Quantum Inflation: A General Approach to Quantum Causal Compatibility

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

[1] *Perimeter Institute for Theoretical Physics, N2L 2Y5 Waterloo, Canada*
[2] *ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*
[3] *Princeton University, Princeton, NJ USA 08544*
[4] *Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, N2L 2Y5*
[5] *ICREA, Passeig Lluis Companys 23, 08010 Barcelona, Spain*
[6] *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 an application in many fields: from the characterization of correlations in quantum networks to the study of quantum effects in thermodynamic and biological processes.

I. INTRODUCTION

Causality is an ubiquitous concept in science. 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 explanation for given statistics, a task known as **causal discovery**. Advances in causal understanding also enables 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 formalise 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 non-trivial inequalities on the observed probabilities. A general method to tackle the causal compatibility problem, known as the **inflation technique**, was obtained in [11]. It consists of a hierarchy of conditions, organised according to their computational cost, that are necessary for a Bayesian network to be 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 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 [13]. Yet, this does not preclude the existence of a method similar to inflation to tackle the question.

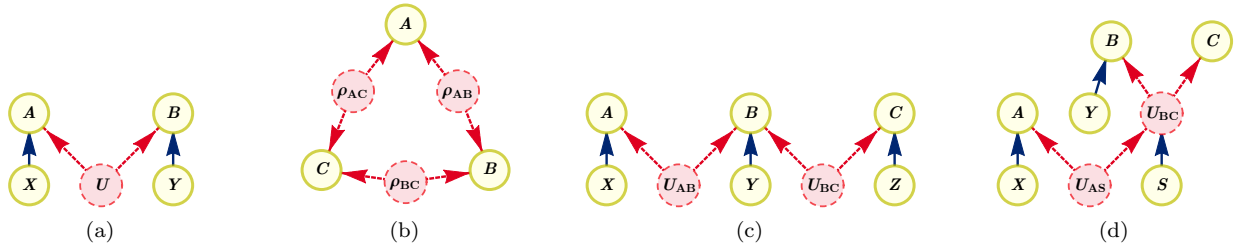


FIG. 1. DAG representation of different causal scenarios. The red, dashed circles are latent variables, and the yellow, single-lined circles denote visible variables. (a) The Bell scenario is one of the simplest causal structures exhibiting a classical-quantum gap, that is, where there exist distributions that can be realized with quantum latent variables (denoted as $U{=}\rho$), but not with classical ones (denoted as $U{=}\lambda$). (b) The triangle scenario also presents a classical-quantum gap. Alternatively, the triangle scenario can also be defined with one visible variable (called *measurement choice*) influencing each of A, B and C. (c) The tripartite-line causal scenario, where two causally independent parties A and C share each some resource with a central party B. (d) Arbitrary 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 is given in Section IV.

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 [14, 15]. Other causal analysis techniques which are fundamentally quantum have been proposed. Notably among these is the quantum entropy vector approach of [16], which is applicable to all causal structures but uses only those constraints on entropies imposed by the causal structure, or the scalar extension of [17], which imposes stronger constraints but cannot be applied to causal structures in which all observed nodes are causally connected, such as the triangle scenario of Fig. 1(b).

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 analogue of the classical inflation technique which avoids the latter's reliance on information broadcast. We first explain in Sec. II how quantum inflation works by means of an example, and provide the details of the construction in Sec. III. Then, we describe how to address quantum causal compatibility in complex scenarios in Sec. IV. Also, quantum inflation can serve as a tool for assessing causal compatiblity with classical models. We detail this construction in Sec. V. Finally, in Sec. VI we apply quantum inflation to characterize correlations achievable in various tripartite quantum causal networks, and after that we conclude in Sec. VII.

II. QUANTUM INFLATION BY EXAMPLE

Consider the quantum causal network depicted in Fig. 2(a), whereby three random variables a, b, c are generated by conducting bipartite measurements over the ends of three bipartite quantum states ρ_{AB}, ρ_{BC}, ρ_{AC}. We are handed the distribution $P_{\text{obs}}(a,b,c)$ of observed variables and asked if it is compatible with this model. How to proceed?

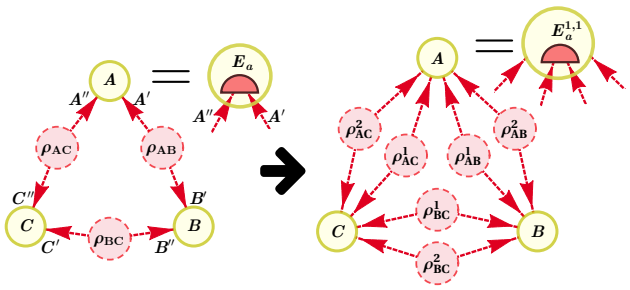
Suppose that there existed indeed bipartite states $\rho_{AB},\rho_{BC},\rho_{AC}$ of systems $A'B',B''C',A''C''$, and commuting measurement operators E_a, F_b, G_c, acting on systems $A'A'',B'B'',C'C''$, respectively, which were able to reproduce the correlations $P_{\text{obs}}(a,b,c)$. Now imagine how the scenario would change if n independent copies $\rho^i_{AB},\rho^i_{BC},\rho^i_{AC}$, $i=1,\ldots,n$ of each of the original states were distributed instead, as depicted in Fig. 2(b). Call ρ the overall quantum state before any measurement is carried out. For any $i,j=1,\ldots,n$ we can, in principle, implement measurement $\{E_a\}_a$ on the i^{th} copy of ρ_{AC} and the j^{th} copy of ρ_{AB}: we denote by $\{E^{i,j}_a\}_a$ the corresponding measurement operators. Similarly, call $\{F^{i,j}_b\}_b$, $(\{G^{i,j}_c\}_c)$ the measurement $\{F_b\}_b$ $(\{G_c\}_c)$ over the states ρ^i_{AB},ρ^j_{BC} $(\rho^i_{BC},\rho^j_{AC})$.

The newly defined operators and their averages under state ρ satisfy non-trivial relations. For example, for $H=E,F,G$ and $i\neq k, j\neq l$ the operators $H^{i,j}_a$ and $H^{k,l}_b$ act on different Hilbert spaces, and hence $\left[H^{i,j}_a,H^{k,l}_b\right]=0$. Similarly, expressions such as $\left\langle E^{1,1}_a E^{1,2}_{a'} F^{2,2}_b\right\rangle_\rho$ and $\left\langle E^{1,2}_a E^{1,1}_{a'} F^{1,2}_b\right\rangle_\rho$ can be shown identical, since one can arrive at the second one from the first one just by exchanging ρ^1_{AB} with ρ^2_{AB}. More generally, for any function $Q(\{E^{i,j}_a,F^{k,l}_b,G^{m,n}_c\})$ of the measurement operators and any three permutations π,π',π'' of the indices $1,\ldots,n$, one should have

$$\begin{aligned}&\left\langle Q(\{E^{i,j}_a,F^{k,l}_b,G^{m,n}_c\})\right\rangle_\rho\\ &\quad=\left\langle Q(\{E^{\pi(i),\pi'(j)}_a,F^{\pi'(k),\pi''(l)}_b,G^{\pi''(m),\pi(n)}_c\})\right\rangle_\rho.\end{aligned}\tag{1}$$

Finally, note that, if we conducted 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 $a^1,\ldots,a^n,b^1,\ldots,b^n,c^1,\ldots,c^n$ would be distributed according to

$$\left\langle\prod_{i=1}^n E^{i,i}_{a^i}F^{i,i}_{b^i}G^{i,i}_{c^i}\right\rangle_\rho=\prod_{i=1}^n P_{\text{obs}}(a^i,b^i,c^i).\tag{2}$$



(a) (b)

FIG. 2. Quantum inflation in the triangle scenario. (a) In the original scenario, by probing systems A',A'' with the quantum measurement $\{E_a\}$, a value a for the random variable A is generated. The values b,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^1_{AB} and ρ^1_{AC}, and the measurements with other superindices are defined in an analogous way.

If the original distribution $P_{\text{obs}}(a,b,c)$ is compatible with the network in Fig. 2(a), then there should exist a Hilbert space $\mathcal{H}$, a state $\rho : \mathcal{H} \to \mathcal{H}$ and operators $\{E_a^{i,j}\}_{i,j,a}$, $\{F_b^{k,l}\}_{k,l,b}$, $\{G_c^{m,n}\}_{m,n,c}$ satisfying the above relations. If such is the case, we say that $P_{\text{obs}}(a,b,c)$ 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 $P_{\text{obs}}(a,b,c)$.

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 non-commutative polynomial optimization (NPO) theory [18]. Originally developed to characterize quantum nonlocality [19, 20], 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 is achieved by means of a hierarchy of semidefinite programming tests [21], see also Appendix A. In our particular case, we are dealing with a feasibility problem. The polynomial operator constraints we wish to enforce on $E_a^{i,j}$,$F_b^{k,l}$,$G_c^{m,n}$ are that they define complete families of projectors, which commute when acting on different quantum systems. The statistical constraints are given by Eqs. (1-2). If for some n we were able to certify, via NPO theory, that $P_{\text{obs}}(a,b,c)$ does not admit an n^{th}-order quantum inflation, then we would have proven that $P_{\text{obs}}(a,b,c)$ does not admit a realization in the quantum network of Fig. 2(a).

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, such as the triangle scenario. To test the incompatibility of a distribution P_{obs}, we 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 and work out how operator averages relate to P_{obs} and to each other. Finally, we would use NPO theory to disprove the existence of a state and operators satisfying the inferred constraints. In Section IV 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.

III. 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 DAG's 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 set of all classical observed parents of such a variable can be understood as the *measurement setting* used to produce this outcome.

The essential premise of quantum inflation is to ask what would happen if multiple copies of the original (unspecified) quantum states were simultaneously available to each party. In this gedankenexperiment 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 non-trivially. Let us therefore denote measurement operators by

$$\hat{O}^{\boldsymbol{s}|k}_{i|m} \equiv \hat{O}\big(\text{Outcome variable}{=}k,\text{Spaces}{=}\boldsymbol{s},\ \text{Setting}{=}m,\text{Outcome}{=}i\big),$$

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. i, the measurement *setting* being used,
4. m, the *outcome* associated with the operator.

For example, using an $n=2$ quantum inflation of the triangle scenario (see Fig. 2) one would find 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 (A''_i) denotes the factor A of the Hilbert space where ρ^i_{AB} (ρ^i_{AC}) acts.

These operators will be regarded as the non-commuting variables of an NPO problem where the polynomial constraints are derived according to rules pertaining to the op-

erators' projective nature and as well as a number of commutation rules. The statistical constraints are then imposed from symmetry under permutations of the state indices and enforcing consistency with the observed probabilities.

Projection rules

For fixed $\boldsymbol{s},k,m$, the non-commuting variables $\{\hat{O}^{\boldsymbol{s}|k}_{i|m}\}_i$ 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

$$\hat{O}^{\boldsymbol{s}|k}_{i|m}=(\hat{O}^{\boldsymbol{s}|k}_{i|m})^\dagger, \hat{O}^{\boldsymbol{s}|k}_{i|m}\hat{O}^{\boldsymbol{s}|k}_{i'|m}=\delta_{ii'}\hat{O}^{\boldsymbol{s}|k}_{i|m}, \forall \boldsymbol{s},k,i,i',m \tag{3a}$$

$$\sum\nolimits_i \hat{O}^{\boldsymbol{s}|k}_{i|m}=\mathbb{1}, \text{ for all } \boldsymbol{s},k,m. \tag{3b}$$

These relations imply, in turn, that each of the non-commuting variables is a bounded operator. Hence, by [18], the hierarchy of SDP programs provided by NPO is complete, i.e., if the said distribution does not admit an n^{th}-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,

$$\left[\hat{O}^{\boldsymbol{s}_1|k_1}_{i_1|m_1}, \hat{O}^{\boldsymbol{s}_2|k_2}_{i_2|m_2}\right]=0 \qquad \text{if } \boldsymbol{s}_1\cap\boldsymbol{s}_2=\emptyset. \tag{4}$$

Later on we consider modifying these commutation rules so as to construct an alternative SDP for constraining the correlations of *classical* causal structures.

Symmetry under permutations of the indices

The critical ingredient that relates the inflated network structure to the original network is that 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

$$\begin{aligned}&\langle \hat{O}^{\boldsymbol{s}_1|k_1}_{i_1|m_1}\cdot\hat{O}^{\boldsymbol{s}_2|k_2}_{i_2|m_2}\cdot\ldots\cdot\hat{O}^{\boldsymbol{s}_n|k_n}_{i_n|m_n}\rangle_\rho \\ &\quad=\left\langle \hat{O}^{\pi(\boldsymbol{s}_1)|k_1}_{i_1|m_1}\cdot\hat{O}^{\pi(\boldsymbol{s}_2)|k_2}_{i_2|m_2}\cdot\ldots\cdot\hat{O}^{\pi(\boldsymbol{s}_n)|k_n}_{i_n|m_n}\right\rangle_\rho.\end{aligned} \tag{5}$$

An example of such statistical constraints imposed in the triangle scenario was given in Eq. (1). Another example, for an inflation level n=3, is the following:

$$\left\langle E_0^{\{A'_1,A''_1\}}E_1^{\{A'_2,A''_2\}}F_0^{\{B''_3,B'_1\}}G_0^{\{C''_1,C'_2\}}\right\rangle_\rho$$

$$\overset{\text{apply } \rho^1_{AB}\leftrightarrow\rho^2_{AB}}{=}\left\langle E_0^{\{A'_2,A''_1\}}E_1^{\{A'_1,A''_2\}}F_0^{\{B''_3,B'_2\}}G_0^{\{C''_1,C'_2\}}\right\rangle_\rho \tag{6a}$$

$$\overset{\text{apply } \rho^1_{AB}\leftrightarrow\rho^3_{AB}}{=}\left\langle E_0^{\{A'_2,A''_1\}}E_1^{\{A'_3,A''_2\}}F_0^{\{B''_3,B'_2\}}G_0^{\{C''_1,C'_2\}}\right\rangle_\rho \tag{6b}$$

$$\overset{\text{apply } \rho^1_{BC}\leftrightarrow\rho^3_{BC}}{=}\left\langle E_0^{\{A'_2,A''_1\}}E_1^{\{A'_3,A''_2\}}F_0^{\{B''_1,B'_2\}}G_0^{\{C''_1,C'_2\}}\right\rangle_\rho, \tag{6c}$$

where, for readability, we identified $\hat{O}^{\boldsymbol{s}|A}_{a|\emptyset}$ ($\hat{O}^{\boldsymbol{s}|B}_{b|\emptyset}$) [$\hat{O}^{\boldsymbol{s}|C}_{c|\emptyset}$] with $E^{\boldsymbol{s}}_a$ ($F^{\boldsymbol{s}}_b$) [$G^{\boldsymbol{s}}_c$].

Consistency with the observed probabilities

Finally, as described by Eq. (2) in Section II, certain product averages can be related to products of the probabilities of the distribution P_{obs} that one wishes to test. More specifically, let n be the order of the considered inflation, and denote by $\vec{j}_k$ the set of Hilbert spaces where variable k acts, with the copy labels of all Hilbert spaces equal to $j\in\{1,\ldots,n\}$ (e.g.: $\vec{1}_A=\{A'_1,A''_1\}$, $\vec{2}_A=\{A'_2,A''_2\}$ in Fig. 2). Then, for any set of indices $\{i_{j,k},m_{j,k}\}_{j,k}$, we have that

$$\left\langle \prod_{j=1}^{n}\prod_{k}\hat{O}^{\vec{j}_k|k}_{i_{j,k}|m_{j,k}}\right\rangle_\rho=\prod_{j=1}^{n}P_{\text{obs}}\left(\bigcap_k(i_{j,k}|m_{j,k},k)\right), \tag{7}$$

where $(i|m,k)$ denotes the event of probing k with setting m and obtaining the result i.

IV. ARBITRARY CAUSAL SCENARIOS

In the previous section we provided a systematic method to characterize the correlations achievable in a subset of quantum causal networks, namely two-layer DAG's where no node has both parents and children. In this section we extend those ideas to characterize correlations in arbitrary causal structures.

A. Classical Exogenous Variables

The first important case that must be dealt with is that of the so-called exogenized causal structures, where all unobserved nodes are root nodes but otherwise classical variables can have both parents and children. We address the case of arbitrary, non-exogenized causal structures in the next section.

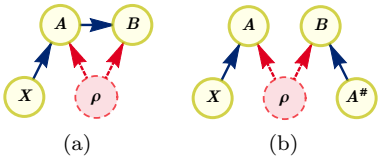


FIG. 3. Example of interruption of generic causal structures. (a) The instrumental causal structure, and (b) its interruption. Note that (a) the random variable A has both parents and children. The constraint to impose in order for the interruption to behave as the original scenario is $P_{(a)}(A{=}a,B{=}b|X{=}x){=}P_{(b)}(A{=}a,B{=}b|X{=}x,A^{\#}{=}a)$.

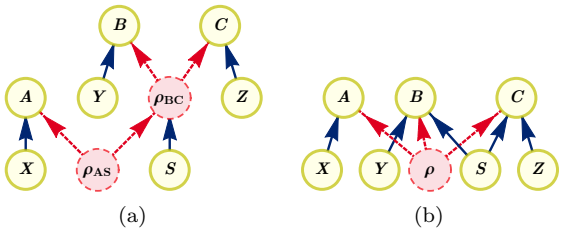


FIG. 4. In (a) there is a causal structure with ρ_{BC} being a non-exogenous unobserved quantum node. In (b) there is a different causal structure, corresponding to the classical latent reduction of the former. While these two graphs would be equivalent if the unobserved nodes were classical, they are demonstrably inequivalent when the unobserved nodes are quantum.

There exists a procedure that can be used to map the correlations of any exogenized causal structure to the correlations of a unique two-layer DAG associated with the original structure. We call such procedure **interruption**. Graphically, interruption modifies a graph as follows: For every observed variable A_i which is neither a root node nor a terminal node, introduce a new variable $A_i^{\#}$ and replace all edges formerly originating from A_i by edges originating from $A_i^{\#}$. In the interruption graph, A_i becomes a terminal node and $A_i^{\#}$ is a root node. Proceeding in this fashion, any causal structure can be converted into a two-layer scenario. Graphical examples of interruption are shown in Fig. 3. Conceptually, interruption has extensive precedent in literature regarding classical causal analysis. It is closely related to the Single-World Intervention Graphs (SWIGS) pioneered by [22], as well as the e-separation technique introduced in [23]. Interruption previously has been used to show Tsirelson inequalities constraining the set of quantum correlations compatible with the Bell scenario can be ported to quantum constraints pertaining to the Instrumental scenario [24], see also the proof of Theorem 25 in Ref. [7].

B. Quantum Exogenous Variables

Thanks to Evan's exogenization procedure [25], classical non-exogenous structures can be transformed into 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.

Unfortunately, when applied to quantum latent variables, exogenization results in a new quantum network that, in general, does not predict the same distributions of observed events as its predecessor. The example in Fig. 4, evidencing this compatibility mismatch, is wholly due to Stefano Pironio.

To make this explicit, in Fig. 4(a) the variable S can serve as 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. No quantum state prepared independently of S can do so, due to the monogamy of quantum correlations [26, 27]. Consequently, it is not possible to reproduce such correlations within the causal network of Fig. 4(b).

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. Thus, in Fig. 4(a) we treat 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,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 n^{th}-order inflation of a causal structure with non-exogenous quantum variables requires taking n copies of the unobserved root nodes. 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" (B,C, in Fig. 4) inherit the copy label j of the Hilbert space S^j.

With this last prescription, the symmetry relabelling rule, Eq. (5), still holds. However, the projection rules (3a-3b) only hold if the non-commuting variable k in question corresponds to an outcome variable in the original graph. If k corresponds to a unitary variable, then the

operator $\hat{O}_m^{\boldsymbol{s}|k}$ must be subject to the constraints

$$\hat{O}_m^{\boldsymbol{s}|k}(\hat{O}_m^{\boldsymbol{s}|k})^\dagger = (\hat{O}_m^{\boldsymbol{s}|k})^\dagger \hat{O}_m^{\boldsymbol{s}|k} = \mathbb{1}. \tag{8}$$

The commutation rule (4) remains valid upon qualifying that the Hilbert spaces listed in $\boldsymbol{s}_1 \cup \boldsymbol{s}_2$ must be simultaneously co-existing in the original graph. For example, in Fig. 4(a), the operators corresponding to the Hilbert spaces associated to the outcome variables B and C co-exist after the transformation U_s is applied over system S. It follows that the corresponding measurement operators $O_{b|y}^{\boldsymbol{s}|B}, O_{c|z}^{\boldsymbol{s}'|C}$ commute. Finally, rule (7) expressing consistency with the observed probabilities 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. Quantum inflation is therefore a technique of full applicability to bound the quantum correlations achievable in any causal scenario.

V. 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 sources can be realized in terms of commuting measurements acting on a quantum state [28]. Therefore, in order to detect correlations incompatible with classical networks, one must generalize the commutation relations in the original quantum inflation method to the constraint that any pair of measurement operators commute. 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 $\boldsymbol{s}_1, \boldsymbol{s}_2, k_1, k_2, m_1, m_2, i_1, i_2$. This defines a hierarchy of constraints that classical correlations compatible with a given causal structure must satisfy.

The associated NPO hierarchy is guaranteed to converge at a finite level. In fact, for 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 article 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 W and GHZ distributions 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.

VI. RESULTS

In the following we demonstrate the power of quantum inflation by reproducing known results and solving open problems in different tripartite causal networks.

A. The Triangle Scenario

The triangle scenario consists of three parties that are influenced in pairs by bipartite latent variables, as depicted in Fig. 1(b). Additionally, each of the visible variables may be influenced by another visible variable, that represent a *measurement choice*.

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

$$P_{\mathrm{W}}(a,b,c) := \begin{cases} \frac{1}{3} & a+b+c=1 \\ 0 & \text{otherwise} \end{cases}. \tag{9}$$

The W distribution was proven in [11] not to be realizable in the triangle scenario where the latent variables are classical. Additionally, it is easy to see that it is realizable with tripartite classical randomness. It can be shown that P_{W} does not admit a second-order quantum inflation. 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 second-order in-

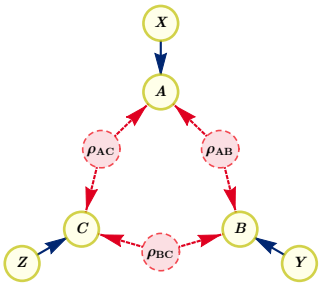


FIG. 5. The quantum triangle scenario with settings. Each of the visible variables A, B and C is now influenced not only by the latent variables, but from additional visible variables X, Y, Z that represent measurement choices.

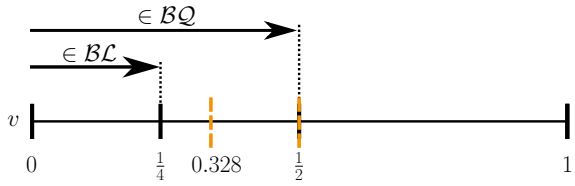


FIG. 6. Summary of results recoverable with quantum inflation in the tripartite-line scenario. Quantum inflation correctly recovers that all $P_{2\text{PR},v}$ with visibility $v>1/2$ are incompatible with the quantum tripartite line scenario already at the NPO hierarchy finite level $\mathcal{S}_2$ 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.

flation and the set of monomials $\mathcal{L}_2$ (see Appendix B for the definition of this set), restricted to operators of length ≤ 3.

B. The Triangle Scenario with Settings

As mentioned before, one can also consider more complicated networks that include additional observable variables to encode for choices of discrete measurement settings. Fig. 5 shows this type of network for the case of the triangle scenario. In this setup we study the Mermin-GHZ distribution, defined by $P_{\text{Mermin}}(a,b,c|x,y,z) :=$

$$\begin{cases} 1/8 & x+y+z=0 \mod 2, \\ \left(1+(-1)^{a+b+c}\right)/8 & x+y+z=1, \\ \left(1-(-1)^{a+b+c}\right)/8 & x+y+z=3. \end{cases} \tag{10}$$

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. Additionally, its noisy version $P_{\text{Mermin},v}$ can be proven not to have a quantum realization for any visibility v higher than $\sqrt{2/3}\approx 0.8165$.

C. The Tripartite-Line Scenario

Quantum inflation is organised 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 of Fig. 1(c), which underlies phenomena such as entanglement swapping. 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, all correlations realizable in the tripartite-line scenario satisfy the following factorization relation

$$\sum_b P_{\text{obs}}(a,b,c|x,y,z) = P_{\text{obs}}(a|x) P_{\text{obs}}(c|z). \tag{11}$$

This scenario has been thoroughly studied in the literature [29, 30]. In fact, it is known that the probability distribution

$$P_{2\text{PR}}(a,b,c|x,y,z) := \left[1+(-1)^{a+b+c+xy+yz}\right]/8, \tag{12}$$

despite satisfying the constraint of Eq. (11), 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, $P_{2\text{PR},v} := vP_{2\text{PR}} + (1-v)/8$, can be realized if the visibility parameter v is sufficiently small [30]. $P_{2\text{PR},v}$ 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 $P_{2\text{PR},v}$ 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 $P_{2\text{PR},v}$ 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 B for a definition of this monomial set). Furthermore, we can also contrast against realizations in terms of classical latent variables by imposing that all measurements in the problem commute. While not being tight, this classical version of quantum inflation witnesses that all $P_{2\text{PR},v}$ with visibility $v>0.328$ cannot be realized in terms of classical hidden variables. This result is obtained 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 B).

VII. 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 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 realizations of famous tripartite distributions in different causal structures, proving that the W and Mermin-GHZ distributions cannot be generated in the triangle scenario with quantum latent variables and bounding their noise resistance. We also showed, in the entanglement swapping scenario, how quantum inflation is capable of recovering known results.

The implementation of quantum inflation comprises 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.

Quantum inflation can find an application in many fields. A clear first application is found in multi-party quantum information protocols [31]. 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.

ACKNOWLEDGMENTS

We acknowledge useful discussions with Stefano Pironio. This work is supported by Fundació Obra Social "la Caixa" (LCF/BQ/ES15/10360001), the ERC CoG QITBOX, the AXA Chair in Quantum Information Science, the Spanish MINECO (QIBEQI FIS2016-80773-P and Severo Ochoa SEV-2015-0522), Fundació Cellex, Generalitat de Catalunya (SGR 1381 and CERCA Programme) and the Austrian Science fund (FWF) stand-alone project P 30947. This research was supported by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

[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. Belén Sainz, R. Kunjwal, and R. W. Spekkens, "Bell Quantified: The Resource Theory of Nonclassicality of Common-Cause Boxes," (2019), arXiv:1903.06311.
[11] E. Wolfe, R. W. Spekkens, and T. Fritz, "The Inflation Technique for Causal Inference with Latent Variables," J. Causal Inference **7** (2019), 1609.00672.
[12] M. Navascués and E. Wolfe, "The Inflation Technique Completely Solves the Causal Compatibility Problem," (2017), arXiv:1707.06476.
[13] W. Slofstra, "The set of quantum correlations is not closed," For. Math., Pi **7**, e1 (2019).
[14] 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).
[15] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, "Cloning and Broadcasting in Generic Probabilistic Theories," (2006), quant-ph/0611295.
[16] R. Chaves, C. Majenz, and D. Gross, "Information–theoretic implications of quantum causal structures," Nat. Comm. **6** (2015).
[17] A. Pozas-Kerstjens, R. Rabelo, Ł. Rudnicki, R. Chaves, D. Cavalcanti, M. Navascués, and A. Acín, "Bounding the sets of classical and quantum correlations in networks," (2019), arXiv:1904.08943.
[18] S. Pironio, M. Navascués, and A. Acín, "Convergent relaxations of polynomial optimization problems with noncommuting variables," SIAM J. Optim. **20**, 2157 (2010).
[19] M. Navascués, S. Pironio, and A. Acín, "Bounding the set of quantum correlations," Phys. Rev. Lett. **98**, 010401 (2007).
[20] 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).
[21] L. Vandenberghe and S. Boyd, "Semidefinite programming," SIAM Rev. **38**, 49 (1996).
[22] 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.
[23] R. J. Evans, "Graphical methods for inequality constraints in marginalized DAGs," in *IEEE International Workshop on Machine Learning for Signal Processing* (2012).
[24] 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).
[25] R. J. Evans, "Margins of discrete Bayesian networks," Annals Stat. **46**, 2623 (2018).
[26] B. Toner, "Monogamy of non-local quantum correlations," Proc. Roy. Soc. A **465**, 59 (2009).
[27] M. P. Seevinck, "Monogamy of correlations versus monogamy of entanglement," Quant. Info. Proc. **9**, 273 (2010).
[28] F. Baccari, D. Cavalcanti, P. Wittek, and A. Acín, "Efficient Device-Independent Entanglement Detection for Multipartite Systems," Phys. Rev. X **7**, 021042 (2017).
[29] C. Branciard, N. Gisin, and S. Pironio, "Characterizing the Nonlocal Correlations Created via Entanglement Swapping," Phys. Rev. Lett. **104**, 170401 (2010).
[30] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, "Bilocal versus nonbilocal correlations in entanglement-swapping experiments," Phys. Rev. A **85**, 1 (2012).
[31] E. Wolfe *et al.*, "Classifying and Bounding Non-Classicality in Quantum Networks," (in preparation).
[32] 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

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

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 $X = (X_1...X_n)$ in $\mathcal{H}$ (where $X_i X_j \neq X_j X_i$ in general) that minimize the expectation value $\langle p(X)\rangle_\rho = \mathrm{Tr}[\rho \cdot p(X)]$ 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. [18], and uses a hierarchy of relaxations where each of the steps is a semidefinite program. 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. [18]) that all non-commuting 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 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 [31].

Appendix B: 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 non-commutative operators in the problem. In this article we use two different hierarchies, that are both asymptotically complete. The levels in the first hierarchy are known as **NPA levels** [19, 20]. 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 $H = E, F, G$. On the other hand, some problems may achieve tighter results at lower levels if one instead considers **local levels** [32]. 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

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

where $H \neq H'$. 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 vary with the specific problem to solve.