Causal Compatibility Inequalities Admitting of 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*
(Dated: June 13, 2017)

It has long been recognized that certain quantum correlations are incompatible with particular assumption about classical causal structure. Given a causal structure of unknown classicality, the presence of such correlations certifies the non-classical nature of the causal structure in a device independent fashion. In the context of two parties sharing a common resource, these non-classical correlations are also known as non-local correlations. Any constraint which is satisfied by all correlations which *are* classically compatible with a given causal structure defines a causal compatibility criterion. Such criteria were recently were recently derived for the Triangle structure [arXiv:1609.00672] in the form of polynomial inequalities, begging the question: do any of those inequalities admit violation by quantum correlations? Numerical investigation suggests 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 new causal compatibility inequalities for the Triangle structure which *do* admit quantum violation. We conclude by considering the possibility of quantum resources potentially qualitatively different from those known previously.

---

* tcfraser@tcfraser.com

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, every aspect of quantum phenomena which fails to be emulated by classical resources has led to a practical exploitation of the non-classical feature in order to solve a some computational of communicational problem [4]. Motivated by past successes, a primal objective of modern quantum information theory is the discovery of new situations wherein quantum mechanics offers an advantage, and certification 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]. 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 EPR paradox [6]. The enumeration of Bell inequalities has since become a widespread systematic way of demonstration the *non-classicality* of a given observation. More recently it has been appreciated that Bell inequalities can understood as consequences of **causal inference** [7]. 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 [8, 9]. Causal compatibility inequalities, such as Bell and Instrumental inequalities, 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 non-communicating parties making measurements on some hidden shared resource λ, where the measurement outcomes (A, B) are presumed to be stochastic functions of the the local choices of measurement settings (S_A, S_B) and the shared resource λ. Quantum non-classicality in the Bell structure has been thoroughly studied since Bell's original work [10]. More complex structures, however, such as the correlation scenarios proposed by Fritz [11, 12], are much less understood. Here we investigate one particular causal structure: the Triangle structure (Fig. 2).

The Triangle structure (Fig. 2) is a causal structure composed of 3 parties labeled A, B, C arranged in a triangular configuration while pair-wise sharing hidden/latent variables X, Y, Z. It has been extensively studied previously (see [13, Fig. 1], [14, Fig. 6], [15, Fig. 8], [16, Fig. 8, App. E], [11, Fig. 3], [17, Fig. 1], . . .). An overview of some milestone results is provided in Section IV. Identifying causal compatibility inequalities for this configuration has been seen as particularly challenging [15]. 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 is the first, therefore, to find causal compatibility inequalities for the Triangle structure that are able to be violated by quantum-accessible distributions. This accomplishment was made possible through the combination of two previous developments. The first, an insight made by Fritz [11] regarding the ability to re-interpret the Bell structure as a portion of the Triangle structure. The second, a new framework for solving causal inference problems developed by Wolfe *et al.* [17] called **The Inflation Technique**. Ultimately, this work serves as a validation that the Inflation Technique is sensitive enough to offer new insights into quantum non-classicality. Moreover, these new inequalities offer an avenue for recognizing previously unknown forms of non-classicality. The authors attempts to find such novel resources were met with only partial success, suggesting clear refinements for future exploration.

The manuscript is organized into the following sections: 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 first conceived of by Fritz [11], which we term the **Fritz distribution**. The Fritz distribution was proven to be non-classical without the use of inequalities — this report fills in the desired missing inequalities. A sample of these inequalities are presented in Section VII, specifically Eqs. (8,9,6). Finally, Section VIII introduces a method for parameterizing the space of quantum-accessible distributions which we used to numerically optimize against our derived inequalities in search of stronger types of non-classicality. Additionally, the Fritz distribution is subjected to noise in order to measure the robustness of the derived inequalities. Section IX concludes with proposed avenues for further research.

Appendix A briefly summaries the Inflation Technique in the specific context of this work. Although the summary presented here is designed to be self-standing, a much more pedagogical introduction is offered by the original work [17]. Appendix A 2 demonstrates how the Inflation Technique was used to derive causal compatibility inequalities for the Triangle structure that are violated 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 [8]. 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 respectively *nodes* and *edges* 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** $\mathsf{An}_{\mathcal{G}}(n) \equiv \bigcup_{i \in \mathbb{N}} \mathsf{Pa}_{\mathcal{G}}^i(n)$ where $\mathsf{Pa}_{\mathcal{G}}^i(n) \equiv \mathsf{Pa}_{\mathcal{G}}\big(\mathsf{Pa}_{\mathcal{G}}^{i-1}(n)\big)$ 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 $\mathsf{Ch}_{\mathcal{G}}^i(n) \equiv \mathsf{Ch}_{\mathcal{G}}\big(\mathsf{Ch}_{\mathcal{G}}^{i-1}(n)\big)$ 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 denoted $\mathsf{Pa}_{\mathcal{G}}(N) = \bigcup_{n \in N} \mathsf{Pa}_{\mathcal{G}}(n)$.

Now that the notion of a causal structure $\mathcal{G}$ is defined along with some standard graph-theoretic notation, it is now possible to precisely define the notion of compatibility between $\mathcal{G}$ and a probability distribution $P_{\mathcal{N}}$ defined over the nodes of $\mathcal{G}$. Intuitively, the absence of certain causal connections within $\mathcal{G}$ should limit the set of possible observations that can be made on $\mathcal{G}$. Specifically, the causal structure $\mathcal{G}$ hypothesizes that if each variable $n \in \mathcal{N}$ is causally influenced by its parents $\mathsf{Pa}_{\mathcal{G}}(n)$, then each n should be conditionally independent of all other variables:

$$P_{\mathcal{N}} = \prod_{n \in \mathcal{N}} P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} = P_{n_1|\mathsf{Pa}_{\mathcal{G}}(n_1)} \times \cdots \times P_{n_k|\mathsf{Pa}_{\mathcal{G}}(n_k)} \tag{1}$$

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 $\mathcal{G}$. The compatibility of a distribution $P_{\mathcal{N}}$ with a causal structure $\mathcal{G}$ pursuant to Equation (1) is equivalent to the requirement that $P_{\mathcal{N}}$ obeys all of the conditional independence constraints implied by the *d-separation relations* between variables/nodes in $\mathcal{G}$ [8]. When one is specified with a joint distribution $P_{\mathcal{N}}$ defined over all nodes of a causal structure $\mathcal{G}$, it is trivially easy to verify whether or not $P_{\mathcal{N}}$ is compatible with $\mathcal{G}$ by simply computing the causal parameters induced by $P_{\mathcal{N}}$ and checking the equality of Eq. (1).

A challenge 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) can not 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 $P_{\mathcal{N}_O} = \prod_{n \in \mathcal{N}_O} P_{n|\mathsf{Pa}_{\mathcal{G}}(n)}$. The complementary, unobservable nodes are termed **latent nodes** $\mathcal{N}_L = \mathcal{N} \backslash \mathcal{N}_O$ representing hidden random variables that are hidden either 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 naive approaches worth mentioning here. The first is to recognize that the conditional independence constraints implied by d-separation relations over the *observable* variables form a family of necessary constraints for $P_{\mathcal{N}_O}$. The failure to satisfy one of these observable independence relations immediately qualifies $P_{\mathcal{N}_O}$ as incompatible with $\mathcal{G}$. Often, the observable d-separation relations implied by a causal structure form an *insufficient* set of constraints for compatible distributions. Such causal structures, which include the Bell structure, have been termed *interesting* [16], and more involved classification techniques must be used [17]. The principle focus of this paper is the Triangle structure and has been shown to be one of the smallest interesting causal structures [16]. An alternative approach requires the cardinality (number of outcomes) of the latent variables to be fixed. In most realistic scenarios, the nature of the latent variables is completely unspecified. More powerful approaches find constraints for $P_{\mathcal{N}}$ in the form of probabilistic inequalities defined over $P_{\mathcal{N}_O}$ and its marginals which constraint the set of compatible probability distributions. We term these constraints **causal compatibility inequalities**[1] If a probability distribution $P_{\mathcal{N}_O}$

―――――

[1] We refer to these inequalities as "causal compatibility inequalities" instead of simply Bell inequalities for two reasons. Firstly, "Bell inequalities" usually are associated specifically with the Bell structure. Secondly, 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*.

happens to *violate* any causal compatibility inequality, then that distribution deemed *incompatible*. In Appendix A 2, a brief discussion of existing techniques for deriving these inequalities will take place.

Unfortunately, a singular inequality cannot prove that a given distribution is *compatible*, only *incompatible*. Instead, a *complete characterization* of compatibility would consist of a complete set of all valid causal compatibility inequalities; satisfaction of a complete set of inequalities does prove compatibility. Currently however, it is unknown how to obtain a complete characterization for any given causal structure, including the Triangle structure.

From the perspective of identifying quantum non-classicality, a causal structure $\mathcal{G}$ adopts the role of a classical hypothesis. Therefore, non-classicality becomes synonymous with incompatibility: if a distribution $P_{\mathcal{N}_O}$ is incompatible, then it is non-classical. Henceforth, we will use these two terms interchangeably. From a resource standpoint, if the non-classicality of $P_{\mathcal{N}_O}$ can be witnessed by an inequality I, but $P_{\mathcal{N}_O}$ can be implemented using quantum states and measurements, then the causal compatibility inequality I represents a physical task where quantum resources out-perform classical resources.

III. BELL STRUCTURE

This section aims to define the Bell structure and summarize the main results associated with its non-classical character. The purpose of this section is to equip an unfamiliar reader with the pertinent background and also to establish a context for which to draw comparisons between the advancements made toward understanding non-classicality of the Bell structure and the analogous results obtained for Triangle structure in the remainder of this work. Readers familiar with the Bell structure (sometimes called Bell scenario/experiment) will experience little obstacles proceeding to Section IV.

The Bell structure (Fig. 1) refers to an iconic causal structure involving two distant observers which make local measurements A, B determined by measurement settings S_A, S_B on a shared latent resource λ [10]. The resulting observed correlations form a conditional probability distribution $P_{AB|S_A S_B}$. Subject to the notions of compatibility presented in Section II, this distribution is compatible with Fig. 1 if and only if there exists some functional specification of λ such that there exists a distribution $P_{AB|S_A S_B, \lambda}$ such that $P_{AB|S_A S_B} = \sum_\lambda P_{AB|S_A S_B, \lambda}$ and also satisfies the following implementation of Eq. (1):

$$P_{AB|S_A S_B, \lambda} = P_{A|S_A, \lambda} P_{B|S_B, \lambda}$$

Any observed distribution $P_{AB|S_A S_B}$ which fails to be explained by the classical causal hypothesis of the Bell structure is appropriately termed non-classical. Often, distributions incompatible with the Bell structure are referred to as non-local because the Bell structure indicates that the latent resource is shared locally and measurements are made locally. In particular, Bell [5] demonstrated that there exists distributions that are non-classical yet are attainable via quantum resources and measurements. If one does not wish to accept fine-tuning arguments, one must accept that quantum correlations can be non-local as no variation to the latent nodes can explain the same correlations [7].[2] A distribution is accessible with quantum resources and measurements if there exists a bipartite quantum state ρ_{AB} of arbitrary dimension and measurement operators $M_{A|S_A}, M_{B|S_B}$ conditional indexed by the measurement settings such that $P_{AB|S_A S_B} = \text{Tr}\big[\rho_{AB} M_{A|S_A} \otimes M_{B|S_B}\big]$. The non-classicality of quantum distributions for the Bell structure can be demonstrated through the use of Bell inequalities such as the CHSH inequality (Eq. (2)) which constrain the correlations between binary variables A and B for classical distributions [18].

$$\langle AB|S_A = 0, S_B = 0\rangle + \langle AB|S_A = 0, S_B = 1\rangle + \langle AB|S_A = 1, S_B = 0\rangle - \langle AB|S_A = 1, S_B = 1\rangle \leq 2 \qquad (2)$$

Non-trivial Bell inequalities such as the CHSH inequality are capable of distinguishing the non-classical nature of distributions beyond the constraints implied by observable d-separation relations; the inequalities presented in Section VII are also of this type.

Finally, must like any causal structure, Bell inequalities act as a measure of the non-classical character of a given distribution. In particular, the antecedent of Eq. (2) has a classical upper-bound of 2 and it can be shown to have a quantum upper-bound of $2\sqrt{2}$ [19]. Furthermore, non-classical distributions that are also non-quantum such as the PR-box correlations are upper-bounded by 4 [20]. In Section VII, newly discovered and analogous inequalities for the Triangle structure are presented. Although their associated quantum and post-quantum upper-bounds are currently unknown, an attempt to numerical compute these is discussed in Section VIII B.

---

[2] This is unlike the Triangle structure wherein a global yet latent common cause is always an acceptable explanation for any correlation and does require fine-tuning. Therefore, we never refer to non-classical correlations in Triangle Structure as non-local.

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

FIG. 2: The casual structure of the Triangle structure. Three variables A, B, C are observable, while X, Y, Z are latent variables.

IV.   TRIANGLE STRUCTURE

Discovering quantum non-classicality in any causal structure is desirable, so why, in particular, is it important to do so in the Triangle structure? As was mentioned in Sections I and II, the **Triangle structure** (Fig. 2) is a causal structure $\mathcal{G}$ consisting of 3 observable variables A, B, C arranged in a triangular configuration while pairwise sharing latent variables X, Y, Z. Following the definition of causal compatibility from Section 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, Z[3]:

$$P_{ABC} = \sum_{X,Y,Z} P_{A|X,Y} P_{B|Y,Z} P_{C|Z,X} P_X P_Y P_Z \tag{3}$$

For certain causal structures, the conditional independence relations provided by *d*-separation relations are a sufficient characterization of compatibility. Recently, Henson *et al.* [16], classified the Triangle structure as an *interesting* causal structure in the sense that conditional independence relations *do not* provide a sufficient characterization of non-classicality.[4] Moreover, Wolfe *et al.* [17] concretely demonstrate that the space of classical distributions on the Triangle structure is non-convex, unlike the Bell structure. It can be argued that convexity of the Bell structure distributions is partially responsible for the wealth of knowledge that is known about the Bell structure. It is also reasonable to speculate that quantum non-classicality in the Triangle structure can be translated directly into a computational advantage for certain computational circuits [21]. Characterizing quantum non-classicality for interesting causal structures remains an unsolved problem, and the Triangle structure is a small, non-trivial case study.

Additionally, Fritz [11] demonstrated that the Triangle structure is the *smallest* correlation scenario[5] in which their exists incompatible, quantum distributions by explicitly mentioning one which we discuss in detail in Section V. Fritz's proof of non-classicality does not involve compatibility inequalities, but instead mentions that inequalities would be helpful for characterizing non-classicality in the Triangle structure [11].

Finally, Wolfe *et al.* [17] proposed the Inflation Technique (summarized in Appendix A) — a novel tool for solving causal inference problems. Using the Inflation Technique, they were able to prove incompatibility between the Triangle structure and the W-type distribution; a quantum-inaccessible distribution whose incompatibility is not witness-able by *any* entropic inequalities or any other known constraints. Consequently, as a an algorithmic framework, the Inflation Technique comes closer to characterizing non-classicality in the Triangle structure than all others.

It is worth noting that a number of causal compatibility inequalities for the Triangle structure have been derived in previous works. For example, Steudel and Ay [13] derived an inequality for a series of causal structures with $n \in \mathbb{N}$ variables where at most $c \in \mathbb{N}$ share a common ancestor. The Triangle structure is a special case with $c = 2$, $n = 3$. Moreover, Henson *et al.* [16] derive a family of entropic inequalities constraining the set of compatible correlations. Most recently, Wolfe *et al.* [17] derive a family of polynomial, 2-outcome, causal compatibility inequalities; 37 nontrivial representatives are printed in [17]. Section VI outlines our initial attempts to utilize these inequalities to search for incompatible distributions that are also quantum-accessible. The subsequent failure of these approaches will motivate the remainder of this work.

---

[3] Here the summation $\sum$ is used to denote a statistical marginalization over the latent variables X, Y, Z with unspecified support.

[4] Note that the Triangle structure possesses *zero* conditional independence relations over the observable nodes A, B, C.

[5] Correlation scenarios differ from the Bell structure in that measurement setting variables are absent in correlation scenarios.

FIG. 3: The Triangle structure re-imagined to mimic the Bell structure. The measurement settings S_A, S_B are latent nodes unlike the Bell structure (Fig. 1).

V. FRITZ DISTRIBUTION

As was originally noticed by Fritz [11], it is possible to construct quantum distributions incompatible with the Triangle structure by utilizing quantum distributions incompatible with the familiar Bell structure. To explain, imagine rearranging the triangle into the configuration depicted in Fig. 3 and contrast it with the Bell structure of Fig. 1. Evidently, under the correct relabeling, large portions of the Triangle structure resemble the Bell structure. The crucial distinction is that the measurement settings S_A, S_B are *observable variables* in the Bell structure but become *latent variables* in the Triangle structure. In order to embed distributions that are incompatible with the Bell structure into the Triangle structure, the latent nodes S_A, S_B need to be observable and independent of the shared latent node λ. This can be accomplished by having C "measure" the measurement settings S_A, S_B and announce them as an outcome. Consequently, any distribution over A, B, S_A, S_B that is incompatible with Bell structure is also incompatible with the Triangle structure provided that C is *perfectly correlated* with S_A, S_B [11]. Understandably there are numerous ways that this can be accomplished, albeit we will focus on elucidating an exemplary method.

The **Fritz distribution**, denoted P_F, is a quantum-accessible distribution incompatible with the Triangle structure [11]. In the Fritz distribution, each of the variables A, B, C have 4 possible outcomes. Explicitly, P_F can be written as:

$$P_F(000) = P_F(110) = P_F(021) = P_F(131) = P_F(202) = P_F(312) = P_F(233) = P_F(323) = \frac{1}{32}\left(2 + \sqrt{2}\right)$$

$$P_F(010) = P_F(100) = P_F(031) = P_F(121) = P_F(212) = P_F(302) = P_F(223) = P_F(333) = \frac{1}{32}\left(2 - \sqrt{2}\right)$$

(4)

Here the notation $P_F(abc) = P_{ABC}(abc) = P(A \mapsto a, B \mapsto b, C \mapsto c)$ 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 $\rho_{AB}, \rho_{BC}, \rho_{CA}$ and measurements M_A, M_B, M_C realized on Fig. 2 using Eq. (10) [11]. The Fritz distribution is best visualized as a $4 \times 4 \times 4$ grid of possible outcomes as depicted in Fig. 4. From this diagram, it can be seen that each of C's outcomes restricts the possible outcomes for A, B into a 2×2 block. If one writes the outcome labels $\{0, 1, 2, 3\}$ in binary $\{00, 01, 10, 11\}$, it can be seen that the left-hand bits for A and B (respectively denoted A_l, B_l) are fixed by the outcome of C. Pursuant to the embedding of Fig. 3 the left-hand bits emulate the measurement settings S_A and S_B and the right-hand bits emulate a two-outcome measurement performed by A, B as they would in Fig. 1. Specifically, C's bits are perfectly correlated with the left-bits of A, B; $C_l = A_l$ and $C_r = B_l$. Consequently, it is possible to define the correlation $\langle A_r B_r \rangle$ between the right-hand bits of A and B:

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

Provided that C is perfectly correlated with A_l and B_l, any Bell inequality for the Bell structure defined over A, B, S_A, S_B can be directly converted to an inequality for the Triangle structure by performing a simple relabeling:

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

$A, B, S_A, S_B \mapsto A_r, B_r, A_l, B_l$. As an example, the famous CHSH inequality [18] (Eq. (2)) constrains the correlation between the right-bits of A and B,

$$\langle A_r B_r | C = 00 \rangle + \langle A_r B_r | C = 01 \rangle + \langle A_r B_r | C = 10 \rangle - \langle A_r B_r | C = 11 \rangle \leq 2 \tag{5}$$

Every compatible distribution P_{ABC} for which C is perfectly correlated A_l, B_l must satisfy Eq. (5). Substituting the Eq. (4) into Eq. (5) yields the maximal violation [19]: $3(1/\sqrt{2}) - (-1/\sqrt{2}) = 2\sqrt{2} \nleq 2$. Therefore, we conclude the Fritz distribution is incompatible with the Triangle structure. Before continuing it is worth noting that Eq. (4) is non-unique. Nonetheless for concreteness, Eq. (4) is taken as *the* Fritz distribution throughout this manuscript.

It is important to understand the domain in which Fritz's proof of incompatibility is valid; the details of which can found in the original work [11]. Specifically, Eq. (5) only acts as a valid causal compatibility inequality for the Triangle structure if there exists perfect correlation between C's outcomes and the measurement pseudo-settings (left-bits) of A and B. As a result, the incompatibility proof employed by Fritz is not applicable to any distribution that significantly deviates from the Fritz distribution presented above. For example, if one combines Eq. (4) with more and more uniform noise, at what point does the resultant distribution transition from incompatibility to compatibility? This question is partially answered in Section VIII A.

Upon reflection, the Fritz distribution can be considered *manufactured*. The phenomenology associated with Bell non-locality or Bell incompatibility are well understood; examining these distributions embedded in the Triangle structure offers no additional perspective onto the types of entanglement resources made accessible by quantum mechanics. Therefore, it is desirable to find incompatible quantum distributions that are qualitatively different than those previously considered for the Bell structure. Pursuant to this, Fritz [11] presented the following problem [11, Problem 2.17]:

Fritz's Problem: Find an example of non-classical quantum correlations in the Triangle structure together with a proof of its non-classicality which does not hinge on Bell's theorem.

Understandably, Fritz's Problem is stated ambiguously as it is not yet entirely clear how to rigorously compare the qualitative features of differing non-classical distributions on differing causal structures (such as the Bell and Triangle structures). The remainder of this paper focuses on our advancement toward solving Fritz's Problem. Section IX concludes by outlining our resolution to Fritz's problem and mentioning areas for future research.

VI. PRELIMINARY RESEARCH

As a preliminary search for quantum incompatibility in the Triangle structure, we performed numerical optimizations using bipartite qubit density matrices and 2-outcome POVM measurements against the compatibility inequalities printed in [17] as well as the entropic inequalities presented by Henson *et al.* [16]. Unfortunately, none of these inequalities could be violated. These early results suggest, although do not prove, that quantum non-classicality does not exist in the Triangle structure for *two-outcome* measurements. Therefore, it is conjectured that all quantum non-classical distributions in the Triangle structure must utilize more than two outcomes in at least one variable.

Pursuant to this conjecture, this work expands the search for quantum-accessible incompatibility in Triangle structure by increasing the cardinality of outcomes.

Moreover, the existence of the Fritz distribution (Section V) directly demonstrates that quantum-accessible incompatibility exists in the Triangle structure and indirectly suggests the existence of causal compatibility inequalities capable of witnessing the non-classicality of the Fritz distribution. The Inflation Technique has been shown to completely solve the classical causal inference problem [22] — indicating that these undiscovered inequalities can be generated using sufficiently complex inflations. The 37 inequalities in [17] were derived using the Spiral inflation (Fig. 9a) and their failure motivated us to consider larger inflations such as those in Figs. 9b and 9c.

These incipient insights were sufficient for focusing the search for new inequalities to 4-outcomes which are violated by the Fritz distribution. Section VII expands on our understanding of non-classicality in the Triangle structure by reporting the discovery of these inequalities which admit quantum-accessible violations. As a by-product, this research supports illuminates the advantage in using the Inflation Technique to find inequalities for interesting causal structures in general.

VII. TRIANGLE STRUCTURE INEQUALITIES

Section II defined causal compatibility inequalities while Section V recalled and discussed the Fritz distribution together with a brief summary 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 non-classicality of any quantum distributions [17]. By leveraging the incompatibility of the Fritz distribution [11] and tools provided by the Inflation Technique [17], we have obtained numerous causal compatibility inequalities for the Triangle structure that are violated by the Fritz distribution. A representative sample of these inequalities are presented here. The methods used to derive these inequalities are largely based on the Inflation Technique [17]. A summary of the Inflation Technique and other methods used will be deferred to Appendix A where unfamiliar readers will find a succinct yet sufficient presentation of the requisites.

The first causal compatibility inequality chosen for presentation is reported below:

$$
\begin{aligned}
I_{\text{WagonWheel}} : \\
+P_{ABC}(000)P_C(3) - 2P_{ABC}(203)P_C(0) - 2P_{ABC}(303)P_C(0) - P_{ABC}(003)P_C(0) \\
-P_{ABC}(013)P_C(0) - P_{ABC}(020)P_C(2) - P_{ABC}(022)P_C(2) - P_{ABC}(023)P_C(0) \\
-P_{ABC}(023)P_C(2) - P_{ABC}(030)P_C(2) - P_{ABC}(031)P_C(2) - P_{ABC}(032)P_C(2) \\
-P_{ABC}(033)P_C(0) - P_{ABC}(033)P_C(2) - P_{ABC}(103)P_C(0) - P_{ABC}(113)P_C(0) \\
-P_{ABC}(123)P_C(0) - P_{ABC}(133)P_C(0) - P_{ABC}(200)P_C(0) - P_{ABC}(200)P_C(1) \\
-P_{ABC}(200)P_C(2) - P_{ABC}(200)P_C(3) - P_{ABC}(201)P_C(0) - P_{ABC}(201)P_C(1) \\
-P_{ABC}(201)P_C(2) - P_{ABC}(201)P_C(3) - P_{ABC}(203)P_C(1) - P_{ABC}(203)P_C(2) \\
-P_{ABC}(203)P_C(3) - P_{ABC}(213)P_C(0) - P_{ABC}(223)P_C(0) - P_{ABC}(300)P_C(0) \\
-P_{ABC}(300)P_C(1) - P_{ABC}(300)P_C(2) - P_{ABC}(300)P_C(3) - P_{ABC}(301)P_C(0) \\
-P_{ABC}(301)P_C(1) - P_{ABC}(301)P_C(2) - P_{ABC}(301)P_C(3) - P_{ABC}(302)P_C(1) \\
-P_{ABC}(303)P_C(1) - P_{ABC}(303)P_C(2) - P_{ABC}(303)P_C(3) - P_{ABC}(313)P_C(0) \le 0
\end{aligned}
\tag{6}
$$

Equation (6) is termed the Wagon Wheel inequality and denoted $I_{\text{WagonWheel}}$ after the identically named inflation (Fig. 9b) utilized to derive it. To reiterate, a summary of the methods used to derive Eq. (6) can be found in Appendices A 1 to A 3.

By construction, every distribution P_{ABC} that is compatible with the Triangle structure (i.e. factorizes according to Eq. (3)) must satisfy $I_{\text{WagonWheel}}$. As a demonstration, the first term in Eq. (6) for the Fritz distribution is $P_{ABC}(000)P_C(3) = \frac{1}{128}\left(2 + \sqrt{2}\right) \simeq 0.0267$ while all of the remaining terms sum to $\simeq 0.0136$. Therefore, the Fritz distribution violates $I_{\text{WagonWheel}}$ with violation $0.0129 \ldots \not\le 0$. Consequently we conclude that the Fritz distribution is incompatible with the Triangle structure, a fact that was previously only demonstrated without the use of causal compatibility inequalities.

It is possible to simplify $I_{\text{WagonWheel}}$ by identifying sets of terms such as $P_{ABC}(300)P_C(0) + P_{ABC}(300)P_C(1) + P_{ABC}(300)P_C(2) + P_{ABC}(300)P_C(3)$ which can be simplified to $P_{ABC}(300)$ by recognizing that $P_C(0) + P_C(1) +$

$P_C(2) + P_C(3) = 1$. Upon doing so, one obtains an equivalent representation for $I_{\text{WagonWheel}}$,

$$
\begin{aligned}
I_{\text{WagonWheel}} : \\
+P_{ABC}(000)P_C(3) + P_{ABC}(021)P_C(2) + P_{ABC}(202) + P_{ABC}(302) + P_{ABC}(233)P_C(0) + P_{AC}(33)P_C(0) \\
-P_{ABC}(303)P_C(0) - P_{ABC}(313)P_C(0) - P_{ABC}(302)P_C(1) - P_{AB}(02)P_C(2) - \\
-P_{AB}(03)P_C(2) - P_{AB}(20) - P_{AB}(30) - P_C(3)P_C(0) \leq 0
\end{aligned}
\tag{7}
$$

In order to understand the importance of $I_{\text{WagonWheel}}$, there are a couple of important concepts to understand. First, the Fritz distribution was originally known to be incompatible with the Triangle structure due to the perfect correlation between C and the measurement pseudo-settings A_l, B_l. The inequality of $I_{\text{WagonWheel}}$ represents a significant advancement in our understanding of non-classicality in the Triangle structure. For example, $I_{\text{WagonWheel}}$ is violated by distributions that do not exhibit perfect correlation between C and A_l, B_l. Furthermore, the existence of an inequality capable of witnessing the non-classicality of the Fritz distribution is guaranteed and the Inflation Technique is capable of producing such an inequality [22], albeit requiring a sufficiently complex inflation of the Triangle structure. In fact, $I_{\text{WagonWheel}}$ was derived using the Wagon-Wheel inflation depicted in Fig. 9b; smaller inflations of the Triangle structure such as the Spiral inflation (Fig. 9a) and others were *not* able to produce causal compatibility inequalities that witness the Fritz distribution. This result demonstrates that the Inflation Technique is the first tool-set powerful enough to produce causal compatibility inequalities such as $I_{\text{WagonWheel}}$ strong enough to witness incompatible quantum distributions for the Triangle structure. More universally, this result shows that the Inflation Technique is a promising tool for understanding quantum non-classical in more general causal structures.

In addition to $I_{\text{WagonWheel}}$, numerous independent inequalities were derived for the Triangle structure that are violated by the Fritz distribution. The majority of these inequalities share most or all of the qualitative features of $I_{\text{WagonWheel}}$, and hence those inequalities are omitted from this manuscript. Instead, we present I_{Web} (Eq. (8)), another causal compatibility inequality for the Triangle structure that is violated by the Fritz distribution and was produced using the Web inflation (Fig. 9c) — a considerably larger inflation than the Wagon-Wheel inflation. Note

that in Eq. (8) $P(abc)$ is shorthand for $P_{ABC}(abc)$.

$$
\begin{aligned}
I_{\text{Web}} : \\
+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) \leq 0
\end{aligned}
\tag{8}
$$

The inequalities presented above, namely $I_{\text{WagonWheel}}$ and I_{Web}, were curated to demonstrate the incompatibility of the Fritz distribution in particular. To reiterate, the Fritz distribution is incompatible with the Triangle structure because it assigns a unique role to one of the variables, specifically C. The asymmetry within the Fritz distribution is reflected in the asymmetry of $I_{\text{WagonWheel}}$ and I_{Web}. In order to discover new forms of non-classicality in the Triangle structure, it is desirable to obtain measures of non-classicality, such as inequalities, that do not admit this bias. In an effort to find a causal compatibility inequality qualitatively distinct from $I_{\text{WagonWheel}}$ or I_{Web}, we developed a technique to derive inequalities with certain symmetry properties such as Eq. (9) which is symmetric with respect to any permutation of the variables A, B, C. These techniques are explained in Appendix A 4. Surprisingly, Eq. (9) is

also violated by the Fritz distribution.

$$I_{\text{SymmetricWeb}}:$$

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

(9)

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, C, i.e. $[P(113)P(330)]_3 \equiv P(113)P(330) + P(131)P(303) + P(311)P(033)$.

Most importantly, these causal compatibility inequalities are a positive resolution of Fritz's problem from Section V. These inequalities are derived without making use of Bell's theorem and instead uses the Inflation Technique to approach non-classicality from a more universal causal inference perspective.

VIII.   NUMERICAL OPTIMIZATIONS & NOISY DISTRIBUTIONS

Causal compatibility inequalities, including those derived in Section VII, offer an avenue into studying quantum non-classicality. Ultimately, the motivation for finding these inequalities is binary: the first is to classify differing distributions by their compatibility with a given causal structure, and the second is to measure the degree in which a given distribution is non-classical. This approach has been well-studied in the Bell structure, where extremal non-classical correlations maximally violate Bell inequalities [20]. However, inequalities are just a singular example of numerous distinct measures of non-classicality [23, Section 5]. Alternatives include measuring the non-local content,
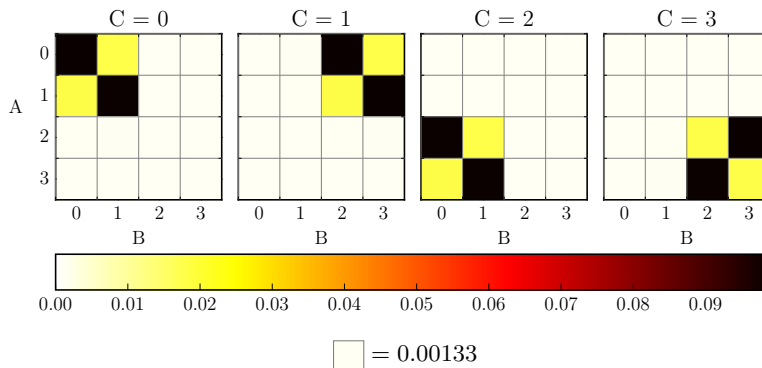
FIG. 5: $\mathcal{N}_{0.085}$: A noisy yet still non-classical variant of the Fritz distribution.

robustness to noise, or even the classical simulation cost of a given distribution. For previously studied causal structures, such as the Bell structure, there exists a uncertainty toward which measure of non-classicality is best. The inequalities derived in this report are likely *not* the best measures of non-classicality simply because they have be curated toward the Fritz distribution. Nonetheless, these inequalities act the *first* measures of non-classicality in the Triangle structure to be found that admit quantum violations. In this section, two measures are used that make use of these compatibility inequalities: robustness to noise and direct numerical violation.

To compare different inequalities simultaneously, our convention is to rearrange each inequality I into the format of Eqs. (8,9,6) where the linear combination of all terms is upper-bound by 0. For each inequality I and distribution P, $I(P)$ denotes the numerical value associated with this upper-bound. Correspondingly, $I(P) > 0$ indicates violation of I while $I(P) \leq 0$ indicates satisfaction of I (with $I(P) = 0$ corresponding to complete saturation of the inequality).

A. Noisy Distributions

In his original work, Fritz [11] was able to find quantum non-classicality in the Triangle structure by embedding Bell non-classicality. As was mentioned in Section V, this embedding required a very idealistic condition to hold: perfect correlations between C and A_l, B_l. From an experimental point of view, it is impossible to use Fritz's original argument to confirm non-classicality because every observable distribution is subject to noise. It is possible to minimize noise by developing more accurate measurement channels but perfect correlations are too idealistic. Causal compatibility inequalities not only permit there to be noise within a set of observations, but they also provide a means of quantifying the amount of noise that is allowed before non-classicality breaks down.

For Triangle structure, we define a completely noisy observation as the uniform probability distribution $\mathcal{U}(abc) = \mathcal{U}_{ABC}(abc) = \frac{1}{64}$ $\forall a, b, c \in \{0, 1, 2, 3\}$, which is trivially compatible with the Triangle structure. The objective of this section is to quantify how much statistical noise can be added to the Fritz distribution P_F before the reported inequalities fail to witness incompatibility. The associated quantifier used was an affine noise parameter ε such that $0 \leq \varepsilon \leq 1$ where $\mathcal{N}_\varepsilon$ denotes an ε-**noisy Fritz distribution**,

$$\mathcal{N}_\varepsilon = (1 - \varepsilon)P_F + \varepsilon\mathcal{U}$$

As ε varies from 0 to 1, more and more noise is added to the Fritz distribution and $\mathcal{N}_\varepsilon$ transitions from an incompatible distribution $\mathcal{N}_0 = P_F$ to a compatible distribution $\mathcal{N}_1 = \mathcal{U}$. Our tests suggest that I_{Web} is most robust to noise and 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.

B. Numerical Optimizations

In order to find quantum distributions that are more non-classical than the Fritz, we performed numerical optimizations against each inequality I by parameterizing the space of quantum-accessible probability distributions that

can be realized on the Triangle structure (Fig. 2) and thus expressed as follows:

$$P_{ABC}(abc) = \text{Tr}[\Pi^\intercal \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA} \Pi M_{A,a} \otimes M_{B,b} \otimes M_{C,c}] \tag{10}$$

Where $\rho_{AB}, \rho_{BC}, \rho_{CA}$ are bipartite density matrices, M_A, M_B, M_C are generic measurements sets and Π is a permutation matrix that aligns the states and measurements appropriately. In order to parameterize all such distributions, we elect to parameterize the states and measurements separately. In order to qualify the scope of Eq. (10) and associated computational complexity of the parameterization, there are a two restrictions that are made with justification. Motivated by the fact that the Fritz distribution (Section 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 projective-operator valued measures (POVMs) for three reasons. First, Section V demonstrates 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 [24], a valid, unbiased parameterization was not found for $k > 2$. Finally, PVMs provide considerable computational advantage over POVMs as they permit Eq. (10) to be re-written as Eq. (11).

$$P_{ABC}(abc) = \langle m_{A,a} m_{B,b} m_{C,c} | \Pi^\intercal \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA} \Pi | m_{A,a} m_{B,b} m_{C,c} \rangle \tag{11}$$

Although there are numerous techniques that can used when parameterizing quantum states and measurements [4, 24–28], a single technique by Spengler *et al.* [29] was found to be most computationally suitable for our purposes. Spengler *et al.* [29] demonstrated that all $d \times d$ unitary matrices U can be parameterized without degeneracy as follows:

$$U = \left[\prod_{m=1}^{d-1} \left(\prod_{n=m+1}^{d} \exp(iP_n \lambda_{n,m}) \exp(i\sigma_{m,n} \lambda_{m,n}) \right) \right] \cdot \left[\prod_{l=1}^{d} \exp(iP_l \lambda_{l,l}) \right] \tag{12}$$

Where the real valued parameters $\lambda = \{\lambda_{n,m} \mid n, m \in 1, \ldots, d\}$ have periodicities $\lambda_{m,n} \in \left[0, \frac{\pi}{2}\right]$ for $m < n$ and $\lambda_{m,n} \in [0, 2\pi]$ for $m \geq n$. Moreover, P_l are one-dimensional projective operators $P_l = |l\rangle\langle l|$ and the $\sigma_{m,n}$ are generalized anti-symmetric σ-matrices $\sigma_{m,n} = -i|m\rangle\langle n| + i|n\rangle\langle m|$ where $1 \leq m < n \leq d$. This parameterization has the useful feature that each of the real-valued parameters $\lambda_{n,m}$ 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 sub-space spanned by $|m\rangle$ and $|n\rangle$ for $m < n$. Analogously, $\exp(iP_n \lambda_{n,m})$ generates the relative phase between $|m\rangle$ and $|n\rangle$ for $m > n$ and $\exp(iP_l \lambda_{l,l})$ fixes the global phase of $|l\rangle$. Finally, although not explicitly mentioned in [29], it is possible to remove the reliance on the computationally expensive matrix exponential operations in Eq. (12) [30].

By parameterizing unitary matrices, it becomes possible to parameterize d-dimensional density matrices and d-element PVMs by recognizing that any orthonormal basis $\{|\psi_j\rangle\}$ (where $1 \leq j \leq d$) 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 $M = \{|m_j\rangle\langle m_j| \mid 1 \leq j \leq d\}$. Since $\{|m_j\rangle\}$ forms an orthonormal basis, one can parameterize M by writing $M = \{U^\dagger|j\rangle\langle j|U \mid 1 \leq j \leq d\}$ and parameterizing U using Eq. (12). This method was inspired by the measurement seeding method for iterative optimization used by Pál and Vértesi [31]. 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 $\text{Tr}(\rho) = \sum_{j=1}^{d} p_j = 1$ the eigenvalues of ρ are parameterized without degeneracy using a tuple of $d-1$ real-valued parameters with periodicity $[0, 2\pi]$ using hyper-spherical coordinates [25, 29]. Additionally, since ρ is Hermitian, the eigenstates $\{|p_j\rangle\}$ form an orthonormal basis and therefore the eigenstates are analogously parameterized using Eq. (12): $\rho = \sum_{j=1}^{d} p_j U^\dagger|j\rangle\langle j|U$. For our purposes, we have set $d = 4$ and fixed $\lambda_{l,l} = 0$ for $1 \leq l \leq d$ in Eq. (12) because the global phase contributions are irrelevant for Eq. (11).

To quantify the degree in which each inequality I could be further violated, we define a relative violation V_F to be the maximum violation relative to the violation achieved by the Fritz distribution, i.e. $V_F \equiv \max_P \{I(P)/I(P_F)\}$. The first inequality in consideration is I_{Web}. At best, numerical optimizations converged to a distribution (Fig. 6) with relative violation $V_F \simeq 1.501$. Almost immediately, it is evident that Fig. 6 closely resembles the Fritz distribution. In fact, Fig. 6 and Fig. 4 share the exact same *possibilistic structure*, i.e. the same set of possible events. This result is not entirely unexpected. As was mentioned in Section VII, these inequalities are derived specifically to prove the incompatibility of the Fritz distribution and the Triangle structure. Consequently, it seems that I_{Web} only has the capacity to demonstrate incompatibility for distributions that resemble the Fritz distribution very closely.

An alternative explanation for qualitative similarities between Fig. 6 and Fig. 4 is related to the sensitively of the initial parameters to our optimization. Above it was shown that our parameter space of choice was 2π-periodic in all entries. Somewhat surprisingly, when supplied with randomly sampled initial parameters, all optimization
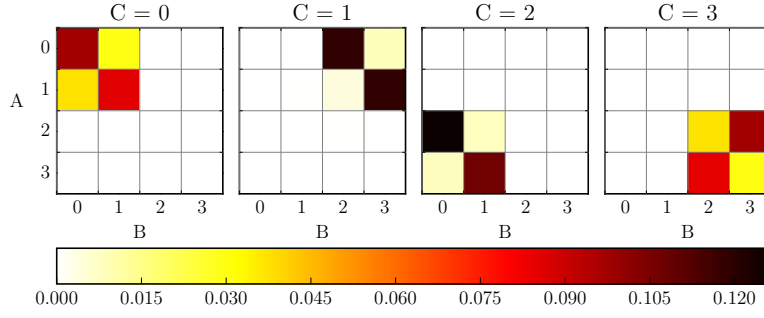
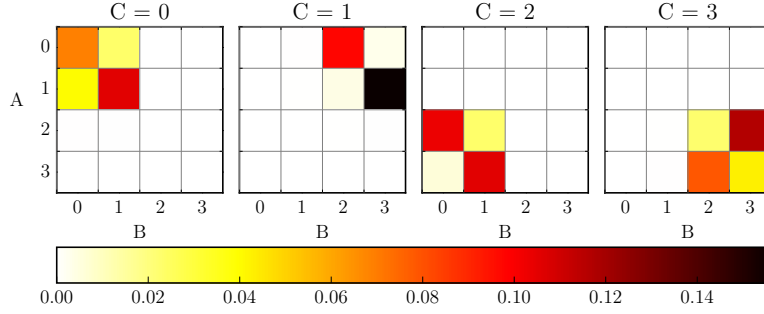FIG. 6: Probability distribution that maximizes violation of I_{Web} ($V_F \simeq 1.501$).

FIG. 7: Probability distribution that maximizes violation of $I_{\text{SymmetricWeb}}$ ($V_F \simeq 2.61$).

methods consistently converged to inequality *saturation* instead of inequality *violation*. Instead, it was observed that only parameters close enough to the *Fritz parameters* (i.e. any non-unique set of parameters that induce the Fritz distribution) were able to achieve violation. Since the inequalities themselves are polynomial and the space of quantum-accessible distributions is non-convex, a number of optimization methods were used in an attempt to eliminate any issues associated with local minima or ill-conditioned convergence. Specifically, the BFGS method [32, p.142] and the Nelder-Mead simplex method [32, p.238] were used along with a method called Basin Hopping [33], which is a hybrid between simulated annealing and gradient descent-based methods. Despite our varied efforts, it remains unclear whether or not our results share the same possibilistic structure as P_{F} because of the inequality itself or the ill-conditioned nature of the optimization. More likely, this phenomena is due to a combination of both effects.

It is also worth reporting that the distribution in Fig. 6 is not accessible when the bipartite states in Eq. (10) are restricted to maximally entangled states. This result resembles a feature of quantum mechanics originally presented by Methot and Scarani [34] which demonstrates that entanglement and non-classicality are different resources.

The second inequality in consideration is the *symmetric* inequality $I_{\text{SymmetricWeb}}$. This inequality was able to
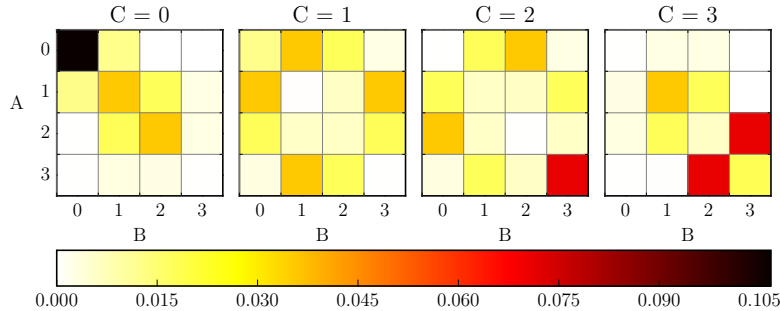

FIG. 8: Symmetrized version of the Fritz distribution ($V_F \simeq 3$).

achieve a greater relative violation than I_{Web} as optimizations converged to a distribution (Fig. 7) with relative violation $V_F \simeq 2.61$. Just as before, the maximum violating distribution shares a possibilistic structure with the Fritz distribution. However, this result is far more surprising. Since the inequality is symmetric with respect to the permutation of parties, one would naively expect that the maximum violating distribution be symmetric as well. If it were not, one could construct a distribution that violates $I_{\text{SymmetricWeb}}$ further by symmetrizing P_F (Fig. 8) by normalizing over all possible permutations of A, B, C. This distribution is also non-classical with relative violation $V_F = 3$. As it turns out, numerical optimizations suggest that the symmetrized Fritz distribution is *not* quantum-accessible in the sense that is cannot be written as Eq. (10). Therefore we conclude that the space of quantum-accessible correlations on the Triangle structure is non-convex; a reiteration of a result mentioned in Wolfe *et al.* [17].

Finally, the maximum relative violation of $I_{\text{WagonWheel}}$ was found to be $V_F = 1$ (up to numerical precision), indicating that the Fritz distribution either represents a local minimum of the parameter space or is in fact *the* maximally violating distribution. Without further investigations, it remains unclear if this behavior is related to the universality of the Fritz distribution or if its an artifact of the methods used to derive $I_{\text{WagonWheel}}$.

IX.   CONCLUSIONS

In Section IV, we elucidated that compatibility in the Triangle structure has been an unsolved problem for nearly a decade. Initially, Branciard *et al.* [15] identified the Triangle structure as a challenging causal structure for studying non-classicality. Afterwards, Henson *et al.* [16] were able to to classify the Triangle structure as *interesting* which partially explains why it remains an enigma. Recently, Fritz [11] discovered the Fritz distribution as the first example of quantum non-classicality in the Triangle structure. In Section V the Fritz distribution was described in detail along with Fritz's Problem [11, Problem 2.17]. Heretofore, researchers have been able to derive causal incompatibility inequalities [13, 16, 17] but unable to show that any are non-trivial in the context of quantum non-classicality.

In Section VII, we presented the first 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 extensive use of the Inflation Technique (Appendix A) applied to the Fritz distribution. In the context of Fritz's Problem, our result corresponds to a resolution to the following *reinterpretation* of Fritz's original Problem:

R.1 Find proofs for causal incompatibility of quantum distributions on the Triangle structure that are *not* reliant on perfect correlations.

Ultimately, this work accomplishes two important tasks. First and foremost, the derivation of inequalities such as Eq. (6) demonstrates the utility of the Inflation Technique specifically for quantum causal inference problems. The Inflation Technique was able to derive causal compatibility inequalities that admit quantum violations in the Triangle structure — a task that no other existing techniques have succeeded at. In this sense, the results of Section VII showcase how the Inflation Technique is capable of generating new proofs of causal incompatibility specifically of quantum distributions for the Triangle structure. Consequently, this work strongly indicates that the Inflation Technique has the potential to solve an assortment of unsolved causal inference problems including those manifest in quantum foundations and quantum information theory. Secondly, Section VIII A demonstrates that the derived causal compatibility inequalities are robust to varying degrees of noise, directly revealing a critical departure from Fritz's original proof of the incompatibility of the Fritz's distribution (Section V) which relies on perfect (non-noisy) correlations. In conclusion **R.1** has been positively solved through the derivation of the inequalities found in Section VII.

Fritz's Problem is concerned with identifying incompatible quantum distributions in the Triangle structure that are qualitatively different than the non-classical distributions of the Bell structure discussed in Section III. When specifically concerned with entanglement resources, Bell structure non-classicality exploits the entanglement of a single bipartite quantum state. In the Triangle structure, the observable nodes A, B, C have access to potentially three different bipartite quantum states. As noted in Section V, the Fritz distribution can be implemented using only a single entangled state.[6] In consideration of our resolution to **R.1**, we propose an alternative reinterpretation of Fritz's Problem that are concerned more with the nature of the entanglement resource being exploited in a incompatible quantum distribution:

R.2 Find incompatible quantum distributions on the Triangle structure that require entanglement in more than one shared state.

–––––––

[6] Of course there exists implementations that make use of three entangled states but at minimum only one is necessary.

The question, as proposed by **R.2**, is: does there exist incompatible quantum distributions that genuinely require the entanglement resources of three states? At first, one might suspect that finding such distributions would affirmatively demonstrate a new form of entanglement resource distinct to the Triangle structure. Unfortunately, this is simply not the case. By increasing the cardinality of the outcomes for A, B, C, it is possible to generate a distribution that is non-classical 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 of which utilizes a distinct party to announce the corresponding measurement pseudo-settings. Such a distribution would be of the form described by **R.2** because under this construction, the removal of any entangled resource would render the distribution quantum-inaccessible. Consequently, it becomes desirable to find non-classical quantum correlations for the Triangle structure which have the fewest number of outcomes. In pursuit of this objective, the authors were able to prove the non-classicality (using the methods described in Appendix A) of a variant of the Fritz distribution where C has only two outcomes: the first corresponding to A, B correlations and the second corresponding to A, B anti-correlations. These advancements suggest the cardinality of measurement outcomes are intimately connected to quantum non-classicality and is a recommended focus for future research.

Using the Inflation technique, it became possible to provide a stronger proof of non-classicality than proposed by Fritz [11]. In particular, noisy variants of the Fritz distribution are provably non-classical (Section VIII A) and there exist quantum distributions quantitatively distinct from the Fritz distribution which admit greater numerical violations (Section VIII B). Despite these advancements, the distributions obtained closely resemble the Fritz distribution indicating that their non-classical nature remains some recycled version of the non-classicality found in the Bell structure. Currently, it remains speculative about how to recognize whether or not a given distribution for the Triangle structure exhibits non-classical correlations that maintain distinct qualitatively advantages from a resource perspective. One possibility is to extend Item **R.2** to include the necessity of entanglement measurements. Ultimately, given any said classifier, it remains unclear whether or not such distributions even exist.

Appendix A: Summary of Inflation Technique

1. Marginal Satisfiability and Inequalities

This section 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** [35]. Additionally, this section aims to accomplish something further. If such as joint distribution $P_{\mathcal{J}}$ exists, how does one find it? Moreover, if a joint distribution does not exist, find an inequality that $\{P_{V_1}, \ldots, P_{V_m}\}$ fails which proves such a joint distribution does not exist. 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 section is presented prior to Appendix A 2 as the marginal problem becomes an integral component of the Inflation Technique in subsequently deriving the inequalities presented in Section VII. Moreover, the Marginal problem is presented here logically independent from the remainder of the manuscript both for procedural clarity and because the marginal problem has applications to numerous areas of mathematics including game theory [36], database theory, knowledge integration of expert system, and of course, quantum information theory [35].

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 with an 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 by [35], a marginal scenario forms an *abstract simplicial complex* where it is required that all subsets of contexts are also contexts: $\forall V \in \mathcal{M}, V' \subset V : V' \in \mathcal{M}$. Throughout

this section, we exclusively consider (without loss of generality) the *maximal* marginal scenario; restricting our focus to the largest contexts in the marginal scenario. 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 asks: given a marginal model $P^{\mathcal{M}} = \{P_V \mid V \in \mathcal{M}\}$ marginal to the joint variables $\mathcal{J}$, does there exist a joint distribution $P_{\mathcal{J}}$ such that each context P_V can be obtained by marginalizing $P_{\mathcal{J}}$?

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

A marginal model $P^{\mathcal{M}}$ is said to be *contextual* if it *does not* admit a joint distribution and *non-contextual* otherwise. Notice that Eq. (A1) is inherently a linear system of constraints which be solved rather efficiently using linear programs. In consideration of this, we will now endeavor to discuss how to cast the 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 $\mathcal{S}(v)$,[7] 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, $\mathcal{S}(v)$ 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 [37] demands the distinction. Specifically for this work, the distinction becomes essential for our discussion and exploit of marginal symmetries in Appendix A 4. As a natural generalization we define the events over a set of random variables $V = \{v_1, \ldots, v_n\}$ in a parallel manner,

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

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 $s \in \mathcal{S}(V)$, then $\mathcal{D}(s) = V$. Under this framework, a probability distribution P_V can be considered as a map from $\mathcal{S}(V)$ 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 $s \in \mathcal{S}(V)$ can be "marginalized" or restricted to a smaller event $s' \in \mathcal{S}(W)$ whenever $W \subseteq V$. For every $W \subseteq V$ and $s \in \mathcal{S}(V)$, the **restriction of s onto W** (denoted $s|_W \in \mathcal{S}(W)$) is the event in $\mathcal{S}(W)$ that *agrees* with each of s's assignments for variables in W: $\forall v \in W : s|_W(v) = s(v)$.

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

The marginalization operation (Eq. (A1)) of the marginal problem is a linear operator. This linear operator maps a joint probability distribution $P_{\mathcal{J}} : \mathcal{S}(\mathcal{J}) \to [0,1]$ into a marginal model $P^{\mathcal{M}} = \{P_V : \mathcal{S}(V) \to [0,1] \mid V \in \mathcal{M}\}$. Since $\mathcal{S}(\mathcal{M})$ and $\mathcal{S}(\mathcal{J})$ are finite, the marginal problem can be represented as a $|\mathcal{S}(\mathcal{M})| \times |\mathcal{S}(\mathcal{J})|$ matrix.

Definition 1. The **incidence matrix** M for a marginal scenario $\mathcal{M} = \{V_1, \ldots, V_k\}$ is a bit-wise matrix where the columns are indexed by *joint* events $j \in \mathcal{S}(\mathcal{J})$ and the rows are events by *marginal* events $m \in \mathcal{S}(\mathcal{M})$. The entries of M are populated whenever a marginal event m is a restriction of the joint event j.

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

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

To illustrate this concretely, consider the following example. Let $\mathcal{J}$ be 3 binary variables $\mathcal{J} = \{A, B, C\}$ and $\mathcal{M}$

---

[7] In the language of sheaf-theory, $\mathcal{S}(v)$ is the *sheaf of events* [37].

be the marginal scenario $\mathcal{M} = \{\{A, B\}, \{B, C\}, \{A, C\}\}$. The incidence matrix for $\mathcal{M}$ is:

$$
M = \begin{array}{c}
(A,B,C) \mapsto \\
(A \mapsto 0, B \mapsto 0) \\
(A \mapsto 0, B \mapsto 1) \\
(A \mapsto 1, B \mapsto 0) \\
(A \mapsto 1, B \mapsto 1) \\
(B \mapsto 0, C \mapsto 0) \\
(B \mapsto 0, C \mapsto 1) \\
(B \mapsto 1, C \mapsto 0) \\
(B \mapsto 1, C \mapsto 1) \\
(A \mapsto 0, C \mapsto 0) \\
(A \mapsto 0, C \mapsto 1) \\
(A \mapsto 1, C \mapsto 0) \\
(A \mapsto 1, C \mapsto 1)
\end{array}
\begin{array}{cccccccc}
(0,0,0) & (0,0,1) & (0,1,0) & (0,1,1) & (1,0,0) & (1,0,1) & (1,1,0) & (1,1,1) \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{array}
\qquad (A3)
$$

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

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

$$\mathcal{P}^{\mathcal{M}} = M \cdot \mathcal{P}^{\mathcal{J}} \qquad (A4)$$

As a quick remark, the particular ordering of the rows and columns of M carries *no importance*, but it *must* be consistent between M, $\mathcal{P}^{\mathcal{J}}$ and $\mathcal{P}^{\mathcal{M}}$. The marginal problem can now be rephrased in the language of the incidence matrix. Suppose one obtains a marginal distribution vector $\mathcal{P}^{\mathcal{M}}$; the marginal problem becomes equivalent to the question: does there exist a joint distribution vector $\mathcal{P}^{\mathcal{J}}$ such that Eq. (A4) holds? This question is naturally framed as the **marginal linear program**:

$$
\begin{aligned}
\text{minimize:} \quad & \emptyset \cdot x \\
\text{subject to:} \quad & x \succeq 0 \\
& M \cdot x = \mathcal{P}^{\mathcal{M}}
\end{aligned}
$$

If this "optimization"[8] is *feasible*, then there exists a vector x than can satisfy Eq. (A4) and is a valid joint distribution vector. Therefore, feasibility of the marginal linear program not only implies that $\mathcal{P}^{\mathcal{J}}$ exists but returns $\mathcal{P}^{\mathcal{J}}$. Moreover if the marginal linear program is *infeasible*, then there *does not* exist a joint distribution $\mathcal{P}^{\mathcal{J}}$. To every linear program, there exists a dual linear program that characterizes the feasibility of the original [38]. Constructing the dual linear program is straightforward [39].

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

The dual marginal linear program not only answers the marginal problem for a specific marginal model $\mathcal{P}^{\mathcal{M}}$ 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. Secondly if the dual constraint $y \cdot M \succeq 0$ holds *and* the primal is feasible, then $y \cdot \mathcal{P}^{\mathcal{M}} = y \cdot M \cdot x \geq 0$. Therefore the *sign* of the dual objective $d \equiv \min(y \cdot \mathcal{P}^{\mathcal{M}})$ classifies a marginal model's contextuality; if $d < 0$ then $y \cdot \mathcal{P}^{\mathcal{M}} \geq 0$ is violated and therefore $\mathcal{P}^{\mathcal{M}}$ is contextual. Likewise if $d \geq 0$ (satisfying $y \cdot \mathcal{P}^{\mathcal{M}}$), then $\mathcal{P}^{\mathcal{M}}$ is non-contextual.[9] This is manifestation of Farkas's lemma [38]. An **infeasibility certificate** [41] is any vector y that satisfies the $y \cdot M \succeq 0$.

———

[8] "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*.

[9] Actually, if $d \geq 0$ 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 [40]. 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 [40]; if the dual is unbounded, then the primal is infeasible.

Most linear program software packages such as Mosek [42], Gurobi [43], CPLEX [44] and CVX/CVX OPT [45, 46] 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 non-contextual marginal models. If y is an infeasibility certificate, then $y \cdot \mathcal{P}^{\mathcal{M}} \geq 0$ 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 [35, 47, 48]: *Given a marginal scenario $\mathcal{M}$, what is the set of all non-contextual marginal models?* Pitowsky [49] demonstrates that the set of non-contextual 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 $\mathcal{P}^{\mathcal{J}}$. Since all joint distributions $\mathcal{P}^{\mathcal{J}}$ are probability distributions, their entries must sum to unity $\sum_{j \in \mathcal{S}(\mathcal{J})} \mathcal{P}_j^{\mathcal{J}} = 1$. This normalization defines the convexity of the polytope; all non-contextual marginal models are convex mixtures of the deterministic marginal models pursuant to Eq. (A4). 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 [10]. From the perspective of a marginal polytope, convex hull algorithms or linear quantifier elimination 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** [17, 47, 50, 51]. Applying the Fourier-Motzkin procedure to completely solve the marginal problem is discussed in more detail in Fritz and Chaves [35]. An excellent survey of existing techniques for solving the marginal problem including Equality Set Projection [51] and Hardy-type hypergraph transversals can be found in Wolfe *et al.* [17]. 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 [50]. For the purposes of this research, solving the marginal problem without reference to a marginal model was intractable. This will become apparent in Appendix A 3 when the Inflation Technique is applied to the Triangle structure, producing marginal models that are very large. Luckily, the Fritz distribution (Section V) 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.

2. Inflation Technique

The **Inflation Technique**, invented by Wolfe *et al.* [17] and inspired by the *do calculus* and *twin networks* of Pearl [8], is a family of causal inference techniques that can be used to determine if a probability distribution $P_{\mathcal{N}_O}$ is compatible or incompatible with a given causal structure $\mathcal{G}$. As a preliminary summary, the Inflation Technique begins by *augmenting* a causal structure with additional copies of its nodes, producing an *inflated* causal structure, and then exposes how causal inference tasks on the inflated causal structure can be used to make inferences on the original causal structure. As a example, four inflations of the Triangle structure are depicted in Fig. 9. 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, A_4 in the inflated causal structure in Fig. 9c. All such copies are deemed equivalent via a **copy-index equivalence** relation denoted '$\sim$'. A copy-index is effectively arbitrary, so we will refer to an arbitrary inflated copy of A as A'.

$$A \sim A_1 \sim A' \not\sim B \sim B_1 \sim B'$$

We preemptively generalize the notion of copy-index equivalence to other mathematical objects like sets, graphs, and groups by saying that $X \sim Y$ if and only if $X = Y$ upon removal of the copy-index. In addition to the common graph-theoretic terminology and notation presented in Section II, we define two more concepts necessary to formally define an inflation of $\mathcal{G}$. First, an **induced subgraph** of $\mathcal{G}$ for a subset of nodes $N \subseteq \mathcal{N}$ is another 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: $\mathsf{AnSub}_{\mathcal{G}}(N) \equiv \mathsf{Sub}_{\mathcal{G}}\big(\mathsf{An}_{\mathcal{G}}(N)\big)$.

Definition 2. An **inflation** of a causal structure $\mathcal{G}$ is another causal structure $\mathcal{G}'$ such that:

$$\forall n' \in \mathcal{N}' : \mathsf{AnSub}_{\mathcal{G}'}(n') \sim \mathsf{AnSub}_{\mathcal{G}}(n) \tag{A5}$$

Verbally, for each inflated node n', the ancestral sub-graph of n' in $\mathcal{G}'$ is equivalent to its counterpart in $\mathcal{G}$ up to copy-index. The motivation behind this definition is that if the ancestry of each node n' in $\mathcal{G}'$ plays the *exact same*

role as the ancestry of its source copy n in $\mathcal{G}$, then every compatible distribution P_N on $\mathcal{G}$ can be used to *create* compatible joint distributions on $\mathcal{G}'$. To do this, first notice that for every distribution P_N compatible with $\mathcal{G}$, there must exist a set of causal parameters $\{P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N}\}$ for $\mathcal{G}$ such that $P_N = \prod_{n \in N} P_{n|\mathsf{Pa}_{\mathcal{G}}(n)}$. Since the inflation criteria (Eq. (A5)) requires that $\mathsf{Pa}_{\mathcal{G}'}(n') \sim \mathsf{Pa}_{\mathcal{G}}(n)$ one can create a set of **inflated causal parameters** by matching causal parameters according to respective copy-indices.

$$\forall n' \in \mathcal{N}' : P_{n'|\mathsf{Pa}_{\mathcal{G}'}(n')} \equiv P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \tag{A6}$$

Which in turn uniquely defines a compatible joint distribution $P_{\mathcal{N}'}$ on $\mathcal{G}'^{10}$.

$$P_{\mathcal{N}'} = \prod_{n' \in \mathcal{N}'} P_{n'|\mathsf{Pa}_{\mathcal{G}'}(n')} \tag{A7}$$

In essence, a compatible joint distribution over $\mathcal{N}$ can be *inflated* to induce compatible joint distributions over $\mathcal{N}'$. Before generalizing this idea, a careful observation needs to be made. For any pair of subsets of nodes $N \subseteq \mathcal{N}$ and $N' \subseteq \mathcal{N}'$ that are equivalent up to copy-index $N \sim N'$, if $\mathsf{AnSub}_{\mathcal{G}}(N) \sim \mathsf{AnSub}_{\mathcal{G}'}(N')$ then any compatible *marginal* distribution P_N, where $N \subset \mathcal{N}$, induces a compatible marginal distribution $P_{N'}$ over N'. In fact, N' must contain no duplicate nodes up to copy index and thus $P_N = P_{N'}$. These subsets of nodes are so essential to the Inflation Technique they are assigned a special name: sets $N' \subseteq \mathcal{N}'$ as called the **injectable sets** of $\mathcal{G}'$ and $N \subset \mathcal{N}$ the **images of the injectable sets** of $\mathcal{G}$.

$$\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}') \equiv \{N' \subseteq \mathcal{N}' \mid \exists N \subseteq \mathcal{N} : N \sim N'\}$$
$$\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}') \equiv \{N \subseteq \mathcal{N} \mid \exists N' \subseteq \mathcal{N}' : N \sim N'\}$$

Injectable sets are incredibly useful for causal compatibility because they permit one to inflate compatibility over a particular subset of *marginal models*. A marginal model is compatible with a given causal structure $\mathcal{G}$ if it admits a joint distribution (see Appendix A 1) that is compatible with $\mathcal{G}$.

Lemma 3. ***The Inflation Lemma*** *[17, Lemma 3] Given a particular inflation $\mathcal{G}'$ of $\mathcal{G}$ and a marginal model $\{P_N \mid N \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\}$ compatible with $\mathcal{G}$ then all marginal models $\{P_{N'} \mid N' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\}$ are compatible with $\mathcal{G}'$ provided that $P_N = P_{N'}$ for all instances where $N \sim N'$.*

Diagrammatically, Lemma 3 can be visualized as follows:

The inflation lemma is the most important result of the causal Inflation Technique. The contrapositive version of Lemma 3 is a powerful tool for deriving causal compatibility constraints. Any causal compatibility constraint, such as a causal compatibility inequality, that constrains injectable marginal models $\{P_{N'} \mid N' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\}$ can be *deflated* into a valid compatibility constraint on marginal models $\{P_N \mid N \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\}$ of the original causal structure by dropping all references to the copy-indices. Additionally, the inflation lemma holds when considering any subset of $\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ (analogously $\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')$). Therefore, one only needs to consider injectable sets that are composed of *observable nodes*.

Heretofore, the Inflation Technique as stated by Lemma 3 does not facilitate the derivation of any non-trivial inequalities. The Inflation lemma simply casts causal inference problems on $\mathcal{G}$ to causal inference problems on $\mathcal{G}'$. Fortunately, an inflated causal structure $\mathcal{G}'$ possesses its own conditional independence relations which can be exploited to turn linear inequalities on $\mathcal{G}'$ into polynomial inequalities on $\mathcal{G}$. After obtaining an inequality for $\mathcal{G}'$, each of its term is generically of the form $P_{N'}$ where N' may not be injectable $N' \notin \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$. At this point, the Inflation lemma is not

---

[10] Of course, there exists compatible joint distributions on $\mathcal{N}'$ that can not be constructed in this manner. The claim is that all joint distributions that are constructed in this way are compatible.
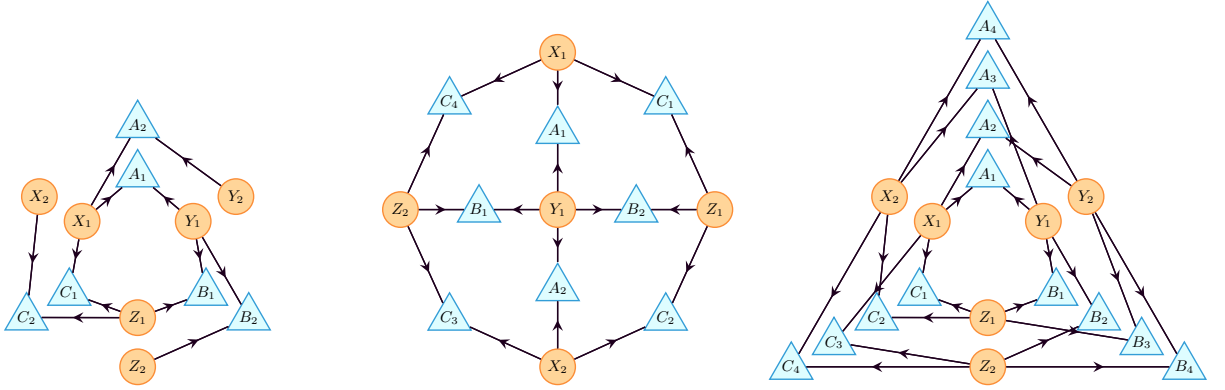
(a) The Spiral inflation. (b) The Wagon-Wheel inflation. (c) The Web Inflation.

FIG. 9: Some inflations of the Triangle structure.

applicable because it requires probabilistic constraints to consist entirely of distributions $P_{N'}$ where N' is injectable. If for example, $N' \notin \mathsf{Inj}_\mathcal{G}(\mathcal{G}')$ could be partitioned into two injectable subsets $N_1' \in \mathsf{Inj}_\mathcal{G}(\mathcal{G}')$ and $N_2' \in \mathsf{Inj}_\mathcal{G}(\mathcal{G}')$ that where **ancestrally independent** in $\mathcal{G}'$,[11]

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

Then the compatible marginal context $P_{N'}$ must be probabilistically separable $P_{N'} = P_{N_1'} P_{N_2'}$ due to Eq. (1). Since N_1' and N_2' are both injectable, the inflation lemma now applies and the original inequality can be deflated.

$$
\begin{array}{ccc}
& N_1' \cup N_2' = N & \\
P_{N'} & \xrightarrow{\quad N_1' \perp N_2' \quad} & P_{N_1'} P_{N_2'} \\
\mid & & \mid \\
N' \notin \mathsf{Inj}_\mathcal{G}(\mathcal{G}') & & N_1', N_2' \in \mathsf{Inj}_\mathcal{G}(\mathcal{G}') \\
\textcolor{red}{\times} & & \downarrow \\
?? & & P_{N_1} P_{N_2}
\end{array}
$$

Not all sets $N' \subseteq \mathcal{N}'$ can be written as the disjoint union of injectable sets, but those that *can* are called the **ai-expressible sets** of $\mathcal{G}'$. A set $N' \subseteq \mathcal{N}'$ is a 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: $\forall i, j : N_i' \perp N_j'$.

A ai-expressible set N' is of primal importance to the Inflation Technique because any distribution over N' will factorize according to its ancestral independence conditions [8][12],

$$P_{N'} = P_{\coprod_i N_i'} = \prod_i P_{N_i'}$$

Application of the inflation lemma to linear inequalities constraining ai-expressible distributions $P_{N'}$ transforms those inequalities into polynomial inequalities over the injectable distributions, allowing one to replace all terms with probability distributions defined over variables in $\mathcal{N}$. Throughout the body of this work, we let $\mathsf{AI}_\mathcal{G}(\mathcal{G}')$ denote the set of all ai-expressible sets. To derive inequalities for the inflated causal structure $\mathcal{G}'$, the ai-expressible sets define a marginal scenario $\mathcal{M}$[13] that are used to derive inequalities using methods discussed in Appendix A 1.

3. Exemplary Inflations of The Triangle structure

Appendix A 2 explains how the Inflation Technique [17] was used to derive novel causal compatibility inequalities for the Triangle structure by constructing inflations of the Triangle structure satisfying Eq. (A5). Explicitly, for a

---

[11] Ancestral independence is equivalent to d-separation conditioned on the empty set $\emptyset$ [8].

[12] In [17], 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. 9.

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

given inflation $\mathcal{G}'$ of the Triangle structure, one constructs a marginal problem (see Appendix A 1) for the marginal scenario $\mathcal{M}$ composed of the maximal ai-expressible sets of $\mathcal{G}'$, i.e. $\mathcal{M} = \mathsf{AI}_{\mathcal{G}}(\mathcal{G}')$. This section acts as a reference for the particular inflations of the Triangle structure that are mentioned throughout this manuscript. Contained within this section is a depiction of each of these inflations in Fig. 9 as well as a their ai-expressible sets. An efficient algorithm for computing the injectable and ai-expressible sets of an inflation can be found in [17].

The first inequality presented in Section VII was Eq. (6). Eq. (6) was derived using the Wagon-Wheel inflation in Fig. 9b. The Wagon-Wheel inflation possesses 4 copies of C (C_1, C_2, C_3, C_4) and 2 copies of A (A_1, A_2) and B (B_1, 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 Eq. (A9). These maximal ai-expressible sets define the following marginal scenario where the joint variables $\mathcal{J}$ are simply the set of observable nodes in the Wagon-Wheel inflation $\mathcal{J} = \mathcal{N}'_O$:

$$\mathcal{M} = \mathsf{AI}_{\mathcal{G}}(\mathcal{G}') = \{\{A_2, B_1, C_3, C_1\}, \{A_1, B_1, C_4, C_2\}, \{A_1, B_2, C_1, C_3\}, \{A_2, B_2, C_2, C_4\}\} \tag{A8}$$

Given that the Fritz distribution is defined over 4-outcomes variables, the variables in the marginal scenario are assigned 4-outcomes as well. This marginal scenario $\mathcal{M}$ then defines an incidence matrix M capable of accommodating the Fritz distribution that has $|\mathcal{S}(\mathcal{M})| = 4 \cdot 4^4 = 1024$ rows and $|\mathcal{S}(\mathcal{J})| = 4^8 = 65,536$ columns. Understandably, 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.

Maximal AI-Expressible Sets	Ancestral Independences	
$\{A_2, B_1, C_3, C_1\}$	$\{A_2, B_1, C_3\} \perp \{C_1\}$	
$\{A_1, B_1, C_4, C_2\}$	$\{A_1, B_1, C_4\} \perp \{C_2\}$	(A9)
$\{A_1, B_2, C_1, C_3\}$	$\{A_1, B_2, C_1\} \perp \{C_3\}$	
$\{A_2, B_2, C_2, C_4\}$	$\{A_2, B_2, C_2\} \perp \{C_4\}$	

To construct the marginal distribution vector $\mathcal{P}^{\mathcal{M}}$ for the Fritz distribution, one begins with $\mathcal{P}^{\mathcal{M}}$ defined in a *symbolic* form,

$$\mathcal{P}^{\mathcal{M}\intercal} = \underbrace{(P_{A_2 B_1 C_3 C_1}(0000), \ldots, P_{A_1 B_1 C_4 C_2}(3232), \ldots, P_{A_2 B_2 C_2 C_4}(3333))}_{1024 \text{ entries}} \tag{A10}$$

Which can be factored into the ancestrally independent injectable sets using Eq. (A9),

$$\mathcal{P}^{\mathcal{M}\intercal} = (P_{A_2 B_1 C_3}(000)P_{C_1}(0), \ldots, P_{A_1 B_1 C_4}(323)P_{C_2}(2), \ldots, P_{A_2 B_2 C_2}(333)P_{C_4}(3)) \tag{A11}$$

And penultimately, each probability distribution in Eq. (A11) is deflated by dropping copy-indices.

$$\mathcal{P}^{\mathcal{M}\intercal} = (P_{ABC}(000)P_C(0), \ldots, P_{ABC}(323)P_C(2), \ldots, P_{ABC}(333)P_C(3)) \tag{A12}$$

This step is permitted because all of the remaining distributions in Eq. (A11) are defined over the injectable sets of the Wagon-Wheel inflation. Finally, each of the elements of $\mathcal{P}^{\mathcal{M}}$ are replaced with numerical values pursuant to the Fritz distribution. For example, $P_{ABC}(323)P_C(2)$ is assigned the following numerical value:

$$P_{ABC}(323)P_C(2) = \frac{1}{32}\left(2 + \sqrt{2}\right) \cdot \frac{1}{4} \simeq 0.0267$$

The same is applied to all other entries of $\mathcal{P}^{\mathcal{M}}$. Finally, this numerical version of $\mathcal{P}^{\mathcal{M}}$ and M are subjected to linear programing software and an infeasibility certificate y was obtained, corresponding precisely to Eq. (6) using Eq. (A12).

The remaining inequalities in Section VII were derived using the Web inflation of Fig. 9c. The maximal ai-expressible

sets of the Wagon-Wheel inflation along with their ancestral dependences can be found in Eq. (A13).

$$
\begin{array}{cc}
\textbf{Maximal AI-Expressible Sets} & \textbf{Ancestral Independences} \\
\{A_1, B_1, C_1, A_4, B_4, C_4\} & \{A_1, B_1, C_1\} \perp \{A_4, B_4, C_4\} \\
\{A_1, B_2, C_3, A_4, B_3, C_2\} & \{A_1, B_2, C_3\} \perp \{A_4, B_3, C_2\} \\
\{A_2, B_3, C_1, A_3, B_2, C_4\} & \{A_2, B_3, C_1\} \perp \{A_3, B_2, C_4\} \\
\{A_2, B_4, C_3, A_3, B_1, C_2\} & \{A_2, B_4, C_3\} \perp \{A_3, B_1, C_2\} \\
\{A_1, B_3, C_4\} & \{A_1\} \perp \{B_3\} \perp \{C_4\} \\
\{A_1, B_4, C_2\} & \{A_1\} \perp \{B_4\} \perp \{C_2\} \\
\{A_2, B_1, C_4\} & \{A_2\} \perp \{B_1\} \perp \{C_4\} \\
\{A_2, B_2, C_2\} & \{A_2\} \perp \{B_2\} \perp \{C_2\} \\
\{A_3, B_3, C_3\} & \{A_3\} \perp \{B_3\} \perp \{C_3\} \\
\{A_3, B_4, C_1\} & \{A_3\} \perp \{B_4\} \perp \{C_1\} \\
\{A_4, B_1, C_3\} & \{A_4\} \perp \{B_1\} \perp \{C_3\} \\
\{A_4, B_2, C_1\} & \{A_4\} \perp \{B_2\} \perp \{C_1\}
\end{array}
\tag{A13}
$$

Analogously to the Wagon-Wheel inflation, these ai-expressible sets form a marginal scenario $\mathcal{M}$ which defines an incidence matrix M that has $|\mathcal{S}(\mathcal{M})| = 4 \cdot 4^6 + 8 \cdot 4^3 = 16,896$ rows, $|\mathcal{S}(\mathcal{J})| = 4^{12} = 16,777,216$ columns and $201,326,592$ non-zero entires.

4. Deriving Symmetric Causal Compatibility Inequalities

Appendix A 2 detailed how to obtain causal compatibility inequalities for any causal structure by constructing a marginal problem (see Appendix A 1) and supplying an incompatible distribution to generate an infeasibility certificate. In Section VII, a causal compatibility inequality symmetric under permutations of the variables A, B, C was presented (see Eq. (9)). This section aims to describe a general technique that can be used to derive inequalities exhibiting this symmetry. In brief, this is accomplished by grouping marginal events $m \in \mathcal{S}(\mathcal{M})$ 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.* [52] discuss computational advantages in considering symmetric versions of marginal polytopes mentioned in Appendix A 1; the number of extremal points typically grows exponentially in $\mathcal{J}$, but only polynomial for the symmetric polytope. They also note a number of interesting inequalities (such as CHSH [18]) can be written in a way that is symmetric under the exchange of parties, demonstrating that non-trivial inequalities can be recovered from facets of a symmetric polytope. Secondly, 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[14].

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 $\{0, 1\}$. Moreover, let $I_{\mathcal{M}} \equiv \{P_{ABC}(010) \leq P_{CD}(00) + P_{AD}(01)\}$ be some inequality constraining the set of non-contextual marginal models $P^{\mathcal{M}}$. Intuitively, the contextuality of a distribution P_{ABCD} should be *independent* of the how one labels the variables in $\mathcal{J}$. Suppose $\varphi \in \mathsf{Perm}(\mathcal{J})$[15] is a permutation of the variables $\mathcal{J}$ such that $\varphi[\{a, b, c, d\}] = \{c, a, d, b\}$. Intuitively φ acts on $I_{\mathcal{M}}$ in the following way,

$$
\begin{aligned}
\varphi[I_{\mathcal{M}}] &= \{P_{\varphi[abc]}(010) \leq P_{\varphi[cd]}(00) + P_{\varphi[ad]}(01)\} \\
&= \{P_{cad}(010) \leq P_{db}(00) + P_{cb}(01)\} \\
&= \{P_{acd}(100) \leq P_{bd}(00) + P_{bc}(10)\}
\end{aligned}
\tag{A14}
$$

Following this example, it is always possible to permute the random variables in marginal scenario to produce new and valid inequalities. In general, permutations over $\mathcal{J}$ can modify the marginal scenario $\varphi[\mathcal{M}] \equiv \{\varphi[V_1], \ldots, \varphi[V_k]\} \neq \mathcal{M}$; this was the case for the above example. Therefore φ has capacity to modify both the terms of the inequality as well as the set of marginal models it constrains $\varphi[I_{\mathcal{M}}] = \varphi[I]_{\varphi[\mathcal{M}]}$. Permutations φ that modify the marginal scenario

---

[14] If the extremal distribution happens to be asymmetric, then one can conclude the space of accessible distributions is non-convex.
[15] The permutation group $\mathsf{Perm}(S)$ over a set S is the set of all bijective maps $\varphi : S \to S$.

have no application within the framework of the Inflation Technique because Lemma 3 requires inflated inequality to constrain 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, $\Phi(\mathcal{M}) \equiv \{\varphi \in \mathsf{Perm}(\mathcal{J}) \mid \forall V \in \mathcal{M} : \varphi[V] \in \mathcal{M}\}$. Obtaining the variable permutation group $\Phi(\mathcal{M})$ is a simple application of a group stabilizer algorithm. After obtaining $\Phi(\mathcal{M})$, one can take known compatibility inequalities $I_{\mathcal{M}}$ can 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 very useful for reducing computational complexity [52], we divert our attention to finding *symmetric* inequalities, i.e. those where $\forall \varphi \in \Phi(\mathcal{M}) : \varphi[I_{\mathcal{M}}] = I_{\mathcal{M}}$. To generate inequalities that exhibit certain symmetries using the methods described in Appendix A 1, 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 $m \in \mathcal{S}(\mathcal{M})$ and joint outcomes $j \in \mathcal{S}(\mathcal{J})$, one can define **group orbits** of Φ in $\mathcal{S}(\mathcal{M})$ and $\mathcal{S}(\mathcal{J})$; respectively $\mathsf{Orb}_\Phi(m) \equiv \{\varphi[m] \mid \varphi \in \Phi\}$, $\mathsf{Orb}_\Phi(j) \equiv \{\varphi[j] \mid \varphi \in \Phi\}$. The action of $\varphi \in \Phi$ on any outcome $f \in \mathcal{S}(V)$ (denoted $\omega[f]$) is defined as $(\varphi[f])(v) \equiv f(\varphi^{-1}[v])$ pursuant to intuitive action used in Eq. (A14). 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.

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

It is possible to analogously define a **symmetric joint distribution vector** $\mathcal{P}_\Phi^{\mathcal{J}}$ indexed by $\mathsf{Orb}_\Phi(j)$, $\left(\mathcal{P}_\Phi^{\mathcal{J}}\right)_{\mathsf{Orb}_\Phi(j)} = \sum_{j' \in \mathsf{Orb}_\Phi(j)} \mathcal{P}_{j'}^{\mathcal{J}}$ and a **symmetric marginal distribution vector** $\mathcal{P}_\Phi^{\mathcal{M}}$ indexed by $\mathsf{Orb}_\Phi(m)$, $\left(\mathcal{P}_\Phi^{\mathcal{M}}\right)_{\mathsf{Orb}_\Phi(m)} = \sum_{m' \in \mathsf{Orb}_\Phi(m)} \mathcal{P}_{m'}^{\mathcal{M}}$. Together, $M_\Phi, \mathcal{P}_\Phi^{\mathcal{J}}$ and $\mathcal{P}_\Phi^{\mathcal{M}}$ define a new, **symmetric marginal problem**: $\mathcal{P}_\Phi^{\mathcal{M}} = M_\Phi \cdot \mathcal{P}_\Phi^{\mathcal{J}}$. The symmetric marginal problem can be solved using the same computational methods discussed in Appendix A 1 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 copy-index.

For the Triangle structure in particular, the variable permutation group is simply the set of permutations on A, B, C: $\Phi(\mathcal{M}) = \mathsf{Perm}(A, B, C)$. For the Web inflation (Fig. 9c), we have obtained $\Phi\left(\mathsf{AI}_\mathcal{G}(\mathcal{G}')\right)$ which is an order 48 group with the following 4 generators:

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

It is easy to verify that $\varphi_1, \varphi_2, \varphi_3, \varphi_4$ are all automorphisms of the web inflation. Moreover they stabilize the ai-expressible sets of Eq. (A13). Importantly,

$$\Phi\left(\mathsf{AI}_\mathcal{G}(\mathcal{G}')\right) \sim \mathsf{Perm}(A, B, C)$$

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 $\mathcal{S}(\mathcal{M})$. Likewise the number of columns is the number of distinct orbits $\mathsf{Orb}_\Phi(j)$ in $\mathcal{S}(\mathcal{J})$. For the Web inflation in particular, M_Φ is $450 \times 358,120$. Using the symmetric incidence matrix and linear programing methods, an infeasibility certificate was found that is capable of witnessing the Fritz distribution. The corresponding deflated inequality is presented in Section VII as Eq. (9).

---

[1] P. W. Shor, SIAM J. Comp. **26**, 1484 (1997).

[2] S. Jordan, "Quantum algorithm zoo," (2016).

[3] C. H. Bennett and G. Brassard, Theor. Comp. Sci. **560**, 7 (2014).

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

[5] J. S. Bell, Physics **1**, 195 (1964).

[6] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).

[7] C. J. Wood and R. W. Spekkens, New J. Phys **17**, 033002 (2012), arXiv:1208.4119.

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

[9] J. Pearl, Stat. Surv. **3**, 96 (2009).

[10] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys **86**, 419 (2013), arXiv:1303.2849.

[11] T. Fritz, New J. Phys **14**, 103001 (2012), arXiv:1206.5115.

[12] T. Fritz, Comm. Math. Phys. **341**, 391 (2014), arXiv:1404.4812.

[13] B. Steudel and N. Ay, "Information-theoretic inference of common ancestors," (2010), arXiv:1010.5720.

[14] R. Chaves, L. Luft, and D. Gross, New J. Phys. **16**, 043001 (2014).

[15] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, Phys. Rev. A **85**, 032119 (2012).

[16] J. Henson, R. Lal, and M. F. Pusey, New J. Phys **16**, 113043 (2014).

[17] E. Wolfe, R. W. Spekkens, and T. Fritz, "The inflation technique for causal inference with latent variables," (2016), arXiv:1609.00672.

[18] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).

[19] B. S. Cirel'son, Lett. Math Phys. **4**, 93 (1980).

[20] D. Rohrlich and S. Popescu, "Nonlocality as an axiom for quantum theory," (1995), arXiv:quant-ph/9508009.

[21] B. M. Terhal and D. P. DiVincenzo, Quant. Inf. Comp. **4**, 134 (2004).

[22] M. Navascués and E. Wolfe, (2017).

[23] J. I. de Vicente, (2014), 10.1088/1751-8113/47/42/424017, arXiv:1401.6941.

[24] D. Petz and L. Ruppert, J. Phys. A **45**, 085306 (2015).

[25] S. R. Hedemann, "Hyperspherical parameterization of unitary matrices," (2013), arXiv:1303.5904.

[26] K. Fujii, K. Funahashi, and T. Kobayashi, Int. J. Geom. Meth. Mod. Phys. **03**, 269 (2006).

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

[28] M. Grasmair, "On the existence of a cholesky factorization," (2014).

[29] C. Spengler, M. Huber, and B. C. Hiesmayr, J. Phys. A **43**, 385306 (2010).

[30] C. Moler and C. V. Loan, SIAM Rev. **45**, 3 (2003).

[31] K. F. Pál and T. Vértesi, Phys. Rev. A **82**, 022116 (2010).

[32] J. Nocedal and S. Wright, *Numerical Optimization (Springer Series in Operations Research and Financial Engineering)* (Springer, 2000).

[33] D. J. Wales and J. P. K. Doye, J. Phys. Chem. A **101**, 5111 (1997).

[34] A. A. Methot and V. Scarani, "An anomaly of non-locality," (2007), arXiv:quant-ph/0601210.

[35] T. Fritz and R. Chaves, IEEE Trans. Info. Theor. **59**, 803 (2011).

[36] N. N. Vorob'ev, Theor. Prob. Applic. **7**, 147 (1962).

[37] S. Abramsky and A. Brandenburger, New J. Phys **13**, 113036 (2011), arXiv:1102.0264.

[38] A. Schrijver, *Theory of Linear and Integer Programming* (Wiley, 1998).

[39] S. Lahaie, "How to take the dual of a linear program," (2008).

[40] Bradley, Hax, and Magnanti, *Applied Mathematical Programming* (Addison-Wesley, 1977) Chap. 4, pp. 143–144.

[41] E. D. Andersen, Comp. Optim. Applic. **20**, 171 (2001).

[42] MOSEK Optimization Suite, "Mosek IEEE Trans. Info. Theor." (2016).

[43] Gurobi Optimizer, "Farkasdual," (2016).

[44] IBM - CPLEX Optimization Studio, "Cplex," (2016).

[45] CVX - Matlab Software for Disciplined Convex Programming, "Dual variables," (2016).

[46] Martin Andersen, Joachim Dahl, and Lieven Vandenberghe, "Linear cone programs," (2016).

[47] S. Abramsky and L. Hardy, Phys. Rev. A **85**, 062114 (2012).

[48] S. Mansfield and T. Fritz, Found. Phys. **42**, 709 (2011).

[49] I. Pitowsky, Math. Prog. **50**, 395 (1991).

[50] G. B. Dantzig and B. C. Eaves, J. Combin. Theor. A **14**, 288 (1973).

[51] C. Jones, E. C. Kerrigan, and J. Maciejowski, *Equality set projection: A new algorithm for the projection of polytopes in halfspace representation*, Tech. Rep. (Cambridge University Engineering Dept, 2004).

[52] J.-D. Bancal, N. Gisin, and S. Pironio, J. Phys. A **43**, 385303 (2010).