Causal compatibility inequalities admitting quantum violations in the triangle structure

Thomas C. Fraser[1,2,*] and Elie Wolfe[1]
[1]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada N2L 2Y5*
[2]*University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*

(Received 10 January 2018; published 8 August 2018)

It has long been recognized that certain quantum correlations are incompatible with a particular assumption about classical causal structure. Given a causal structure of unknown classicality, the presence of such correlations certifies the nonclassical nature of the causal structure in a device-independent fashion. In structures where all parties share a common resource, these nonclassical correlations are also known as nonlocal correlations. Any constraint satisfied by all correlations which are classically compatible with a given causal structure defines a causal compatibility criterion. Such criteria were recently derived for the triangle structure (E. Wolfe *et al.*, arXiv:1609.00672) in the form of polynomial inequalities, begging the question of whether any of those inequalities admit violation by quantum correlations. Numerical investigation suggests that they do not, and we further conjecture that the set of correlations admitted by the classical triangle structure is equivalent to the set of correlations admitted by its quantum generalization whenever the three observable variables are binary. Our main contribution in this work, however, is the derivation of causal compatibility inequalities for the triangle structure which do admit quantum violation. This provides a robust-to-noise witness of quantum correlations in the triangle structure. We conclude by considering the possibility of quantum resources potentially qualitatively different from those known previously.

DOI: 10.1103/PhysRevA.98.022113

I. INTRODUCTION

In recent decades, the technical utility of quantum mechanics has become abundantly clear. In the realm of computation, quantum algorithms, such as Shor's algorithm [1] and numerous others [2], scale exponentially better than their classical counterparts. In the realm of secure communication, quantum protocols, a popular example being quantum key distribution [3], are able to provide privacy even against hypothetical adversaries with unlimited computational power, a desideratum which classical protocols are unable to fulfill. Throughout history, numerous quantum phenomena which fail to be emulated by classical physics have been identified as resources for solving computational or communication problems [4]. Motivated by past successes, a primal objective of modern quantum information theory is to discover new situations wherein quantum mechanics offers an advantage and to certify that the quantum advantage is genuine.

From a foundational prospective, the most robust demonstrations of quantum phenomena with no classical emulation have involved Bell inequalities [5,6]. Originally, Bell inequalities were derived as a way to show that no hidden variable theory could ever account for quantum mechanics; in this sense Bell inequalities are a response to the famous Einstein-Podolsky-Rosen paradox [7]. The enumeration of Bell inequalities has since become a widespread systematic method for demonstrating the nonclassicality of a given observation. More recently it has been appreciated that Bell inequalities can be understood as consequences of causal inference [8]. Causal inference is concerned with classifying observations into those which can and cannot be explained by a hypothesized causal structure. The abstract nature of causal inference is responsible for its presence in numerous scientific fields including machine learning and biology [9,10]. Causal compatibility inequalities, such as Bell and instrumental inequalities [11–13], characterize the space of observations that are compatible with a hypothesized causal structure, albeit the characterization offered by practically derivable inequalities is often only an approximation. To derive the traditional Bell inequalities from causal inference one starts with a (classical) causal structure known as the Bell structure, as depicted in Fig. 1. The fundamental Bell structure involves noncommunicating parties making measurements on some hidden shared resource λ, where the measurement outcomes (A and B) are presumed to be stochastic functions of the local choices of measurement settings (S_A and S_B) and the shared resource λ. Quantum nonclassicality in the Bell structure has been thoroughly studied since Bell's original work [6]. More complex structures, however, such as the correlation scenarios proposed by Fritz [14,15], are much less understood. Here we investigate one particular correlation scenario named the triangle structure (Fig. 2).

The triangle structure (Fig. 2) is a causal structure comprised of three parties labeled A, B, and C arranged in a triangular configuration while pairwise sharing hidden (latent) variables X, Y, and Z. It has been extensively studied previously (see, e.g., [16], Fig. 1; [17], Fig. 6; [18], Fig. 8; [19], Fig. 8 and Appendix E; [14], Fig. 3; [20], Fig. 4; and [21], Fig. 1). An overview of some milestone results is provided in Sec. IV. Identifying causal compatibility inequalities for this configuration has been seen as particularly challenging

*tfraser@perimeterinstitute.ca

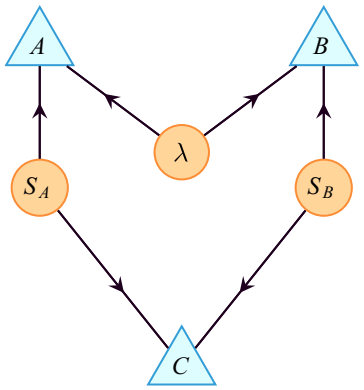

FIG. 1. Bell structure consisting of two observers A and B together with measurement settings S_A and S_B, respectively. The shared latent variable is labeled λ.

[18]. Further identifying causal compatibility inequalities of such high resolution that such inequalities can be violated by quantum-accessible distributions has remained out of reach for the triangle structure.

This work finds causal compatibility inequalities for the triangle structure that are known to be violated by quantum-accessible distributions. This accomplishment was made possible through the combination of two previous developments: first, the insight of Fritz [14] regarding the ability to reinterpret the Bell structure as a portion of the triangle structure, and second, the framework for solving causal inference problems developed by Wolfe *et al.* [21] called the inflation technique. Ultimately, this work serves as a validation that the inflation technique is efficient and sensitive enough at low orders to offer insights into quantum nonclassicality [22]. Moreover, these inequalities offer an avenue for recognizing previously unknown forms of nonclassicality. The authors' attempts to find such novel resources were met with only partial success, suggesting the need for both conceptual refinements and future exploration.

The first half of the paper, namely, Secs. II–V, is entirely a review of previous works. Section II recalls important notions from causal inference theory and sets up the notation to be used. Section III offers a summary of the popular Bell structure and associated inequalities. Section IV discusses the triangle structure and provides an overview of existing research, identifying its stark differences from the Bell structure and motivating why the triangle structure is worth studying. Section V defines and discusses a singularly quantum correlation conceived of by Fritz [14], which we term the Fritz distribution. In the same work [14], the Fritz distribution was proven to be nonclassical without the use of inequalities.

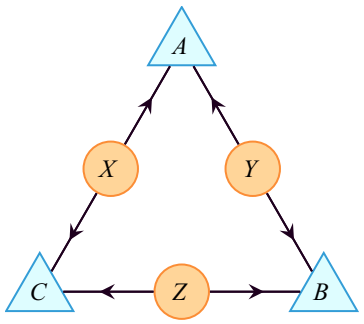

FIG. 2. Triangle structure consisting of three observable variables A, B, and C and three latent variables X, Y, and Z.

The second half of the paper, namely, Secs. VI–IX, presents the main contributions made by this research project. Specifically, in Sec. VII we improve upon the results of Fritz [14] by offering a direct proof of the incompatibility of the Fritz distribution using inequalities generated by the inflation technique [21]. A sample of such inequalities is presented in Sec. VII, specifically the inequalities (12), (13), and (15). Aside from confirming the utility of the inflation technique, this paper explores the importance of having derived these inequalities. First, an inequality-based proof has the advantage of being robust to experimental noise. In Sec. VII B the Fritz distribution is subjected to noise in order to measure the robustness of the derived inequalities. Second, we numerically optimize our derived inequalities over quantum-accessible distributions (using qubits) in an effort to find the maximum violations achievable by quantum theory. The culmination of these analyses naturally prompts a discussion, found in Sec. IX, regarding the fundamental problem of recognizing and classifying nonclassicality in the triangle structure. Section X summarizes.

Appendix B briefly summaries the inflation technique in the specific context of this work. Although the summary presented in Appendix B is designed to be self-standing, a much more pedagogical introduction is offered by the original work [21]. Appendix C demonstrates how the inflation technique was used to derive the causal compatibility inequalities for the triangle structure which admit violation by quantum-accessible distributions.

II. CAUSAL COMPATIBILITY

The task of causal inference is to determine the set of potentially observable probability distributions compatible with some hypothesis about causal relationships [9]. If an observed distribution can be explained by the hypothesized causal mechanism, then the distribution is said to be compatible with said causal mechanism. In order to define compatibility rigorously, we first need to formally define the notion of a causal hypothesis.

A hypothesis of causal mechanism is formally referred to as a causal structure and can be represented as a directed acyclic graph. A directed graph $\mathcal{G}$ is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of nodes and edges, respectively, where each edge $e \in \mathcal{E}$ connects a pair of nodes $n, m \in \mathcal{N}$ with a directed arrow $e = \{n \to m\}$. A directed graph is acyclic if there are no paths following the directions of the edges starting from and returning to the same node. The nodes $\mathcal{N}$ of a causal structure represent random variables while the edges $\mathcal{E}$ represent a casual influence from one variable to another pursuant to the prescribed direction.

Henceforth, we will utilize a number of familiar notions from graph theory and denote them accordingly. The parents of a node $n \in \mathcal{N}$ are all nodes which point directly into n, i.e., $\mathsf{Pa}_{\mathcal{G}}(n) \equiv \{m \mid m \to n\}$. Similarly defined are the children of a node $\mathsf{Ch}_{\mathcal{G}}(n) \equiv \{m \mid n \to m\}$. Recursively defined are the ancestors of a node $\mathsf{An}_{\mathcal{G}}(n) \equiv \bigcup_{i \in \mathbb{N}} \mathsf{Pa}^i_{\mathcal{G}}(n)$, where $\mathsf{Pa}^i_{\mathcal{G}}(n) \equiv$

$\mathsf{Pa}_{\mathcal{G}}(\mathsf{Pa}_{\mathcal{G}}^{i-1}(n))$ and $\mathsf{Pa}_{\mathcal{G}}^{0}(n) = n$, and the descendants of a node $\mathsf{De}_{\mathcal{G}}(n) \equiv \bigcup_{i\in\mathbb{N}} \mathsf{Ch}_{\mathcal{G}}^{i}(n)$, where $\mathsf{Ch}_{\mathcal{G}}^{i}(n) \equiv \mathsf{Ch}_{\mathcal{G}}(\mathsf{Ch}_{\mathcal{G}}^{i-1}(n))$ and $\mathsf{Ch}_{\mathcal{G}}^{0}(n) = n$. Finally, we extend this notation to a subset of nodes $N \subseteq \mathcal{N}$ by performing a union over elements. As an example, the parents of the nodes $N \subseteq \mathcal{N}$ are defined $\mathsf{Pa}_{\mathcal{G}}(N) = \bigcup_{n\in N} \mathsf{Pa}_{\mathcal{G}}(n)$.

Let us now formalize the notion of compatibility between a causal structure $\mathcal{G}$ and a probability distribution $P_{\mathcal{N}}$ defined over the nodes of $\mathcal{G}$. A causal structure $\mathcal{G}$ hypothesizes that each variable $n \in \mathcal{N}$ is only directly influenced by its parents $\mathsf{Pa}_{\mathcal{G}}(n)$. Therefore, if each variable is conditioned on its parentage, the probability distribution $P_{\mathcal{N}}$ should factorize accordingly:

$$P_{\mathcal{N}} = \prod_{n\in\mathcal{N}} P(n \mid \mathsf{Pa}_{\mathcal{G}}(n)) = P(n_1 \mid \mathsf{Pa}_{\mathcal{G}}(n_1)) \times \cdots \times P(n_k \mid \mathsf{Pa}_{\mathcal{G}}(n_k)). \quad (1)$$

If a given distribution $P_{\mathcal{N}}$ defined over all of the nodes $\mathcal{N}$ of $\mathcal{G}$ satisfies Eq. (1), then $P_{\mathcal{N}}$ is said to be compatible with $\mathcal{G}$. The conditional distributions in Eq. (1) (i.e., $\{P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N}\}$) are referred to as a set of causal parameters for $\mathcal{G}$. If a distribution $P_{\mathcal{N}}$ cannot be factorized according to Eq. (1), $P_{\mathcal{N}}$ is said to be incompatible with $\mathcal{G}$. When one is specified with a joint distribution $P_{\mathcal{N}}$ defined over all nodes of a causal structure $\mathcal{G}$, it is possible to completely determine whether or not $P_{\mathcal{N}}$ is compatible with $\mathcal{G}$ by computing the causal parameters induced by $P_{\mathcal{N}}$ and checking the equality of Eq. (1). A challenge, however, is presented when one is supplied with a partial observation, i.e., a joint distribution $P_{\mathcal{N}_O}$ where $\mathcal{N}_O \subset \mathcal{N}$ is some subset of variables referred to as the observable nodes $\mathcal{N}_O$. In such cases, $P_{\mathcal{N}_O}$ does not induce a unique set of causal parameters for $\mathcal{G}$ and Eq. (1) cannot be verified by direction calculation. Instead, compatibility between $\mathcal{G}$ and $P_{\mathcal{N}_O}$ depends on the existence or non-existence of a set of causal parameters for $\mathcal{G}$ such that $P_{\mathcal{N}_O} = \sum_{n\notin\mathcal{N}_O} P_{\mathcal{N}}$ where $P_{\mathcal{N}}$ is again given by Eq. (1). The complementary, unobservable nodes are termed latent nodes $\mathcal{N}_L = \mathcal{N} \setminus \mathcal{N}_O$ and should be understood as hidden random variables that either are unknowable by some fundamental process or cannot be measured due to other limitations.

There are several approaches to tackling the compatibility problem when dealing with latent variables; there are two common approaches worth mentioning here. The first is to recognize that many equality constraints are implied by the causal structure, including conditional independence relations and so-called Verma constraints among others; see Refs. [23,24] for thorough treatments. The failure to satisfy an equality constraint immediately disqualifies $P_{\mathcal{N}_O}$ from being compatible with $\mathcal{G}$. Equality constraints are easily derived given a causal structure, and checking equality-constraint satisfaction is the minimalistic algorithm which powers the overwhelming majority of practical causal inference hypothesis testing in the fields of machine learning and artificial intelligence. In quantum theory, however, we require strong, more sensitive, causal inference techniques. This is because the equality constraints satisfied by compatible classical correlations are also all satisfied by quantum correlations [19]. Our focus therefore in on deriving inequality constraints (over $P_{\mathcal{N}_O}$ and its marginals) implied by a causal structure, which we term causal compatibility inequalities.[1] For some causal structures, the equality constraints associated with it are sufficient to perfectly characterize the distributions genuinely compatible with it; for others, however, inequality constraints are also important. Causal structures for which inequality constraints are relevant have been termed interesting [19] and such structures include the instrumental structure, the Bell structure, and the triangle structure studied here, among infinitely many others. Herein we use the inflation technique [21] to find causal compatibility inequalities; Appendix B discusses the inflation technique as applied in this work.

If a probability distribution $P_{\mathcal{N}_O}$ happens to violate any causal compatibility inequality, then that distribution is deemed incompatible. Conversely, a singular inequality can only be used to prove that a given distribution is incompatible; a single inequality cannot certify compatibility. A complete characterization of compatibility consists of a complete set of all valid causal compatibility inequalities such that satisfaction of the entire set certifies compatibility. Currently, however, it is unknown how to obtain a complete characterization for all causal structures, including the triangle structure.

From the perspective of identifying quantum nonclassicality, a causal structure $\mathcal{G}$ adopts the role of a classical hypothesis. Therefore, nonclassicality becomes synonymous with incompatibility: If a distribution $P_{\mathcal{N}_O}$ is incompatible, then it is nonclassical. Henceforth, we will use these two terms interchangeably. From a resource standpoint, if the nonclassicality of $P_{\mathcal{N}_O}$ with $\mathcal{G}$ can be witnessed by an inequality I, but nevertheless $P_{\mathcal{N}_O}$ can be implemented using quantum states and measurements while otherwise respecting the causal relations of $\mathcal{G}$, then the causal compatibility inequality I represents a task or game where quantum resources outperform classical resources relative to $\mathcal{G}$.

III. BELL STRUCTURE

This section aims to define the Bell causal structure and to review some of the traditional witnesses used to assess the classicality (or lack thereof) of distributions relative to it. The purpose of this section is to equip readers with the pertinent background and also to draw comparisons between the advancements made toward understanding nonclassicality of the Bell structure versus analogous results obtained for triangle structure later in this work.

The bipartite Bell structure (Fig. 1) refers to an iconic causal structure involving two distant parties who observe the outcomes of local measurements as random variables A and B determined by their individual measurement settings S_A and S_B and where the parties are also presumed to be commonly informed by some shared latent resource λ [6]. The observed correlations naturally form a conditional probability distribution $P_{AB|S_AS_B}$. Subject to the notions of compatibility

[1] We refer to these inequalities as causal compatibility inequalities instead of Bell inequalities for two reasons. First, Bell inequalities usually are associated specifically with the Bell structure. Second, the inequalities derived in this work are fundamentally distinct from a typical Bell inequality in that these inequalities are polynomial over $P_{\mathcal{N}_O}$ instead of linear.

presented in Sec. II, $P_{AB|S_AS_B}$ is compatible with Fig. 1 if and only if there exists some distribution P_λ and causal parameters $P_{A|S_A,\lambda}$ and $P_{B|S_B,\lambda}$ such that[2]

$$P_{AB|S_AS_B} \text{ is classically compatible with Fig. 1} \iff P_{AB|S_AS_B} = \sum_\lambda P_{A|S_A,\lambda} P_{B|S_B,\lambda} P_\lambda. \quad (2)$$

Any observed distribution $P_{AB|S_AS_B}$ which fails to be explained by the classical causal hypothesis of the Bell structure as defined by Eq. (2) is appropriately termed nonclassical. Often, distributions incompatible with the Bell structure are referred to as nonlocal because the Bell structure markedly lacks causal influence from one party to another. In particular, Bell [5] demonstrated that there exist distributions that are nonclassical yet are attainable via local measurements on a shared quantum resource.

Let us now contrast the classical definition of compatibility with its quantum generalization. Quantum correlation arise from a quantumified version of the causal structure, in which latent variables are replaced with quantum systems. Concretely, a distribution is accessible with quantum resources and measurements for the quantumified Bell structure if and only if there exists a bipartite quantum state ρ_{AB} of arbitrary dimension, as well as Hilbert-space-localized measurement sets $M_{A|S_A}$ and $M_{B|S_B}$,[3] with indices conditional upon the measurement settings, such that

$$P_{AB|S_AS_B} \text{ is a quantum realization of Fig. 1} \iff P_{AB|S_AS_B} = \mathrm{Tr}[\rho_{AB} M_{A|S_A} \otimes M_{B|S_B}]. \quad (3)$$

The nonclassicality of quantum distributions for the Bell structure can be demonstrated through the use of Bell inequalities which constrain the correlations between binary variables A and B for classically compatible distributions. A notable example is the Clauser-Horne-Shimony-Holt (CHSH) inequality [25]

$$\langle AB|S_A=0,S_B=0\rangle + \langle AB|S_A=0,S_B=1\rangle + \langle AB|S_A=1,S_B=0\rangle - \langle AB|S_A=1,S_B=1\rangle \leqslant 2. \quad (4)$$

Nontrivial Bell inequalities such as the CHSH inequality are capable of witnessing the nonclassical nature of quantum distributions; the inequalities presented in Sec. VII are also of this type. Equality constraints, as previously mentioned, never have that sort of high-resolution discernment sensitivity.

IV. TRIANGLE STRUCTURE

As was mentioned in Secs. I and II, the triangle structure (Fig. 2) is a causal structure $\mathcal{G}$ consisting of three observable variables A, B, and C arranged in a triangular configuration while pairwise sharing latent variables X, Y, and Z. Following the definition of causal compatibility from Sec. II, a distribution $P_{\mathcal{N}_O} = P_{ABC}$ is compatible with the triangle structure if and only if there exists a choice of causal parameters $\{P_{A|X,Y}, P_{B|Y,Z}, P_{C|Z,X}, P_X, P_Y, P_Z\}$ such that P_{ABC} is a marginalization of P_{ABCXYZ} over X, Y, and Z,[4] i.e.,

$$P_{ABC} \text{ is classically compatible with Fig. 2} \iff P_{ABC} = \sum_{X,Y,Z} P_{A|X,Y} P_{B|Y,Z} P_{C|Z,X} P_X P_Y P_Z. \quad (5)$$

By contrast, the quantum realization of the triangle structure is defined with substantially greater freedom, namely,

$$P_{ABC} \text{ is a quantum realization of Fig. 1} \iff P_{ABC} = \mathrm{Tr}[\Pi^\intercal \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA} \Pi M_A \otimes M_B \otimes M_C], \quad (6)$$

where ρ_{AB}, ρ_{BC}, and ρ_{CA} are bipartite density matrices, M_A, M_B, and M_C are generic measurements sets, and Π is a permutation matrix to align the underlying tensor structure of the states and measurements appropriately.

The triangle structure serves as an excellent test case for furthering our understanding of quantum nonclassicality in network causal structures. It maintains superficial simplicity (only three observable variables) while introducing many challenging features not found in the study of the Bell structure. For example, the spaces of both classical and quantum distributions on the triangle structure are nonconvex [14,21], unlike for the Bell structure. The convexity of the Bell structure's distributions is arguably responsible for the wealth of knowledge about it, including its complete characterization of classicality [6]. Importantly, Fritz [14] explicitly demonstrated the existence of (at least) one incompatible but quantum distribution for the triangle structure, so it is known to possess quantum nonclassicality. It seems reasonable to assume that quantum nonclassicality in the triangle structure should be translatable into a computational advantage for certain computational circuits [26]; novel instances of nonclassicality are expected to correspond to novel information-theoretical quantum advantages. A fundamental limitation of Fritz's proof of nonclassicality, however, is that it does not involve causal compatibility inequalities and hence does not advance our repertoire of inequality constraints for the triangle structure. Some inequality constraints for the triangle structure have been derived in previous works. For example, Steudel and Ay [16] derived an inequality distinguishing the distributions compatible with the triangle structure from those compatible with structures in which all the observable variables share a common latent ancestor. Henson *et al.* [19] derived a family of entropic inequalities for the triangle structure, which was then expanded somewhat by Weilenmann and Colbeck [20]. Recently, Wolfe *et al.* [21] derived a variety of especially sensitive, polynomial causal compatibility inequalities for the triangle structure. In particular, the inequalities of [21] expose a previously unclassified (as assessed by all formerly known constraints)

[2]Here the summation $\sum_\lambda$ is used to denote a statistical marginalization over the latent variable λ with unspecified support.

[3]Note that a measurement set $M_A = \{M_A^1, M_A^2, \ldots, M_A^k\}$ services a shorthand notation in the sense $P_A = \mathrm{Tr}[\rho M_A] \Rightarrow P_A(a) = \mathrm{Tr}[\rho M_A^a]$.

[4]Here the summation $\sum_{X,Y,Z}$ is used to denote a statistical marginalization over the latent variables X, Y, and Z with unspecified support.

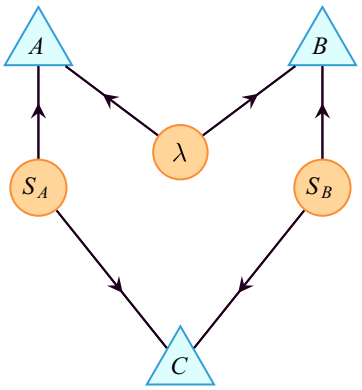

FIG. 3. Triangle structure reimagined to mimic the Bell structure. The measurement settings S_A and S_B are latent nodes unlike the Bell structure (Fig. 1).

distribution called the w distribution[5] as incompatible with the triangle structures. Remarkably, none of the existing causal compatibility inequalities for the triangle structure were known to admit violation by any quantum distribution. That is to say, it was unknown if any inequality known for the triangle structure might be useful for distinguishing quantum distributions from their classical counterparts. Section VI reports our attempt to utilize these aforementioned inequalities to search for incompatible distributions that are also quantum accessible. The subsequent failure of these approaches effectively motivates the remainder of this work. In summary, the triangle structure is a desirable case study because it is known to admit nonclassical distributions using quantum resources, but yet no inequality heretofore could separate its classical distributions from its quantum distributions. This failure represents a gap in our understanding of quantum nonclassicality and prompts the discovery of alternative inequalities.

V. FRITZ DISTRIBUTION

As realized by Fritz [14], one may construct quantum distributions incompatible with the triangle structure by recasting quantum distributions incompatible with the familiar Bell structure into a settings-free tripartite format. To explain, imagine rearranging the triangle structure into the configuration depicted in Fig. 3 so that it closely resembles the Bell structure (Fig. 1). Evidently, under the correct relabeling, large portions of the triangle structure resemble the Bell structure. The crucial distinction is that S_A and S_B are random variables representing the recorded measurement settings in the Bell structure, whereas those S_A and S_B are latent variables in the triangle structure, which get reported as auxiliary outcomes for Alice and Bob.

The analysis of nonclassicality changes, however, when S_A and S_B are not freely chosen by the observers but rather by a process outside of the individual party's control. Relaxing the assumption of measurement independence opens up a possible loophole, namely, the possibility that the auxiliary outcomes S_A and S_B of Alice and Bob might be manipulated via dependence on their shared latent variable λ. This loophole is closed by having the third party in the triangle structure, Charlie, also report the latent variables S_A and S_B as a multivariate outcome. In this manner, the perfect correlation of C's record of S_A and S_B with the records of S_A reported by A and of S_B reported by B testifies to the independence of S_A and S_B from λ. Consequently, any distribution over A, B, S_A, and S_B that is incompatible with the Bell structure is also incompatible with the triangle structure provided that C is perfectly correlated with S_A and S_B [14].

The exemplifying quantum distribution corresponding to a recasting of a nonclassical Bell structure distribution into the triangle structure is the Fritz distribution [14]. In the Fritz distribution, denoted by P_{F}, each of the variables A, B, and C is taken to have four possible outcomes $\{0, 1, 2, 3\}$. Explicitly, P_{F} can be written as

$$\begin{aligned} P_{\mathrm{F}}(000) &= P_{\mathrm{F}}(110) = P_{\mathrm{F}}(021) = P_{\mathrm{F}}(131) = P_{\mathrm{F}}(202) \\ &= P_{\mathrm{F}}(312) = P_{\mathrm{F}}(233) = P_{\mathrm{F}}(323) = \tfrac{1}{32}(2+\sqrt{2}), \\ P_{\mathrm{F}}(010) &= P_{\mathrm{F}}(100) = P_{\mathrm{F}}(031) = P_{\mathrm{F}}(121) = P_{\mathrm{F}}(212) \\ &= P_{\mathrm{F}}(302) = P_{\mathrm{F}}(223) = P_{\mathrm{F}}(333) = \tfrac{1}{32}(2-\sqrt{2}). \end{aligned} \tag{7}$$

In Eqs. (7), the notation $P_{\mathrm{F}}(abc) = P(A \mapsto a, B \mapsto b, C \mapsto c)$ is used as shorthand. The Fritz distribution is quantum accessible in the sense that P_{F} can be implemented using a set of quantum states ρ_{AB}, ρ_{BC}, and ρ_{CA} and measurements M_A, M_B, and M_C realized on Fig. 2 using Eq. (6). When expressing the outcomes $\{0, 1, 2, 3\}$ as pairs of binary digits $\{00, 01, 10, 11\}$, it can be seen that the left-hand bits for A and B (respectively denoted by A_l and B_l) are fixed by the outcome of C,

$$\begin{aligned} P_{\mathrm{F}}(00,00,00) &= P_{\mathrm{F}}(01,01,00) = P_{\mathrm{F}}(00,10,01) \\ &= P_{\mathrm{F}}(01,11,01) = \tfrac{1}{32}(2+\sqrt{2}), \\ P_{\mathrm{F}}(10,00,10) &= P_{\mathrm{F}}(11,01,10) = P_{\mathrm{F}}(10,11,11) \\ &= P_{\mathrm{F}}(11,10,11) = \tfrac{1}{32}(2+\sqrt{2}), \\ P_{\mathrm{F}}(00,01,00) &= P_{\mathrm{F}}(01,00,00) = P_{\mathrm{F}}(00,11,01) \\ &= P_{\mathrm{F}}(01,10,01) = \tfrac{1}{32}(2-\sqrt{2}), \\ P_{\mathrm{F}}(10,01,10) &= P_{\mathrm{F}}(11,00,10) = P_{\mathrm{F}}(10,10,11) \\ &= P_{\mathrm{F}}(11,11,11) = \tfrac{1}{32}(2-\sqrt{2}). \end{aligned} \tag{8}$$

In Eqs. (8), the notation $P_{\mathrm{F}}(a_l a_r, b_l b_r, c_l c_r) = P(A_l \mapsto a_l, A_r \mapsto a_r, B_l \mapsto b_l, B_r \mapsto b_r, C_l \mapsto c_l, C_r \mapsto c_r)$ is used as shorthand. This observation can be difficult to verify using Eqs. (8), but becomes easier after organizing the possible outcomes into a $4 \times 4 \times 4$ grid as depicted in Fig. 4. From this diagram, it can be seen that each of C's outcomes restricts the possible outcomes for A and B into a 2×2 block. Effectively, C's bits are perfectly correlated with the left-hand bits of A and B; $C_l = A_l$ and $C_r = B_l$. Therefore, pursuant to the embedding of Fig. 3, the left-hand bits A_l and B_l emulate the measurement settings S_A and S_B, whereas the right-hand bits A_r and B_r emulate the outcomes which would be obtained by A and B back in the Bell structure (Fig. 1). Provided that C is perfectly correlated with A_l and B_l, any Bell inequality

[5] Although the w distribution is nonclassical, it is also nonquantum. The nonquantum nature of the w distribution has been demonstrated by Navascués and Wolfe [27].

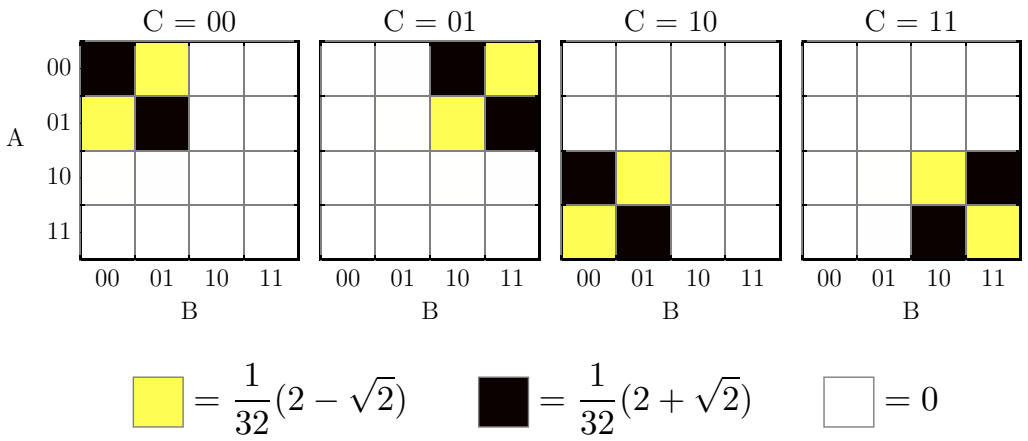

FIG. 4. Fritz distribution visualized using a $4 \times 4 \times 4$ grid. The four outcomes of A, B, and C are written in binary as a doublet of bits to illustrate that certain bits act as measurement pseudosettings.

for the Bell structure defined over A, B, S_A, and S_B can be directly converted to an inequality for the triangle structure by performing the surjective relabeling

$$A_r \leftarrow A, \quad B_r \leftarrow B, \quad A_l \leftarrow S_A, \quad B_l \leftarrow S_B,$$
$$C_l \leftarrow S_A, \quad C_r \leftarrow S_B. \tag{9}$$

As an example, the famous CHSH inequality (4) [25] is translated into a constraint on the correlation between the right-hand bits of A and B under Eqs. (9),

$$C = C_l C_r = A_l B_l \Longrightarrow \langle A_r B_r | C = 00\rangle + \langle A_r B_r | C = 01\rangle + \langle A_r B_r | C = 10\rangle - \langle A_r B_r | C = 11\rangle \leqslant 2, \tag{10}$$

where the correlation $\langle A_r B_r \rangle$ between the right-hand bits of A and B is defined as

$$\langle A_r B_r \rangle = P_{A_r B_r}(00) + P_{A_r B_r}(11) - P_{A_r B_r}(01) - P_{A_r B_r}(10). \tag{11}$$

Therefore, every compatible distribution P_{ABC} for which C is perfectly correlated with A_l and B_l must satisfy the inequality (10). Substituting P_{F} into the inequality (10) yields the traditional maximal quantum violation [28] $3(1/\sqrt{2}) - (-1/\sqrt{2}) = 2\sqrt{2} \not\leqslant 2$.

It is important to understand the domain in which Fritz's proof of incompatibility is valid; its proof relies on the perfect correlation between C's outcomes and the measurement pseudosettings (left-hand bits) of A and B. For example, if one combines Eq. (7) with slight uniform noise, what can be said with confidence regarding if the resulting modified distribution is classical or not? At what point does the resulting distribution transition from incompatibility to compatibility? These questions are partially answered in Sec. VII B.

Plainly, P_{F} is a valid but manufactured example. The phenomenology associated with Bell nonlocality or Bell incompatibility are well understood; examining these distributions embedded in the triangle structure offers no additional insight into the types of nonclassical resources made accessible by quantum mechanics. The goal, therefore, is to find incompatible quantum distributions that are qualitatively different from those previously considered for the Bell structure [29]. Recognizing this, Fritz [14] presented the following problem (see [14], Problem 2.17).

Fritz's problem. Find an example of nonclassical quantum correlations in the triangle structure together with a proof of its nonclassicality which does not hinge on Bell's theorem.

Fritz's problem is concerned with how to find and recognize nonclassical quantum distributions specifically for the triangle structure. The original proof of nonclassicality essentially recycled Bell's theorem and was limited by the requirement of perfect correlations [14]. Fritz's problem, as originally stated, does not require that the type of nonclassicality be novel to the triangle structure, rather only that the proof should avoid Bell's theorem. Section VI delineates our initial, failed attempts at resolving Fritz's problem; Sec. VII reports our eventual success, via the discovery of different causal compatibility inequalities.

Though not explicit, we read in the spirit of Fritz's problem a desideratum for the discovery of a truly different form of nonclassicality for the triangle structure. Such a discovery would presumably lead to an understanding of different advantages of quantum resources in network structures; this related problem has attracted attention and conjecture [29], but remains open. We attempt to make progress on this problem by leveraging the causal compatibility inequalities derived herein, but Sec. VIII delineates how the effort is plagued by instabilities in our numerical optimization which we have not yet overcome. Consequently, Sec. IX discusses potential avenues for rigorously reformulating Fritz's problem in order to best capture this desire for novel quantum nonclassicality.

VI. PRELIMINARY RESEARCH

As a preliminary search for quantum incompatibility in the triangle structure, we performed numerical optimizations (in search of violation) against the previously published compatibility inequalities of Wolfe *et al.* [21], as well as against the entropic inequalities of Henson *et al.* [19]. For historical context, the entropic inequalities of [19] have already been independently investigated for quantum incompatibility by Weilenmann and Colbeck [20] using a variety of computational methods. Unfortunately, these methods failed to identify quantum-accessible distributions capable of violating any of

the entropic inequalities considered in [20]. Additionally, the inequalities presented in [21] have not been previously investigated numerically.

Explicitly, we parametrized the subset of quantum distributions accessible by bipartite qubit density matrices and two-outcome positive-operator-valued measure (POVM) measurements.[6] This preliminary investigation did not yield interesting solutions, as none of the triangle structure inequalities in [21] or [19] were violated.[7] Unfortunately, these early results are inconclusive for two reasons. First, an exhaustive search would need to consider the possibility of larger Hilbert spaces for the shared quantum states. Second, it is known that the inequalities [21] are incomplete; there exist other constraints on two-outcome distributions for the triangle structure yet to be discovered.

Nonetheless, having failed to observe a quantum-classical gap in the triangle structure for binary-outcome measurements, a continued search for nonclassical distributions in the triangle structure must expand the gamut of inequalities to optimize against to include inequalities referencing strictly more than two outcomes.

VII. TRIANGLE STRUCTURE INEQUALITIES

Section II introduced the notion of causal compatibility inequalities and Sec. V discussed the Fritz distribution P_{F} together with the initial inequality-free proof of its incompatibility with the triangle structure. Heretofore, there were no known causal compatibility inequalities for the triangle structure that were capable of witnessing the nonclassicality of any quantum distributions [21]. By leveraging the incompatibility of the Fritz distribution [14] and tools provided by the inflation technique [21], we have obtained numerous causal compatibility inequalities for the triangle structure that are violated by the Fritz distribution. A representative trio of these P_{F}-incompatibility-witnessing inequalities are presented here: $I_{\text{wagon wheel}}$ per the inequality (12), notable for its simplicity; I_{web} per the inequality (13), which best witnesses the nonclassicality of P_{F} in the presence of noise; and $I_{\text{symmetric web}}$ per the inequality (15), which is symmetric with respect to all permutations of the three parties.

Readers who are unfamiliar with the inflation technique [21] and wish to understand in detail how the following inequalities are derived are recommended to consult Appendixes A–C, wherein they will find a succinct yet sufficient presentation of the requisites needed for this paper. To briefly summarize, Appendix B reviews the basics of the inflation technique and formally defines the notion of an inflation of a causal structure. Appendix C demonstrates how to use the inflation technique and a given probability distribution, such as the Fritz distribution, to cast the causal compatibility problem as a particular kind of linear program known as a marginal problem, defined in Appendix A.

The inflation technique is known to completely solve the causal compatibility problem through increasing orders of inflations [22]. Consequently, the derivation of the inequalities (12), (13), and (15) is guaranteed by [22] for sufficiently large inflations. Nonetheless, the following inequalities were obtained using the relatively small inflations found in Fig. 8.[8] This efficiency is not universally guaranteed; for comparison, we remark that, unlike the Fritz distribution, the conjectured incompatibility of the four-outcome distribution proposed in [29] is not confirmed by inflation technique at the same level.

We emphasize that each of these causal compatibility inequalities independently provides a positive resolution of Fritz's problem. The inequalities are derived without making use of Bell's theorem; rather, they follow from the inflation technique's broader perspective on nonclassicality as a special case of causal inference. Afterward, Sec. IX returns to the topic of Fritz's problem and the status of our resolution.

A. Wagon-wheel inequality

The first causal compatibility inequality chosen for presentation, the wagon-wheel inequality $I_{\text{wagon wheel}}$, is reported below:

$$\begin{aligned}
&+P_{A_lB_l}(11) - P_{A_lB_lC_lC_r}(1111) + P_{A_lB_l}(00)P_{C_lC_r}(11)\\
&\quad + P_{C_lC_r}(01)P_{C_lC_r}(10) - P_{C_lC_r}(11)P_{A_lA_rB_lB_rC_lC_r}(000000)\\
&\quad - P_{C_lC_r}(11)P_{A_lA_rB_lB_rC_lC_r}(010100)\\
&\quad - P_{C_lC_r}(10)P_{A_lA_rB_lB_rC_lC_r}(001001)\\
&\quad - P_{C_lC_r}(10)P_{A_lA_rB_lB_rC_lC_r}(011101)\\
&\quad - P_{C_lC_r}(01)P_{A_lA_rB_lB_rC_lC_r}(100110)\\
&\quad - P_{C_lC_r}(01)P_{A_lA_rB_lB_rC_lC_r}(110010)\\
&\quad + P_{C_lC_r}(00)P_{A_lA_rB_lB_rC_lC_r}(101111)\\
&\quad + P_{C_lC_r}(00)P_{A_lA_rB_lB_rC_lC_r}(111011) \leqslant 0. \qquad (12)
\end{aligned}$$

The inequality (12) is termed the wagon-wheel inequality and is denoted by $I_{\text{wagon wheel}}$ after the (identically named) inflated structure of Fig. 8(b) used to derive it. To reiterate, a summary of the methods used to derive the inequality (12) can be found in Appendixes A–C. Moreover, note that the inequality (12) is reported using the same two-bit notation discussed in Sec. V such that $P_{A_lA_rB_lB_rC_lC_r}(a_la_rb_lb_rc_lc_r) = P_{ABC}(abc)$.

Aside from increasing outcome cardinality, we are also forced to consider larger inflations than those analyzed in [21]. This is because we found that the smaller inflations considered there, such as the spiral inflation depicted in Fig. 8(a), were simply unable to witness the incompatibility of the Fritz distribution, even when analyzed explicitly using four possible outcomes for every observable variable.

By construction, every distribution P_{ABC} that is compatible with the triangle structure per Eq. (5) must satisfy $I_{\text{wagon wheel}}$. On the other hand, the Fritz distribution violates $I_{\text{wagon wheel}}$ with violation $\frac{1}{16} \not\leqslant 0$. Consequently, we affirm that the Fritz

[6] In Sec. VIII we discuss methods for conducting similar parametrizations of quantum states and measurements.

[7] We also directly checked all two-outcome coarse grainings of the Fritz distribution against the inequalities in [19,21], with no violation.

[8] Note that the adjectives *large* and *small* used to describe an inflation only become well defined in the context of the hierarchy proposed in [22]. For example, the web inflation in Fig. 8(c) is second in the hierarchy of [22], whereas the wagon-wheel inflation in Fig. 8(b) is somewhere between the first and second order.

distribution is incompatible with the triangle structure, a fact previously only demonstrated without inequalities.

B. Web inequality

We next present the inequality (13), another causal compatibility inequality for the triangle structure that is violated by the Fritz distribution. The web inflation shown in Fig. 8(c) was used to produce the eponymous I_{web} per the inequality (13). The web inflation is considerably larger than the wagon-wheel inflation; it is computationally more demanding to work with, albeit capable of yielding strictly stronger inequalities. For brevity, we employ the shorthand $P(abc)$ in lieu of $P_{ABC}(abc)$ in presenting the inequality I_{web}:

$$
\begin{aligned}
&+P(000)P(202)+P(000)P(212)+P(202)P(233)+P(302)P(312)\\
&-2P(123)P(210)-2P(123)P(310)-2P(130)P(213)-2P(133)P(210)-2P(133)P(310)-P(000)P(003)-P(000)P(013)-P(000)P(023)\\
&-P(000)P(033)-P(000)P(103)-P(000)P(113)-P(000)P(123)-P(000)P(133)-P(000)P(203)-P(000)P(213)-P(003)P(010)\\
&-P(003)P(020)-P(003)P(030)-P(003)P(100)-P(003)P(110)-P(003)P(120)-P(003)P(130)-P(003)P(200)-P(003)P(210)\\
&-P(003)P(220)-P(003)P(230)-P(003)P(300)-P(003)P(310)-P(003)P(320)-P(003)P(330)-P(010)P(013)-P(010)P(021)\\
&-P(010)P(023)-P(010)P(033)-P(010)P(100)-P(010)P(103)-P(010)P(113)-P(010)P(123)-P(010)P(133)-P(010)P(203)\\
&-P(010)P(213)-P(013)P(020)-P(013)P(030)-P(013)P(100)-P(013)P(110)-P(013)P(120)-P(013)P(130)-P(013)P(200)\\
&-P(013)P(210)-P(013)P(220)-P(013)P(230)-P(013)P(300)-P(013)P(310)-P(013)P(320)-P(013)P(330)-P(020)P(023)\\
&-P(020)P(033)-P(020)P(103)-P(020)P(113)-P(020)P(123)-P(020)P(133)-P(020)P(203)-P(020)P(210)-P(020)P(211)\\
&-P(020)P(212)-P(020)P(223)-P(020)P(233)-P(020)P(310)-P(020)P(311)-P(020)P(312)-P(020)P(313)-P(021)P(100)\\
&-P(021)P(121)-P(021)P(210)-P(021)P(211)-P(021)P(213)-P(021)P(310)-P(021)P(311)-P(021)P(313)-P(022)P(210)\\
&-P(022)P(211)-P(022)P(212)-P(022)P(213)-P(022)P(310)-P(022)P(311)-P(022)P(312)-P(022)P(313)-P(023)P(030)\\
&-P(023)P(100)-P(023)P(110)-P(023)P(120)-P(023)P(130)-P(023)P(200)-P(023)P(211)-P(023)P(212)-P(023)P(213)\\
&-P(023)P(220)-P(023)P(230)-P(023)P(300)-P(023)P(311)-P(023)P(312)-P(023)P(313)-P(023)P(320)-P(023)P(330)\\
&-P(030)P(033)-P(030)P(103)-P(030)P(113)-P(030)P(123)-P(030)P(133)-P(030)P(203)-P(030)P(210)-P(030)P(211)\\
&-P(030)P(212)-P(030)P(223)-P(030)P(233)-P(030)P(310)-P(030)P(311)-P(030)P(312)-P(030)P(313)-P(031)P(210)\\
&-P(031)P(211)-P(031)P(213)-P(031)P(223)-P(031)P(310)-P(031)P(311)-P(031)P(313)-P(032)P(210)-P(032)P(211)\\
&-P(032)P(212)-P(032)P(213)-P(032)P(310)-P(032)P(311)-P(032)P(312)-P(032)P(313)-P(033)P(100)-P(033)P(110)\\
&-P(033)P(120)-P(033)P(130)-P(033)P(200)-P(033)P(211)-P(033)P(212)-P(033)P(213)-P(033)P(220)-P(033)P(230)\\
&-P(033)P(300)-P(033)P(311)-P(033)P(312)-P(033)P(313)-P(033)P(320)-P(033)P(330)-P(100)P(103)-P(100)P(113)\\
&-P(100)P(123)-P(100)P(133)-P(100)P(203)-P(100)P(213)-P(103)P(110)-P(103)P(120)-P(103)P(130)-P(103)P(200)\\
&-P(103)P(210)-P(103)P(220)-P(103)P(230)-P(103)P(300)-P(103)P(310)-P(103)P(320)-P(103)P(330)-P(110)P(113)\\
&-P(110)P(123)-P(110)P(133)-P(110)P(203)-P(110)P(213)-P(113)P(120)-P(113)P(130)-P(113)P(200)-P(113)P(210)\\
&-P(113)P(220)-P(113)P(230)-P(113)P(300)-P(113)P(310)-P(113)P(320)-P(113)P(330)-P(120)P(123)-P(120)P(133)\\
&-P(120)P(203)-P(120)P(210)-P(120)P(211)-P(120)P(212)-P(120)P(223)-P(120)P(233)-P(120)P(310)-P(120)P(311)\\
&-P(120)P(312)-P(120)P(313)-P(121)P(131)-P(121)P(210)-P(121)P(211)-P(121)P(213)-P(121)P(310)-P(121)P(311)\\
&-P(121)P(312)-P(121)P(313)-P(122)P(210)-P(122)P(211)-P(122)P(212)-P(122)P(213)-P(122)P(310)-P(122)P(311)\\
&-P(122)P(312)-P(122)P(313)-P(123)P(130)-P(123)P(200)-P(123)P(211)-P(123)P(212)-P(123)P(213)-P(123)P(220)\\
&-P(123)P(230)-P(123)P(300)-P(123)P(311)-P(123)P(312)-P(123)P(313)-P(123)P(320)-P(123)P(330)-P(130)P(133)\\
&-P(130)P(203)-P(130)P(210)-P(130)P(211)-P(130)P(212)-P(130)P(223)-P(130)P(233)-P(130)P(310)-P(130)P(311)\\
&-P(130)P(312)-P(130)P(313)-P(131)P(210)-P(131)P(211)-P(131)P(213)-P(131)P(310)-P(131)P(311)-P(131)P(313)\\
&-P(132)P(210)-P(132)P(211)-P(132)P(212)-P(132)P(213)-P(132)P(310)-P(132)P(311)-P(132)P(312)-P(132)P(313)\\
&-P(133)P(200)-P(133)P(211)-P(133)P(212)-P(133)P(213)-P(133)P(220)-P(133)P(230)-P(133)P(300)-P(133)P(311)\\
&-P(133)P(312)-P(133)P(313)-P(133)P(320)-P(133)P(330)-P(200)P(203)-P(200)P(213)-P(200)P(223)-P(200)P(233)\\
&-P(200)P(303)-P(200)P(313)-P(200)P(323)-P(200)P(333)-P(203)P(210)-P(203)P(220)-P(203)P(230)-P(203)P(300)\\
&-P(203)P(310)-P(203)P(320)-P(203)P(330)-P(210)P(213)-P(210)P(223)-P(210)P(233)-P(210)P(303)-P(210)P(313)\\
&-P(210)P(323)-P(210)P(333)-P(212)P(212)-P(212)P(333)-P(213)P(220)-P(213)P(230)-P(213)P(300)-P(213)P(310)\\
&-P(213)P(320)-P(213)P(330)-P(220)P(223)-P(220)P(233)-P(220)P(303)-P(220)P(313)-P(220)P(323)-P(220)P(333)\\
&-P(223)P(230)-P(223)P(300)-P(223)P(310)-P(223)P(320)-P(223)P(330)-P(230)P(233)-P(230)P(303)-P(230)P(313)\\
&-P(230)P(323)-P(230)P(333)-P(233)P(300)-P(233)P(310)-P(233)P(320)-P(233)P(330)-P(233)P(333)-P(300)P(303)\\
&-P(300)P(313)-P(300)P(323)-P(300)P(333)-P(303)P(310)-P(303)P(320)-P(303)P(330)-P(310)P(313)-P(310)P(323)\\
&-P(310)P(333)-P(313)P(320)-P(313)P(330)-P(320)P(323)-P(320)P(333)-P(323)P(330)-P(330)P(333)\le 0.
\end{aligned}
\tag{13}
$$

As mentioned in Sec. V, this original proof of the nonclassicality of P_{F} required an idealistic condition to hold: perfect correlations of C with A_l and B_l. It is impossible to use Fritz's original argument to confirm nonclassicality

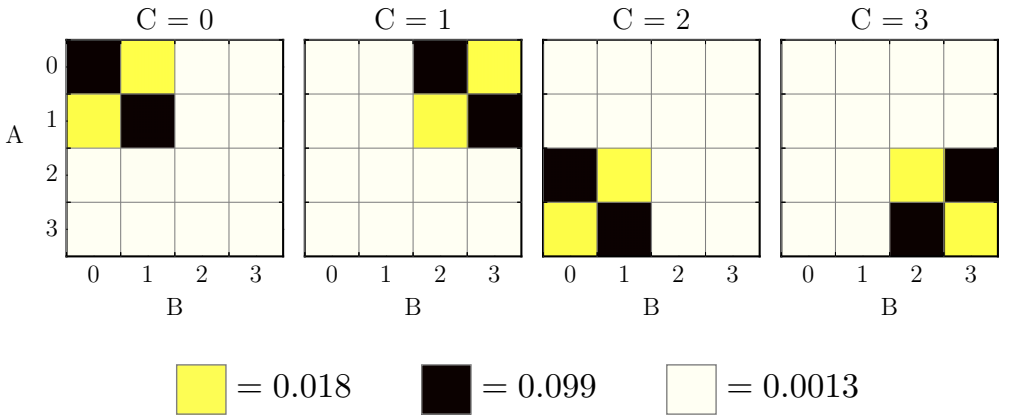

FIG. 5. Distribution $\mathcal{N}_{0.085}$, a noisy yet still nonclassical variant of the Fritz distribution.

from an experimental point of view, because every laboratory-achievable distribution is subject to some amount of noise. One can minimize noise, e.g., by developing high-accuracy measurement channels, but perfect correlations are unattainable. Causal compatibility inequalities such as I_{web} permit there to be noise within a set of observations before the ability to certify nonclassicality breaks down.

Here we quantify how much statistical noise can be added to the Fritz distribution P_{F} before I_{web} fails to witness incompatibility; we define the ε-noisy Fritz distribution as

$$\mathcal{N}_\varepsilon = (1-\varepsilon)P_{\text{F}} + \varepsilon \mathcal{U}_{ABC}, \tag{14}$$

where $\mathcal{U}_{ABC}(abc) \equiv \frac{1}{64}\ \forall a, b, c \in \{0, 1, 2, 3\}$. As ε varies from 0 to 1, noise is added to the Fritz distribution and $\mathcal{N}_\varepsilon$ transitions from an incompatible distribution $\mathcal{N}_0 = P_{\text{F}}$ to a compatible distribution $\mathcal{N}_1 = \mathcal{U}_{ABC}$. We find that I_{web} demonstrates that the Fritz distribution remains incompatible with the triangle structure up to a noise parameter of $\varepsilon \simeq 0.085$; the associated distribution $\mathcal{N}_{0.085}$ is plotted in Fig. 5. Of course, there remains the possibility that another inequality will be able to withstand a larger degree of noise than I_{web}; an exhaustive search has not been conducted.

C. Symmetric web inequality

In Sec. VIII we will consider numerically optimizing our inequalities over quantum strategies (seeking violations) towards the desired end goal of discovering different forms of nonclassicality in the triangle structure. Of course, if some inequality achieves its optimal quantum violation on a distribution qualitatively similar to P_{F}, then such an inequality is unlikely to lead us to discover nonclassicality of a type different from P_{F}. Indeed, we will show that the previous pair of $I_{\text{wagon wheel}}$ and I_{web}, which were both specifically curated to demonstrate the incompatibility of P_{F}, apparently achieves maximum quantum violation on distributions very similar to P_{F} itself.

In order for the numerical optimization to avoid distributions similar to P_{F} we require an objective function which is not tailor designed to witness P_{F}. Finding such an inequality proved extremely challenging; while we could generate thousands of inequalities using the inflation technique [21], it appeared than only a vanishingly small fraction of those inequalities admitted quantum violation whatsoever. We therefore modified the inflation technique to give us an inequality which not only witnesses the nonclassicality of P_{F} but is also symmetric with respect to any permutation of the variables A, B, and C, resulting in the inequality (15). The modification of the inflation techniques which forces symmetry is explained in Appendix D. As P_{F} is strongly asymmetric due to the special role of C, we can hope that a symmetric inequality (even one which does witness P_{F}) might achieve its optimal quantum violation on a distribution qualitative distinct from P_{F}. The details of our numerical finding are discussed in Sec. VIII. We have the symmetric web inequality $I_{\text{symmetric web}}$,

$$
\begin{aligned}
&+2[P(312)P(312)]_6 + 3[P(000)P(323)]_3 + 3[P(131)P(323)]_3 + [P(000)P(121)]_3 + [P(110)P(333)]_3 + [P(202)P(202)]_3 \\
&-12[P(002)P(031)]_6 - 12[P(010)P(330)]_6 - 12[P(032)P(210)]_6 - 12[P(101)P(331)]_6 - 12[P(122)P(321)]_6 - 12[P(130)P(312)]_6 \\
&-12[P(202)P(220)]_6 - 16[P(002)P(120)]_6 - 16[P(002)P(200)]_6 - 16[P(002)P(220)]_3 - 16[P(002)P(230)]_6 - 16[P(002)P(302)]_6 \\
&-16[P(002)P(303)]_6 - 16[P(002)P(330)]_3 - 16[P(003)P(032)]_6 - 16[P(003)P(033)]_6 - 16[P(003)P(231)]_6 - 16[P(003)P(300)]_6 \\
&-16[P(003)P(310)]_6 - 16[P(003)P(320)]_6 - 16[P(003)P(322)]_6 - 16[P(012)P(202)]_6 - 16[P(012)P(210)]_6 - 16[P(012)P(313)]_6 \\
&-16[P(013)P(120)]_6 - 16[P(013)P(130)]_6 - 16[P(013)P(203)]_6 - 16[P(013)P(220)]_6 - 16[P(020)P(101)]_3 - 16[P(020)P(103)]_6 \\
&-16[P(020)P(312)]_6 - 16[P(021)P(311)]_6 - 16[P(021)P(332)]_6 - 16[P(022)P(203)]_6 - 16[P(022)P(302)]_6 - 16[P(022)P(303)]_6 \\
&-16[P(022)P(311)]_3 - 16[P(022)P(313)]_6 - 16[P(023)P(103)]_6 - 16[P(023)P(222)]_6 - 16[P(023)P(232)]_6 - 16[P(030)P(201)]_6 \\
&-16[P(031)P(301)]_6 - 16[P(031)P(332)]_6 - 16[P(032)P(311)]_6 - 16[P(033)P(202)]_6 - 16[P(033)P(300)]_3 - 16[P(033)P(312)]_6 \\
&-16[P(033)P(331)]_6 - 16[P(033)P(333)]_3 - 16[P(102)P(130)]_6 - 16[P(102)P(220)]_6 - 16[P(102)P(230)]_6 - 16[P(102)P(310)]_6 \\
&-16[P(103)P(110)]_6 - 16[P(103)P(133)]_6 - 16[P(113)P(130)]_6 - 16[P(113)P(220)]_3 - 16[P(113)P(321)]_6 - 16[P(120)P(201)]_6 \\
&-16[P(120)P(223)]_6 - 16[P(120)P(303)]_6 - 16[P(120)P(313)]_6 - 16[P(123)P(230)]_6 - 16[P(123)P(302)]_6 - 16[P(123)P(330)]_6 \\
&-16[P(130)P(203)]_6 - 16[P(130)P(223)]_6 - 16[P(131)P(303)]_3 - 16[P(132)P(202)]_6 - 16[P(133)P(202)]_6 - 16[P(201)P(330)]_6 \\
&-16[P(203)P(330)]_6 - 16[P(210)P(213)]_6 - 16[P(222)P(303)]_3 - 16[P(223)P(310)]_6 - 16[P(231)P(301)]_6 - 16[P(232)P(330)]_6 \\
&-16[P(233)P(303)]_6 - 16[P(300)P(331)]_6 - 16[P(301)P(330)]_6 - 16[P(302)P(331)]_6 - 16[P(313)P(320)]_6 - 20[P(132)P(301)]_6 \\
&-20[P(133)P(301)]_6 - 24[P(033)P(310)]_6 - 24[P(102)P(231)]_6 - 24[P(103)P(331)]_6 - 24[P(130)P(303)]_6 - 2[P(010)P(031)]_6 \\
&-2[P(100)P(100)]_3 - 2[P(100)P(223)]_6 - 2[P(110)P(223)]_3 - 2[P(121)P(312)]_6 - 2[P(212)P(302)]_6 - 3[P(000)P(333)]_1 \\
&-3[P(031)P(312)]_6 - 3[P(223)P(333)]_3 - 3[P(302)P(312)]_6 - 4[P(000)P(022)]_3 - 4[P(000)P(212)]_3 - 4[P(000)P(301)]_6 \\
&-4[P(001)P(031)]_6 - 4[P(003)P(211)]_6 - 4[P(010)P(132)]_6 - 4[P(010)P(311)]_6 - 4[P(011)P(211)]_3 - 4[P(012)P(021)]_6 \\
&-4[P(022)P(122)]_3 - 4[P(022)P(222)]_3 - 4[P(022)P(232)]_6 - 4[P(031)P(233)]_6 - 4[P(100)P(122)]_3 - 4[P(100)P(230)]_6 \\
&-4[P(101)P(131)]_3 - 4[P(110)P(123)]_6 - 4[P(110)P(320)]_6 - 4[P(111)P(311)]_3 - 4[P(111)P(321)]_6 - 4[P(122)P(122)]_3 \\
&-4[P(122)P(220)]_6 - 4[P(122)P(232)]_6 - 4[P(123)P(131)]_6 - 4[P(132)P(332)]_6 - 4[P(201)P(201)]_6 - 4[P(201)P(311)]_6 \\
&-4[P(202)P(311)]_6 - 4[P(202)P(321)]_6 - 4[P(212)P(310)]_6 - 4[P(212)P(331)]_6 - 4[P(223)P(312)]_6 - 4[P(302)P(333)]_6 \\
&-4[P(311)P(323)]_6 - 5[P(010)P(333)]_3 - 5[P(031)P(100)]_6 - 5[P(223)P(223)]_3 - 5[P(233)P(333)]_3 - 8[P(000)P(030)]_3 \\
&-8[P(000)P(032)]_6 - 8[P(003)P(111)]_3 - 8[P(003)P(331)]_3 - 8[P(012)P(030)]_6 - 8[P(012)P(110)]_6 - 8[P(012)P(221)]_6 \\
&-8[P(013)P(021)]_6 - 8[P(013)P(022)]_6 - 8[P(013)P(311)]_6 - 8[P(020)P(100)]_6 - 8[P(020)P(110)]_6 - 8[P(020)P(112)]_6 \\
&-8[P(020)P(210)]_6 - 8[P(022)P(121)]_6 - 8[P(022)P(130)]_6 - 8[P(023)P(112)]_6 - 8[P(023)P(213)]_6 - 8[P(030)P(210)]_6 \\
&-8[P(030)P(311)]_6 - 8[P(032)P(111)]_6 - 8[P(033)P(121)]_6 - 8[P(033)P(131)]_6 - 8[P(100)P(132)]_6 - 8[P(102)P(321)]_6 \\
&-8[P(110)P(300)]_6 - 8[P(111)P(130)]_6 - 8[P(112)P(123)]_6 - 8[P(112)P(331)]_3 - 8[P(113)P(200)]_6 - 8[P(113)P(331)]_3 \\
&-8[P(120)P(220)]_6 - 8[P(120)P(302)]_6 - 8[P(120)P(310)]_6 - 8[P(121)P(211)]_6 - 8[P(121)P(301)]_6 - 8[P(121)P(303)]_3 \\
&-8[P(121)P(320)]_6 - 8[P(122)P(201)]_6 - 8[P(122)P(203)]_6 - 8[P(123)P(200)]_6 - 8[P(123)P(203)]_6 - 8[P(123)P(222)]_6 \\
&-8[P(123)P(313)]_6 - 8[P(130)P(212)]_6 - 8[P(131)P(310)]_6 - 8[P(132)P(310)]_6 - 8[P(132)P(312)]_6 - 8[P(132)P(322)]_6 \\
&-8[P(133)P(200)]_3 - 8[P(133)P(213)]_6 - 8[P(133)P(222)]_3 - 8[P(133)P(300)]_3 - 8[P(133)P(312)]_6 - 8[P(133)P(322)]_3 \\
&-8[P(201)P(313)]_6 - 8[P(210)P(320)]_6 - 8[P(211)P(221)]_6 - 8[P(211)P(320)]_6 - 8[P(213)P(322)]_6 - 8[P(220)P(223)]_3 \\
&-8[P(221)P(312)]_6 - 8[P(222)P(312)]_6 - 8[P(231)P(312)]_6 - 8[P(232)P(313)]_3 - 8[P(233)P(313)]_6 - 8[P(302)P(332)]_6 \\
&-8[P(311)P(321)]_6 - [P(000)P(121)]_3 - [P(000)P(323)]_3 - [P(110)P(333)]_3 - [P(121)P(333)]_3 - [P(131)P(323)]_3 \\
&-[P(312)P(312)]_6 \leq 0.
\end{aligned}
\tag{15}
$$

Note that $P(abc)$ is shorthand for $P_{ABC}(abc)$ and $[P(113)P(330)]_3$ is shorthand for a sum the over permutations of A, B, and C, e.g., $[P(113)P(330)]_3 \equiv P(113)P(330) + P(131)P(303) + P(311)P(033)$.

VIII. NUMERICAL OPTIMIZATION

In order to find quantum distributions that are more nonclassical than the Fritz distribution, we perform numerical optimizations against each inequality I by parametrizing the space of quantum-accessible probability distributions that can be realized on the triangle structure (Fig. 2) and thus expressed in the form of Eq. (6). In order to parametrize all such distributions, we elect to parametrize the states and measurements separately. In order to qualify the scope of Eq. (6) and associated computational complexity of the parametrization, there are a two restrictions that are made with justification. Motivated by the fact that the Fritz distribution (Sec. V) only requires qubit states, the states ρ are taken to be bipartite qubit states which are more computationally feasible compared to n-dimensional states whereby the joint density matrix $\rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA}$ becomes an $n^6 \times n^6$ matrix. Additionally, we restrict our focus to projective-valued measures (PVMs) instead of POVMs for three reasons. First, Fritz [14] demonstrates via the Fritz distribution that PVMs are sufficient for generating incompatible quantum distributions in the triangle structure. Second, although generating k-outcome POVM measurements is possible using rejection sampling techniques [30], a valid unbiased parametrization is not found for $k > 2$. Finally, PVMs provide considerable computational advantage over POVMs as they permit Eq. (6) to be rewritten as

$$
\begin{aligned}
P_{ABC}(abc) = \langle m_{A,a} m_{B,b} m_{C,c} | \Pi^{\intercal} \rho_{AB} \otimes \rho_{BC} \\
\otimes \rho_{CA} \Pi | m_{A,a} m_{B,b} m_{C,c} \rangle.
\end{aligned}
\tag{16}
$$

Although there are numerous techniques that can used when parametrizing quantum states and measurements [4,30–34], a single technique by Spengler *et al.* [35] was found to be most computationally suitable for our purposes. Spengler *et al.* [35] demonstrated that all $d \times d$ unitary matrices U can be parameterized without degeneracy as

$$U = \left[\prod_{m=1}^{d-1}\left(\prod_{n=m+1}^{d}\exp(i\,P_n\lambda_{n,m})\exp(i\sigma_{m,n}\lambda_{m,n})\right)\right] \times \left(\prod_{l=1}^{d}\exp(i\,P_l\lambda_{l,l})\right), \tag{17}$$

where the real-valued parameters $\lambda = \{\lambda_{n,m} \mid n, m \in 1,\ldots,d\}$ have periodicities $\lambda_{m,n} \in [0, \frac{\pi}{2}]$ for $m < n$ and $\lambda_{m,n} \in [0, 2\pi]$ for $m \geqslant n$. Moreover, P_l are one-dimensional projective operators $P_l = |l\rangle\langle l|$ and the $\sigma_{m,n}$ are generalized antisymmetric σ matrices $\sigma_{m,n} = -i|m\rangle\langle n| + i|n\rangle\langle m|$, where $1 \leqslant m < n \leqslant d$. This parametrization has the useful feature that each of the real-valued parameters $\lambda_{n,m}$ has a direct and intuitive physical affect on each element of a computational basis $\{|1\rangle,\ldots,|d\rangle\}$. Explicitly, $\exp(i\sigma_{m,n}\lambda_{m,n})$ applies a rotation to the subspace spanned by $|m\rangle$ and $|n\rangle$ for $m < n$. Analogously, $\exp(i\,P_n\lambda_{n,m})$ generates the relative phase between $|m\rangle$ and $|n\rangle$ for $m > n$ and $\exp(i\,P_l\lambda_{l,l})$ fixes the global phase of $|l\rangle$. Finally, although not explicitly mentioned in [35], it is possible to remove the reliance on the computationally expensive matrix exponential operations [36] in Eq. (17) and replace them with elementary trigonometric functions in the parameters $\lambda_{m,n}$.

By parametrizing unitary matrices, it becomes possible to parametrize d-dimensional density matrices and d-element PVMs by recognizing that any orthonormal basis $\{|\psi_j\rangle\}$ (where $1 \leqslant j \leqslant d$) can be transformed into the computational basis $\{|j\rangle\}$ by a unitary transformation U, i.e., $U|\psi_j\rangle = |j\rangle$. First consider a d-element PVM $M = \{|m_j\rangle\langle m_j| \mid 1 \leqslant j \leqslant d\}$. Since $\{|m_j\rangle\}$ forms an orthonormal basis, one can parametrize M by writing $M = \{U^\dagger|j\rangle\langle j|U \mid 1 \leqslant j \leqslant d\}$ and parametrizing U using Eq. (17). This method was inspired by the measurement seeding method for iterative optimization used by Pál and Vértesi [37]. Analogously, this argument can be extended to full-rank d-dimensional density matrices ρ by performing a spectral decomposition $\rho = \sum_{j=1}^{d} p_j|p_j\rangle\langle p_j|$ into eigenvalues $\{p_j\}$ and eigenstates $\{|p_j\rangle\}$. Since $\mathrm{Tr}(\rho) = \sum_{j=1}^{d} p_j = 1$ the eigenvalues of ρ are parametrized without degeneracy using a tuple of $d-1$ real-valued parameters with periodicity $[0, 2\pi]$ using hyperspherical coordinates [31,35]. Additionally, since ρ is Hermitian, the eigenstates $\{|p_j\rangle\}$ form an orthonormal basis and therefore the eigenstates are analogously parametrized using Eq. (17): $\rho = \sum_{j=1}^{d} p_j U^\dagger|j\rangle\langle j|U$. For our purposes, we have set $d = 4$ and fixed $\lambda_{l,l} = 0$ for $1 \leqslant l \leqslant d$ in Eq. (17) because the global phase contributions are irrelevant for Eq. (16).

Since the inequalities we seek to optimally violate are polynomial and since the space of quantum-accessible distributions is nonconvex, we employ a number of optimization methods consecutively in an attempt to avoid pitfalls associated identifying local minima and ill-conditioned convergence. Specifically, the Broyden-Fletcher-Goldfarb-Shanno method [38, p.142] and the Nelder-Mead simplex method (see [38], p. 238) were used along with a method called basin hopping [39], which is a hybrid between simulated annealing and gradient-descent-based methods.

Nevertheless, we strongly caution against misinterpreting the numerical optima present in Sec. VIII as if our findings represent genuine global maxima of violation. Evidence for the unreliability of our numerical methods is the following: When supplied with randomly sampled initial parameters, all optimization methods consistently converged to saturating the target inequalities, instead of violating it, even though we know that all the inequalities we consider do admit quantum violation, namely, by P_{F}. To achieve inequality violation, we found ourselves forced to initialize the numerical optimizer with parameters that were nearby (in parameter space) to parameters which generate the Fritz distribution P_{F}. Upon doing so, it was observed that the numerical methods converged invariably to parameters which generated distributions similar to P_{F} (see Fig. 6, for instance), consequently making interpreting the results somewhat murky. Could it be that the inequality's global maximum is in fact not far from P_{F}? Or is this a limitation of the ill-conditioned nature of the optimization? The phenomenon is likely due to a combination of both effects.

Numerical optimization results

Our best numerical optimization of $I_{\text{wagon wheel}}$ was found to be the Fritz distribution itself, visualized in Fig. 4. Therefore, P_{F} is either a local minimum of the parameter space or in fact the maximally violating distribution of $I_{\text{wagon wheel}}$; it remains unclear to us if this behavior is related to the global optimality of the Fritz distribution or if it is an artifact of the methods used to derive $I_{\text{wagon wheel}}$, though we suspect the former.

Our best numerical optimization of I_{web} is visualized in Fig. 6. Almost immediately, it is evident that Fig. 6 closely resembles the Fritz distribution; in fact Figs. 6 and 4 share their possibilistic structure. A distribution's possibilistic structure is the subset of events which it assigns a non-zero probability. That Fig. 6 and Fig. 4 have the same possibilistic structure is not entirely unexpected, as was mentioned in Sec. VII, as I_{web} was derived specifically to prove the incompatibility of the Fritz distribution with the triangle structure. Consequently, it seems that I_{web} best witnesses incompatibility of those distributions that closely resemble the P_{F}. It could be, however, that the sharing of possibilistic structure is an artifact of unreliable convergence, such that the true optimally violating I_{web} quantum distribution might have a different structure.

It is perhaps curious, however, that violation achieved by the distribution in Fig. 6 is not accessible when the bipartite states in Eq. (6) are restricted to maximally entangled states. This finding of more violation with less entanglement resembles a feature of quantum mechanics originally presented by Methot and Scarani [40] demonstrating that entanglement and nonclassicality are different resources.

Our best numerical optimization of $I_{\text{symmetric web}}$ is visualized in Fig. 7. Counterintuitively, the distribution in Fig. 7, which achieves a greater violation of $I_{\text{symmetric web}}$ than P_{F}, shares its possibilistic structure with the Fritz distribution.

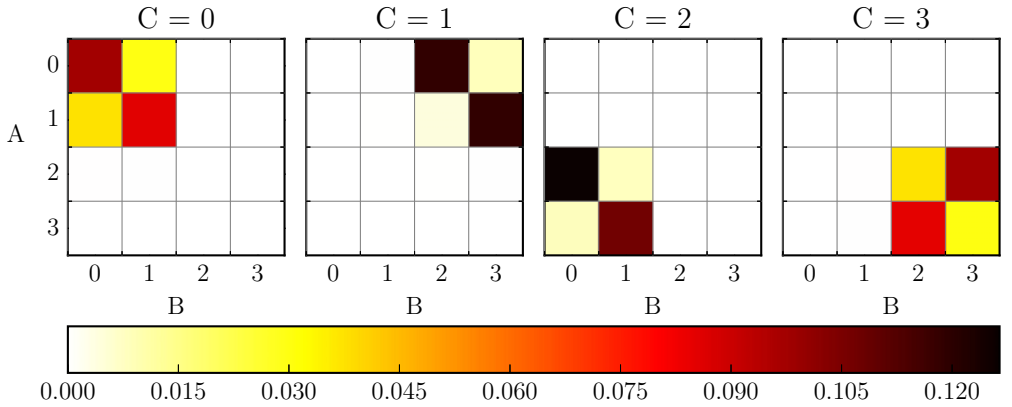

FIG. 6. Quantum probability distribution of the triangle structure that maximizes violation of I_{web}. Notice that this distribution has precisely the same possibilistic structure as the Fritz distribution, as might be expected, since I_{web} was specifically generated to witness the nonclassicality of P_{F}.

Additionally, since $I_{\text{symmetric web}}$ is symmetric with respect to the permutation of parties, one might expect that the maximum violating distribution be symmetric as well. To the contrary, our numerical optimization failed to find any symmetric quantum distributions capable of violating $I_{\text{symmetric web}}$.

IX. REVISITING FRITZ'S PROBLEM

In Sec. V we defined the Fritz distribution and summarized the inequality-free proof of its incompatibility with the triangle structure due to [14]. In addition, Sec. V discusses Fritz's problem as a quest to find quantum distributions incompatible with the triangle structure that are also qualitatively distinct from those distributions which are incompatible with the Bell structure. In light of the perhaps unsatisfactory results of Sec. VIII, the purpose of this section is to revisit Fritz's problem and to attempt to rigorously formulate a criterion for quantum nonclassicality that is genuine to the triangle structure.

When specifically concerned with entanglement resources, Bell structure nonclassicality exploits the entanglement of a single bipartite quantum state. Unlike the Bell structure in Fig. 1, the observable nodes of the triangle structure in Fig. 2 have access to potentially three different bipartite quantum states. As noted in [14], the Fritz distribution can be implemented using only a single entangled state.[9] Therefore, it becomes reasonable to propose that any quantum nonclassicality in the triangle structure that requires entanglement in at least two shared states constitutes nonclassicality distinct from Bell nonclassicality. At first, one might suspect that finding such distributions would affirmatively demonstrate a form of entanglement resource unique to the triangle structure. Unfortunately, this is not the case. By increasing the cardinality of each variable A, B, and C to $4^3 = 64$, it is possible to generate a distribution that is nonclassical yet quantum and requires all three of the shared resources to be entangled. This can be accomplished by superimposing three copies of statistically independent Fritz distributions to the triangle structure, each

[9]Of course there exist implementations that make use of three entangled states, but at minimum, one is necessary.

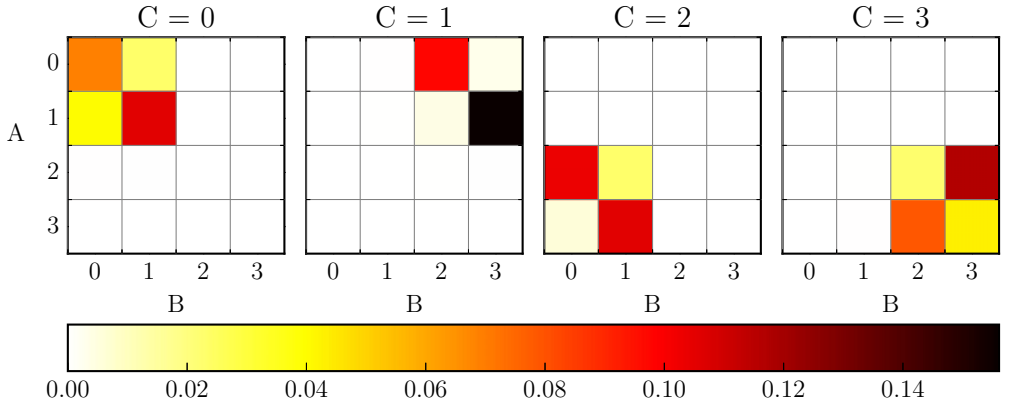

FIG. 7. Quantum probability distribution of the triangle structure that maximizes violation of $I_{\text{symmetric web}}$. Notice that this distribution has precisely the same (asymmetric) possibilistic structure as the Fritz distribution, even though $I_{\text{symmetric web}}$ itself is symmetric with respect to permutations of the variables A, B, and C.

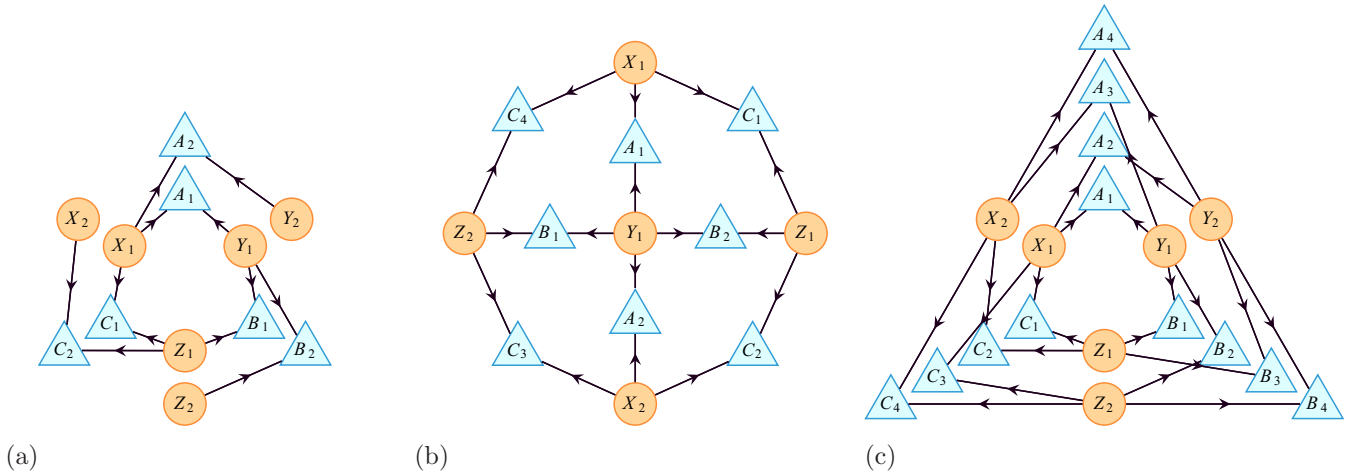

FIG. 8. Some inflations of the triangle structure: (a) spiral inflation, (b) wagon-wheel inflation, and (c) web inflation.

of which utilizes a distinct party to announce the corresponding measurement pseudosettings. Such a distribution would require entanglement in all three quantum states; under this construction, the removal of any entangled resource would render the distribution quantum inaccessible.[10] Consequently, demanding the necessity of entanglement in every shared state is insufficient for finding novel nonclassicality.

In consideration of this particular construction, perhaps the correct assessment for novel nonclassicality must incorporate a restriction on the cardinality of the observed variables. Pursuant to this objective, we were able to prove the nonclassicality of a variant of the Fritz distribution where C has only two outcomes: the first corresponding to A and B correlations and the second corresponding to A and B anticorrelations.[11] Nonetheless, such a distribution fails to deviate significantly from Bell nonclassicality.

It is worth noting that there are quantum-accessible distributions which are conjectured, but not proven, to be incompatible with the triangle structure. For instance, the distribution proposed in [29] required the use of entangled measurements and thus if proven incompatible might constitute a different form of quantum nonclassicality. Unfortunately, as was previously noted in Sec. VII, the proposed distribution in [29] satisfies all inequalities generated by the inflations considered in Fig. 8. Presently, computational limitations prevent us from considering inflations larger than those in Fig. 8, although we remain optimistic.

In truth, properly defining the novelty of quantum correlations and also recognizing the resourcefulness of those correlations in the triangle structure, or any other causal structure, is a deep and meaningful, albeit unsettled, problem. Ultimately, if a satisfactory classifier of novelty is constructed, it remains unclear whether or not novel distributions even exist.

X. CONCLUSION

In Sec. IV we elucidated that establishing or rejecting compatibility with the triangle structure has been a challenging problem for nearly a decade [14,16,18,19,29]. Though some causal compatibility inequalities were known, those inequalities did not appear useful for the purpose of witnessing quantum nonclassicality. Recently, Fritz [14] gave an example of quantum nonclassicality in the triangle structure, relying on an inequality-free proof that is not robust to any amount of noise.

In Sec. VII we presented examples of causal compatibility inequalities capable of having quantum violations in the sense that they are violated by quantum-accessible distributions. This result was made possible through the inflation technique [21] (Appendix B) applied to the Fritz distribution. Moreover, the inequalities in Sec. VII were derived using the inflations in Fig. 8, each of which is low in the hierarchy proposed by [22], thus revealing the relative efficiency of the inflation technique.

In Sec. VII B it was demonstrated that these causal compatibility inequalities are robust to noise, directly revealing a critical departure from Fritz's original proof of the incompatibility of Fritz's perfect-correlation example. In Sec. VIII we found quantum distributions quantitatively distinct from Fritz's example of a recycled Bell theorem, such that these optimized distributions more strongly violate certain causal compatibility inequalities.

Despite these advancements, the distributions we discovered hew closely to the Fritz distribution, indicating that their nonclassical nature remains some recycled version of the nonclassicality found in the Bell structure. Section IX discusses potential proposals certifying the genuineness of nonclassicality in the triangle structure. Presently, the existence, and subsequent suitable classification of fundamentally novel nonclassicality, remains speculative and certainly warrants future research.

[10] This observation was original provided by Navascués [41].

[11] An inequality-free proof of the incompatibility of this distribution, using arguments analogous to those presented in [14], can be found in [20]. Assessing the compatibility of this variant using the methods in Appendixes A–C was first suggested to the authors by Rosset [42].

ACKNOWLEDGMENTS

This research was supported in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported 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 Research, Innovation and Science. The authors are grateful to Miguel Navascués and Denis Rosset for their insightful discussions.

APPENDIX A: MARGINAL SATISFIABILITY AND INEQUALITIES

This Appendix aims to explain how to solve the following decision problem: Given a collection of probability distributions $\{P_{V_1}, \ldots, P_{V_m}\}$ where each set of variables $V_i \subseteq \mathcal{J}$ is a subset of some complete set of variables $\mathcal{J}$, does there exist a joint distribution $P_{\mathcal{J}}$ such that each P_{V_i} can be obtained by marginalizing $P_{\mathcal{J}}$ over the variables not in V_i, i.e., $P_{V_i} = \sum_{\mathcal{J}\setminus V_i} P_{\mathcal{J}}$? Colloquially, this problem is referred to as the marginal problem [43]. Additionally, this section aims to accomplish something further: If such a joint distribution $P_{\mathcal{J}}$ exists, how does one find it? If not, how does one find an inequality whose violation by $\{P_{V_1}, \ldots, P_{V_m}\}$ proves the nonexistence of a joint distribution? This is accomplished by illustrating how the marginal problem can be expressed as a linear program in which the solution to the marginal problem is encoded in the feasibility or infeasibility of said linear program. This Appendix is presented prior to Appendix B as the marginal problem becomes an integral component of the inflation technique in subsequently deriving the inequalities presented in Sec. VII. Moreover, the marginal problem is presented here logically independent from the remainder of the paper both for procedural clarity and because the marginal problem has applications in numerous areas of mathematics including game theory [44], database theory, knowledge integration of expert system, and of course quantum information theory [43].

To begin, several pieces of nomenclature will be introduced to facilitate discussions. First, the set $\mathcal{M} = \{V \mid V \subseteq \mathcal{J}\}$ of subsets of $\mathcal{J}$ is referred to as the marginal scenario and each element $V \in \mathcal{M}$ is termed a (marginal) context of $\mathcal{M}$. The complete set of marginal distributions is referred to as the marginal model and is denoted by a superscript $P^{\mathcal{M}} \equiv \{P_V \mid V \in \mathcal{M}\}$. A marginal model acts as the most general description of a family of observations that can be made over $\mathcal{J}$. Strictly speaking, as defined in [43], a marginal scenario forms an abstract simplicial complex where it is required that all subsets of contexts are also contexts: $V' \subset V : V' \in \mathcal{M}\ \forall V \in \mathcal{M}$. Throughout this Appendix we exclusively consider (without loss of generality) the maximal marginal scenario, restricting our focus to the largest marginal contexts. Additionally, all marginal scenarios are taken to be complete in the sense that the marginal scenario covers the complete set of observable variables, i.e., $\mathcal{J} = \bigcup_{V\in\mathcal{M}} V$. Finally, we henceforth assume that each variable $v \in \mathcal{J}$ has a finite cardinality.

The marginal problem is, given a marginal model $P^{\mathcal{M}} = \{P_V \mid V \in \mathcal{M}\}$, marginal to the joint variables $\mathcal{J}$, whether there exists a joint distribution $P_{\mathcal{J}}$ such that each context P_V can be obtained by marginalizing $P_{\mathcal{J}}$:

$$P_V = \sum_{\mathcal{J}\setminus V} P_{\mathcal{J}}\ \forall V \in \mathcal{M}. \tag{A1}$$

A marginal model $P^{\mathcal{M}}$ is said to be contextual if it does not admit a joint distribution and noncontextual otherwise. Notice that Eq. (A1) is inherently a linear system of constraints which can be solved efficiently using linear programs. In consideration of this, we will now endeavor to discuss how to cast Eq. (A1) as a matrix multiplication equation so that it becomes possible to discuss existing methods for deriving constraints on the set of contextual marginal models.

To every discrete random variable v there corresponds a prescribed set of outcomes O_v. We also define the set of all events over v, denoted by $\mathcal{E}(v)$,[12] to be the set of all functions $s : \{v\} \to O_v$, each representing the event that a measurement on v was made where $s(v) \in O_v$ was observed. Evidently, $\mathcal{E}(v)$ and O_v have a one-to-one correspondence and this distinction can be confounding. There is rarely any harm in referring synonymously to either as outcomes. Nonetheless, a sheaf-theoretic treatment of contextuality [45] demands the distinction. Specifically for this work, the distinction becomes essential for our discussion and exploitation of marginal symmetries in Appendix D. As a natural generalization we define the events over a set of random variables $V = \{v_1, \ldots, v_n\}$ in a parallel manner,

$$\mathcal{E}(V) \equiv \{s : V \to O_V \mid \forall i : s(v_i) \in O_{v_i}\}. \tag{A2}$$

Each event s can be compactly represented as a set of mappings over each element of V, i.e., $s = \{v_i \mapsto s(v_i)\}_{i=1}^{k}$. The domain $\mathcal{D}(s)$ of an event s is the set of random variables it valuates, i.e., if $s \in \mathcal{E}(V)$, then $\mathcal{D}(s) = V$. Under this framework, a probability distribution P_V can be considered as a map from $\mathcal{E}(V)$ to the unit interval $[0, 1]$. The marginal problem inherently depends on the concept of probabilistic marginalization. This concept can be understood at the level of events; one event $s \in \mathcal{E}(V)$ can be marginalized or restricted to a smaller event $s' \in \mathcal{E}(W)$ whenever $W \subseteq V$. For every $W \subseteq V$ and $s \in \mathcal{E}(V)$, the restriction of s onto W [denoted by $s|_W \in \mathcal{E}(W)$] is the event in $\mathcal{E}(W)$ that agrees with each of s's assignments for variables in W: $s|_W(v) = s(v)\ \forall v \in W$.

For every marginal scenario $\mathcal{M} = \{V_1, \ldots, V_k\}$, it is useful to put special emphasis on the joint events $\mathcal{E}(\mathcal{J})$ which represent all possible global events over the entire set of joint variables. Similarly, we define the context events for a particular context $V \in \mathcal{M}$ as $\mathcal{E}(V)$. Finally, we elect to define the marginal events as the disjoint union over all context events and by an abuse of notation we will define this union as $\mathcal{E}(\mathcal{M}) = \coprod_{V\in\mathcal{M}} \mathcal{E}(V)$. Each marginal section $m \in \mathcal{E}(\mathcal{M})$ has a domain $\mathcal{D}(m) = V$ for some $V \in \mathcal{M}$. By construction, each marginal event $m \in \mathcal{E}(\mathcal{M})$ is a restriction of some joint event $j \in \mathcal{E}(\mathcal{J})$.

The marginalization operation of the marginal problem is a linear operation, mapping a joint probability distribu-

[12] In the language of sheaf theory, $\mathcal{E}(v)$ is the sheaf of events [45].

tion $P_{\mathcal{J}} : \mathcal{E}(\mathcal{J}) \to [0, 1]$ into a marginal model $P^{\mathcal{M}} = \{P_V : \mathcal{E}(V) \to [0, 1] \mid V \in \mathcal{M}\}$. Since $\mathcal{E}(\mathcal{M})$ and $\mathcal{E}(\mathcal{J})$ are finite, the marginal problem can be represented as a $|\mathcal{E}(\mathcal{M})| \times |\mathcal{E}(\mathcal{J})|$ matrix.

Definition. The incidence matrix M for a marginal scenario $\mathcal{M} = \{V_1, \ldots, V_k\}$ is a bitwise matrix where the columns are indexed by joint events $j \in \mathcal{E}(\mathcal{J})$ and the rows are events by marginal events $m \in \mathcal{E}(\mathcal{M})$. The entries of M are populated whenever a marginal event m is a restriction of the joint event j,

$$M_{m,j} \equiv \begin{cases} 1 & \text{for } m = j|_{\mathcal{D}(m)} \\ 0 & \text{otherwise.} \end{cases} \tag{A3}$$

The incidence matrix has $|\mathcal{E}(\mathcal{J})|$ columns, $|\mathcal{E}(\mathcal{M})| = \sum_i |\mathcal{E}(V_i)|$ rows, and $k|\mathcal{E}(\mathcal{J})|$ nonzero entries.

To illustrate this concretely, consider the following example. Let $\mathcal{J}$ be three binary variables $\mathcal{J} = \{A, B, C\}$ and $\mathcal{M}$ be the marginal scenario $\mathcal{M} = \{\{A, B\}, \{B, C\}, \{A, C\}\}$. The incidence matrix for $\mathcal{M}$ is

$$M = \begin{array}{c} \\ (A \mapsto 0, B \mapsto 0) \\ (A \mapsto 0, B \mapsto 1) \\ (A \mapsto 1, B \mapsto 0) \\ (A \mapsto 1, B \mapsto 1) \\ (B \mapsto 0, C \mapsto 0) \\ (B \mapsto 0, C \mapsto 1) \\ (B \mapsto 1, C \mapsto 0) \\ (B \mapsto 1, C \mapsto 1) \\ (A \mapsto 0, C \mapsto 0) \\ (A \mapsto 0, C \mapsto 1) \\ (A \mapsto 1, C \mapsto 0) \\ (A \mapsto 1, C \mapsto 1) \end{array} \begin{array}{c} \begin{array}{cccccccc} (A, B, C) \mapsto & (0,0,0) & (0,0,1) & (0,1,0) & (0,1,1) & (1,0,0) & (1,0,1) & (1,1,0) & (1,1,1) \end{array} \\ \begin{pmatrix} \mathbf{1} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{1} \\ \mathbf{1} & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & \mathbf{1} \end{pmatrix} \end{array}. \tag{A4}$$

The incidence matrix acts (to the right) on a vector representing the joint probability $P_{\mathcal{J}}$ distribution and outputs a vector representing the marginal model $P^{\mathcal{M}}$. The joint distribution vector $P^{\mathcal{J}}$ for a probability distribution $P_{\mathcal{J}}$ is the vector indexed by joint events $j \in \mathcal{E}(\mathcal{J})$ whose entries are populated by the probabilities that $P_{\mathcal{J}}$ assigns to each joint event: $P^{\mathcal{J}}_j = P_{\mathcal{J}}(j)$. Analogously, the marginal distribution vector $P^{\mathcal{M}}$ for a marginal model $P^{\mathcal{M}}$ is the vector whose entries are probabilities over the set of marginal outcomes $\mathcal{E}(\mathcal{M})$: $P^{\mathcal{M}}_m = P_{\mathcal{D}(m)}(m)$.

By design, the marginal and joint distribution vectors are related via the incidence matrix M. Given a joint distribution vector $P^{\mathcal{J}}$, one can obtain the marginal distribution vector $P^{\mathcal{M}}$ by multiplying M by $P^{\mathcal{J}}$,

$$P^{\mathcal{M}} = M \cdot P^{\mathcal{J}}. \tag{A5}$$

As a quick remark, the particular ordering of the rows and columns of M carries no importance, but it must be consistent between M, $P^{\mathcal{J}}$, and $P^{\mathcal{M}}$. The marginal problem can now be rephrased in the language of the incidence matrix. Suppose one obtains a marginal distribution vector $P^{\mathcal{M}}$; the marginal problem becomes equivalent to the question of whether there exists a joint distribution vector $P^{\mathcal{J}}$ such that Eq. (A5) holds. This question is naturally framed as the marginal linear program:

$$\begin{aligned} &\text{minimize } \emptyset \cdot x \\ &\text{subject to } x \succeq 0, \\ &\qquad M \cdot x = P^{\mathcal{M}}. \end{aligned} \tag{A6}$$

If this "optimization"[13] is feasible, then there exists a vector x than can satisfy Eq. (A5) and is a valid joint distribution vector. Therefore, feasibility of the marginal linear program not only implies that $P^{\mathcal{J}}$ exists but returns $P^{\mathcal{J}}$. Moreover, if the marginal linear program is infeasible, then there does not exist a joint distribution $P^{\mathcal{J}}$. For every linear program there exists a dual linear program that characterizes the feasibility of the original [46]. Constructing the dual linear program is straightforward [47]:

$$\begin{aligned} &\text{minimize} y \cdot P^{\mathcal{M}} \\ &\text{subject to} y \cdot M \succeq 0. \end{aligned} \tag{A7}$$

The dual marginal linear program not only answers the marginal problem for a specific marginal model $P^{\mathcal{M}}$ but as a by-product provides an inequality that witnesses its contextuality. If this is not obvious, first notice that the dual problem is never infeasible; by choosing y to be trivially the null vector $\emptyset$ of appropriate size, all constraints become satisfied. Second, if the dual constraint $y \cdot M \succeq 0$ holds and the primal is feasible, then $y \cdot P^{\mathcal{M}} = y \cdot M \cdot x \geqslant 0$. Therefore, the sign of the dual objective $d \equiv \min(y \cdot P^{\mathcal{M}})$ classifies a marginal model's contextuality; if $d < 0$ then $y \cdot P^{\mathcal{M}} \geqslant 0$ is violated and therefore $P^{\mathcal{M}}$ is contextual. Likewise if

[13] "Optimization" is presented in quotes here because the minimization objective is trivially always zero ($\emptyset$ denotes the null vector of all zero entries). The primal value of the linear program is of no interest; all that matters is its feasibility.

$d \geqslant 0$ (satisfying $y \cdot P^{\mathcal{M}}$), then $P^{\mathcal{M}}$ is noncontextual.[14] This is a manifestation of Farkas's lemma [46]. An infeasibility certificate [49] is any vector y that satisfies $y \cdot M \succeq 0$. Most linear program software packages such as MOSEK [50], GUROBI [51], CPLEX [52], and CVX/CVX OPT [53,54] are capable of producing infeasibility certificates. Furthermore, for every y satisfying $y \cdot M \succeq 0$ there corresponds a certificate inequality that constraints the set of noncontextual marginal models. If y is an infeasibility certificate, then $y \cdot P^{\mathcal{M}} \geqslant 0$ is satisfied by all contextual marginal models.

The marginal problem can sometimes take on a more general variant that does not begin with a specific marginal model [43,55,56]: Given a marginal scenario $\mathcal{M}$, what is the set of all noncontextual marginal models? Pitowsky [57] demonstrates that the set of noncontextual marginal models forms a convex polytope called the marginal polytope. The extremal rays of a marginal polytope directly correspond to the columns of M which further correspond to deterministic joint distributions $P^{\mathcal{J}}$. Since all joint distributions $P^{\mathcal{J}}$ are probability distributions, their entries must sum to unity $\sum_{j \in \mathcal{E}(\mathcal{J})} P^{\mathcal{J}}_j = 1$. This normalization defines the convexity of the polytope; all noncontextual marginal models are convex mixtures of the deterministic marginal models pursuant to Eq. (A5). The marginal polytope is a beneficial tool for understanding contextuality. First, the facets of a marginal polytope correspond to a finite set of linear inequalities that are complete in the sense that all contextual distributions violate at least one facet inequality [6]. From the perspective of a marginal polytope, linear quantifier elimination or convex hull algorithms can be used to compute a representation of the complete set of linear inequalities and completely solve the marginal problem. A popular tool for linear quantifier elimination is Fourier-Motzkin elimination [21,55,58]. Applying the Fourier-Motzkin procedure to completely solve the marginal problem is discussed in more detail by Fritz and Chaves [43]. An excellent survey of existing techniques for solving the marginal problem including equality set projection [59] and Hardy-type hypergraph transversals can be found in the work of Wolfe *et al.* [21]. In conclusion, there are a number of computational tools available to solve the marginal problem completely whenever no marginal model is provided.

Each of the above-mentioned techniques suffers from computational complexity limitations. For example, the Fourier-Motzkin procedure is in the worst case doubly exponential in the number of initial inequalities [58]. For the purposes of this research, solving the marginal problem without reference to a marginal model was intractable. This will become apparent in Appendix C when the inflation technique is applied to the triangle structure, producing considerably large marginal models. Luckily, the Fritz distribution allows one to avoid the complexity issues of the complete marginal problem and instead focus on the original problem of determining whether or not a particular marginal model admits a joint distribution or not.

[14] Actually, if $d \geqslant 0$ then it is exactly $d = 0$ due to the existence of the trivial $y = \emptyset$. This observation is an instance of the complementary slackness property of [48]. Moreover, if $d < 0$, then it is unbounded $d = -\infty$. This latter point becomes clear upon recognizing that for any y with $d < 0$, another $y' = \alpha y$ can be constructed (with $\alpha > 1$) such that $d' = \alpha d < d$. Since a more negative d' can always be found, it must be that d is unbounded. This is a demonstration of the fundamental unboundedness property of [48]; if the dual is unbounded, then the primal is infeasible.

APPENDIX B: INFLATION TECHNIQUE

The inflation technique by Wolfe *et al.* [21] and inspired by the do-calculus and twin networks of Pearl [9] is a family of causal inference techniques that can be used to determine if an observable probability distribution P_{N_o} is compatible or incompatible with a given causal structure $\mathcal{G}$. As a precursor, the inflation technique begins by augmenting a causal structure $\mathcal{G}$ with additional copies of its nodes, producing an inflated causal structure $\mathcal{G}'$ called an inflation, and then exposes how causal inference tasks on the inflation can be used to make inferences on the original causal structure. For reference, a few inflations of the triangle structure are depicted in Fig. 8. Copies of nodes in the inflated causal structure are distinguished by an additional subscript called the copy index. For example, node A of Fig. 2 has copies A_1, A_2, A_3, and A_4 in the inflated causal structure in Fig. 8(c). All such copies are deemed equivalent via a copy-index equivalence relation denoted by $\sim$. A copy index is effectively arbitrary, so we will refer to an arbitrary inflated copy of A as A', i.e., $A \sim A_1 \sim A' \not\sim B \sim B_1 \sim B'$.[15]

In addition to the common graph-theoretic terminology and notation presented in Sec. II, two related concepts need introductions. First, an induced subgraph of $\mathcal{G}$ for a subset of nodes $N \subseteq \mathcal{N}$ is the graph composed of nodes N and all edges e of the original graph that are contained in N: $\mathsf{Sub}_{\mathcal{G}}(N) \equiv (N, \{e = \{n \to m\} \mid n, m \in N\})$. An ancestral subgraph of $\mathcal{G}$ for a subset of nodes $N \subseteq \mathcal{N}$ is the induced subgraph due to the ancestry of N: $\mathsf{AnSub}_{\mathcal{G}}(N) \equiv \mathsf{Sub}_{\mathcal{G}}(\mathsf{An}_{\mathcal{G}}(N))$.

The inflation technique begins with distribution P_N defined over some observable nodes N of $\mathcal{G}$ and the *a priori* assumption that it is compatible with $\mathcal{G}$ pursuant to the definitions provided in Sec. II. Regarding $\mathcal{G}$ as a causal hypothesis, the observable correlation P_N can only be influenced by the ancestry of N in $\mathcal{G}$. Consequently, for any set of nodes N' of an inflation $\mathcal{G}'$ where the ancestral subgraph $\mathsf{AnSub}_{\mathcal{G}'}(N')$ happens to be homomorphic to the ancestral subgraph $\mathsf{AnSub}_{\mathcal{G}}(N)$, one can conclude that the distribution $P_{N'}$, induced by P_N,[16] is compatible with $\mathcal{G}'$ using the same latent explanations for P_N in $\mathcal{G}$. This observation is known as the inflation lemma (see [21], Lemma 3).

To formalize the inflation lemma, we define the injectable sets of $\mathcal{G}'$, denoted by $\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$, as all sets of nodes in $\mathcal{G}'$ whose ancestral subgraphs are homomorphic (via copy-index removal) to an ancestral subgraph in $\mathcal{G}$: $\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}') \equiv \{N' \subseteq \mathcal{N}' \mid \exists N \subseteq \mathcal{N} : \mathsf{AnSub}_{\mathcal{G}'}(N') \sim \mathsf{AnSub}_{\mathcal{G}}(N)\}$. Analogously

[15] Note that we preemptively generalize the notion of copy-index equivalence to other mathematical objects such as sets, graphs, and groups by saying that $X \sim Y$ if and only if X is equivalent to Y upon removal of the copy index.

[16] The inflated distribution $P_{N'}$ assigns the same probability to all events of P_N whenever the events are equivalent under the removal of copy indices.

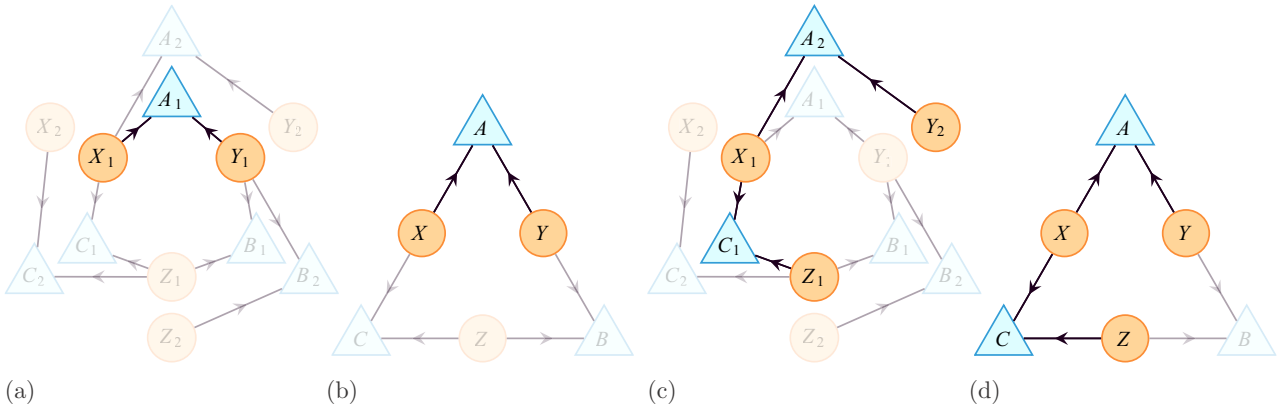

FIG. 9. Some injectable sets $\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ of the spiral inflation $\mathcal{G}'$ and their corresponding images $\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')$ in the triangle structure $\mathcal{G}$: (a) $\mathsf{AnSub}_{\mathcal{G}'}(\{A_1\})$, (b) $\mathsf{AnSub}_{\mathcal{G}}(\{A\})$, (c) $\mathsf{AnSub}_{\mathcal{G}'}(\{A_2, C_1\})$, and (d) $\mathsf{AnSub}_{\mathcal{G}}(\{A, C\})$.

defined are the images of the injectable sets in $\mathcal{G}$: $\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}') \equiv \{N \subseteq \mathcal{N} \mid \exists N' \subseteq \mathcal{N}' : \mathsf{AnSub}_{\mathcal{G}'}(N') \sim \mathsf{AnSub}_{\mathcal{G}}(N)\}$.

To illustrate these concepts, consider the spiral inflation $\mathcal{G}'$ depicted in Fig. 8(a). The ancestral subgraph of $\{A_1\}$ in the spiral inflation [denoted by $\mathsf{AnSub}_{\mathcal{G}'}(\{A_1\})$] is highlighted in Fig. 9(a) and is clearly homomorphic to the ancestral subgraph of its image $\{A\}$ in the triangle structure $(\mathcal{G})$ [Fig. 9(b)]. Additionally, the set $\{A_2, C_1\}$ is an injectable set of $\mathcal{G}'$ because $\mathsf{AnSub}_{\mathcal{G}'}(\{A_2, C_1\})$ [highlighted in Fig. 9(c)] is homomorphic (via copy-index removal) to the $\mathsf{AnSub}_{\mathcal{G}}(\{A, C\})$ in the triangle structure $(\mathcal{G})$ [Fig. 9(d)].

The injectable sets of an inflation are of principle importance to the inflation technique. Given a distribution $P_{\mathcal{N}_o}$ defined over the observable nodes of $\mathcal{G}$, one can compute a marginal model defined over the injectable sets $P^{\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')} = \{P_{N'} \mid N' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\}$ and their images $P^{\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')} = \{P_N \mid N \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\}$. The contrapositive of the inflation lemma casts the compatibility problem for $P^{\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')}$ (which exhibits equivalent compatibility as $P_{\mathcal{N}_o}$) into a compatibility problem for $P^{\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')}$ of any nontrivial inequalities. If $P^{\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')}$ is found to be incompatible with $\mathcal{G}'$, then $P^{\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G})}$ (and also $P_{\mathcal{N}_o}$) must be incompatible with $\mathcal{G}$. Fortunately, the inflated causal structure $\mathcal{G}'$ possesses its own d-separation relations, which enforces conditional independence equality constraints on $P^{\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')}$. A useful subset of d-separation relations are those corresponding to unconditional d separations, known as ancestrally independent sets [9]. Two sets N_1' and N_2' are ancestrally independent $(N_1' \perp N_2')$ if they have distinct ancestry in $\mathcal{G}'$,

$$N_1' \perp N_2' \iff \mathsf{An}_{\mathcal{G}'}(N_1') \cap \mathsf{An}_{\mathcal{G}'}(N_2') = \emptyset. \tag{B1}$$

Ancestral independence implies an unconditional probabilistic independence: If $N_1' \perp N_2'$, then $P_{N_1' \cup N_2'} = P_{N_1'} P_{N_2'}$. If $N_1', N_2' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$, then the associated probabilistic constraint is applicable to $P^{\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')}$. This notion generalizes to more than two ancestrally independent (AI) sets. A set $N' \subseteq \mathcal{N}'$ is an AI-expressible set if it can be decomposed into the disjoint union of injectable sets $N' = \coprod_i N_i' \mid N_i' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ and all pairs N_i', N_j' are ancestrally independent: $N_i' \perp N_j' \; \forall i, j$. The analogous probabilistic constraint is $P_{N'} = P_{\coprod_i N_i'} = \prod_i P_{N_i'}$ [9].[17] Throughout this work, we let $\mathsf{AIExpr}_{\mathcal{G}}(\mathcal{G}')$ denote the set of all AI-expressible sets.[18] Efficient algorithms for computing the injectable and AI-expressible sets of an inflation can be found in [21]. Table I and II tabulate the AI-expressible sets and the associated ancestral separations for the wagon-wheel and web inflations [respectively Figs. 8(b) and 8(c)].

Unlike $P^{\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')}$, which is noncontextual by construction, $P^{\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')}$ contains overlapping marginals, meaning its contextuality remains unknown and must be determined using any of the techniques discussed in Appendix A. More importantly, the inflation technique introduces constraints relating to the AI-expressible sets. In practice, the equality constraints implied by $\mathsf{AIExpr}_{\mathcal{G}}(\mathcal{G}')$ permit one to construct a marginal model defined over the AI-expressible sets $\mathsf{AIExpr}_{\mathcal{G}}(\mathcal{G}')$ resulting in greater resolution in classifying the compatibility of $P_{\mathcal{N}_o}$. In summary, the inflation technique partially transforms the compatibility problem into a marginal problem, wherein one can solve the marginal problem (Appendix A) to either determine the compatibility of an observable distribution $P_{\mathcal{N}_o}$ with $\mathcal{G}$ or

[17] In [21], this concept is generalized into terms which can be factorized via d-separation conditions and the corresponding inflated sets are termed expressible sets. This generalization was not required for this work in particular because the ancestral independence relations formed a generating set of d-separation conditions for all of the inflations considered in Fig. 8.

[18] Analogously to a marginal scenario $\mathcal{M}$, the AI-expressible sets $\mathsf{AIExpr}_{\mathcal{G}}(\mathcal{G}')$ form an abstract simplicial complex. Therefore, in practice, it is completely sufficient to focus on the maximally AI-expressible sets.

TABLE I. Maximally AI-expressible sets for the wagon-wheel inflation.

$\mathsf{AIExpr}_{\mathcal{G}}(\mathcal{G}')$	Ancestral separations
$\{A_2, B_1, C_3, C_1\}$	$\{A_2, B_1, C_3\} \perp \{C_1\}$
$\{A_1, B_1, C_4, C_2\}$	$\{A_1, B_1, C_4\} \perp \{C_2\}$
$\{A_1, B_2, C_1, C_3\}$	$\{A_1, B_2, C_1\} \perp \{C_3\}$
$\{A_2, B_2, C_2, C_4\}$	$\{A_2, B_2, C_2\} \perp \{C_4\}$

TABLE II. Maximally AI-expressible sets for the web inflation.

$\mathsf{AIExpr}_{\mathcal{G}}(\mathcal{G}')$	Ancestral separations
$\{A_1, B_1, C_1, A_4, B_4, C_4\}$	$\{A_1, B_1, C_1\} \perp \{A_4, B_4, C_4\}$
$\{A_1, B_2, C_3, A_4, B_3, C_2\}$	$\{A_1, B_2, C_3\} \perp \{A_4, B_3, C_2\}$
$\{A_2, B_3, C_1, A_3, B_2, C_4\}$	$\{A_2, B_3, C_1\} \perp \{A_3, B_2, C_4\}$
$\{A_2, B_4, C_3, A_3, B_1, C_2\}$	$\{A_2, B_4, C_3\} \perp \{A_3, B_1, C_2\}$
$\{A_1, B_3, C_4\}$	$\{A_1\} \perp \{B_3\} \perp \{C_4\}$
$\{A_1, B_4, C_2\}$	$\{A_1\} \perp \{B_4\} \perp \{C_2\}$
$\{A_2, B_1, C_4\}$	$\{A_2\} \perp \{B_1\} \perp \{C_4\}$
$\{A_2, B_2, C_2\}$	$\{A_2\} \perp \{B_2\} \perp \{C_2\}$
$\{A_3, B_3, C_3\}$	$\{A_3\} \perp \{B_3\} \perp \{C_3\}$
$\{A_3, B_4, C_1\}$	$\{A_3\} \perp \{B_4\} \perp \{C_1\}$
$\{A_4, B_1, C_3\}$	$\{A_4\} \perp \{B_1\} \perp \{C_3\}$
$\{A_4, B_2, C_1\}$	$\{A_4\} \perp \{B_2\} \perp \{C_1\}$

derive compatibility inequalities for $\mathcal{G}$, the latter of which is discussed in Appendix C.

APPENDIX C: DERIVING INEQUALITIES FROM THE INFLATION TECHNIQUE

Appendix B summarizes how the inflation technique [21] can cast the compatibility problem into the marginal problem by leveraging inflations of the triangle structure. Explicitly, for a given inflation $\mathcal{G}'$ of the triangle structure, one constructs a marginal problem (Appendix A) for the marginal scenario $\mathcal{M}$ composed of the maximal AI-expressible sets of $\mathcal{G}'$, i.e., $\mathcal{M} = \mathsf{AIExpr}_{\mathcal{G}}(\mathcal{G}')$.

Recall that the first inequality (12) presented in Sec. VII was derived using the wagon-wheel inflation [Fig. 8(b)]. The wagon-wheel inflation possesses four copies of C (C_1, C_2, C_3, and C_4) and two copies of A (A_1 and A_2) and B (B_1 and B_2) arranged in the shape of a wagon wheel. The maximal AI-expressible sets of the wagon-wheel inflation along with their ancestral dependences can be found in Table I. These maximal AI-expressible sets define a marginal scenario where the joint variables $\mathcal{J}$ are the set of observable nodes in the wagon-wheel inflation $\mathcal{J} = \mathcal{N}'_O$:

$$\mathcal{M} = \mathsf{AIExpr}_{\mathcal{G}}(\mathcal{G}') = \{\{A_2, B_1, C_3, C_1\}, \{A_1, B_1, C_4, C_2\}, \{A_1, B_2, C_1, C_3\}, \{A_2, B_2, C_2, C_4\}\}. \tag{C1}$$

Given that the Fritz distribution is defined over four-outcome variables, the variables in the marginal scenario are assigned four outcomes as well. This marginal scenario $\mathcal{M}$ then defines an incidence matrix M capable of accommodating the Fritz distribution that has $|\mathcal{E}(\mathcal{M})| = 4 \times 4^4 = 1024$ rows and $|\mathcal{E}(\mathcal{J})| = 4^8 = 65\,536$ columns; this matrix is not reproduced here. The sheer size of the incidence matrix used here makes complete solutions of the marginal problem using tools such as linear-quantifier elimination intractable. To construct the marginal distribution vector $P^{\mathcal{M}}$ for the Fritz distribution, one begins with $P^{\mathcal{M}}$ defined in a symbolic form

$$P^{\mathcal{M}\intercal} = \underbrace{(P_{A_2B_1C_3C_1}(0000), \ldots, P_{A_1B_1C_4C_2}(3232), \ldots, P_{A_2B_2C_2C_4}(3333))}_{1024 \text{ entries}}, \tag{C2}$$

which can be factored into the ancestrally independent injectable sets using Table I,

$$P^{\mathcal{M}\intercal} = (P_{A_2B_1C_3}(000)P_{C_1}(0), \ldots, P_{A_1B_1C_4}(323)P_{C_2}(2), \ldots, P_{A_2B_2C_2}(333)P_{C_4}(3)). \tag{C3}$$

Penultimately, each probability distribution in Eq. (C3) is deflated by dropping copy indices,

$$P^{\mathcal{M}\intercal} = (P_{ABC}(000)P_C(0), \ldots, P_{ABC}(323)P_C(2), \ldots, P_{ABC}(333)P_C(3)). \tag{C4}$$

This step is permitted because all of the remaining distributions in Eq. (C3) are defined over the injectable sets of the wagon-wheel inflation. Finally, each of the elements of $P^{\mathcal{M}}$ are replaced with numerical values pursuant to the Fritz distribution. For example, $P_{ABC}(323)P_C(2)$ is assigned the following numerical value:

$$P_{ABC}(323)P_C(2) = \tfrac{1}{32}(2 + \sqrt{2}) \times \tfrac{1}{4} \simeq 0.0267. \tag{C5}$$

The same is applied to all other entries of $P^{\mathcal{M}}$. Finally, this numerical version of $P^{\mathcal{M}}$ and M are subjected to linear programming software and an infeasibility certificate y was obtained, corresponding precisely to the inequality (12) using Eq. (C4) in bitwise notation.

The remaining inequalities in Sec. VII were derived using the web inflation of Fig. 8(c). The maximal AI-expressible sets of the wagon-wheel inflation along with their ancestral dependences can be found in Table II. Analogously to the wagon-wheel inflation, these AI-expressible sets form a marginal scenario $\mathcal{M}$ which defines an incidence matrix M that has $|\mathcal{E}(\mathcal{M})| = 4 \times 4^6 + 8 \times 4^3 = 16\,896$ rows, $|\mathcal{E}(\mathcal{J})| = 4^{12} = 16\,777\,216$ columns, and 201 326 592 nonzero entries.

APPENDIX D: DERIVING SYMMETRIC CAUSAL COMPATIBILITY INEQUALITIES

Appendix B detailed how to obtain causal compatibility inequalities for any causal structure by constructing a corresponding marginal problem (as defined in Appendix A) and supplying an incompatible distribution to generate an infeasibility certificate. The inequality (15) presented the causal compatibility inequality $I_{\text{symmetric web}}$ which is symmetric under permutations of the variables A, B, and C. This section aims to describe a general technique that can be used to derive $I_{\text{symmetric web}}$ and other inequalities also exhibiting this symmetry. In brief, this is accomplished by grouping marginal

events $m \in \mathcal{E}(\mathcal{M})$ of a marginal scenario $\mathcal{M}$ into orbits under the action of variable permutations.

Exploiting symmetries of the marginal scenario is useful for a few distinct reasons. First, Bancal *et al.* [60] discuss computational advantages in considering symmetric versions of marginal polytopes mentioned in Appendix A; the number of extremal points typically grows exponentially in $\mathcal{J}$, but only polynomially for the symmetric polytope. They also note that a number of interesting inequalities (such as the CHSH inequality [25]) can be written in a way that is symmetric under the exchange of parties, demonstrating that nontrivial inequalities can be recovered from facets of a symmetric polytope. Second, numerical optimizations against symmetric inequalities lead to one of two interesting cases: Either the extremal distribution is symmetric itself or it is not. The latter case generates a family of incompatible distributions obtained by applying symmetry operations on the extremal distribution.[19]

To clarify which symmetries we have in mind, first consider the marginal scenario $\mathcal{M} = \{\{A, B, C\}, \{C, D\}, \{A, D\}\}$ where each variable in $\mathcal{J} = \{A, B, C, D\}$ has binary outcomes $\{0, 1\}$. Here $I_{\mathcal{M}} \equiv \{P_{ABC}(010) \leqslant P_{CD}(00) + P_{AD}(01)\}$ is an inequality constraining the set of noncontextual marginal models $P^{\mathcal{M}}$. Now the contextuality of a distribution P_{ABCD} should be invariant under those permutations of the variable labels in $\mathcal{J}$ which map the marginal scenario to itself, but not all variable relabelings preserve $\mathcal{M}$. An example of a permutation $\varphi \in \mathsf{Perm}(\mathcal{J})$[20] which does not preserve $\mathcal{M}$ is $\varphi[\{a, b, c, d\}] = \{c, a, d, b\}$. The action of φ on $I_{\mathcal{M}}$ is

$$\begin{aligned}\varphi[I_{\mathcal{M}}] &= \{P_{\varphi[abc]}(010) \leqslant P_{\varphi[cd]}(00) + P_{\varphi[ad]}(01)\} \\ &= \{P_{cad}(010) \leqslant P_{db}(00) + P_{cb}(01)\} \\ &= \{P_{acd}(100) \leqslant P_{bd}(00) + P_{bc}(10)\}, \end{aligned} \tag{D1}$$

which yields a valid albeit irrelevant inequality, as the resulting inequality no longer pertains to $\mathcal{M}$.

Permutations φ that modify the marginal scenario have no application within the framework of the inflation technique (Appendix B) because the inflation lemma only holds when the inflated inequality constrains injectable sets. Therefore, the desired set of symmetries for our purposes is a subgroup of $\mathsf{Perm}(\mathcal{J})$ that takes $\mathcal{M}$ to $\mathcal{M}$.

The variable permutation group $\Phi(\mathcal{M})$ for a marginal scenario $\mathcal{M}$ is the joint permutation subgroup that stabilizes the marginal scenario, $\Phi(\mathcal{M}) \equiv \{\varphi \in \mathsf{Perm}(\mathcal{J}) \mid \forall V \in \mathcal{M} : \varphi[V] \in \mathcal{M}\}$. In general, the variable permutation group $\Phi(\mathcal{M})$ can be obtained using group stabilizer algorithms. After obtaining $\Phi(\mathcal{M})$, one can take known compatibility inequalities $I_{\mathcal{M}}$ and create a whole family of inequalities $\{\varphi[I_{\mathcal{M}}] \mid \varphi \in \Phi(\mathcal{M})\}$ that are valid for the same marginal scenario $\varphi[I_{\mathcal{M}}] = \varphi[I]_{\varphi[\mathcal{M}]} = \varphi[I]_{\mathcal{M}}$.

Although useful for reducing computational complexity [60], we divert our attention to finding symmetric inequalities, i.e., those where $\varphi[I_{\mathcal{M}}] = I_{\mathcal{M}}\ \forall \varphi \in \Phi(\mathcal{M})$. To generate inequalities that exhibit certain symmetries using the methods described in Appendix A, it is sufficient to perform a change of basis on the incidence matrix M for a given marginal scenario $\mathcal{M}$. Through repeated action of $\varphi \in \Phi$ on marginal outcomes $m \in \mathcal{E}(\mathcal{M})$ and joint outcomes $j \in \mathcal{E}(\mathcal{J})$, one can define group orbits of Φ in $\mathcal{E}(\mathcal{M})$ and $\mathcal{E}(\mathcal{J})$: $\mathsf{Orb}_\Phi(m) \equiv \{\varphi[m] \mid \varphi \in \Phi\}$ and $\mathsf{Orb}_\Phi(j) \equiv \{\varphi[j] \mid \varphi \in \Phi\}$, respectively. The action of $\varphi \in \Phi$ on any outcome $f \in \mathcal{E}(V)$ (denoted by $\omega[f]$) is defined as $(\varphi[f])(v) \equiv f(\varphi^{-1}[v])$ pursuant to intuitive action used in Eq. (D1). Using these group orbits, it is possible to contract the incidence matrix M of a marginal scenario into a symmetrized version. The symmetric incidence matrix M_Φ for a marginal scenario $\mathcal{M}$ and the variable permutation group Φ is a contracted version of the incidence matrix M for $\mathcal{M}$. Each row of M_Φ corresponds to a marginal orbit $\mathsf{Orb}_\Phi(m)$. Analogously, each column of M_Φ corresponds to a joint orbit $\mathsf{Orb}_\Phi(j)$. The entries of M_Φ are integers and correspond to summing over the rows and columns of M that belong to each orbit,

$$(M_\Phi)_{\mathsf{Orb}_\Phi(m),\mathsf{Orb}_\Phi(j)} = \sum_{\substack{j' \in \mathsf{Orb}_\Phi(j) \\ m' \in \mathsf{Orb}_\Phi(m)}} M_{m',j'}. \tag{D2}$$

It is possible to analogously define a symmetric joint distribution vector $P^{\mathcal{J}}_\Phi$ indexed by $\mathsf{Orb}_\Phi(j)$, $(P^{\mathcal{J}}_\Phi)_{\mathsf{Orb}_\Phi(j)} = \sum_{j' \in \mathsf{Orb}_\Phi(j)} P^{\mathcal{J}}_{j'}$, and a symmetric marginal distribution vector $P^{\mathcal{M}}_\Phi$ indexed by $\mathsf{Orb}_\Phi(m)$, $(P^{\mathcal{M}}_\Phi)_{\mathsf{Orb}_\Phi(m)} = \sum_{m' \in \mathsf{Orb}_\Phi(m)} P^{\mathcal{M}}_{m'}$. Together, M_Φ, $P^{\mathcal{J}}_\Phi$, and $P^{\mathcal{M}}_\Phi$ define a symmetric marginal problem $P^{\mathcal{M}}_\Phi = M_\Phi \cdot P^{\mathcal{J}}_\Phi$. The symmetric marginal problem can be solved using the same computational methods discussed in Appendix A and will produce symmetric inequalities.

In the context of the inflation technique, a variable symmetry $\Phi(\mathcal{M}')$ over an inflated marginal scenario $\mathcal{M}'$ does not always correspond to a variable symmetry under deflation $\Phi(\mathcal{M})$. In order to derive deflated inequalities that are symmetric under an exchange of parties, it is required that $\Phi(\mathcal{M}') \sim \Phi(\mathcal{M})$ are equivalent up to the copy index.

For the triangle structure in particular, the variable permutation group is the set of permutations on A, B, and C: $\Phi(\mathcal{M}) = \mathsf{Perm}(A, B, C)$. For the web inflation [Fig. 8(c)], we have obtained $\Phi(\mathsf{AIExpr}_{\mathcal{G}}(\mathcal{G}'))$, an order 48 group with the following four generators:

φ_1	φ_2	φ_3	φ_4
$A_1 \to A_4$	$A_1 \to A_1$	$A_1 \to C_1$	$A_1 \to A_1$
$A_2 \to A_3$	$A_2 \to A_3$	$A_2 \to C_2$	$A_2 \to A_2$
$A_3 \to A_2$	$A_3 \to A_2$	$A_3 \to C_3$	$A_3 \to A_3$
$A_4 \to A_1$	$A_4 \to A_4$	$A_4 \to C_4$	$A_4 \to A_4$
$B_1 \to B_4$	$B_1 \to C_1$	$B_1 \to A_1$	$B_1 \to B_2$
$B_2 \to B_3$	$B_2 \to C_3$	$B_2 \to A_2$	$B_2 \to B_1$
$B_3 \to B_2$	$B_3 \to C_2$	$B_3 \to A_3$	$B_3 \to B_4$
$B_4 \to B_1$	$B_4 \to C_4$	$B_4 \to A_4$	$B_4 \to B_3$
$C_1 \to C_4$	$C_1 \to B_1$	$C_1 \to B_1$	$C_1 \to C_3$
$C_2 \to C_3$	$C_2 \to B_3$	$C_2 \to B_2$	$C_2 \to C_4$
$C_3 \to C_2$	$C_3 \to B_2$	$C_3 \to B_3$	$C_3 \to C_1$
$C_4 \to C_1$	$C_4 \to B_4$	$C_4 \to B_4$	$C_4 \to C_2$.

(D3)

[19] If the extremal distribution happens to be asymmetric, then one can conclude that the space of accessible distributions is nonconvex.

[20] The permutation group $\mathsf{Perm}(S)$ over a set S is the set of all bijective maps $\varphi : S \to S$.

Note that φ_1, φ_2, φ_3, and φ_4 are all automorphisms of the web inflation. Moreover, they stabilize the AI-expressible sets of Table II. Importantly,

$$\Phi(\mathsf{AIExpr}_{\mathcal{G}}(\mathcal{G}')) \sim \mathsf{Perm}(A, B, C). \tag{D4}$$

To see this, φ_1 and φ_4 become the identity element in $\mathsf{Perm}(A, B, C)$ upon removal of the copy index, leaving φ_2 to generate reflections and φ_3 to generate rotations.

The symmetric incidence matrix M_Φ for the web inflation is considerably smaller than M. The number of rows of M_Φ is a number of distinct orbits $\mathsf{Orb}_\Phi(m)$ in $\mathcal{E}(\mathcal{M})$. Likewise the number of columns is the number of distinct orbits $\mathsf{Orb}_\Phi(j)$ in $\mathcal{E}(\mathcal{J})$. For the web inflation in particular, M_Φ is $450 \times 358\,120$. Using the symmetric incidence matrix and linear programming methods, an infeasibility certificate was found that is capable of witnessing the Fritz distribution. The corresponding deflated inequality is presented in Sec. VII as the inequality (15).

[1] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM J. Comput. **26**, 1484 (1997).

[2] S. Jordan, Quantum algorithm zoo, 2016, available at https://math.nist.gov/quantum/zoo/

[3] C. H. Bennett and G. Brassard, Quantum cryptography: Public key distribution and coin tossing, Theor. Comput. Sci. **560**, 7 (2014).

[4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2011), p. 109.

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

[6] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, Rev. Mod. Phys. **86**, 419 (2013).

[7] A. Einstein, B. Podolsky, and N. Rosen, Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777 (1935).

[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 (2012).

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

[10] J. Pearl, Causal inference in statistics: An overview, Stat. Surv. **3**, 96 (2009).

[11] J. Pearl, in *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence* (Kaufmann, Burlington, 1995), pp. 435–443.

[12] B. Bonet, in *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence* (Kaufmann, Burlington, 2001), pp. 48–55.

[13] R. J. Evans, in *Proceedings of the 2012 IEEE International Workshop on Machine Learning for Signal Processing* (IEEE, Piscataway, 2012), pp. 1–6.

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

[15] T. Fritz, Beyond Bell's theorem II: Scenarios with arbitrary causal structure, Commun. Math. Phys. **341**, 391 (2014).

[16] B. Steudel and N. Ay, Information-theoretic inference of common ancestors, Entropy **17**, 2304 (2015).

[17] R. Chaves, L. Luft, and D. Gross, Causal structures from entropic information: Geometry and novel scenarios, New J. Phys. **16**, 043001 (2014).

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

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

[20] M. Weilenmann and R. Colbeck, Non-Shannon inequalities in the entropy vector approach to causal structures, Quantum **2**, 57 (2018).

[21] E. Wolfe, R. W. Spekkens, and T. Fritz, The inflation technique for causal inference with latent variables, arXiv:1609.00672.

[22] M. Navascués and E. Wolfe, The inflation technique solves completely the classical inference problem, arXiv:1707.06476.

[23] R. J. Evans, Graphs for margins of Bayesian Networks, Scandinavian J. Stat. **43**, 625 (2016).

[24] T. S. Richardson, J. M. Robins, and I. Shpitser, in *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence* (AUAI Press, 2012), p. 13.

[25] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed Experiment to Test Local Hidden-Variable Theories, Phys. Rev. Lett. **23**, 880 (1969).

[26] B. M. Terhal and D. P. DiVincenzo, Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games, Quantum Inf. Comput. **4**, 134 (2004).

[27] M. Navascués and E. Wolfe (private communication).

[28] B. S. Cirel'son, Quantum generalizations of Bell's inequality, Lett. Math. Phys. **4**, 93 (1980).

[29] N. Gisin, The elegant joint quantum measurement and some conjectures about N-locality in the triangle and other configurations, arXiv:1708.05556.

[30] D. Petz and L. Ruppert, Optimal quantum-state tomography with known parameters, J. Phys. A: Math. Theor. **45**, 085306 (2015).

[31] S. R. Hedemann, Hyperspherical parameterization of unitary matrices, arXiv:1303.5904.

[32] K. Fujii, K. Funahashi, and T. Kobayashi, Jarlskog's parametrization of unitary matrices and qudit theory, Int. J. Geom. Methods Mod. Phys. **03**, 269 (2006).

[33] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).

[34] M. Grasmair (unpublished).

[35] C. Spengler, M. Huber, and B. C. Hiesmayr, A composite parameterization of unitary groups, density matrices and subspaces, J. Phys. A: Math. Theor. **43**, 385306 (2010).

[36] C. Moler and C. V. Loan, Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later, SIAM Rev. **45**, 3 (2003).

[37] K. F. Pál and T. Vértesi, Maximal violation of a bipartite three-setting, two-outcome Bell inequality using infinite-dimensional quantum systems, Phys. Rev. A **82**, 022116 (2010).

[38] J. Nocedal and S. Wright, Numerical Optimization, *Springer Series in Operations Research and Financial Engineering* (Springer, Berlin, 2000).
[39] D. J. Wales and J. P. K. Doye, Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms, J. Phys. Chem. A **101**, 5111 (1997).
[40] A. A. Methot and V. Scarani, An anomaly of non-locality, Quantum Inf. Comput. **7**, 157 (2007).
[41] M. Navascués (private communication).
[42] D. Rosset (private communication).
[43] T. Fritz and R. Chaves, Entropic inequalities and marginal problems, IEEE Trans. Inf. Theory **59**, 803 (2011).
[44] N. N. Vorob'ev, Consistent families of measures and their extensions, Theor. Prob. Appl. **7**, 147 (1962).
[45] S. Abramsky and A. Brandenburger, The sheaf-theoretic structure of non-locality and contextuality, New J. Phys. **13**, 113036 (2011).
[46] A. Schrijver, *Theory of Linear and Integer Programming* (Wiley, New York, 1998).
[47] S. Lahaie, How to take the dual of a linear program, 2008 (unpublished), available at http://www.cs.columbia.edu/coms6998-3/lpprimer.pdf
[48] S. P. Bradley, A. C. Hax, and T. L. Magnanti, *Applied Mathematical Programming* (Addison-Wesley, Reading, 1977), Chap. 4, pp. 143–144.
[49] E. D. Andersen, Comput. Optim. Appl. **20**, 171 (2001).
[50] MOSEK optimization suite, 2016.
[51] Gurobi Optimizer, FarkasDual, 2018, available at https://www.gurobi.com/documentation/8.0/refman/farkasdual.html.
[52] IBM, CPLEX optimization studio, CPLEX, 2016, available at http://www-01.ibm.com/support/docview.wss?uid=swg21400058.
[53] CVX, MATLAB software for disciplined convex programming, Dual variables, 2016, available at http://cvxr.com/cvx/doc/basics.html#dual-variables.
[54] M. Andersen, J. Dahl, and L. Vandenberghe, Linear cone programs, 2016, available at http://cvxopt.org/userguide/coneprog.html#linear-cone-programs
[55] S. Abramsky and L. Hardy, Logical Bell inequalities, Phys. Rev. A **85**, 062114 (2012).
[56] S. Mansfield and T. Fritz, Hardy's Non-locality paradox and possibilistic conditions for non-locality, Found. Phys. **42**, 709 (2011).
[57] I. Pitowsky, Correlation polytopes: Their geometry and complexity, Math. Prog. **50**, 395 (1991).
[58] G. B. Dantzig and B. C. Eaves, Fourier-Motzkin elimination and its dual, J. Combin. Theor. A **14**, 288 (1973).
[59] C. Jones, E. C. Kerrigan, and J. Maciejowski, Equality set projection: A new algorithm for the projection of polytopes in halfspace representation, Cambridge University Engineering Department Report No. CUED/F-INFENG/TR.463 (Cambridge University Engineering Dept. 2004), https://infoscience.epfl.ch/record/169768/.
[60] J.-D. Bancal, N. Gisin, and S. Pironio, Looking for symmetric Bell inequalities, J. Phys. A: Math. Theor. **43**, 385303 (2010).