (12). Minimizing or maximizing such a linear combination of expectation values can also be cast as a noncommutative polynomial optimization problem, that can be similarly tackled with NPO. In many polynomial optimization problems, no information about the distribution P_{obs} is specified in advance; in Sec. VII C we show an explicit example of how Eq. (12) is exploited in the optimization of polynomial Bell operators. In other cases, such as the quantum network security analysis in Sec. VIII, we consider polynomial optimization subject to partial specification of the identifiable monomials, for instance, if one has constraints pertaining to marginals of P_{obs}.

In fact, this idea can be generalized even further to carry optimizations over polynomials of arbitrary operator averages, i.e., not necessarily those averages related to products of observed probabilities (we show a simple example in Sec. VII C 2). Moreover, the quantum inflation technique can also be applied for studying classical causal inference. We explain how to do so in Sec. VI, but, before that, we describe how to extend the ideas above to arbitrary quantum causal networks.

V. ARBITRARY CAUSAL SCENARIOS

In the previous section, we provide a systematic method to characterize the correlations achievable in a subset of quantum causal networks, namely, two-layer DAGs, where no node has both parents and children and where every observable node has at most one observable parent. Let us denote such DAGs to have a **network form**. In this section, we generalize all previous methodology to characterize correlations in arbitrary causal structures composed of the elements described in Sec. II. Namely, we explain how an arbitrary causal structure composed of these elements can be reduced to network form cases, which we can already solve.

A. Visible nodes with parents and children: Maximal interruption

The first generalization we consider extends our techniques for network form DAGs to arbitrary so-called **latent exogenous** causal structures, where all unobserved nodes are root nodes but otherwise classical variables can have both parents and children.

We use the term **maximal interruption** to refer to our procedure for mapping the correlations of any latent-exogenous causal structure to the correlations of a unique network form structure. The procedure is based on recursive application of two simple rules.

First, suppose a directed edge is found to originate from an *endogenous* (nonroot) observable node, such as node A in Fig. 5(a). Such a node is playing two distinct roles simultaneously: On the one hand, it acts like a setting variable relative to its children. On the other hand, it acts like an outcome variable relative to its parents. The standard NPA semidefinite programming hierarchy [28, 29] presumes that every variable is exclusively either a setting or an outcome, however. Interruption splits such dual-role nodes in two, yielding a gedanken experiment graph with an effective outcome node as well as an effective setting node; see Fig. 5(b). Postse-

lecting on the effective outcome variable, for example, A in Fig. 5(b), taking a value that matches that of the effective setting variable, $A_1^{\#}$ in Fig. 5(b), allows us to characterize the set of distributions compatible with the original graph as a projection of the space of distributions compatible with the interrupted graph. This idea has precedent in classical causal networks in the single-world intervention graphs pioneered by Ref. [34] (see also the node-splitting procedure of Ref. [35]), as well as the e-separation technique introduced in Ref. [36]. Interruption of the sort illustrated in Fig. 5 is already exploited in quantum information research in Refs. [31–33].

The other rule relevant to maximal interruption involves the case where one observable node has multiple children, such as node S in Fig. 6(a). If every outcome
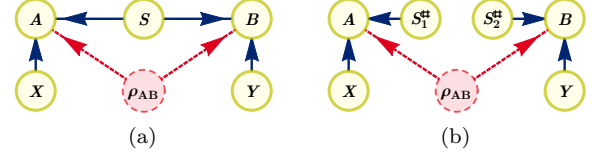
FIG. 6. (a) A Bell-like causal structure with a setting common to both parties and (b) its maximal interruption, which is a Bell causal structure assigning distinct outcome variables wholly distinct input variables. Quantum correlations in the common-setting scenario (a) can be explored by applying the standard NPA semidefinite programming hierarchy [28, 29] to the scenario in (b) and then translating back to (a) according to the postselection relation $P_{6(a)}(A{=}a,B{=}b|X{=}x,Y{=}y,S{=}s) = P_{6(b)}(A{=}a,B{=}b|X{=}x,Y{=}y,S_1^{\#}{=}S_2^{\#}{=}s)$.
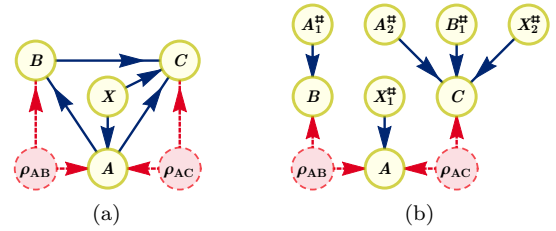
FIG. 7. An example causal structure (a) and its maximal interruption (b). Note that in (a) the observable nodes A and B both possess at least one parent node as well as at least one child node. Additionally, the observable nodes A and X have multiple child nodes. Constraints on distributions extended to the interrupted graph (such as those obtained via quantum inflation) are translated into constraints on the original distribution using the postselection relation $P_{7(a)}(A{=}a,B{=}b,C{=}c|X{=}x) = P_{7(b)}(A{=}a,B{=}b,C{=}c|X_1^{\#}{=}X_2^{\#}{=}x,A_1^{\#}{=}A_2^{\#}{=}a,B_1^{\#}{=}b)$. The maximally interrupted DAG in (b) has the two-layer network structure considered in Sec. IV.
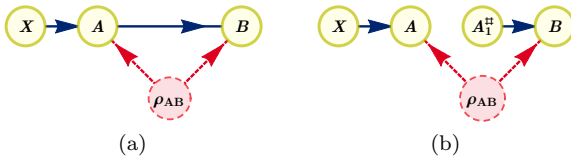

FIG. 5. (a) The instrumental causal structure and (b) its maximal interruption, which is isomorphic to the standard causal structure of a Bell experiment. Tsirelson-type inequalities pertaining to the interrupted graph are translated into constraints for the quantum instrumental scenario [31–33] via the postselection relation $P_{5(a)}(A{=}a,B{=}b|X{=}x) = P_{5(b)}(A{=}a,B{=}b|X{=}x,A_1^{\#}{=}a)$.

node is considered a "party" in the interpretation of Bell scenario causal structures, then such nodes can be thought of setting variables which are common to more than one party. Think of Fig. 6(a) as a Bell experiment in which Alice's (respectively, Bob's) setting consists of one bit determined by a source of private randomness X (respectively, Y), as well as one bit determined by a shared random number S. Without loss of generality, we can characterize the set of correlations that Alice and Bob can observe in this scenario by instead characterizing compatibility relative to Fig. 6(b) and then postselecting on $S_1^\# = S_2^\#$. That is, we can characterize the set of correlations that Alice and Bob *would* observe if the S bit for Alice and the S bit for Bob were independent—i.e., Fig. 6(b)—such that the set of distributions compatible with Fig. 6(a) is the projection of the set of distributions compatible with the independent-settings scenario onto the subspace where $S_1^\# = S_2^\#$.

Maximal interruption, then, is the combination of both these two forms of node splitting. Graphically, maximal interruption modifies a graph $\mathcal{G}$ as follows: For every observed variable V which is neither a root node nor a terminal node, as well as for any V having multiple children, we introduce k new variables $\{V_i^\#\}_i$ where k denotes the number of edges outgoing from V. We then replace each edge formerly originating from V by an edge originating from the corresponding $V_i^\#$. In the resulting (partially) *interrupted* graph $\mathcal{G}'$, V is a terminal node and every $V_i^\#$ is a root node. The correlations in $\mathcal{G}'$ are related to those in the original graph $\mathcal{G}$ by postselection, namely

$$P_{\mathcal{G}}(...,V{=}v) = P_{\mathcal{G}'}(...,V{=}v|V_1^\#{=}...{=}V_k^\#{=}v). \quad (13)$$

Proceeding in this fashion, any latent exogenous causal structure can be converted into a network form. A graphical example of this conversion is shown in Fig. 7.

Distributions over the nodes of the interrupted graph $\mathcal{G}'$ constitute *extensions* of distributions pertaining to the original graph. That is, observed statistics P pertaining to $\mathcal{G}$ only partially specify an extension P' to $\mathcal{G}'$. Critically, a distribution P is compatible with a graph $\mathcal{G}$ if and only if there exists some **valid extension** P' of P which is compatible with the interrupted graph $\mathcal{G}'$. $P'_{AB|A^\#}$ is said to constitute a valid extension of the original distribution P_{AB} when $P'_{AB|A^\#}$ is compatible with the interruption graph $\mathcal{G}'$ and recovers the original distribution P_{AB} under postselection. Formally, we have the following lemma.

Fundamental Lemma of Interruption

$$P_{AB} \in \mathsf{CompatibleDistributions}\big[\mathcal{G}\big]$$
$$\text{iff} \quad \exists_{P'} : P'_{AB|A^\#} \in \mathsf{ValidExtensions}_{\mathcal{G}'}\big[P_{AB}\big],$$
$$\text{i.e.} \quad P_{AB}(ab) = P'_{AB|A^\#}(ab\,|a) \qquad (14)$$
$$\text{and} \quad P'_{AB|A^\#} \in \mathsf{CompatibleDistributions}\big[\mathcal{G}'\big].$$

This result allows us to witness causal incompatibility relative to $\mathcal{G}$ by witnessing causal incompatibility relative to $\mathcal{G}'$. Since $\mathcal{G}'$ is of network form, we can readily apply quantum inflation to it.

B. Latent nodes with parents

The remaining case that needs to be considered is that of latent nonexogenous causal structures, where latent nodes can have parents. The Evans exogenization procedure [37] allows *classical* latent nonexogenous structures to be transformed into latent exogenous causal structures with the same predictive power. The procedure consists in replacing all arrows from a parent node to a latent node with arrows from the parent node to the children of the latent node. This operation is repeated for all parents of all latent nodes such that finally all latent variables become parentless.

When applied to quantum latent variables, however, exogenization results in a new quantum network that, in general, does not predict the same distributions of observed events as its predecessor. This result is again related to the impossibility of broadcasting quantum information [17]. The example in Fig. 8, evidencing this compatibility mismatch, is wholly due to Stefano Pironio.

To make the issue explicit, in the original network shown in Fig. 8(a) the variable S has the interpretation of a setting, which adjusts the state ρ_{BC} before it is sent to B and C. Thus, it is possible for $P(A,B|X,Y,S{=}0)$ to maximally violate a Bell inequality for A and B, and $P(A,C|X,Z,S{=}1)$ to maximally violate a Bell inequality for A and C. If one applies the Evans exogenization procedure to the original network, one obtains the network in Fig. 8(b). However, no quantum state prepared independently of S can maximally violate a Bell inequality between A and B and at the same time maximally violate another Bell inequality between A and C, due to
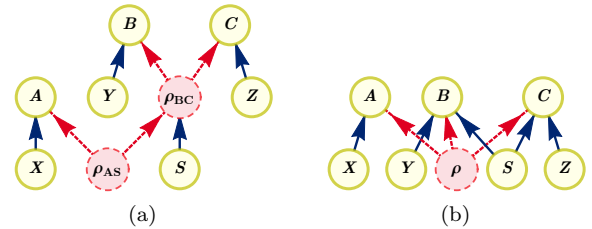

(a)   (b)

FIG. 8. In (a), there is a causal structure with ρ_{BC} being a nonexogenous unobserved quantum node. In (b), there is a different causal structure, corresponding to the classical latent reduction of the former. While these two graphs are equivalent if the unobserved nodes are classical, they are demonstrably inequivalent when the unobserved nodes are quantum.

the monogamy of quantum correlations [38, 39]. Consequently, it is not possible to reproduce such correlations within the causal network in Fig. 8(b), a product of applying Evans exogenization to Fig. 8(a).

One way to deal with this predicament is to regard observable variables with unobserved children as random variables indicating the classical control of a quantum channel. If a nonroot latent node has latent parents, then it is understood as a controlled channel on the Hilbert space defined by the latent parents. If, instead, a nonroot latent node has exclusively observable parents, then the latent node is regarded as a controlled source of quantum state preparation, which we model as a source of quantum states followed by a quantum channel controlled by the observable parents. In the same way, latent variables with latent children are treated as uncontrolled quantum channels (we defer to Appendix B the discussion that illustrates the need to account for this type of channels). Thus, in Fig. 8(a), we treat the latent variable ρ_{AS} as a joint quantum state and the root variable S as the classical control for a quantum channel acting on the BC subsystems. That is, one understands $\rho_{ABC|S=s} = \hat{U}_s \rho_{AS} \hat{U}_s^\dagger$, where, for all values of s, $\hat{U}_s$ is a unitary operator that commutes with any operator acting solely on A's subsystem. As such, the joint distribution of the values of the visible variables A, B, and C conditioned on the root visible nodes can be understood as generated by

$$P(A=a,B=b,C=c|X=x,Y=y,Z=z,S=s)$$
$$= \left\langle \hat{U}_s^\dagger \hat{A}_x^a \hat{B}_y^b \hat{C}_z^c \hat{U}_s \right\rangle_{\rho_{AS}} = \left\langle \hat{A}_x^a \hat{U}_s^\dagger \hat{B}_y^b \hat{C}_z^c \hat{U}_s \right\rangle_{\rho_{AS}}.$$

This interpretation can be made without loss of generality, since the subspace S of the complete Hilbert space AS can be understood as containing the subspaces corresponding to B and C.

As in the exogenous case, an nth-order inflation of a causal structure with nonexogenous quantum variables requires taking n copies of the unobserved root nodes and of the unobserved nodes with only visible parents. Each unitary operator $\hat{U}_s$ in the original causal structure gives rise to operators of the form $\hat{O}_s^{\{S^j\}|U}$ in the inflated graph, where j denotes the copy of the Hilbert space where $\hat{U}_s$ acts. The unitary or outcome operators associated to the descendants of any such "unitary node" (for instance, B and C, in Fig. 8) inherit the copy label j of the Hilbert space S^j.

With this last prescription, the symmetry relabeling rule [Eq. (10)] still holds. However, the projection rules [Eqs. (8a) and (8b)] hold only if the noncommuting variable k in question corresponds to an outcome variable in the original graph. If k corresponds to a unitary variable, then the operator $\hat{O}_m^{s|k}$ must be subject to the constraints

$$\hat{O}_m^{s|k}(\hat{O}_m^{s|k})^\dagger = (\hat{O}_m^{s|k})^\dagger \hat{O}_m^{s|k} = \mathbb{1}. \quad (15)$$

The commutation rule (9) remains valid upon qualifying that the Hilbert spaces listed in $\boldsymbol{s}_1 \cup \boldsymbol{s}_2$ must be si-

multaneously coexisting in the original graph. For example, in Fig. 8(a), the operators corresponding to the Hilbert spaces associated to the outcome variables B and C coexist after the transformation U_s is applied over system S. It follows that the corresponding measurement operators $O_{b|y}^{\boldsymbol{s}|B}$ and $O_{c|z}^{\boldsymbol{s}'|C}$ commute. Finally, rule (12) expressing consistency with identifiable monomials must also be amended to take into account that descendants of a unitary variable must be bracketed by the corresponding unitary and its adjoint.

Note that the aforementioned operator and statistical constraints are all polynomial, and, thus, they can all be enforced in the framework of NPO theory.

Interruption, classical exogenization, and the treatment for quantum exogenous variables hereby presented cover all possible nontrivial causal influences in arbitrary quantum causal structures composed of preparations, transformations, and measurements as described in Sec. II. Quantum inflation is, therefore, a technique of full applicability to bound the quantum correlations achievable in general causal scenarios.

VI. SDP FOR CLASSICAL COMPATIBILITY

The quantum inflation technique can be easily adapted for solving the problem of causal compatibility with an arbitrary *classical* causal structure. It is known that any correlation achievable with only classical latent variables can be realized in terms of commuting measurements acting on a quantum state [40]. Therefore, in order to detect correlations incompatible with classical structures, one must generalize the commutation relations in the original quantum inflation method to the constraint that any pair of measurement operators commutes. That is,

$$\hat{O}_{i_1|m_1}^{\boldsymbol{s}_1|k_1} \cdot \hat{O}_{i_2|m_2}^{\boldsymbol{s}_2|k_2} = \hat{O}_{i_2|m_2}^{\boldsymbol{s}_2|k_2} \cdot \hat{O}_{i_1|m_1}^{\boldsymbol{s}_1|k_1},$$

for all $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, k_1, k_2, m_1, m_2, i_1, and i_2. This generalization defines a hierarchy of constraints that classical correlations compatible with a given causal structure must satisfy.

Contrary to the quantum case, the NPO hierarchy with commuting operators associated to a fixed inflation level is guaranteed to converge at a finite level. In fact, for NPO hierarchy levels higher than $N \cdot m \cdot (d-1)$—where N is the number of parties, m is the number of settings per party, and d is the number of outcomes per measurement—application of the commutation relations allows one to reduce any product of the operators involved into one of shorter length. For a fixed inflation level, the problem solved at the highest level of the NPO hierarchy is analogous to the linear program solved in classical inflation [11] at the same inflation level.

In contrast with the original classical inflation technique, the classical variant of the quantum inflation

technique described in this work uses semidefinite programming and exhibits far more efficient scaling with the inflation hierarchy than the original linear programming approach [12]. One must note that this gain in efficiency comes at the expense of introducing further relaxations in the problem. Nevertheless, this classical variant of quantum inflation is capable of recovering a variety of seminal results of classical causal compatibility, such as the incompatibility of the so-called W and GHZ distributions (described in the results sections) with classical realizations in the triangle scenario. It also identifies the distribution described by Fritz [6], known to have a quantum realization in the triangle scenario, as incompatible with classical realizations. For all these results, the relaxed SDP formulation is far less memory demanding than the raw linear programming formulation. Furthermore, the SDP approach is the only method that can be used when using inflation to assess causal compatibility in the presence of terminal nodes which can take continuous values.

In conclusion, not only can quantum inflation be leveraged to obtain results for networks with classical sources, but we argue that it is the most suitable technique to be used for addressing causal compatibility with classical realizations in large networks.

VII. APPLICATIONS TO QUANTUM CORRELATIONS IN NETWORKS

We illustrate in this section various problems akin to quantum information science where quantum inflation provides a means to its solution or to approximations of it. We focus here on broad classes of problems (determining whether a distribution can be generated in a quantum network, finding witnesses of network nonlocality, and characterizing the quantum correlations achievable in a particular network) that underlie many practical problems and which we illustrate in concrete examples of relevance. Except those in Secs. VII A 1 and VII A 3, all the results described are novel, and quantum inflation outperforms alternative methods for characterizing quantum network correlations.[2] Moreover, to date, quantum inflation is the only technique capable of making strong quantitative statements about genuine multipartite quantum nonlocality under local operations and shared randomness (as illustrated in Ref. [41]), in that it is *the* natural framework for studying the device-independent analog of genuine network multipartite entanglement [42].

————————

[2] In the particular case in Sec. VII A 1, we still answer an open problem, although we later find that the same result can also be derived by using nonfanout classical inflations of the triangle scenario. Note that nonfanout classical inflations yield GPT-valid—and, hence, also quantum-valid—constraints [11].
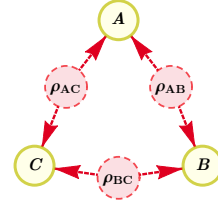

FIG. 9. The quantum triangle scenario, without settings. Each of the observable variables A, B, and C are dependent only on the sources of bipartite entanglement. This figure is essentially a reproduction of Fig. 1(b), the only difference being some added specificity regarding the quantum nature of the latent nodes.

A. Quantum causal compatibility

We begin by showing examples of specific probability distributions that we are able to identify as incompatible with various tripartite quantum causal networks. While quantum inflation applies to networks with observed variables of arbitrary cardinality, we hereafter consider scenarios involving exclusively two-output observed variables. We adopt the standard notations and label these outputs as $\{0,1\}$, unless otherwise specified.

1. The triangle scenario

We start with the simplest version of the triangle scenario consisting of three parties that are influenced in pairs by bipartite latent variables, as depicted in Fig. 9. The first example we study in this scenario is the so-called W distribution. This distribution is defined by the task of all parties outputting the outcome 0 except one, which should output 1. Explicitly, it is

$$P_{\mathrm{W}}(a,b,c) := \begin{cases} \frac{1}{3} & a+b+c=1, \\ 0 & \text{otherwise.} \end{cases} \qquad (16)$$

The W distribution is proven in Ref. [11] not to be realizable in the triangle scenario when the latent variables are classical. Additionally, it is easy to see that it is realizable with tripartite classical randomness. However, the question of whether the W distribution is realizable in the quantum triangle scenario remained open. It can be shown that P_{W} does not admit a second-order quantum inflation, and a witness that certifies this incompatibility can be found in Eq. (21). Therefore, a quantum realization of the W distribution in the triangle scenario is impossible.

Quantum inflation is robust and certifies that, when mixing the W distribution with white noise, the resulting distribution $P_{\mathrm{W},v}(a,b,c) := vP_{\mathrm{W}}(a,b,c)+(1-v)/8$ does not have a quantum realization in the triangle for all visibilities v higher than $3(2-\sqrt{3}) \approx 0.8039$. This result is obtained by solving the NPO program associated to a
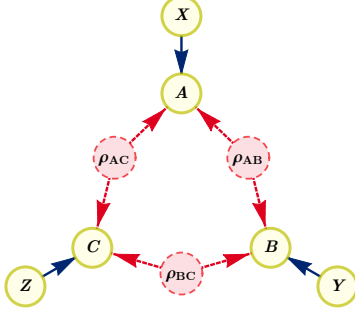
FIG. 10. The quantum triangle scenario with settings, sometimes denoted $\mathcal{Q}^\triangle$ for brevity. Each of the observable variables A, B, and C are now dependent not only on the sources of bipartite entanglement, but on additional observable classical variables X, Y, and Z representing measurement choices.

second-order inflation and the set of monomials $\mathcal{L}_2$ (see Appendix C for the definition of this set), restricted to operators of length ≤ 3.

2. The triangle scenario with settings

One can also consider more complicated networks that include additional observable variables to encode for choices of discrete measurement settings. Figure 10 shows this type of network for the case of the triangle scenario. In this setup we study the Mermin-GHZ distribution, defined by $P_{\text{Mermin}}(a,b,c|x,y,z) :=$

$$\begin{cases} 1/8 & x+y+z=0 \mod 2, \\ \left(1+(-1)^{a+b+c}\right)/8 & x+y+z=1, \\ \left(1-(-1)^{a+b+c}\right)/8 & x+y+z=3. \end{cases} \quad (17)$$

Quantum inflation also allows one to prove that the Mermin-GHZ distribution is not compatible with a quantum realization in the triangle scenario with inputs, by showing that P_{Mermin} does not admit a second-order quantum inflation; such incompatibility is certified by the violation of Eq. (22). Additionally, its noisy version $P_{\text{Mermin},v} := vP_{\text{Mermin}} + (1-v)/8$ can be proven not to have a quantum realization for any visibility v higher than $\sqrt{2/3} \approx 0.8165$.

3. The tripartite-line scenario

Quantum inflation is organized as an infinite hierarchy of necessary conditions. There are, however, situations in which it recovers the quantum boundary at a finite step. An example of these situations is provided by the tripartite-line scenario in Fig. 11, which underlies phenomena such as entanglement swapping. The tripartite-line scenario constitutes the essential causal unit employed in quantum repeaters and long-distance
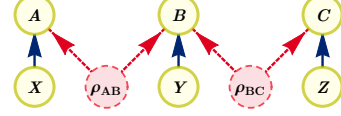

FIG. 11. The tripartite-line causal scenario, where two causally independent parties A and C each share some quantum entanglement with a central party B, sometimes denoted $\mathcal{Q}^{\curlyvee}$ for brevity. This figure is a reproduction of Fig. 3(a).

entanglement distribution networks. The main characteristic of this structure is that there is no causal connection between the extreme variables A and C. As a consequence of this characteristic, all correlations realizable in the tripartite-line scenario satisfy the following factorization relation

$$\sum_b P_{\text{obs}}(a,b,c|x,y,z) = P_{\text{obs}}(a|x)P_{\text{obs}}(c|z). \quad (18)$$

This scenario is thoroughly studied in the literature [25, 26]. In fact, it is known that the probability distribution

$$P_{\text{2PR}}(a,b,c|x,y,z) := \left[1+(-1)^{a+b+c+xy+yz}\right]/8, \quad (19)$$

despite satisfying the constraint of Eq. (18), cannot be realized in the tripartite-line scenario in terms of classical or quantum latent variables. However, it is known that its mixture with white noise, $P_{\text{2PR},v} := vP_{\text{2PR}} + (1-v)/8$, can be realized if the visibility parameter v is sufficiently small [26]. $P_{\text{2PR},v}$ admits a realization in terms of quantum latent variables for any $0 \leq v \leq 1/2$ and in terms of classical latent variables for any $0 \leq v \leq 1/4$.

Quantum inflation correctly recovers that all $P_{\text{2PR},v}$ with visibility $v > 1/2$ are incompatible with the quantum tripartite-line scenario. It does so by certifying that, for any $v > 1/2$, the corresponding $P_{\text{2PR},v}$ does not admit a second-order inflation, and this infeasibility is found already at the NPO hierarchy level corresponding to the set $\mathcal{S}_2$ (see Appendix C for a definition of this monomial set). That is, in this scenario, a finite order of quantum inflation, namely, the second, already singles out the whole region of quantum violation.

Furthermore, we can also contrast against realizations in terms of classical latent variables by following the predicament in Sec. VI, that is, by imposing that all operators in the problem commute. This classical version of quantum inflation, when analyzing compatibility with a third-order inflation, solving the NPO problem associated to the corresponding set of monomials $\mathcal{L}_1$ (its definition can be found in Appendix C), witnesses that all $P_{\text{2PR},v}$ with visibility $v > 0.328$ cannot be realized in terms of classical hidden variables. Lower values of this bound can, in principle, be achieved by considering higher inflation orders and larger monomial
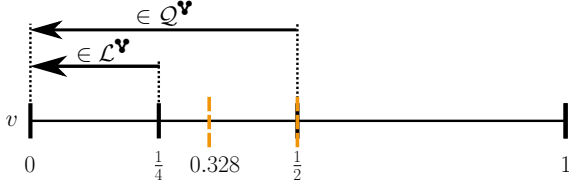
FIG. 12. Summary of results recoverable with quantum inflation in the tripartite-line scenario in Fig. 3(a) (orange dashed lines). $\mathcal{Q}^{\text{🐝}}$ and $\mathcal{L}^{\text{🐝}}$ denote, respectively, the sets of quantum and classical correlations that can be generated in the tripartite-line scenario. Quantum inflation correctly recovers that all $P_{2\text{PR},v}$ with visibility $v > 1/2$ are incompatible with the quantum tripartite line scenario already at the NPO hierarchy finite level $\mathcal{S}_2$ when assessing compatibility with a second-order inflation. When imposing that all measurements commute, it witnesses that all $P_{2\text{PR},v}$ with visibility $v > 0.328$ cannot be realized in terms of classical hidden variables.

sets. The required computational capabilities for solving these problems, however, fall outside those provided by standard tabletop computers. We defer the interested reader to Sec. X for a brief discussion on the computational cost of implementing quantum inflation, and to Ref. [43, Chap. 5] for a fuller analysis.

B. Witnesses of quantum causal incompatibility

A feature of semidefinite programming that has been widely employed in quantum information theory is the fact that the solution certificates of SDPs can be interpreted as Bell-like inequalities or witnesses that are capable of identifying correlations not attainable using the considered resources [29, 40]. The certificates obtained when using quantum inflation can have a similar interpretation as witnesses of quantum causal incompatibility. For instance, the certificate that provides the value of $v_{\max} = 6 - 3\sqrt{3} \approx 0.8039$ for the W distribution in the quantum triangle scenario gives rise to the inequality

W's Certificate

$$
\begin{pmatrix}
-2P_{\inf}(a^{\{A'_1,A''_1\}}=0) \\
+P_{\inf}(a^{\{A'_1,A''_1\}}=0,a^{\{A'_2,A''_2\}}=0) \\
-2P_{\inf}(a^{\{A'_1,A''_1\}}=0,b^{\{B'_1,B''_1\}}=0) \\
+4P_{\inf}(a^{\{A'_1,A''_1\}}=0,b^{\{B'_2,B''_2\}}=0) \\
-4P_{\inf}(a^{\{A'_1,A''_1\}}=0,b^{\{B'_1,B''_1\}}=0, \\
\qquad\qquad a^{\{A'_2,A''_2\}}=0,b^{\{B'_2,B''_2\}}=0) \\
+2P_{\inf}(a^{\{A'_1,A''_1\}}=0,a^{\{A'_2,A''_2\}}=0,b^{\{B'_1,B''_1\}}=0) \\
+2P_{\inf}(a^{\{A'_1,A''_1\}}=0,b^{\{B'_1,B''_1\}}=0,b^{\{B'_2,B''_2\}}=0) \\
-2P_{\inf}(a^{\{A'_1,A''_2\}}=0,b^{\{B'_1,B''_2\}}=0,c^{\{C'_1,C''_2\}}=0) \\
+[A\to B\to C\to A]+[A\to C\to B\to A]
\end{pmatrix} \le 0, \quad (20)
$$

where $+[A \to B \to C \to A]$ and $+[A \to C \to B \to A]$ mean repeating every term in the sum under the cyclic and anticyclic permutations of subsystems, respectively—thereby implicitly tripling the coefficients of any cyclic-invariant terms. This certificate is defined in terms

of a quantum distribution compatible with a second-order inflation of the triangle scenario. Now, by using the rules of Eq. (12) for consistency with identifiable monomials, this certificate can be transformed into a witness of tripartite distributions whose violation signals the distribution as being incompatible with a realization in the triangle scenario with quantum latent variables (note that we switch to the expectation-value picture[3]):

W's Witness

$$
\begin{pmatrix}
\langle A\rangle+\langle A\rangle^2-\langle A\rangle\langle AB\rangle-\langle A\rangle\langle AC\rangle \\
-2\langle BC\rangle+\langle B\rangle\langle C\rangle-\langle BC\rangle^2-\langle A\rangle\langle B\rangle\langle C\rangle \\
+[A\to B\to C\to A]+[A\to C\to B\to A]
\end{pmatrix} \le 3. \quad (21)
$$

In fact, the W distribution of Eq. (16) attains a value of $3 + 8/9$ for this witness.

Remarkably, the witness obtained is polynomial in the elements of the probability distribution. This result is in stark contrast with witnesses obtained through other techniques, which are either linear in the elements of the probability distribution or in the variables' entropies.

Exploiting SDP duality also enables us to obtain a polynomial witness for the infeasibility of the Mermin-GHZ distribution box discussed in Sec. VII A 2, namely,

Mermin Box Polynomial Witness

$$
\begin{pmatrix}
\begin{pmatrix}\langle A_0B_0C_0\rangle^2 \\ +\langle A_0B_1C_1\rangle^2 \\ +\langle A_1B_0C_1\rangle^2 \\ +\langle A_1B_1C_0\rangle^2\end{pmatrix} + 3\begin{pmatrix}\langle A_1B_1C_1\rangle^2 \\ +\langle A_0B_0C_1\rangle^2 \\ +\langle A_0B_1C_0\rangle^2 \\ +\langle A_1B_0C_0\rangle^2\end{pmatrix} \\
+2\begin{pmatrix}\langle A_0B_1C_1\rangle(\langle A_1B_0C_1\rangle-\langle A_0B_0C_0\rangle) \\ +\langle A_1B_0C_1\rangle(\langle A_1B_1C_0\rangle-\langle A_0B_0C_0\rangle) \\ +\langle A_1B_1C_0\rangle(\langle A_0B_1C_1\rangle-\langle A_0B_0C_0\rangle)\end{pmatrix} \\
+6\begin{pmatrix}\langle A_0B_0C_1\rangle(\langle A_1B_0C_0\rangle-\langle A_1B_1C_1\rangle) \\ +\langle A_0B_1C_0\rangle(\langle A_0B_0C_1\rangle-\langle A_1B_1C_1\rangle) \\ +\langle A_1B_0C_0\rangle(\langle A_0B_1C_0\rangle-\langle A_1B_1C_1\rangle)\end{pmatrix}
\end{pmatrix} \le 32. \quad (22)
$$

We apply this witness to the 45 nonlocal extremal points of the tripartite nonsignaling scenario [44] and notice that it recognizes 15 out of these 45 points as not realizable in the quantum triangle scenario.[4]

C. Optimization in quantum causal scenarios

As explained in Sec. IV, the constraints (8), (9), (10), and (12) characterize relaxations of the set of quantum correlations compatible with a given causal structure. This characterization can be employed to easily bound optimal values of polynomials of the measurement operators in the problem via NPO theory [27]. We provide in what follows several examples of this procedure.

----

[3] This result is obtained by performing the substitutions $p(i=0)=\frac{1}{2}(1+\langle I\rangle)$, and $p(i=0,j=0)=\frac{1}{4}(1+\langle I\rangle+\langle J\rangle+\langle IJ\rangle)$.

[4] In the ordering of Ref. [44], the boxes incompatible with the quantum triangle scenario are numbers 2, 13, 21, 22, 27, 29, 30, 31, 34, 36, 39, 41, 43, 44, and 46.

1. Optimization of linear functionals

As a first application, we derive upper bounds to the maximum value that certain Bell-like operators can achieve in the quantum triangle scenario. The results herein are obtained by considering second-order inflations, solving the associated NPO problems with the set of monomials $\mathcal{S}_2 \cup \mathcal{L}_1$. Including the identity operator, we find that each party has nine possible operators at this SDP level, such that the resulting moment matrix involved in the calculations has size 873×873.

Mermin's Inequality [45]

$$
\begin{pmatrix}
\langle A_1 B_0 C_0 \rangle \\
+ \langle A_0 B_1 C_0 \rangle \\
+ \langle A_0 B_0 C_1 \rangle \\
- \langle A_1 B_1 C_1 \rangle
\end{pmatrix}
\leq
\begin{cases}
2 & \mathcal{L}^\triangle, \mathcal{L}^\varrho \\
3.085^* & \mathcal{Q}^\triangle \\
4 & \mathcal{Q}^\varrho, \mathcal{NS}^\triangle, \mathcal{NS}^\varrho
\end{cases},
\quad (23)
$$

Svetlichny's Inequality [46]

$$
\begin{pmatrix}
\langle A_1 B_0 C_0 \rangle \\
+ \langle A_0 B_1 C_0 \rangle \\
+ \langle A_0 B_0 C_1 \rangle \\
- \langle A_1 B_1 C_1 \rangle \\
- \langle A_0 B_1 C_1 \rangle \\
- \langle A_1 B_0 C_1 \rangle \\
- \langle A_1 B_1 C_0 \rangle \\
+ \langle A_0 B_0 C_0 \rangle
\end{pmatrix}
\leq
\begin{cases}
4 & \mathcal{L}^\triangle, \mathcal{L}^\varrho \\
4.405^* & \mathcal{Q}^\triangle \\
4\sqrt{2} & \mathcal{Q}^\varrho \\
8 & \mathcal{NS}^\triangle, \mathcal{NS}^\varrho
\end{cases}.
\quad (24)
$$

In Eqs. (23) and (24), the triangle $\triangle$ denotes the causal triangle scenario, while ϱ refers to the standard tripartite Bell scenario where all parties receive shares of a joint quantum state. Clearly, for any set $\mathcal{X} = \{\mathcal{L}, \mathcal{Q}, \mathcal{NS}\}$, the bound obtained in $\mathcal{X}^\triangle$ is smaller than or equal to the one obtained in $\mathcal{X}^\varrho$. The asterisk means that the values given are upper bounds; that is, quantum inflation shows that Mermin's and Svetlichny's inequalities cannot exceed 3.085 or 4.405, respectively, in the quantum triangle scenario, but, at the moment, it is unknown whether those values are attainable.

It is known that both the algebraic maximum of 4 for Mermin's inequality and the algebraic maximum of 8 for Svetlichny's inequality can be achieved in the triangle structure, if one considers that the latent nodes distribute nonsignaling resources (see, for instance, [47, Sec. III C]). This result means that no difference between the triangle and the Bell scenarios can be identified with these two inequalities when the latent variables represent nonsignaling resources, i.e., $\max_{\mathcal{NS}^\triangle} = \max_{\mathcal{NS}^\varrho}$ for these two inequalities. Consequently, our finding here that $\max_{\mathcal{Q}^\triangle} < \max_{\mathcal{Q}^\varrho}$ for these two inequalities cannot be recovered by means of the GPT-valid (nonfanout) variant of the inflation method considered in Ref. [11, Sec. V D].

In a similar manner, the maximum values attainable by any linear function in the classical triangle and Bell scenarios also coincide, because, even though $\mathcal{L}^\triangle \subsetneq \mathcal{L}^\varrho$, the *extremal* correlations in $\mathcal{L}^\varrho$ are also members of the set $\mathcal{L}^\triangle$, as the local deterministic strategies are product correlations.

Remarkably, we see a divergence in linear-function optimization over $\triangle$ and ϱ when considering quantum latent variables. In the quantum triangle scenario, it is not possible to saturate such inequalities up to the bounds attainable in the quantum Bell scenario. This constraint has the implication that having access to tripartite quantum resources allows for improved performance in information tasks relative to having access to arbitrary bipartite quantum states and unlimited shared randomness [42].

Recall that, while polynomial functions may differ over the triangle scenarios with and without shared randomness, *linear*-function optimization is independent of the presence or absence of shared randomness.[5] That is, even though $\mathcal{Q}^\triangle \subsetneq \mathcal{Q}^{\triangle+\mathrm{SR}} \subsetneq \mathcal{Q}^\varrho$, it is the case that $\mathcal{Q}^{\triangle+\mathrm{SR}}$ is the convex hull of $\mathcal{Q}^\triangle$. The results above, thus, demonstrate that the Mermin-GHZ box of Eq. (17) is not achievable in $\mathcal{Q}^{\triangle+\mathrm{SR}}$, a result relevant to Ref. [41].
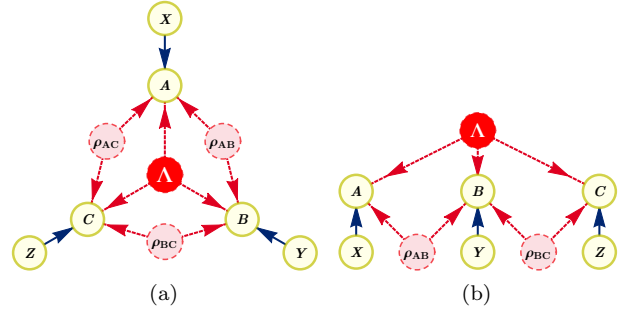

FIG. 13. (a) The quantum triangle scenario with settings and shared randomness, sometimes denoted $\mathcal{Q}^{\triangle+\mathrm{SR}}$ for brevity. (b) The quantum tripartite-line scenario with shared randomness, sometimes denoted $\mathcal{Q}^{\vee+\mathrm{SR}}$ for brevity. Such "hybrid" DAGs denote the presence of quantum sources along with the presence of a classical source of shared randomness. The classical latent node is denoted Λ, and is depicted in white text on a dark red background to distinguish it from the other quantum latent nodes. Though $\mathcal{Q}^{\triangle+\mathrm{SR}}$ is inequivalent to $\mathcal{Q}^\triangle$ depicted in Fig. 10, and similarly $\mathcal{Q}^{\vee+\mathrm{SR}}$ is inequivalent to $\mathcal{Q}^\vee$ depicted in Fig. 11, nevertheless both causal structures in each pair are equivalent up to *linear*-function optimization.

———————

[5] Linear-function optimization problems are agnostic to the presence or absence of shared randomness only when there are no additional constraints on P_{obs}, as is the case with our maximizing the Mermin or Svetlichny functionals here. However, constrained optimization is *not* agnostic to the presence or absence of shared randomness, even if the objective function is linear in P_{obs}. In the absence of shared randomness, constraints on P_{obs} are incorporated into the quantum inflation semidefinite program via Eq. (12), i.e., the rule regarding consistency with identifiable monomials. However, Eq. (12) does not hold in the presence of shared randomness; it is replaced by the weaker Eq. (25). The cryptography example in Sec. VIII explicitly demonstrates the dramatic impact of adding shared randomness to a linear-function optimization problem involving constraints on P_{obs}.

The values 3.085 and 4.405 given for $\mathcal{Q}^\triangle$ in Eqs. (23) and (24) represent upper bounds to the real maximum values achievable. We can readily establish lower bounds of $2\sqrt{2}$ and 4, respectively, by considering explicit bipartite quantum correlations for Alice and and Bob and taking Charlie to always answer $+1$. It is an open question whether, when increasing the inflation and NPO hierarchies, one will find that the values for $\mathcal{Q}^\triangle$ will collapse to those lower bounds.

It should be noted that the quantum inflation technique can readily be adapted to witness the incompatibility of a given distribution with a causal structure involving *classical shared randomness* in addition to quantum latent nodes, such as the quantum triangle scenario with settings and shared randomness or the quantum tripartite-line scenario with shared randomness depicted in Fig. 13. In particular, Eqs. (8)–(10) hold regardless of the presence or absence of shared randomness. However, Eq. (12) is valid only in the absence of shared randomness. The analog of Eq. (12) for hybrid causal structures with shared randomness simply replaces higher-degree monomials with degree-1 monomials, i.e.,

$$\left\langle \prod_k \hat{O}_{i_k|m_k}^{\vec{\mathbf{1}}_k|k} \right\rangle_\rho = P_{\text{obs}}\left(\bigcap_k (i_k|m_k,k) \right). \qquad (25)$$

2. *Optimization of nonlinear functionals*

Quantum inflation can also be used to optimize nonlinear witnesses f. The essential idea can be seen already in Sec. VII B, where Eq. (12) relates linear functions in an inflation to polynomials in the corresponding original scenario. Note that, for this optimization to be possible, in general one must consider at least an order-q inflation when optimizing a polynomial of degree q.

Consider, for instance, minimizing the 2-norm between a distribution $P_{\text{obs}}(a,b,c)$ achievable in the quantum triangle scenario and that of the W distribution, that is:

$$
\begin{aligned}
f(P_{\text{obs}}) &\equiv \sum_{a,b,c=0,1} |P_{\text{obs}}(a,b,c) - P_{\text{W}}(a,b,c)|^2 \\
&= \sum_{a,b,c=0,1} \begin{pmatrix} P_{\text{obs}}(a,b,c)^2 + P_{\text{W}}(a,b,c)^2 \\ -2P_{\text{obs}}(a,b,c)P_{\text{W}}(a,b,c) \end{pmatrix},
\end{aligned} \qquad (26)
$$

where one can readily verify that $\sum_{a,b,c} P_{\text{W}}(a,b,c)^2 = 1/3$.

One can estimate the minimum value of this function using second-order inflation or higher, via Eq. (12). Explicitly, the constraints from Eq. (12) that apply to this problem are

$$P_{\text{obs}}(a,b,c)^2 = \left\langle E_a^{1,1} E_a^{2,2} F_b^{1,1} F_b^{2,2} G_c^{1,1} G_c^{2,2} \right\rangle,$$

and $P_{\text{obs}}(a,b,c) P_{\text{W}}(a,b,c) = P_{\text{W}}(a,b,c)\left\langle E_a^{1,1} F_b^{1,1} G_c^{1,1} \right\rangle.$

While obtaining nontrivial bounds on $f(P_{\text{obs}})$ is, in principle, possible, in practice, it appears to require levels of the NPO hierarchy too computationally expensive.

VIII. APPLICATIONS TO QUANTUM CRYPTOGRAPHY IN NETWORKS

In this section, we showcase a concrete application of quantum inflation: the security analysis of multipartite quantum key distribution protocols where the parties not all share a joint quantum state but are instead arranged in a network. Because entangled systems can be transmitted only over relatively short distances and only a small number of parties can practically share an entangled state distributed by a single source, multisource networks are a promising solution for long-distance quantum key distribution. However, the fact that the underlying causal structure is no longer the Bell scenario has a profound impact in the associated security proofs [48].

Imagine that two parties, A and C, attempt to establish a secret key. They are too far apart, so they employ a quantum repeater B in order to establish correlations. However, a malicious party E may be eavesdropping what in reality are tripartite sources that leak one of the subsystems generated,[6] as in Fig. 14(a). The $ABC|XZ$ marginal within the eavesdropped quantum repeater
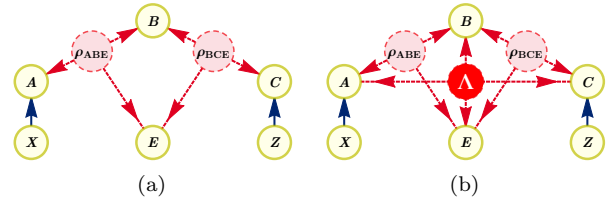

(a)   (b)

FIG. 14. (a) A and C employ the repeater B in order to generate a secret shared key. While A and C have a choice of which measurement to perform in the states they receive—denoted by X and Z—B performs a four-outcome Bell state measurement. Unbeknowns to A, B, and C, a fourth party E may be eavesdropping the sources in an attempt to extract information about the secret key that A and C are establishing. (b) The eavesdropped network scenario with shared randomness. Some tripartite distributions which are cryptographically secure under the assumption of no shared randomness become insecure if hidden dependence on shared randomness cannot be excluded on physical grounds.

---

[6] These "third subsystems" need not strictly be additional particles generated at each source but can, for instance, be the fields surrounding them [49]. Eve is allowed to perform a joint measurement on her two shares, one received from each source (call them E' and E'', using the notation in Sec. III), producing the value of the random variable E.

scenario is essentially a special case of the tripartite-line scenario in Fig. 3(a). An important question to answer is how reliably can E guess the key generated by A and C by manipulating the leaked subsystems.

Reference [48] tackles this problem, finding a trade-off between the violation of previously known network inequalities and the amount of information that E can infer about the outcomes of A and C upon performing a measurement, under the assumption that the eavesdropper is bounded only by nonsignaling constraints. Since no quantum constraints are considered, the obtained bounds on the information achievable by the eavesdropper are very conservative. In fact, if A, B, and C observe statistics compatible with a quantum entanglement swapping experiment, the information bounds derived in Ref. [48] turn out to be trivial. To derive tighter bounds for the more realistic case of quantum eavesdroppers, one needs to use tools bounding the set of allowed quantum correlations in the considered network, something that quantum inflation provides. In the following, we analyze the same scenario using the semidefinite relaxations of the set of quantum correlations achievable in the scenario in Fig. 14(a) provided by quantum inflation,[7] finding tighter bounds on the eavesdropper predictability on the measurement outcomes by the honest parties A and C.

We are interested in bounding how much information E can infer about A's and C's outcomes in the network in Fig. 14(a) from their choice of measurements. We follow Ref. [48] and estimate this information via the total variation distance $\Delta_{\mathrm{E,E|AC}}$ between the distribution of E's outcome with and without conditioning on the inputs and outputs of A and C, that is,

$$\Delta_{\mathrm{E,E|AC}}^{a,c,x,z} := \frac{1}{2}\sum_{e=1}^{|E|}\left|P_{\mathrm{E|ACXZ}}(e|a,c,x,z)-P_{\mathrm{E}}(e)\right|, \quad (27)$$

when parties A, B, and C observe some specific probability distribution $P_{\mathrm{ABC|XZ}}(a,b,c|x,z)$. For the sake of example, we focus on bounding the total variation distance conditioned upon $a=c=x=z=0$, i.e., $\Delta_{\mathrm{E,E|AC}}^{0,0,0,0}$. That is, we want to maximize $\Delta_{\mathrm{E,E|AC}}^{0,0,0,0}$ over all the probability distributions $\{P_{\mathrm{ABCE|XZ}}\}$ that can be quantumly realized in the network in Fig. 14(a), and whose marginal on A, B, and C matches the given $P_{\mathrm{ABC|XZ}}$.

Superficially, it might appear that the unspecified cardinality of $|E|$ constitutes an obstacle to exploring the space of all $\{P_{\mathrm{ABCE|XZ}}\}$ compatible with the network in Fig. 14(a). We bypass this concern, however, by exploiting an alternative formulation of total variation distance [50], namely,

$$\Delta_{\mathrm{E,E|AC}}^{a,c,x,z} = P_{\mathrm{E|ACXZ}}(\mathbf{0}^\star|a,c,x,z)-P_{\mathrm{E}}(\mathbf{0}^\star), \qquad (28)$$

where $\mathbf{0}^\star$ is the *bin* (coarse graining over a subset of possible values of E) which maximizes $P_{\mathrm{E|ACXZ}}(\mathbf{0}^\star|a,c,x,z)-P_{\mathrm{E}}(\mathbf{0}^\star)$, i.e.,

$$\mathbf{0}^\star := \bigvee e' : P_{\mathrm{E|ACXZ}}(e'|a,c,x,z)\geq P_{\mathrm{E}}(e'). \qquad (29)$$

Note that $\mathbf{0}^\star$ is an implicit function of the values $\{a,c,x,z\}$ which identify the particular total variation distance $\Delta_{\mathrm{E,E|AC}}^{a,c,x,z}$. Since we are concerned only about binning E into $e\in\mathbf{0}^\star$ versus $e\notin\mathbf{0}^\star$, we can effectively model E as dichotomic for the purposes of optimization, allowing us to set up the following SDP:

$$\begin{aligned} \max \quad & \frac{P_{\mathrm{ACE|XZ}}(\mathbf{0}^\star,0,0|0,0)}{P_{\mathrm{AC|XZ}}(0,0|0,0)}-P_{\mathrm{E}}(\mathbf{0}^\star) \\ & = \frac{\left\langle A_{0|0}^1 C_{0|0}^1 E_{\mathbf{0}^\star}^{1,1}\right\rangle}{P_{\mathrm{AC|XZ}}(0,0|0,0)}-\left\langle E_{\mathbf{0}^\star}^{1,1}\right\rangle \\ \mathrm{s.t.} \quad & \Gamma^{(n,I)}\succeq 0, \text{ and Eqs. (8), (9), (10).} \\ & \text{Eq. (12) is also imposed, but only on} \\ & \text{the marginal } P_{\mathrm{ABC|XZ}}, \text{ i.e. excluding} \\ & \text{identifiable monomials involving } E. \end{aligned} \quad (30)$$

Note that we distinguish between *identifiable monomials*—which are defined by the order of inflation via Eq. (12)—versus *known probabilities*—which may be defined by a *partial* specification of the global probability distribution over the observable nodes in a given causal structure. That is, one may find that a *strict subset* of all identifiable monomials are preassigned numerical values when tackling constrained optimization problems, such as is the case with Eq. (30). In Eq. (30), $\Gamma^{(n,I)}$ is the moment matrix associated to the set of operators $\left\{A_{a|x}^i, B_b^{j,k}, C_{c|z}^l, E_e^{m,o}\right\}$ with $1\leq i,j,k,l,m,o\leq I$, for some I^{th}-order inflation[8] and some NPO hierarchy level n.

In Fig. 15, we depict upper bounds for $\Delta_{\mathrm{E,E|AC}}^{0,0,0,0}$ when the parties A, B, and C measure the standard distribution in repeater networks (see Appendix D for its definition) with some visibility v. Importantly, when A, B, and C observe the distribution corresponding to the sources sending noiseless singlets (i.e., $v=1$), we certify that the distance between the distributions cannot exceed *zero*, and, thus, there is no better strategy for E to guess A's and C's outcomes than making a uniformly random guess. Thus, for $v=1$, we prove that the key established between A and C is completely secure, as expected. Note that, in contrast with Ref. [48], we allow E to be a quantum eavesdropper, meaning that Eve is

---

[7] This scenario can, in principle, also be numerically characterized via the scalar extension method in Ref. [19].

[8] Recall that, in the eavesdropped repeater of Fig. 14(a), the indices i, j, and m correspond to copies of the state ρ_{ABE} and the indices k, l, and o correspond to copies of the state ρ_{BCE}.
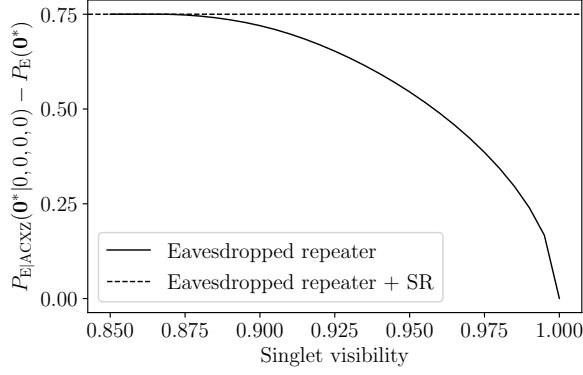
FIG. 15. Upper bounds to the total variation distance $\Delta^{0,0,0,0}_{\mathrm{E,E|AC}} := P_{\mathrm{E|ACXZ}}(\mathbf{0}^\star|0,0,0,0) - P_{\mathrm{E}}(\mathbf{0}^\star)$ when the parties A, B, and C observe the distribution depicted in Eq. (D1) with singlet visibility v. The solid curve tracks the upper bound in the eavesdropped repeater network depicted in Fig. 14(a), whereas the dashed line tracks the worst case bound, when all the parties are coordinated through classical shared randomness, as in Fig. 14(b). The computations for the eavesdropped repeater network are carried out by considering its second-order quantum inflation and the moment matrix generated by operator sequences associated to the corresponding local level $\mathcal{L}_1$ not exceeding length two. This computation results in a moment matrix of size 857×857. On the other hand, the bound $\Delta^{a,c,x,z}_{\mathrm{E,E|AC}} \leq \min\{1, 2(2-\mathcal{R})\}$ offered by Ref. [48] merely implies the trivial $\Delta^{0,0,0,0}_{\mathrm{E,E|AC}} \leq 1$ for this example, as $2(2-\mathcal{R})$ evaluates to $2(2-v\sqrt{2})$, and $1 \leq 2(2-v\sqrt{2})$ for all $v \in [0,1]$.

allowed to perform a joint entangled measurement on her subsystems of both ρ_{ABE} and ρ_{BCE}.

For comparison, imagine if it were impossible to ensure the absence of any classical shared randomness between the parties. Dropping the no-shared-randomness assumptions means that we now consider the network underlying the process to be that of Fig. 14(b). In that causal scenario, we can readily establish that $\max_P \Delta^{0,0,0,0}_{\mathrm{E,E|AC}} = 3/4$ when $P_{\mathrm{ABC|XZ}}$ is the standard distribution in repeater networks (D1), for any visibility v. This result follows from the fact that such $P_{\mathrm{ABC|XZ}}$ can be simulated classically with tripartite shared randomness; i.e., $P_{\mathrm{ABC|XZ}}$ does not violate any tripartite Bell inequality. Since $P_{\mathrm{AC|XZ}}$ *could* be a deterministic function of Λ, it follows that there exists a causal model for Fig. 14(b) in which the eavesdropper E can perfectly predict the outcomes of A, B, and C leveraging her access to Λ. Perfect predictability (i.e., *zero* security) leads to $P_{\mathrm{E|ACXZ}}(\mathbf{0}^\star|0,0,0,0) = 1$ and $P_{\mathrm{E}}(\mathbf{0}^\star) = P_{\mathrm{AC}}(0,0)$ and, more generally, to the absolute maximal possible total variation distance $\Delta^{a,c,x,z}_{\mathrm{E,E|AC}} = 1 - P_{\mathrm{AC|XZ}}(0,0|0,0) = 3/4$.

In summary, our analysis demonstrates the practical value for quantum cryptography applications of being able to bound the quantum correlations achievable in

networks by means of quantum inflation. It also highlights the critical role that network structure assumptions play in establishing security proofs, as just the addition of shared randomness transforms the distribution from being perfectly secure to completely insecure.

IX. BOUNDING CAUSAL EFFECTS UNDER QUANTUM CONFOUNDING

We now turn to explore an application of quantum inflation unrelated to quantum information theory. It is the quantum version of *mediation analysis*, which uses observational data and a promise of causal structure to estimate causal relationships and predict counterfactual intervention. Informally, mediation analysis quantifies the relative strengths of different causal pathways which could play a role in establishing some observed correlations. While this aspect has not been explored in the context of quantum mechanics (with the very recent exception of Ref. [33]), mediation analysis is a widespread tool in medical and biochemical studies [51–53]. To quantify causal strengths in disciplines concerned with microscopic phenomena, it is crucial to account for causes mediated by quantum carriers, something possible using quantum inflation.

Reichenbach's common causal principle states that all nonspurious statistical correlations must admit some causal explanation, and this principle can be generalized to include quantum causal explanations as well [8, 10, 54, 55]. Often, however, one may wish to distinguish between various causal possibilities for establishing statistical correlation. Moreover, even knowledge of the true causal structure is still inferior to a complete understanding of functional relationships between observable variables.

Consider the multiple possible causal explanations for correlation between variables A and B in Fig. 16. We have

- the *direct effect* of A on B, associated with the *edge* $A \to B$,
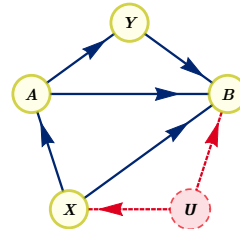- the *indirect effect* of A on B, associated with the *direct path* $A \to Y \to B$,

FIG. 16. A graph admitting multiple causal explanations for correlation between A and B.

- the *common dependence* of A and B on X, associated with the *forking path* $A \leftarrow X \rightarrow B$, and
- the common dependence of A and B on U, associated with the forking path $A \leftarrow X \leftarrow U \rightarrow B$.

Of course, in reality, the true causal explanation of a correlation typically involves *all* these routes. The subfield of causal inference known as **mediation analysis** provides definitions of effect strength parameters for quantifying the relative significance of different causal pathways [56–59]. Such measures are generally intended to distinguish between

- the *total causal effect* of A on B, associated with *all* paths originating at A and terminating at B, and
- everything else, namely the *total common cause dependence* shared between A and B, associated with *all* their common ancestors.

Latent **confounding** is the possibility of partially explaining a pair of variables' observed correlation in terms of the functional dependence on their unobserved common causes. Estimating (or identifying) the strength of causal effects in the presence of latent confounding is the subject of extensive research [57–64]; our contribution in this section is to unlock the possibility of effect estimation in the presence of quantum latent confounding by using quantum inflation.

The most well-studied measure of effect strength is the **average causal effect** (ACE) which quantifies how much one variable B functionally depends on another A. It is defined as the variation in the expectation value of B under intervention of different values for A. Formally,

$$\text{ACE}[a_1, a_2, B] := \langle B \textbf{ do } a_1^\# \rightarrow B \rangle - \langle B \textbf{ do } a_2^\# \rightarrow B \rangle$$
$$\text{with} \qquad \text{ACE}[A, B] := \max_{a_1, a_2} \text{ACE}[a_1, a_2, B]. \qquad (31)$$

Note that the distribution of B under *intervention* on A has a meaning distinct from conditioning on A. When A and B share some common causal ancestry, typically

$$P(B|\{A=a\}) \neq P(B \textbf{ do } \{A^\#=a\} \rightarrow B), \qquad (32)$$

where the notation $\{A^\#=a\} \rightarrow B$ indicates that the value of A *as seen by B* is been artificially set to a, independent of the actual value of A that may be observed. For a review of do-conditionals and distinction between passive (observational) and active (interventional) conditioning, we refer the reader to Refs. [58, 60–62]. The fundamental lemmas of mediation analysis constrain the possible values of interventional do-conditionals from knowledge of both the underlying causal structure and the distribution under passive observation. Every do-conditional can be expressed in terms of extending the original distribution to a particular interruption of the original graph.

When estimating the do-conditional $\{A^\#=a^\#\} \rightarrow B$, the interrupted graph $\mathcal{G}'$ is formed by replacing the $A \rightarrow B$ edge set in $\mathcal{G}$ with $A^\# \rightarrow B$, such that $\mathcal{G}'$ contains the additional variables $A^\#$. Whenever P_{ABC} is defined over the observed variables $\{A, B, C\}$, then the extended

distribution $P'_{ABC|A^\#}$ further pertains to the variables $A^\#$ as well.

The fundamental principle of mediation analysis is that every viable do-conditional constitutes some extended distribution or suitable marginal thereof. Accordingly, the potential range of a do-conditional is defined by variation over all *valid extensions* of the original distributions.

Note that the interrupted graph $\mathcal{G}'$ representing the intervention on $\mathcal{G}$ has a vertex set consisting of all nodes of $\mathcal{G}$—say, ABC—along with the auxiliary intervention variables $A^\#$. When the do-conditional pertains only to a *marginal* of the interrupted graph $\mathcal{G}'$, the do-conditional P' refers to a strict subset of the observed variables in $\mathcal{G}'$. Without loss of generality, therefore, we consider characterizing the set of do-conditionals of the form $P_{AB \textbf{ do } A^\# \rightarrow B}$ which are compatible with some observed distribution P_{ABC} relative to the intervention $A^\# \rightarrow B$ depicted in $\mathcal{G}'$.

Fundamental Lemma of Mediation Analysis:

$$P_{AB \textbf{ do } A^\# \rightarrow B} \in \mathsf{CompatibleDoConditionals}_{\mathcal{G}'}\big[P_{ABC}\big]$$
$$\text{iff} \quad \exists_{P'} : P'_{ABC|A^\#} \in \mathsf{ValidExtensions}_{\mathcal{G}'}\big[P_{ABC}\big]$$
$$\text{such that} \qquad\qquad\qquad\qquad\qquad\qquad (33)$$
$$P_{AB \textbf{ do } A^\# \rightarrow B}(ab \textbf{ do } a^\#) = \sum_c P'_{ABC|A^\#}(abc|a^\#).$$

Recall that $P'_{ABC|A^\#}$ is said to be valid extension of the original distribution P_{ABC} to $\mathcal{G}'$ only if it is compatible with the interrupted graph and recovers the original distribution under postselection. That is,

$$P'_{ABC|A^\#} \in \mathsf{ValidExtensions}_{\mathcal{G}'}\big[P_{ABC}\big]$$
$$\text{when} \quad P_{ABC}(abc) = P'_{ABC|A^\#}(abc|a) \qquad (34)$$
$$\text{and} \quad P'_{ABC|A^\#} \in \mathsf{CompatibleDistributions}\big[\mathcal{G}'\big].$$

which pretty much recapitulates Eq. (14), except here we distinguish between the nodes pertinent to the conditional (AB) and all other observed nodes (C).

For the purposes of this work, we emphasize that mediation analysis explicitly maps the problem of bounding causal effects to the problem of causal compatibility relative to an interrupted graph. In particular, mediation analysis implies that the quantum inflation technique can provide upper and lower bounds for the *quantum* ACE, i.e., can constrain causal effects in the presence of quantum confounding. For instance, quantum ACE can be lower (respectively, upper) bounded by minimizing (respectively, maximizing) $\text{ACE}[a_1, a_2, B]$ over extended distributions which are quantum compatible with the graph interruption corresponding to replacing the edge $A \rightarrow B$ with the edge $A^\# \rightarrow B$.

An important result in mediation analysis is that some do-conditionals are constrained to a single point value (called *identifiable*) regardless of the particular
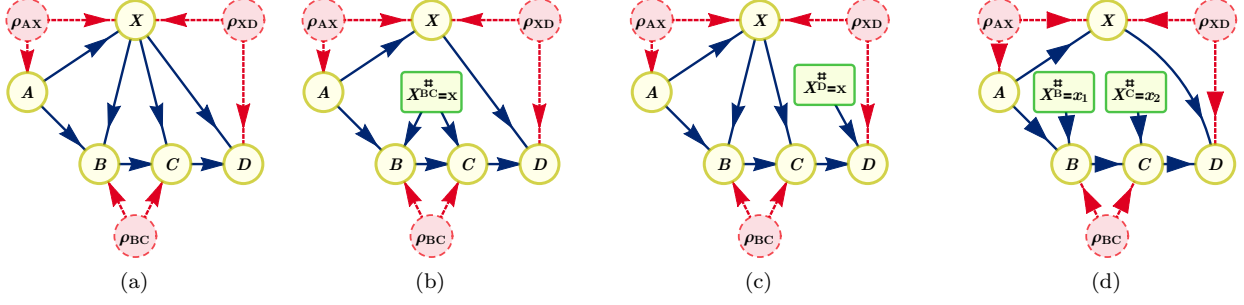
FIG. 17. A survey of hypothetical interventions on edges emanating from the node X in a pedagogically motivated example causal structure.
(a) The example causal structure without any interventions.
(b) The do-conditional $P(ABCDX \textbf{ do } \{X^\# = x^\#\} \rightarrow \{BC\})$ where both edges $X \rightarrow B$ and $X \rightarrow C$ are intervened upon with the same value of X being sent down both intervened edges. This do-conditional is identifiable and has a unique value computable from observational statistics alone, namely $P(ABCDX = x \textbf{ do } \{X^\# = x^\#\} \rightarrow \{BC\}) = P(ABCDX = x)P(BC|A, X = x^\#)/P(BC|A, X = x)$.
(c) The do-conditional $P(ABCDX \textbf{ do } \{X^\# = x^\#\} \rightarrow D)$ where the edge $X \rightarrow D$ is intervened upon. This do-conditional is nonidentifiable (as X and D share a latent parent) and, hence, is interesting to estimate using quantum inflation.
(d) The do-conditional $P(ABCDX \textbf{ do } \{X^\# = x_1^\#\} \rightarrow B, \{X^\# = x_2^\#\} \rightarrow C)$ where the pair of edges $X \rightarrow B$ and $X \rightarrow C$ are intervened upon with a different value of X being sent down the two intervened edges. This do-conditional is nonidentifiable (as B and C share a latent parent) and, hence, is interesting to estimate using quantum inflation.

observed statistics. The criteria for determining identifiable do-conditionals (and efficient algorithms for computing their values) is the subject of Refs. [58–62]. Nonidentifiable do-conditionals can typically be constrained only to lie within certain numerical spans [57, 63, 64].[9] Visual examples of some identifiable and nonidentifiable do-conditionals are provided in Fig. 17.

We note that the criteria for determining identifiability are valid independent of the (non)classicality of the latent variables. Consequently, quantum and classical mediation analysis can differ only in terms of the inequalities they imply for ranges of nonidentifiable do-conditionals. Reference [65] highlights such a quantum-classical gap regarding the nonidentifiable do-conditional of the instrumental scenario. Chaves *et al.* [65] explicitly construct a quantum causal model for the instrumental scenario where the quantum average causal effect is zero even though the resulting observational distribution would necessarily imply a strictly positive classical ACE.

To illustrate the utility of quantum inflation for mediation analysis, consider the quantum triangle scenario supplemented by a directed edge $A \rightarrow B$, illustrated in Fig. 18(a). The GHZ distribution

$$P_{\text{GHZ}}(a,b,c) := \begin{cases} \frac{1}{2} & a = b = c \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

is compatible with this graph. To achieve this correlation, however, we anticipate that B must *functionally* depend on A. However, $P_{B \textbf{ do } A^\# \rightarrow B}$ is not identifiable, as A and B share an unobserved common parent. We employ the quantum inflation technique to verify that the $\text{ACE}[A, B] = 1$ for the GHZ distribution without assuming classicality, thus teaching us that the quantum ACE is *not* less than the classical ACE in this example. This result means that, when reproducing the GHZ distribution in the quantum triangle scenario with signaling per Fig. 18(a), the functional dependence of B on the signal from A cannot be reduced relative to the strong dependence of B on A required to reproduce the GHZ distribution in the classical triangle scenario with signaling. This result is in stark contrast to the example in Ref. [65], where an instrumental scenario correlation requires reduced functional dependence between the observable variables when the latent node is quantum versus when it is classical.

Authors' note: Following the initial publication of the manuscript on the arXiv, the authors became aware of the related work in Ref. [33] which explores quantifying direct causal influence in the quantum Instrumental scenario.

X. CONCLUSIONS

We introduced the quantum inflation technique, a systematic method to discern whether an observable distribution is compatible with a causal explanation involving quantum degrees of freedom. The technique is
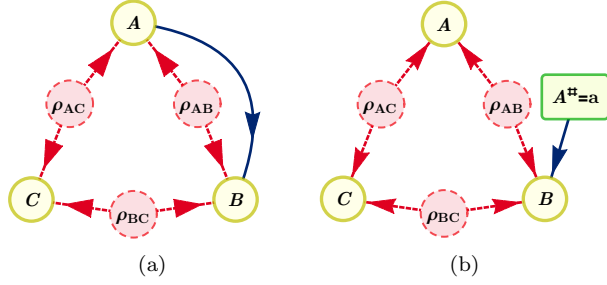
_____

[9] Fine-tuned instances of the observable distribution can constrain a nonidentifiable do-conditional to the point where it can take only some extremal fixed value.

FIG. 18. (a) The triangle scenario supplemented with a directed edge from $A \to B$. (b) The relevant interrupted graph used to both define and compute the ACE of A on B.

of general applicability, in that it can be employed to analyze correlations achievable by any quantum causal structure with, potentially, visible-to-visible, latent-to-visible, visible-to-latent, or latent-to-latent connections. Furthermore, we discussed how a slight modification allows also to study causal realizations in terms of classical latent variables.

We used quantum inflation to study correlations achievable in different quantum causal structures. First, we proved that the W and Mermin-GHZ distributions cannot be generated in the triangle scenario with quantum latent variables, bounded their noise resistance, and showed how quantum inflation is capable of recovering known results in the entanglement swapping scenario. Moreover, exploiting the dual formulation of semidefinite programs, we derived polynomial quantum causal witnesses for the triangle scenario with and without inputs. Second, we showed how quantum inflation can be employed as a tool for constrained polynomial optimization. We illustrated this fact in a variety of ways. We bounded the maximum values achievable by Mermin's and Svetlichny's inequalities in the quantum triangle scenario, finding significant gaps relative to the values achievable when one has access to arbitrary three-way entanglement. We also bounded the amount of information that an eavesdropper can obtain when two distant parties generate secret key using a quantum repeater. Finally, we illustrated the applicability of quantum inflation for mediation analysis in the presence of quantum common causes.

The implementation of quantum inflation synthesizes two different hierarchies: the one of inflations and, for each inflation, the NPO hierarchy used to determine whether a distribution admits such an inflation. While asymptotic convergence has been proven for the latter; that of the former is an open question. Nevertheless, we have identified situations in which tight results can be obtained at finite steps of the hierarchies. Moreover, recent results [66] show that a convergent NPO hierarchy can be defined for the analysis of quantum correlations in sequential Bell scenarios. Given that many nonexoge-

nous causal structures can be mapped to networks linking such sequential scenarios (though not all, as shown in Appendix B), Ref. [66]'s proof becomes an interesting alternative to quantum inflation for analyzing such special cases.

The number of copies of a given operator increases polynomially with the quantum inflation hierarchy level, where the exponent is given by the number of sources that feed the original operator. Moreover it is known [29] that, for a fixed inflation level, the complexity of each level in the corresponding NPO hierarchy, $\mathcal{S}_n$, is exponential in the level index n. For these reasons, in terms of practical applicability, standard tabletop computers are typically incapable of handling inflation and NPO hierarchy levels beyond 3. The optimal choice of which hierarchy to prioritize under scarce computational resources appears to vary depending on the particular problem at hand. We direct the reader to Ref. [43, Chap. 5] for a fuller discussion on the impacts of varying the different hierarchy level choices and an account of some intuitions in this matter. Despite this computational load, the SDPs associated to quantum inflation present a high degree of symmetry, and it is expected that exploiting these symmetries explicitly will reduce the computational barriers currently associated with high levels of the inflation hierarchy.

Quantum inflation can find an application in many fields. Clear first applications are generalizations of entanglement theory and quantum information protocols to networks [41, 42]. From a more general perspective, and due to the central role that causality has in science, we expect quantum inflation to become a fundamental tool for analyzing causality in any situation where a quantum behavior is presumed.

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

[2] S. Morgan and C. Winship, *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (Cambridge University Press, 2007).

[3] I. Shpitser and J. Pearl, "Complete identification methods for the causal hierarchy," J. Mach. Learn. Res. **9**, 1941 (2008).

[4] J. S. Bell, "On the problem of hidden variables in quantum mechanics," Rev. Mod. Phys. **38**, 447 (1966).

[5] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," Rev. Mod. Phys. **86**, 419 (2014).

[6] T. Fritz, "Beyond Bell's theorem: correlation scenarios," New J. Phys. **14**, 103001 (2012).

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

[8] C. J. Wood and R. W. Spekkens, "The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning," New J. Phys. **17**, 033002 (2015).

[9] R. Chaves, R. Kueng, J. Bohr Brask, and D. Gross, "Unifying framework for relaxations of the causal assumptions in Bell's theorem," Phys. Rev. Lett. **114**, 140403 (2015).

[10] E. Wolfe, D. Schmid, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, "Quantifying Bell: the Resource Theory of Nonclassicality of Common-Cause Boxes," Quantum **4**, 280 (2020).

[11] E. Wolfe, R. W. Spekkens, and T. Fritz, "The Inflation Technique for causal inference with latent variables," J. Causal Inference **7**, 20170020 (2019).

[12] M. Navascués and E. Wolfe, "The inflation technique completely solves the causal compatibility problem," J. Causal Inference **8**, 70 (2020).

[13] D. Rosset, N. Gisin, and E. Wolfe, "Universal bound on the cardinality of local hidden variables in networks," Quant. Info. & Comp. **18**, 910 (2018).

[14] W. Slofstra, "The set of quantum correlations is not closed," For. Math., Pi **7**, e1 (2019).

[15] Z. Ji, A. Natarajan, T. Vidick, J. Wright, and H. Yuen, "MIP*=RE," arXiv:2001.04383 (2020).

[16] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, "Noncommuting mixed states cannot be broadcast," Phys. Rev. Lett. **76**, 2818 (1996).

[17] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, "Cloning and broadcasting in generic probabilistic theories," quant-ph/0611295 (2006).

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

[19] A. Pozas-Kerstjens, R. Rabelo, L. Rudnicki, R. Chaves, D. Cavalcanti, M. Navascués, and A. Acín, "Bounding the sets of classical and quantum correlations in networks," Phys. Rev. Lett. **123**, 140503 (2019).

[20] J. Åberg, R. Nery, C. Duarte, and R. Chaves, "Semidefinite tests for quantum network topologies," Phys. Rev. Lett. **125**, 110505 (2020).

[21] R. O. Ness, K. Sachs, and O. Vitek, "From correlation to causality: Statistical approaches to learning regulatory relationships in large-scale biomolecular investigations," J. Proteome Res. **15**, 683 (2016).

[22] A. J. Hartemink, D. K. Gifford, T. S. Jaakkola, and R. A. Young, "Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks," in *Biocomputing* (World Scientific, 2001) pp. 422–433.

[23] N. Lambert, Y.-N. Chen, Y.-C. Cheng, C.-M. Li, G.-Y. Chen, and F. Nori, "Quantum biology," Nat. Phys. **9**, 10 (2013).

[24] J. Cao, R. J. Cogdell, D. F. Coker, H.-G. Duan, J. Hauer, U. Kleinekathöfer, T. L. C. Jansen, T. Mančal, R. J. D. Miller, J. P. Ogilvie, V. I. Prokhorenko, T. Renger, H.-S. Tan, R. Tempelaar, M. Thorwart, E. Thyrhaug, S. Westenhoff, and D. Zigmantas, "Quantum biology revisited," Sci. Adv. **6** (2020).

[25] C. Branciard, N. Gisin, and S. Pironio, "Characterizing the nonlocal correlations created via entanglement swapping," Phys. Rev. Lett. **104**, 170401 (2010).

[26] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, "Bilocal versus nonbilocal correlations in entanglement-swapping experiments," Phys. Rev. A **85**, 032119 (2012).

[27] S. Pironio, M. Navascués, and A. Acín, "Convergent relaxations of polynomial optimization problems with non-commuting variables," SIAM J. Optim. **20**, 2157 (2010).

[28] M. Navascués, S. Pironio, and A. Acín, "Bounding the set of quantum correlations," Phys. Rev. Lett. **98**, 010401 (2007).

[29] M. Navascués, S. Pironio, and A. Acín, "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations," New J. Phys. **10**, 073013 (2008).

[30] L. Vandenberghe and S. Boyd, "Semidefinite programming," SIAM Rev. **38**, 49 (1996).

[31] T. Van Himbeeck, J. Bohr Brask, S. Pironio, R. Ramanathan, A. B. Sainz, and E. Wolfe, "Quantum violations in the Instrumental scenario and their relations to the Bell scenario," Quantum **3**, 186 (2019).

[32] I. Agresti, G. Carvacho, D. Poderini, L. Aolita, R. Chaves, and F. Sciarrino, "Experimental connection between the Instrumental and Bell inequalities," Proc. MDPI AG **12**, 27 (2019).

[33] M. Gachechiladze, N. Miklin, and R. Chaves, "Quantifying causal influences in the presence of a quantum common cause," Phys. Rev. Lett. **125**, 230401 (2020).

[34] T. S. Richardson and J. M. Robins, *Single World Intervention Graphs (SWIGs) : A Unification of the Counterfactual and Graphical Approaches to Causality* (Now Publishers Inc, 2013) Working Paper #128, Center for Stat. & Soc. Sci., U. Washington.

[35] J. Barrett, R. Lorenz, and O. Oreshkov, "Quantum causal models," arXiv:1906.10726 (2019).

[36] R. J. Evans, "Graphical methods for inequality constraints in marginalized DAGs," in *IEEE International Workshop on Machine Learning for Signal Processing* (2012).

[37] R. J. Evans, "Margins of discrete Bayesian networks," Annals Stat. **46**, 2623 (2018).

[38] B. Toner, "Monogamy of non-local quantum correlations," Proc. Roy. Soc. A **465**, 59 (2009).

[39] M. P. Seevinck, "Monogamy of correlations versus monogamy of entanglement," Quant. Info. Proc. **9**, 273 (2010).

[40] F. Baccari, D. Cavalcanti, P. Wittek, and A. Acín, "Efficient device-independent entanglement detection for multipartite systems," Phys. Rev. X **7**, 021042 (2017).

[41] D. Schmid, T. C. Fraser, R. Kunjwal, A. B. Sainz, E. Wolfe, and R. W. Spekkens, "Understanding the interplay of entanglement and nonlocality: motivating and developing a new branch of entanglement theory," arXiv:2004.09194 (2020).

[42] M. Navascués, E. Wolfe, D. Rosset, and A. Pozas-Kerstjens, "Genuine Network Multipartite Entanglement," Phys. Rev. Lett. **125**, 240505 (2020).

[43] A. Pozas-Kerstjens, *Quantum information outside quantum information*, Ph.D. thesis, Universitat Politècnica de Catalunya (2019).

[44] S. Pironio, J.-D. Bancal, and V. Scarani, "Extremal correlations of the tripartite no-signaling polytope," J. Phys. A **44**, 065303 (2011).

[45] N. D. Mermin, "Quantum mysteries revisited," Amer. J. Phys. **58**, 731 (1990).

[46] G. Svetlichny, "Distinguishing three-body from two-body nonseparability by a Bell-type inequality," Phys. Rev. D **35**, 3066 (1987).

[47] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, "Nonlocal correlations as an information-theoretic resource," Phys. Rev. A **71**, 022101 (2005).

[48] C. M. Lee and M. J. Hoban, "Towards device-independent information processing on general quantum networks," Phys. Rev. Lett. **120**, 020504 (2018).

[49] L. P. Thinh, J.-D. Bancal, and E. Martín-Martínez, "Certified randomness from a two-level system in a relativistic quantum field," Phys. Rev. A **94**, 022321 (2016).

[50] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times* (American Mathematical Society, 2017).

[51] Y.-T. Huang and W.-C. Pan, "Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators," Biometrics **72**, 402 (2016).

[52] J. Hutton, T. Fatima, T. J. Major, R. Topless, L. K. Stamp, T. R. Merriman, and N. Dalbeth, "Mediation analysis to understand genetic relationships between habitual coffee intake and gout," Arthritis Res. Ther. **20** (2018).

[53] M. B. Sohn and H. Li, "Compositional mediation analysis for microbiome studies," Ann. Appl. Stat. **13**, 661 (2019).

[54] E. G. Cavalcanti and R. Lal, "On modifications of Reichenbach's principle of common cause in light of Bell's theorem," J. Phys. A **47**, 424018 (2014).

[55] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, "Quantum common causes and quantum causal models," Phys. Rev. X **7**, 031021 (2017).

[56] D. Janzing, D. Balduzzi, M. Grosse-Wentrup, and B. Schölkopf, "Quantifying causal influences," Ann. Statist. **41**, 2324 (2013).

[57] C. H. Miles, P. Kanki, S. Meloni, and E. J. Tchetgen Tchetgen, "On partial identification of the pure direct effect," arXiv:1509.01652 (2015).

[58] D. Malinsky, I. Shpitser, and T. Richardson, "A potential outcomes calculus for identifying conditional path-specific effects," arXiv:1903.03662 (2019).

[59] R. Bhattacharya, R. Nabi, and I. Shpitser, "Semiparametric inference for causal effects in graphical models with hidden variables," arXiv:2003.12659 (2020).

[60] I. Shpitser and E. J. Tchetgen Tchetgen, "Causal inference with a graphical hierarchy of interventions," Ann. Statist. **44**, 2433 (2016).

[61] I. Shpitser and E. Sherman, "Identification of personalized effects associated with causal pathways," in *Uncertainty in Artificial Intelligence*, Vol. 2018 (2018).

[62] M. J. Stensrud, J. G. Young, V. Didelez, J. M. Robins, and M. A. Hernán, "Separable effects for causal inference in the presence of competing events," arXiv:1901.09472 (2019).

[63] Z. Cai, M. Kuroki, J. Pearl, and J. Tian, "Bounds on direct effects in the presence of confounded intermediate variables," Biometrics **64**, 695 (2008).

[64] C. Kang and J. Tian, "Inequality constraints in causal models with hidden variables," in *Uncertainty in Artificial Intelligence* (2006).

[65] R. Chaves, G. Carvacho, I. Agresti, V. Di Giulio, L. Aolita, S. Giacomini, and F. Sciarrino, "Quantum violation of an instrumental test," Nat. Phys. **14**, 291 (2018).

[66] J. Bowles, F. Baccari, and A. Salavrakos, "Bounding sets of sequential quantum correlations and device-independent randomness certification," Quantum **4**, 344 (2020).

[67] T. Moroder, J.-D. Bancal, Y.-C. Liang, M. Hofmann, and O. Gühne, "Device-independent entanglement quantification and related applications," Phys. Rev. Lett. **111**, 030501 (2013).

Appendix A: Non-commutative polynomial optimization

A generic NPO problem can be cast as

$$p^* = \min_{(\mathcal{H}, X, \rho)} \langle p(X) \rangle_\rho$$
$$\text{such that} \quad q_i(X) \succeq 0 \quad \forall i = 1...m_q,$$

that is, finding a Hilbert space $\mathcal{H}$, a positive-semidefinite operator $\rho : \mathcal{H} \to \mathcal{H}$ with trace one, and a list of bounded operators $X = (X_1, \ldots, X_n)$ in $\mathcal{H}$ (where $X_i X_j \neq X_j X_i$ in general) that minimize the expectation value $\langle p(X) \rangle_\rho = \text{Tr}[\rho \cdot p(X)]$ of the polynomial operator $p(X)$ given some polynomial constraints $q_i(X) \succeq 0$, where $q_i(X) \succeq 0$ means that the operator $q_i(X)$ should be positive semidefinite. One can also add to the optimization statistical constraints of the form $\langle r_j(X) \rangle_\rho \geq 0$, for $j = 1,...,m_r$. Note that the NPO formalism can also accommodate equality constraints of the form $q(X) = 0$ or $\langle r_j(X) \rangle_\rho = 0$, since they are equivalent to the constraints $q(X), -q(X) \succeq 0$ and $\langle r_j(X) \rangle_\rho, \langle -r_j(X) \rangle_\rho \geq 0$, respectively.

The procedure for solving these problems is described in Ref. [27], and uses a hierarchy of relaxations where each of the hierarchy's steps is an SDP instance. The solutions to these problems form a monotonically increasing sequence of lower bounds on the global minimum p^*:

$$p_1 \leq p_2 \leq \cdots \leq p_\infty \leq p^*.$$

If the constraints $\{q_i(X) \succeq 0\}_i$ imply (explicitly or implicitly; see Ref. [27]) that all noncommuting variables $X_1,...,X_n$ are bounded (and they do so in all the NPO problems considered in this work), then the sequence of lower bounds is asymptotically convergent. That is, $p_\infty = p^*$.

In our case, given some observed correlations, we deal with a feasibility problem about the existence of a quantum state and measurements subject to polynomial operator and statistical constraints arising from the causal networks and the observed correlations. This feasibility problem can be mapped into an optimization problem in different ways. For instance, while not being the most practical procedure, the easiest way of doing it is by considering a constant polynomial $p(X) = 1$ as the function to be optimized. This problem has a solution equal to 1 provided that the polynomial and statistical constraints are simultaneously satisfiable. Any step in the hierarchy is therefore a necessary SDP test to be satisfied for the causal model to be compatible with the observed correlations. Note that the same formalism can be used to optimize polynomials of the operators, such as, for instance, Bell-like inequalities, over quantum correlations compatible with a given causal structure [41].

Appendix B: The treatment of latent variables with latent parents

The study of network form quantum causal structures with quantum inflation is, essentially, solving the problem of *classifying entanglement networks* in a device-independent manner [41, 42]. Generalizing to latent nonexogenous scenarios can, thus, be seen as a further generalization of the problem of classification of entanglement generation in general causal structures. Consider, for instance, four parties and two three-way sources of entanglement. One can imagine two distinct ways to generate the final four-partite quantum state from the initial pair of sources. In the first case, depicted in Fig. 19(a), each of the four parties applies arbitrary (local) quantum channels to the states they receive from the sources. In the second case, depicted in Fig. 19(b), a nonlocal quantum channel spanning Bob and Charlie's systems is applied prior to the final assignment of Hilbert spaces to parties.

One might naively think that this further generality does not add much—after all, Bob and Charlie can share arbitrary entanglement in either scenario—but that is not the case. Only the causal structure depicted in Fig. 19(b) can give rise to the state

$$\rho_{19(b)} := \frac{|AB00\rangle\langle AB00| + |A0C1\rangle\langle A0C1|}{2}, \quad \text{(B1)}$$

$$\text{where} \quad |AB00\rangle = \frac{|0000\rangle + |1100\rangle}{\sqrt{2}}$$

$$\text{and} \quad |A0C1\rangle = \frac{|0001\rangle + |1011\rangle}{\sqrt{2}}.$$

The state $\rho_{19(b)}$ describes maximal $A-B$ entanglement and $C = |0\rangle$ when $D = |0\rangle$ mixed with maximal $A-C$ entanglement and $B = |0\rangle$ when $D = |1\rangle$. Here, measuring ρ'_D in the computational basis will steer the ρ'_{ABC} toward either maximal entanglement between A and B or maximal entanglement between A and C. This steer-
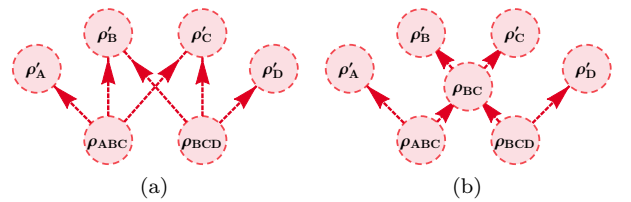

FIG. 19. Inequivalent state generating procedures. (a) depicts a two-layer state generating procedure, whereas (b) allows for the application of an arbitrary Bob-Charlie channel, generating the intermediate state ρ_{BC}, prior to distributing Hilbert spaces to the four parties. One can confirm that (a) generates a strict subset of the states realizable in (b), and quantum inflation can distinguish these two scenarios in a device-independent manner.

ing is possible only if the $B-C$ entanglement structure is determined by a node which shares causal ancestry with A as well with D. There is no node in Fig. 19(a) that satisfies these requirements, which implies that the state $\rho_{19(b)}$ cannot be generated by such a causal structure. On the other hand, ρ_{BC} in Fig. 19(b) is a node which does satisfy the requirements, and thus, both structures are not equivalent.

Alternatively, this result can also be understood through monogamy of nonlocality: Access to $\rho_{19(b)}$ enables maximally violating the CHSH inequality by either players A and B or by players A and C, where nonlocal parties are flagged by the value of a measurement on ρ'_D. This result precisely replicates the example in Fig. 8, except that the switch setting S in Fig. 8(a) is replaced by measurement on ρ'_D in Fig. 19(b) in the computational basis.

An explicit realization of $\rho_{19(b)}$ in Fig. 19(b) is as follows. Let ρ_{ABC} and ρ_{BCD} be, respectively, projectors onto the pure states $(|00\rangle + |11\rangle)_{AB_1}$ and $(|00\rangle+|11\rangle)_{B_2D}\otimes|0\rangle_C$. Then, perform a controlled swap on qubits B_2 and C depending on B_1's computational basis value, and trace out the subsystem B_1.

In Sec. V B, we prescribe encoding all quantum channels in the causal structure as unspecified unitaries that can be incorporated into the generating monomials. Though Fig. 8 there illustrates classically controlled quantum channels, uncontrolled quantum channels can equally well be incorporated into the elementary monomials. Consider applying quantum inflation to Fig. 19. Inflation imbues every unitary with a pair of indices corresponding to which copies of the root quantum states the given unitary pertains to. For the unrestricted case in Fig. 19(a), we find that misaligning the copy index pertaining to one of the root quantum states leads to a pair of operators which need not commute, namely,

$$U_{BC}^{i,j}\cdot U_{BC}^{i,k} \neq U_{BC}^{i,k}\cdot U_{BC}^{i,j} \quad \text{for} \quad j\neq k. \qquad (B2)$$

On the other hand, the local unitaries relevant to Fig. 19(b) *do* commute even in the presence of misaligned copy indices, namely

$$U_B^{i,j}\cdot U_C^{i,k} = U_C^{i,k}\cdot U_B^{i,j} \quad \text{for} \quad j\neq k. \qquad (B3)$$

One can readily see, therefore, how inflation levels $n\geq 2$ give rise to SDPs with *more constrained* moment matrices when quantum inflation is applied to Fig. 19(a) instead of Fig. 19(b). This example is a device-independent distinction afforded by the prescription of Sec. V B which cannot be recovered using the sequential measurements framework of Ref. [66].

Appendix C: Monomial sets for NPO problems

The hierarchies of semidefinite programs that bound the solutions of NPO problems can be described in terms of sets of products of the noncommutative operators in the problem. In this work, we use two different hierarchies that are both asymptotically complete. The levels in the first hierarchy are known as **NPA levels** [28, 29]. The NPA level n, $\mathcal{S}_n$, is associated to the set of all products of operators in the problem, of length no larger than n. For example, the set $\mathcal{S}_2$ associated to the inflations of the quantum triangle scenario discussed in the main text is

$$\mathcal{S}_2 = \{\mathbb{1}\}\cup\{H_p^{i,j}\}_{p,i,j}\cup\{H_p^{i,j}(H')_q^{k,l}\}_{p,q,i,j,k,l},$$

where $H_p, H'_q = E, F, G$. On the other hand, some problems may achieve tighter results at lower levels if one instead considers **local levels** [67]. The local level n, $\mathcal{L}_n$, is built from the products of operators that contain at most n operators of a same party. For instance, in the quantum triangle scenario, the set $\mathcal{L}_1$ associated to its inflations is

$$\begin{aligned}\mathcal{L}_1 = &\{\mathbb{1}\}\cup\{H_p^{i,j}\}_{p,i,j}\cup\{H_p^{i,j}(H')_q^{k,l}\}_{p,q,i,j,k,l}\\&\cup\{E_a^{i,j}F_b^{k,l}G_c^{m,n}\}_{a,b,c,i,j,k,l,m,n},\end{aligned}$$

where $H_p \neq H'_q$. While both hierarchies are asymptotically complete, they satisfy the relation $\mathcal{S}_n\subset\mathcal{L}_n\not\subset\mathcal{S}_{n+1}$ (in fact, the smallest set of the NPA hierarchy that contains $\mathcal{L}_n$ is $\mathcal{S}_{pn}$, where p is the number of parties), and, thus, the use of finite levels of one or the other hierarchy may be more or less convenient depending on the specific problem to solve.

Appendix D: Quantum key distribution in the tripartite-line scenario

The distribution observed by A, B, and C that allows them to guarantee that a quantum eavesdropper E cannot guess the outputs of the extremes A and C with a confidence higher than that shown in Fig. 15 is generated in a standard quantum repeater scenario, where the sources distribute maximally entangled bipartite quantum states and the repeater party B performs a Bell state measurement in the standard basis. Following Ref. [48], we choose the sources to distribute singlets $|\psi^-\rangle = (|01\rangle-|10\rangle)/\sqrt{2}$ and the parties A and C to perform the measurements $(\sigma_z+\sigma_x)/\sqrt{2}$ for $x=z=0$ and $(\sigma_z-\sigma_x)/\sqrt{2}$ for $x=z=1$. In such a case, the distribution observed by A, B and C is $P_{\text{ABC}|\text{XZ}}(a,b_1b_0,c|x,z) :=$

$$\frac{2+(-1)^{a+c}\left[(-1)^{b_0}+(-1)^{b_1+x+z}\right]}{32}.$$

As shown in Sec. VIII, if A, B, and C observe this distribution, they are guaranteed that an eavesdropper E cannot predict the outputs of A and C better than randomly guessing them.

If the sources are not perfect and instead distribute noisy singlets, $\rho = v|\psi^-\rangle\langle\psi^-| + (1-v)\mathbb{1}/4$, then the

distribution observed by A, B, and C takes the form $P_{\text{ABC}|\text{XZ},v}(a,b_1b_0,c|x,z):=$

$$v^2 P_{\text{ABC}|\text{XZ}}(a,b_1b_0,c|x,z)+\frac{1-v^2}{16}$$
$$=\frac{2+v^2(-1)^{a+c}\big[(-1)^{b_0}+(-1)^{b_1+x+z}\big]}{32}. \qquad \text{(D1)}$$

If A, B and C observe this distribution, they can be sure only that an eavesdropper cannot guess the outputs of A and C with a probability higher than that depicted in Fig. 15.