Causal Compatibility Inequalities Admitting of Quantum Violations in the Triangle Scenario

Thomas C. Fraser*

Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada
University of Waterloo, Waterloo, Ontario, Canada
(Dated: January 17, 2017)

Quantum correlations are often incompatible with a classical assumption of causal structure. This non-classicality is often known as quantum non-locality, and it is witnessed through the violation of causal compatibility inequalities, such as Bell inequalities. Such inequalities were recently derived for the triangle scenario [arXiv:1609.00672], begging the question: can these inequalities be violated by quantum correlations? Here we answer this affirmatively, and discuss specific triangle scenario inequalities and quantum configurations which manifest non-classical correlations. Numerical optimizations reveal quantum resources potentially qualitatively different from those known previously.

* tcfraser@tcfraser.com

CONTENTS

I. INTRODUCTION

In recent decades, the technical utility of quantum mechanics has become abundantly clear; quantum computational algorithms such as Shor's algorithm [1] and numerous others [2] maintain advantage over their classical counterparts. Additionally, classical communications technologies stand to benefit from the nature of quantum mechanics — a popular example being Quantum Key Distribution [3]. Any time classical resources are unable to emulate quantum mechanics, there represents an opportunity to exploit the non-classical features of quantum mechanics in order to solve a variety of computational and communicational problems [4]. Albeit, not all tasks stand to benefit from quantum resources [5]. Presently, discovering situations or tasks where quantum mechanics offers an advantage, and proving that advantage is genuine, is a primal objective of theoretical quantum information theory.

From a foundational prospective, the majority of attempts to characterize precisely where quantum mechanics and classical mechanics differ have involved Bell inequalities [6]. Originally, Bell inequalities were derived in response to the famous EPR paradox [7] to rule out hidden variable theories for quantum mechanics. Studying and discovering Bell-type inequalities is useful because a complete set of Bell inequalities provides a systematic way of classifying the *classicality* of a given observation. These inequalities are best understood from the perspective of **causal inference** — a problem which aims to determine which observations can or can not 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]. Inequalities such as Bell inequalities constrain the space of observations that are compatible with a hypothesized causal structure. In the case of the original Bell inequalities, the classical causal model in question is known as the Bell scenario [10] and is depicted in fig. 1. The Bell scenario involves two parties (A, B) who each make measurements decided by random settings (S_A, S_B) on a shared hidden resource λ. Quantum non-classicality in the Bell scenario and related structures have been studied thoroughly since Bell's original work [11]. However, novel scenarios such as the correlation scenarios proposed by Fritz [12, 13] are arguably less understood compared to the Bell scenario. Here we investigate one particular causal scenario: the triangle scenario (fig. 2).

The triangle scenario (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 studied extensively in literature before (see [14, Fig. 1], [15, Fig. 6], [16, Fig. 8], [17, Fig. 8, Appendix E], [12, Fig. 3], [18, Fig. 1], ...). A healthy portion of which the existing literature will be discussed in section III. For instance Branciard *et al.* [16] noted that characterizing locality in the triangle scenario remained an *open problem* and that identifying causal compatibility inequalities for this configuration *seemed challenging.* The subject of this document is to expound on the developments made towards understanding quantum non-classicality in the triangle scenario.

As a preliminary summary, causal compatibility inequalities were derived for the triangle scenario that are able to be violated by specific quantum-accessible distributions. This was accomplished by combining two previous works. The first, an insight made by Fritz [12] regarding the ability to re-interpret the Bell scenario as a portion of the triangle scenario. The second, a new framework for solving causal inference problems developed by Wolfe *et al.* [18] called **The Inflation Technique**. Ultimately, this work serves as a strong indication that the Inflation Technique offers new insights into specifically quantum non-classicality. Moreover, these new inequalities offer an avenue for finding and identifying new forms of non-classicality unknown previously. This report discusses attempts to find such resources with partial success and outlines refinements that can be made for future exploration.

The following paper 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 discusses the triangle scenario and its existing research in detail; identifying its stark differences from the Bell scenario and motivates why the triangle scenario was chosen as a case study. Section IV defines and discusses a singular quantum correlation first discovered by Fritz [12] which we term the **Fritz distribution**. The Fritz distribution was proven to be non-classical without the use of inequalities — this report derives inequalities that do. Section V briefly summaries the Inflation Technique in the context of this work. Although the summary presented here is sufficient, the original work [18] offers a more concrete and pedagogical introduction. Section VI demonstrates how the inflation technique can be used to derive causal compatibility inequalities for the triangle scenario that are violated by quantum-accessible distributions. A sample of these inequalities are presented at the end of section VI as well as appendices C 1 and C 2. Finally, section VII mentions a method for parameterizing the space of quantum-accessible distributions which is 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.

II. CAUSAL COMPATIBILITY

The task of causal inference aims to determine whether or not a given family of observed probability distributions can be explained by a proposed causal mechanism [8]. If a set of observed distributions can be explained by the hypothesized causal mechanism, then the observations are said to be *compatible* with the causal mechanism. In order to define compatibility rigorously, we first need to formally define what is meant by "set of observations".

A. Marginal Models

Given a set of observable variables $\mathcal{J} = \{v_1, \ldots, v_k\}$ where each $v \in \mathcal{J}$ is a statistical random variable, one can conceivably make a measurement over of these variables and obtain a joint probability distribution $P_\mathcal{J} = P_{v_1, \ldots, v_n}$. More generally however, and depending on the context, it is possible to make a measurement on a portion of the system $V \subset \mathcal{J}$ in order to obtain a marginal distribution P_V. With complete generality however, one could in principle make a measurement on a collection of portions of the system $\mathcal{M} = \{V_1, \ldots, V_m \mid \forall i : V_i \subseteq \mathcal{J}\}$ obtaining a collection of marginal probability distributions $\{P_{V_1}, \ldots, P_{V_m} \mid \forall i : V_i \in \mathcal{M}\}$.

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 $P^\mathcal{M}$:

$$P^\mathcal{M} = \{P_V \mid V \in \mathcal{M}\}$$

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 [19], 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 document, we only 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 $\mathcal{J}$.

$$\mathcal{J} = \bigcup_{V \in \mathcal{M}} V$$

B. Causal Structure

A **causal structure** abstractly represents the hypothesized causal mechanism. Mathematically, a causal structure is a directed acyclic graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ of nodes $\mathcal{N}$ and edges $\mathcal{E}$ as defined in appendix A 1. The nodes of a causal structure are random variables and are classified into two disjoint classes: the **latent nodes** $\mathcal{N}_L$ representing the set of nodes that are unobservable and the **observed nodes** $\mathcal{N}_O$. The latent nodes of a causal structure represent completely hidden variables that are hidden by some fundamental process or cannot be measured due to other limitations. The observed nodes represent random variables that are measurable and in practice will be observed by the marginal model $P^\mathcal{M}$. The edges $\mathcal{E}$ of a causal structure represent casual influence between variables within the causal structure.

A number of causal structures are depicted throughout this document (see figs. 1 and 2 as well as fig. 5). As an example, the triangle scenario has 3 observable nodes and 3 latent nodes,

$$\mathcal{N}_O = \{A, B, C\} \qquad \mathcal{N}_L = \{X, Y, Z\}$$

along with 6 edges indicating that the latent nodes $\{X, Y, Z\}$ cause and influence the observed nodes $\{A, B, C\}$.

$$\mathcal{E} = \{\{X \to A\}, \{Y \to A\}, \{Y \to B\}, \{Z \to B\}, \{Z \to C\}, \{X \to C\}\}$$

C. Compatibility

Now that the notions of a causal structure $\mathcal{G}$ and a marginal model $P^\mathcal{M}$ are defined, it is now possible to precisely define the notion of compatibility between the two. Intuitively, the causal connections within $\mathcal{G}$ should restrict and

limit the set of possible observations that can be made on $\mathcal{G}$. Namely, if each variable $n \in \mathcal{N}$ is caused by its parents $\mathsf{Pa}_\mathcal{G}(n)$, then n should be independent of all other variables that are not its descendants. This is known as the **local Markov property**:

$$n \perp\!\!\!\perp \mathcal{N} \setminus \mathsf{De}_\mathcal{G}(n) \mid \mathsf{Pa}_\mathcal{G}(n) \quad \forall n \in \mathcal{N} \tag{1}$$

Where $\perp\!\!\!\perp$ denotes statistical independence. With this as motivation, we define a set of **causal parameters** for a causal structure $\mathcal{G}$ to be a specification of a conditional distribution for every variable $n \in \mathcal{N}$ given its parents:

$$\left\{ P_{n|\mathsf{Pa}_\mathcal{G}(n)} \mid n \in \mathcal{N} \right\} \tag{2}$$

It is important to note that a set of causal parameters is not unique; the numerical nature of set of causal parameters is a choice. Altogether, a marginal model $P^\mathcal{M}$ is **compatible** with a causal structure $\mathcal{G}$ if there exists a choice of causal parameters such that each context $P_V \in P^\mathcal{M}$ can be *recovered* from the following series of operations:

1. Obtain a joint probability distribution over *all* variables of the causal structure using the chosen casual parameters:[1]

$$P_\mathcal{N} = \prod_{n \in \mathcal{N}} P_{n|\mathsf{Pa}_\mathcal{G}(n)} \tag{3}$$

   This step ensures that the local Markov property (eq. (2)) holds for all variables $n \in \mathcal{N}$.

2. Marginalize over the latent variables of $\mathcal{G}$ to obtain a joint distribution over the observed variables $\mathcal{N}_O$:

$$P_{\mathcal{N}_O} = \sum_{\mathcal{N}_L} P_\mathcal{N} \tag{4}$$

3. Finally marginalize over the observed nodes not in V to obtain P_V:

$$P_V = \sum_{\mathcal{N}_O \setminus V} P_{\mathcal{N}_O} \tag{5}$$

It is important to note that compatibility between $P^\mathcal{M}$ and $\mathcal{G}$ requires that it is possible to recover *each* context $P_V \in P^\mathcal{M}$ from the *same* set of causal parameters. If it is not possible for $P^\mathcal{M}$ to be recovered for *any* choice of causal parameters, then $P^\mathcal{M}$ is **incompatible**.

D. Inequalities

As defined in section II C, determining compatibility is equivalent to determining if a set of causal parameters exists with the above-mentioned properties. In practice however, checking causal parameters systemically is intractable, especially when the cardinality (number of outcomes) of the latent variables is unspecified. Instead, **causal compatibility inequalities**[2] are probabilistic inequalities defined over the marginal contexts $P_V \in P^\mathcal{M}$ within the marginal model. If a marginal model $P^\mathcal{M}$ happens to *violate* any causal compatibility inequality, then that marginal model is definitely *incompatible*. In section VI, a brief discussion of existing techniques for deriving these inequalities will take place.

Unfortunately, a singular inequality can not prove that a given marginal model 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 scenario.

From the perspective of identifying quantum non-classicality, the causal structure $\mathcal{G}$ acts as the classical hypothesis for a particular marginal model $P^\mathcal{M}$. Therefore, non-classicality becomes synonymous with incompatibility: if a marginal model $P^\mathcal{M}$ is incompatible, then it is non-classical. Henceforth, we will use these two terms interchangeably. From a resource standpoint, if the non-classicality of $P^\mathcal{M}$ can be witnessed by an inequality I, but $P^\mathcal{M}$ can be implemented using quantum states and measurements, then the causal compatibility inequality I represents a concrete task where quantum resources out-perform classical resources.

---

[1] Here $\prod$ denotes the product distribution $\prod_{n \in \mathcal{N}} P_{n|\mathsf{Pa}_\mathcal{G}(n)} = P_{n_1|\mathsf{Pa}_\mathcal{G}(n_1)} \times \cdots \times P_{n_k|\mathsf{Pa}_\mathcal{G}(n_k)}$.

[2] We refer to these inequalities as "causal compatibility inequalities" instead of simply Bell inequalities for two reasons. First, "Bell inequalities" usually are associated specifically with the Bell scenario/Bell causal structure. Second, the inequalities derived in this work are fundamentally distinct from a typical Bell inequality in that these inequalities are *polynomial* over $P^\mathcal{M}$ instead of *linear*.
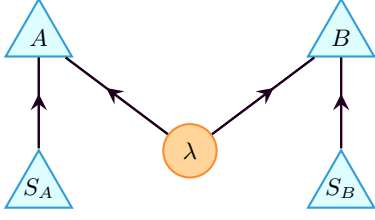
FIG. 1: The Bell scenario consisting of two observers A, B together with measurement settings S_A and S_B respectively. The shared hidden variable is labeled λ.
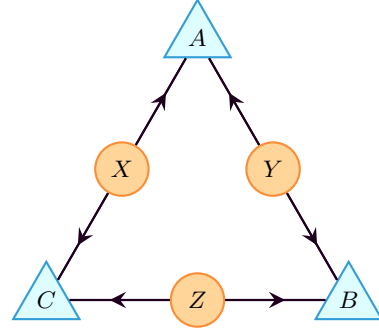

FIG. 2: The casual structure of the triangle scenario. Three variables A, B, C are observable, while X, Y, Z are latent variables.

III. TRIANGLE SCENARIO

Discovering quantum non-classicality in any causal structure is desirable, so why, in particular, is it important to do so in the triangle scenario? As was mentioned in sections I and II, the **triangle scenario** (fig. 2) is a causal structure $\mathcal{G}$ consisting of 3 observable variables A, B, C arranged in a triangular configuration while pair-wise 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 scenario if and only if there exists a choice of causal parameters $\left\{P_{A|X,Y}, P_{B|Y,Z}, P_{C|Z,X}, P_X, P_Y, P_Z\right\}$ such that the joint distribution $P_{\mathcal{N}} = P_{ABCXYZ}$ is a product over the causal parameters,

$$P_{ABCXYZ} = P_{A|X,Y} P_{B|Y,Z} P_{C|Z,X} P_X P_Y P_Z$$

And also, the given distribution P_{ABC} is a marginalization of P_{ABCXYZ} over X, Y, Z:

$$P_{ABC} = \sum_{X,Y,Z} P_{A|X,Y} P_{B|Y,Z} P_{C|Z,X} P_X P_Y P_Z \tag{6}$$

For certain causal structures, the conditional independence relations provided by d-separation relations are a sufficient characterization of compatibility. Recently, Henson *et al.* [17], classified the triangle scenario as an *interesting* causal structure in the sense that conditional independence relations *do not* provide a sufficient characterization of non-classicality.[3] Moreover, Wolfe *et al.* [18] concretely demonstrate that the space classical distributions on the triangle scenario is non-convex, unlike the Bell scenario. It can be argued that convexity of the Bell scenario distributions is partially responsible for the wealth of knowledge that is known about the Bell scenario. It is also reasonable to speculate that quantum non-classicality in the triangle can be translated directly into a computational advantage for certain computational circuits [20]. Characterizing quantum non-classicality for interesting causal structures remains an unsolved problem, and the triangle scenario is a small, non-trivial case study.

Additionally, Fritz [12] demonstrated that the triangle scenario is the *smallest* correlation scenario[4] in which their exists incompatible, quantum distributions by explicitly mentioning one which we have elected to call the Fritz distribution (see section IV). Fritz's proof of non-classicality does not involve compatibility inequalities, but instead mentions that inequalities would be helpful for characterizing non-classical in the triangle scenario.

Finally Wolfe *et al.* [18] proposed the Inflation Technique (summarized in section V), a novel tool for solving causal inference problems. Using the inflation technique, they were able to prove incompatibility between the triangle scenario and the W-type distribution; a quantum non-accessible distribution whose incompatibility is not witness-able by *any* entropic inequalities or any other known constraints. Therefore as a an algorithmic framework, the Inflation

---

[3] Note that the triangle scenario possesses *zero* conditional independence relations over the observable nodes A, B, C.

[4] Correlation scenarios differ from the Bell scenario in that measure setting variables are non-existence in correlation scenarios.
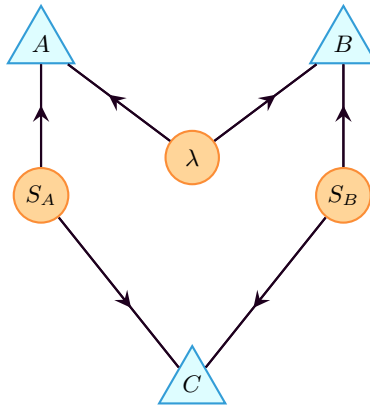
FIG. 3: The triangle scenario re-imagined to mimic the Bell scenario. The measurement settings S_A, S_B are latent nodes unlike the Bell scenario (fig. 1).

Technique comes closer to characterizing non-classicality in the triangle scenario that all others.

It is worth noting that a number of causal compatibility inequalities for the triangle scenario have been derived in previous works. For example, Steudel and Ay [14] 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 scenario is a special case with $c = 2$, $n = 3$. Moreover, Henson *et al.* [17] derive a family of entropic inequalities constraining the set of compatible correlations. Most impressively, Wolfe *et al.* [18] derive a family of polynomial, 2-outcome, causal compatibility inequalities; 37 non-trivial representatives are printed in [18].

As a preliminary search for quantum incompatibility in the triangle scenario, we performed numerical optimizations using bipartite qubit density matrices and 2-outcome POVM measurements against the compatibility inequalities printed in [18] as well as the entropic inequalities presented by Henson *et al.* [17]. Unfortunately, none of these inequalities could be violated. These early results suggest, although do not prove, that quantum non-classical does not exist in the triangle scenario for *two-outcome* measurements. We make *no claim* that this is a universal truth and simply take it as motivation to explore incompatibility in triangle scenario for a degree of outcomes.

In conclusion, the triangle scenario is a causal structure whose non-classicality remains an open problem. The triangle scenario is a desirable case study because there is a quantum distribution that is known to be incompatible, but no previously known inequalities were able to witness its incompatibility. This failure represents a gap in our understanding of quantum non-classicality, where recent developments (namely the Inflation Technique) have made partial progress. This document expands on our understanding of non-classicality in the triangle scenario and also our understanding on how to find inequalities using the Inflation Technique for interesting causal structures in general.

IV.   FRITZ DISTRIBUTION

As was originally noticed by Fritz [12], it is possible to construct quantum distributions incompatible with the triangle scenario by utilizing quantum distributions incompatible with the familiar Bell scenario. To explain, imagine rearranging the triangle into the configuration depicted in fig. 3 and contrast it with the Bell scenario of fig. 1. Evidently, under the correct relabeling, large portions of the Triangle Scenario resemble the Bell Scenario. The crucial distinction is that the measurement settings S_A, S_B are *observable variables* in the Bell scenario but become *latent variables* in the triangle scenario. In order to embed distributions that are incompatible with the Bell scenario into the triangle scenario, 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 scenario is also incompatible with the triangle scenario provided that C is *perfectly correlated* with S_A, S_B. 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 known to be incompatible with the
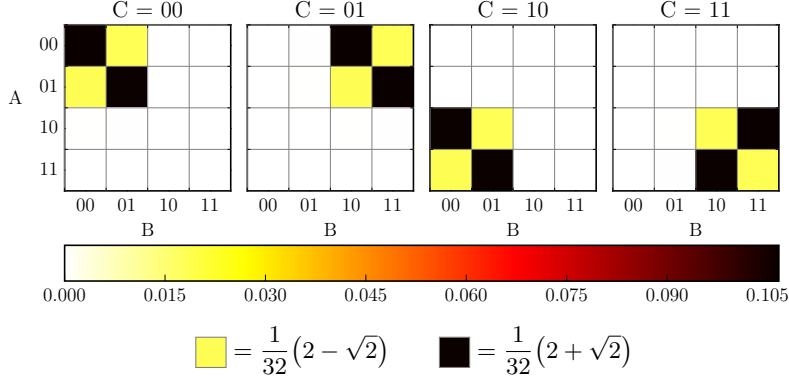
FIG. 4: The Fritz distribution visualized using a $4 \times 4 \times 4$ grid. The 4 outcomes of A, B, C are written in binary as a doublet of bits to illustrate that certain bits act as measurement pseudo-settings.

triangle scenario [12]. In the Fritz distribution, each of the variables A, B, C have 4 possible outcomes. Explicitly, P_F can be written as:

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

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

(7)

Here the notation $P_F(abc) = P_{ABC}(abc) = P(A = a, B = b, C = c)$ is used as shorthand. 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 between the right-hand bits of A and B as the probability that A_r and B_r are equal less the probability that they differ:

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

Provided that C is perfectly correlated with A_l and B_l, any Bell inequality for the Bell scenario defined over A, B, S_A, S_B can be directly converted to an inequality for the triangle scenario using the following mapping:

$$A, B, S_A, S_B \mapsto A_r, B_r, A_l, B_l$$

As an example, the famous CHSH inequality [21],

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

Can be written for the right-bits of A and B,

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

(8)

Therefore, every compatible distribution P_{ABC} where C is perfectly correlated A_l, B_l must satisfy eq. (8). Substituting the Fritz distribution (eq. (7)) into eq. (8) yields maximal violation [22],

$$3\left(\frac{1}{\sqrt{2}}\right) - \left(-\frac{1}{\sqrt{2}}\right) = 2\sqrt{2} \nleq 2$$

Therefore, the Fritz distribution is incompatible with the triangle scenario as promised.

The Fritz distribution P_F is of special interest because it can be realized as a set of quantum states and measurements on the quantum causal model depicted in fig. 9. The shared resource between A and B (graphically denoted λ in fig. 3) is the maximally entangled Bell state:

$$\left|\Phi^+\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \qquad \rho_{AB} = \left|\Phi^+\right\rangle\left\langle\Phi^+\right|$$

While the states shared with C are the following mixed states:

$$\rho_{BC} = \rho_{CA} = \frac{|00\rangle\langle00| + |11\rangle\langle11|}{2}$$

The measurements employed by A, B and C are separable, projective measurements,

$$M_A = \{|0\psi_1\rangle\langle0\psi_1|, |0\psi_5\rangle\langle0\psi_5|, |1\psi_3\rangle\langle1\psi_3|, |1\psi_7\rangle\langle1\psi_7|\}$$
$$M_B = \{|\psi_60\rangle\langle\psi_60|, |\psi_20\rangle\langle\psi_20|, |\psi_01\rangle\langle\psi_01|, |\psi_41\rangle\langle\psi_41|\}$$
$$M_C = \{|00\rangle\langle00|, |10\rangle\langle10|, |01\rangle\langle01|, |11\rangle\langle11|\}$$

Where $|\psi_n\rangle$ is shorthand for,

$$|\psi_n\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{in/4}|1\rangle\right)$$

Using the above states and measurements, the Fritz distribution P_F can be recovered using the following:

$$P_{ABC} = \mathrm{Tr}[\Pi^\intercal \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA} \Pi M_A \otimes M_B \otimes M_C]$$

Where Π is a 64×64 permutation matrix to be elaborated in section VII C 4.

Before continuing it is worth noting that eq. (7) is non-unique. Any distribution that is equal to eq. (7) via a permutation of outcomes or exchange of parties is should also be referred to as a Fritz distribution. Moreover, the quantum realization presented here is not unique either. Nonetheless for concreteness, eq. (7) is taken as *the* Fritz distribution throughout this paper. Moreover, the quantum implementation of the Fritz distribution presented above is non-unique also. For example, it is possible to implement the Fritz distribution using maximally entangled Bell states for each shared resource $\rho_{AB}, \rho_{BC}, \rho_{CA}$ provided there is an appropriate adjustment to the measurements M_A, M_B, M_C. Ultimately however, only ρ_{AB} is required to be entangled in order to reproduce the Fritz; after-all, C is simply classically announcing A_r and B_r as he receives them.

Additionally, it is important to understand the domain in which Fritz's proof of incompatibility is valid. Specifically, eq. (8) can only be applied to the triangle scenario where there exists perfect correlation between C's outcomes and the measurement pseudo-settings (left-bits) of A and B. The details of this restriction are discussed and proven in Fritz's original work [12]. As such, 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. (7) with more and more uniform noise, at what point does the resultant distribution transition from incompatibility to compatibility? This problem is discussed more in section VII A.

Upon reflection, the Fritz distribution can rightly be considered slightly manufactured. The phenomenology associated with Bell non-locality or Bell incompatibility are well understood; examining these distributions under a triangle scenario embedding offers no additional perspective onto the types of 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 scenario. Fritz [12] presented the following problem: *Find an example of non-classical quantum correlations in the triangle scenario together with a proof of its non-classicality which does not hinge on Bell's Theorem.*[5] The problem is understandably stated ambiguously as it is not yet entirely clear how to distinguish between non-classical distributions on differing causal structures. In order to remove some of the ambiguity, consider the following refinements of the problem:

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

———

[5] In the way Fritz defines, *non-classical* correlations are the class of *incompatible* correlations.

R.2 Find incompatible quantum distributions on the triangle scenario that satisfy all Bell scenario inequalities under Fritz-type embeddings.

R.3 Find incompatible quantum distributions that require entanglement in more than one shared state.

The remainder of this document focuses on a positive solution to **R.1** but leaves **R.2** and **R.3** partially unanswered.

V. INFLATION TECHNIQUE

The **Inflation Technique**, invented by Wolfe *et al.* [18] 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 marginal model $P^{\mathcal{M}}$ 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, inflations of the triangle scenario are depicted in fig. 5. 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 *Web inflation* of fig. 5d. 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. Equipped with the common graph-theoretic terminology and notation of theorem 15, an inflation can be formally defined as follows:

Definition 1. An **inflation** of a causal structure $\mathcal{G}$ is another causal structure $\mathcal{G}'$ such that:

$$\forall n' \in \mathcal{N}' : \mathsf{AnSub}_{\mathcal{G}'}(n') \sim \mathsf{AnSub}_{\mathcal{G}}(n) \tag{9}$$

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 joint distribution $P_{\mathcal{N}}$ on $\mathcal{G}$ can be used to *create* compatible joint distributions on $\mathcal{G}'$.

To do this, first notice that every compatible joint distribution $P_{\mathcal{N}}$ uniquely defines a set of causal parameters for $\mathcal{G}$. It is possible to obtain these causal parameters by computing the single-variable marginal distributions of $P_{\mathcal{N}}$,

$$\forall n \in \mathcal{N} : P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} = \sum_{\mathcal{N} \setminus n} P_{\mathcal{N}} \tag{10}$$

Now it is possible to make use of the inflation criteria (eq. (9)) which at least enforces that $\mathsf{Pa}_{\mathcal{G}'}(n') \sim \mathsf{Pa}_{\mathcal{G}}(n)$. From this, one can create a set of **inflated causal parameters** using eq. (10),

$$\forall n' \in \mathcal{N}' : P_{n'|\mathsf{Pa}_{\mathcal{G}'}(n')} \equiv P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \tag{11}$$

Which in turn uniquely defines a compatible joint distribution $P_{\mathcal{N}'}$ on $\mathcal{G}'$[6].

$$P_{\mathcal{N}'} = \prod_{n' \in \mathcal{N}'} P_{n'|\mathsf{Pa}_{\mathcal{G}'}(n')} \tag{12}$$

Together, eqs. (10) to (12) form the **weak inflation lemma**: compatible joint distributions over $\mathcal{N}$ induce compatible joint distributions over $\mathcal{N}'$. Before generalizing to the inflation lemma, 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 $\mathsf{AnSub}_{\mathcal{G}}(N) \sim$

---

[6] 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.

$\mathsf{AnSub}_{\mathcal{G}'}(N')$ 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 allow one to extend the weak inflation lemma to *the* inflation lemma:

Lemma 2. ***The Inflation Lemma*** *[18, 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 2 can be visualized as follows:

$$\underbrace{\left\{P_N \mid N \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\right\}}_{\text{compatible with } \mathcal{G}} \longrightarrow \left\{P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N}\right\}$$

$$\Big\downarrow \text{define}$$

$$\underbrace{\left\{P_{N'} \mid N' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\right\}}_{\text{compatible with } \mathcal{G}'} \longleftarrow \left\{P_{n'|\mathsf{Pa}_{\mathcal{G}'}(n')} \mid n' \in \mathcal{N}'\right\}$$

The inflation lemma is the most important result of the causal inflation technique. The contrapositive version of lemma 2 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 2 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 $\mathcal{G}$.

Suppose that one were to obtain an inequality on $\mathcal{G}'$ that contained probabilistic terms such as $P_{N'}$ where $N' \notin \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$. At this point, the Inflation lemma is not 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 $\mathcal{G}'$,

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

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.

$$P_{N'} \xrightarrow[\substack{N_1' \perp N_2'}]{N_1' \cup N_2' = N} P_{N_1'}P_{N_2'}$$

$$\begin{array}{ccc} \Big| & & \Big| \\ N' \notin \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}') & & N_1', N_2' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}') \\ \textcolor{red}{\times} & & \downarrow \\ ?? & & P_{N_1}P_{N_2} \end{array}$$

Not all sets $N' \subseteq \mathcal{N}'$ can be written as the disjoint union of injectable sets, but those that can are called the **pre-injectable sets**. A set $N' \subseteq \mathcal{N}'$ is a pre-injectable set if it can be decomposed into the disjoint union of injectable
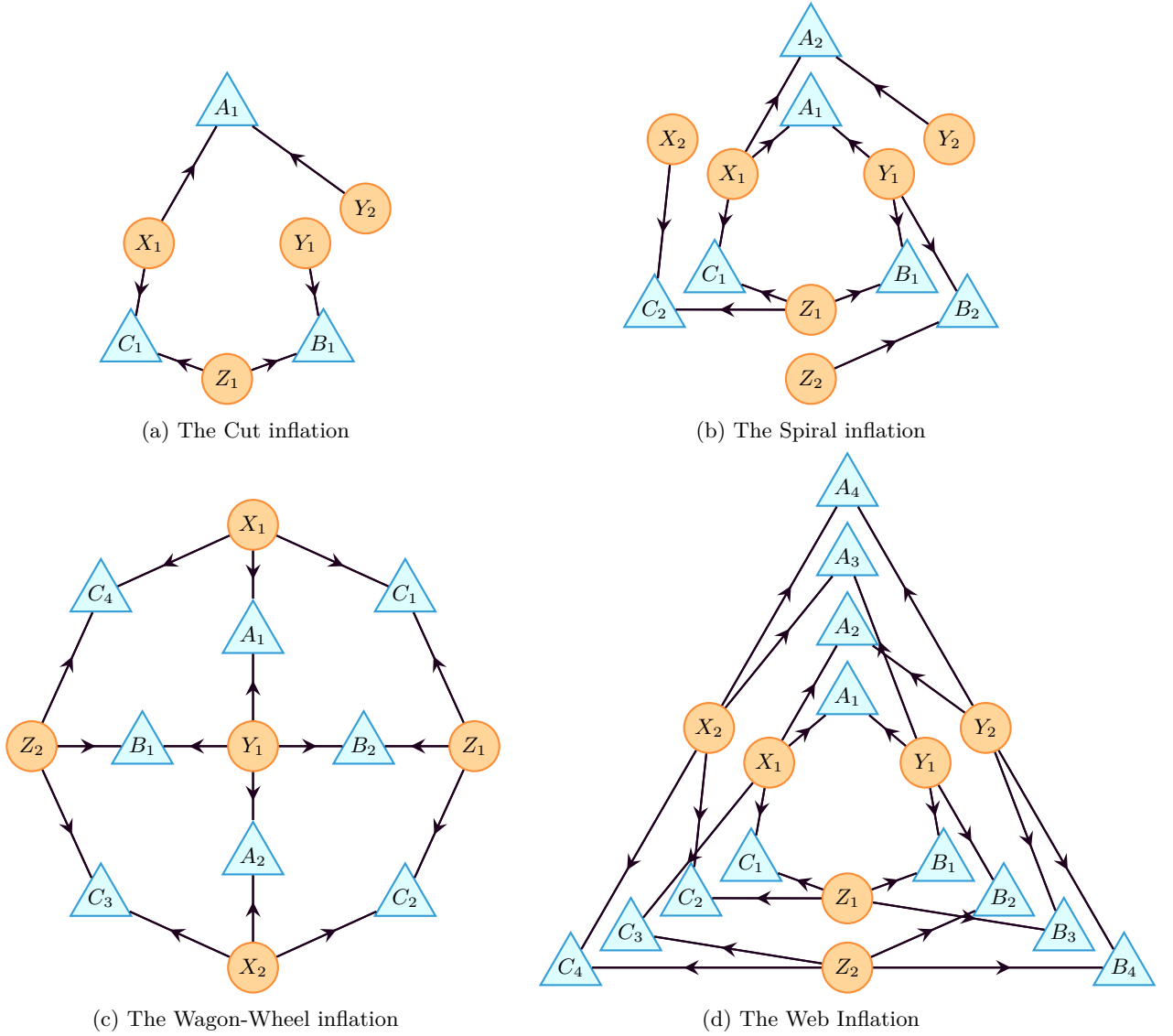
(a) The Cut inflation

(b) The Spiral inflation

(c) The Wagon-Wheel inflation

(d) The Web Inflation

FIG. 5: Some inflations of the triangle scenario.

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 pre-injectable set N' is of primal importance to the inflation technique because any distribution over N' will factorize according to graphical d-separation conditions [8],

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

Application of the inflation lemma to linear inequalities constraining pre-injectable 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 $\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')$ denote the set of all pre-injectable sets. Naturally, the pre-injectable sets will define a marginal scenario $\mathcal{M}$ that in section VI will be used to derive compatibility inequalities[7].

---

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

VI.   DERIVING INEQUALITIES

In section II, we defined the notion of a causal compatibility inequality without providing a means for deriving/computing them. This section aims to delineate algorithms and methods that can be used to derive causal compatibility inequalities for generic causal inference problems. Recalling the definition of causal compatibility presented in section II C, a marginal model $P^{\mathcal{M}}$ is compatible with a causal structure $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ if it can be recovered from a series of operations. First choose a set of causal parameters $\left\{ P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N} \right\}$ and then construct a joint distribution $P_{\mathcal{N}} = \prod_{n \in \mathcal{N}} P_{n|\mathsf{Pa}_{\mathcal{G}}(n)}$ over all nodes $\mathcal{N}$ (this step is eq. (3)). Afterwards, each marginal distribution $P_V \in \mathcal{M}$ should be recoverable from a marginalization over $P_{\mathcal{N}}$ (this step is the combination of eqs. (4) and (5)). Intuitively, these two requirements for a compatible marginal model $P^{\mathcal{M}}$ provide two necessary components of compatibility:

1. The marginal model $P^{\mathcal{M}}$ must admit a joint distribution $P_{\mathcal{N}}$ (eqs. (4) and (5)).

2. The joint distribution $P_{\mathcal{N}}$ must satisfy the Markov separability criteria (eq. (3)).

In practice, this is the natural order in which to classify the compatibility of a marginal model $P^{\mathcal{M}}$. First check if the marginal model admits a joint distribution and if and only if it does, determine whether or not the local Markov property can be satisfied for some choice of causal parameters. Intuitively, there are two reasons why a given marginal model $P^{\mathcal{M}}$ would be deemed incompatible. One possibility is that the marginal model $P^{\mathcal{M}}$ does not admit a joint distribution $P_{\mathcal{N}}$. Alternatively, it could be possible that a joint distribution $P_{\mathcal{N}}$ exists but can not factor into causal parameters as per eq. (3).

Colloquially, the first necessary condition, namely the existence (or non-existence) of a joint distribution, is referred to as **the marginal problem**. The marginal problem asks: given a marginal model $P^{\mathcal{M}} = \{P_V \mid V \in \mathcal{M}\}$ marginal to the joint variables $\mathcal{J} = \bigcup_{V \in \mathcal{M}} V$, does there exist a joint distribution $P_{\mathcal{J}}$ such that each context P_V can be obtained by marginalizing $P_{\mathcal{J}}$?

$$\forall V \in \mathcal{M} : P_V = \sum_{\mathcal{J} \setminus V} P_{\mathcal{J}} \tag{13}$$

Note that we have changed notation from the nodes of a causal structure $\mathcal{N}$ to a joint set of variables $\mathcal{J}$. This is done simply because the marginal problem is well-defined independent of a given causal structure. In fact, if a marginal model $P^{\mathcal{M}}$ does not admit a joint distribution, then it cannot be compatible with *any* causal structure whatsoever. A marginal model is said to be *contextual* if it *does not* admit a joint distribution and *non-contextual* otherwise. The marginal problem has applications to numerous areas of mathematics including game theory [23], database theory, knowledge integration of expert system, and of course, quantum information theory [19].

The marginal problem is distinct from the second necessary condition, which we term the **Markov separability problem** in numerous ways. First of all, the marginal problem as applied to causal inference tasks is only concerned with the observable variables $\mathcal{N}_O$ of $\mathcal{G}$. This differs from the Markov separability problem which depends directly on the entire graphical structure of $\mathcal{G}$ and thus must include the latent variables $\mathcal{N}_L$. The purpose of latent variables is to have unspecified behavior which makes checking the local Markov property exceedingly difficult in comparison to the marginal problem. Secondly, eq. (13) is inherently a linear system of constraints which be solved rather efficiently using linear programs (to be discussed in section VI A). The local Markov property concerns itself with probabilistic independence constraints which are polynomial. Fortunately, the Inflation Technique (summarized in section V) is able to cast (at least partially) the Markov separability problem as a marginal problem over an inflated causal structure. As such the Inflation Technique in many ways can be considered a relaxation technique which turns polynomial constraints into linear constraints.

In consideration of these observations, we will now endeavor to discuss how causal compatibility inequalities can be derived by casting the marginal problem as a linear program. After doing, it becomes possible to discuss existing methods for deriving constraints on the set of contextual marginal models.

A.   The Marginal Problem as a Linear Program

Heretofore, our discussions of compatibility and the marginal problem have been general in the sense of making no reference to the nature the random variables $v \in \mathcal{J}$. In particular, whether the outcomes of v are discrete or

continuous, finite or unbounded, or otherwise. In the case of infinite outcomes, the marginal problem can be used to derive entropic inequalities defined over the mutual information between variables $v \in \mathcal{J}$ [19]. Entropic inequalities however have been shown to be less constraining than inequalities defined over variables with finite cardinality [18]. Throughout this document, we restrict our attention to finite outcomes. To assist with the subsequent analysis, we recall some familiar notions of probability theory.

1. Events & Outcomes

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 $\mathcal{S}(v)$[8] 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.

$$\mathcal{S}(v) \equiv \{s : \{v\} \to O_v\}$$

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 [24] demands the distinction. Specifically for this work, the distinction becomes essential for our discussion and exploit of marginal symmetries in section VI C. As a natural generalization we define the event over a collection of random variables $V = \{v_1, \ldots, v_n\}$ in a parallel manner,

$$\mathcal{S}(V) \equiv \{s : V \to O_V \mid \forall i : s(v_i) \in O_{v_i}\} \tag{14}$$

Each event s can be compactly represented as a set of mappings over each element of V.

$$s = \{v_1 \mapsto s(v_1), \ldots, v_k \mapsto s(v_k)\} = \{v_i \mapsto s(v_i)\}_{i=1}^k$$

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 $[0, 1]$.

Example 3. Let $s \in \mathcal{S}(V)$ be an event over V, a probability distribution P_V over V will assign a non-negative probability $P_V(s)$ to s. Let $V = \{v_1, v_2\}$ with each $v \in V$ having 3 distinct outcomes $O_v = \{0, 1, 2\}$. Furthermore let $f \in \mathcal{S}(V)$ be the event $(v_1 \mapsto 2, v_2 \mapsto 1)$. The following notations are interchangeable,

$$P(s) = P_{\mathcal{D}(s)}(s) = P_V(s) = P_{v_1, v_2}(s) = P(v_1 \mapsto s_1(v_1), v_2 \mapsto s_2(v_2)) = P(v_1 = 2, v_2 = 1) = P_{v_1 v_2}(21)$$

Each of which should be read as "the probability that v_1 got outcome 2 and v_2 got outcome 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 $W \subseteq V$.

Definition 4. 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 of all context events and by an abuse of notation we will denote this disjoint union as $\mathcal{S}(\mathcal{M})$.

$$\mathcal{S}(\mathcal{M}) = \coprod_{V \in \mathcal{M}} \mathcal{S}(V)$$

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 $j \in \mathcal{S}(\mathcal{J})$.

---

[8] In the language of sheaf-theory, $\mathcal{S}(v)$ is the *sheaf of events* [24].

2. Linear Programs

The marginalization operation (eq. (13)) 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 casted as a $|\mathcal{S}(\mathcal{M})| \times |\mathcal{S}(\mathcal{J})|$ matrix.

Definition 5. 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:

Example 6. Let $\mathcal{J}$ be 3 binary variables $\mathcal{J} = \{A, B, C\}$ and $\mathcal{M}$ be the marginal scenario $\mathcal{M} = \{\{A, B\}, \{B, C\}, \{A, C\}\}$. The incidence matrix becomes:

$$M = \begin{array}{c} \\ (A,B,C) \mapsto \\ (A\mapsto0,B\mapsto0) \\ (A\mapsto0,B\mapsto1) \\ (A\mapsto1,B\mapsto0) \\ (A\mapsto1,B\mapsto1) \\ (B\mapsto0,C\mapsto0) \\ (B\mapsto0,C\mapsto1) \\ (B\mapsto1,C\mapsto0) \\ (B\mapsto1,C\mapsto1) \\ (A\mapsto0,C\mapsto0) \\ (A\mapsto0,C\mapsto1) \\ (A\mapsto1,C\mapsto0) \\ (A\mapsto1,C\mapsto1) \end{array} \begin{array}{cccccccc} (0,0,0) & (0,0,1) & (0,1,0) & (0,1,1) & (1,0,0) & (1,0,1) & (1,1,0) & (1,1,1) \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{array} \tag{15}$$

The incidence matrix needs to act (to the right) on a vector representing the joint probability $P_{\mathcal{J}}$ distribution and output a vector representing the marginal model $P^{\mathcal{M}}$.

Definition 7. The **joint distribution vector** $\mathcal{P}^{\mathcal{J}}$ for a probability distribution $P_{\mathcal{J}}$ is the vector whose entries are the positive, real-valued probabilities that $P_{\mathcal{J}}$ assigns to each joint event $j \in \mathcal{S}(\mathcal{J})$. $\mathcal{P}^{\mathcal{J}}$ shares the same indices as the *column* indices of M.

$$\forall j \in \mathcal{S}(\mathcal{J}) : \mathcal{P}_j^{\mathcal{J}} = P_{\mathcal{J}}(j)$$

Definition 8. Analogously, **marginal distribution vector** $\mathcal{P}^{\mathcal{M}}$ for a marginal model $P^{\mathcal{M}} = \{P_{V_1}, \ldots, P_{V_k}\}$ is the vector whose entries are probabilities over the set of marginal outcomes $\mathcal{S}(\mathcal{M})$. $\mathcal{P}^{\mathcal{M}}$ shares the same indices as the *row* indices of M.

$$\forall m \in \mathcal{S}(\mathcal{M}) : \mathcal{P}_m^{\mathcal{M}} = P_{\mathcal{D}(m)}(m)$$

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 $\mathcal{P}^{\mathcal{J}}$.

$$\mathcal{P}^{\mathcal{M}} = M \cdot \mathcal{P}^{\mathcal{J}} \tag{16}$$

Because the marginal distribution vector $\mathcal{P}^{\mathcal{M}}$ is simply another representation of the marginal model $P^{\mathcal{M}}$, we will henceforth use the language interchangeably. This is likewise done so for the joint distribution (vector) $P_{\mathcal{J}}$ ($\mathcal{P}^{\mathcal{J}}$). 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. (16) holds?*

The marginal problem can sometimes take on a more general variant that does not begin with a specific marginal model [19, 25, 26]: *Given a marginal scenario $\mathcal{M}$, what is the set of all non-contextual marginal models?* Pitowsky [27] 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 correspond to each of the $|\mathcal{S}(\mathcal{J})|$ *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. (16). 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 [11]. From the perspective of a marginal polytope, convex hull algorithms can be used to compute a representation of the complete set of linear inequalities and completely solve the marginal problem. Alternatively, one can completely characterize the separation between the space of contextual and non-contextual distributions by the system of inequalities associated with eq. (16). Explicitly,

$$
\begin{aligned}
\forall m \in \mathcal{S}(\mathcal{M}): \quad & \mathcal{P}_m^{\mathcal{M}} - \sum_j M_{m,j} \mathcal{P}_j^{\mathcal{J}} \geq 0 \\
\forall m \in \mathcal{S}(\mathcal{M}): \quad & -\mathcal{P}_m^{\mathcal{M}} + \sum_j M_{m,j} \mathcal{P}_j^{\mathcal{J}} \geq 0 \\
\forall j \in \mathcal{S}(\mathcal{J}): \quad & \mathcal{P}_j^{\mathcal{J}} \geq 0
\end{aligned}
\tag{17}
$$

Geometrically eq. (17) is a family of $2|\mathcal{S}(\mathcal{M})| + |\mathcal{S}(\mathcal{J})|$ half-planes bounding $\mathbb{R}^{|\mathcal{S}(\mathcal{M})|} \times \mathbb{R}^{|\mathcal{S}(\mathcal{J})|}$. However testing the contextuality of a marginal model $\mathcal{P}^{\mathcal{M}}$ requires a collection of inequalities bounding just $\mathbb{R}^{|\mathcal{S}(\mathcal{M})|}$. Transforming eq. (17) into the desired system of inequalities is the task of linear-quantifier elimination. A popular tool for linear quantifier elimination is **Fourier-Motzkin elimination** [18, 25, 28, 29]. Applying the Fourier-Motzkin procedure to completely solve the marginal problem is discussed in more detail in Fritz and Chaves [19]. An excellent survey of existing techniques for solving the marginal problem including Equality Set Projection [29] and Hardy-type hypergraph transversals can be found in Wolfe *et al.* [18]. 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 act worst case doubly exponential in the number of initial inequalities [28]. For the purposes of this research, solving the marginal problem without reference to a marginal model was intractable. This will become apparent in section VI B when the inflation technique is applied to the triangle scenario, producing marginal models that are very large. Moreover, algorithms to partially solve the marginal problem such as Hardy-type hypergraph transversals were used to determine hundreds of thousands of causal compatibility inequalities for the triangle scenario; none of which were able to be violated quantumly using numerical optimizations. Therefore, simpler approaches to the marginal problem needed to be considered.

Luckily, the Fritz distribution (section IV) allows one to avoid the complexity issues of the complete marginal problem and instead focus on the original problem of determining where or not a particular marginal model admitted a joint distribution or not. This question is naturally framed as the following linear problem:

Definition 9. The Marginal Linear Program is the following 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"[9] is *feasible*, then there exists a vector x than can satisfy eq. (16) and is a valid joint distribution vector. Therefore feasibility implies that $\mathcal{P}^{\mathcal{J}}$ exists and is precisely x. Moreover if the marginal linear program is *infeasible*, then there *does not* exist a joint distribution $\mathcal{P}^{\mathcal{J}}$.

---

[9] "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*.

To every linear program, there exists a dual linear program that characterizes the feasibility of the original [30]. Constructing the dual linear program is straightforward [31].

Definition 10. The Dual Marginal Linear Program:

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

Where y is a real valued vector with the same length as $\mathcal{P}^{\mathcal{M}}$.

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 the following must hold:

$$y \cdot \mathcal{P}^{\mathcal{M}} = y \cdot M \cdot x \geq 0 \tag{18}$$

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 eq. (18) is violated and therefore $\mathcal{P}^{\mathcal{M}}$ is contextual. Likewise if $d \geq 0$ (satisfying eq. (18)), then $\mathcal{P}^{\mathcal{M}}$ is non-contextual.[10] This is manifestation of Farkas's lemma [30].

An **infeasibility certificate** [33] is any vector y that satisfies the constraint of theorem 10 and also permits violation of eq. (18) for some marginal distribution vector $\mathcal{P}^{\mathcal{M}}$.

$$y \in \mathbb{R}^{|\mathcal{S}(\mathcal{M})|} : y \cdot M \succeq 0, \quad y \cdot \mathcal{P}^{\mathcal{M}} < 0$$

Most linear program software packages such as Mosek [34], Gurobi [35], CPLEX [36] and CVX/CVX OPT [37, 38] 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 all non-contextual marginal models $\mathcal{P}^{\mathcal{M}}$ must satisfy eq. (18) making $y \cdot \mathcal{P}^{\mathcal{M}} \geq 0$ a valid causal compatibility inequality.

B. Triangle Scenario Inequalities

Equipped with the Fritz distribution (section IV) and the inflation technique (section V), it is now possible to derive causal compatibility inequalities for the triangle that are violated by the Fritz distribution; the procedure for doing so is as follows. First select an inflation $\mathcal{G}'$ of the triangle scenario $\mathcal{G}$ and identify the maximal pre-injectable sets $\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')$. These sets define the marginal scenario to be used in the Inflation Technique. Next, inflate the Fritz distribution into the pre-injectable sets as described by the Inflation lemma (lemma 2). Finally, obtain infeasibility certificates using some linear programing software and obtain a valid causal compatibility inequality that witnesses the Fritz distribution's incompatibility.

As a first demonstration, we make use of the Wagon-Wheel inflation (fig. 5c) to find these inequalities. 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. By inspection[11], the maximal pre-injectable sets of the Wagon-Wheel inflation are as follows:

Maximal Pre-injectable Sets	Ancestral Independences	
$\{A_2, B_1, C_3, C_1\}$	$\{A_2, B_1, C_3\} \perp \{C_1\}$	
$\{A_1, B_1, C_4, C_2\}$	$\{A_1, B_1, C_4\} \perp \{C_2\}$	(19)
$\{A_1, B_2, C_1, C_3\}$	$\{A_1, B_2, C_1\} \perp \{C_3\}$	
$\{A_2, B_2, C_2, C_4\}$	$\{A_2, B_2, C_2\} \perp \{C_4\}$	

---

[10] 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 [32]. 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 [32]; if the dual is unbounded, then the primal is infeasible.

[11] Computing the injectable and pre-injectable sets computationally is possible and a technique is discussed in Wolfe *et al.* [18].

These maximal pre-injectable sets define a marginal scenario where $\mathcal{G}'$ is the Wagon-Wheel inflation of the triangle scenario $\mathcal{G}$.

$$\mathcal{M} = \mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}') = \{\{A_2, B_1, C_3, C_1\}, \{A_1, B_1, C_4, C_2\}, \{A_1, B_2, C_1, C_3\}, \{A_2, B_2, C_2, C_4\}\} \tag{20}$$

The complete set of variables $\mathcal{J}$ is simply the set of observable nodes in the Wagon-Wheel inflation,

$$\mathcal{J} = \mathcal{N}_O' = \{A_1, A_2, B_1, B_2, C_1, C_2, C_3, C_4\}$$

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 capable of accommodating the Fritz distribution. Analogously to theorem 6, this marginal scenario defines an incidence matrix M that has $|\mathcal{S}(\mathcal{M})| = \sum_{V \in \mathcal{M}} |\mathcal{S}(V)| = 4 \cdot 4^4 = 1024$ rows and $|\mathcal{S}(\mathcal{J})| = \prod_{v \in \mathcal{J}} |\mathcal{S}(v)| = 4^{|\mathcal{J}|} = 4^8 = 65,536$ 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.

The marginal distribution vector $\mathcal{P}^{\mathcal{M}}$ is indexed by 1024 marginal outcomes $m \in \mathcal{S}(\mathcal{M})$ of the following *symbolic* form,

$$\mathcal{P}^{\mathcal{M}\mathsf{T}} = \underbrace{(P_{A_2 B_1 C_3 C_1}(0000), \dots, P_{A_1 B_1 C_4 C_2}(3232), \dots, P_{A_2 B_2 C_2 C_4}(3333))}_{1024 \text{ entries}} \tag{21}$$

Which can be factored into the ancestrally independent injectable sets using eq. (19),

$$\mathcal{P}^{\mathcal{M}\mathsf{T}} = (P_{A_2 B_1 C_3}(000) P_{C_1}(0), \dots, P_{A_1 B_1 C_4}(323) P_{C_2}(2), \dots, P_{A_2 B_2 C_2}(333) P_{C_4}(3)) \tag{22}$$

And finally, each probability distribution in eq. (22) is deflated by dropping copy-indices.

$$\mathcal{P}^{\mathcal{M}\mathsf{T}} = (P_{ABC}(000) P_C(0), \dots, P_{ABC}(323) P_C(2), \dots, P_{ABC}(333) P_C(3)) \tag{23}$$

This step is permitted because all of the remaining distributions in eq. (22) are defined over the injectable sets of the Wagon-Wheel inflation. Numerically however, each of the elements of $\mathcal{P}^{\mathcal{M}}$ are replaced with numerical values corresponding to the Fritz distribution upon deflation. Pursuant to eq. (7), $P_{ABC}(323) P_C(2)$, for example, 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. Below is an example of an infeasibility inequality obtained over the entries of eq. (23).

$$\begin{aligned}
2P_{ABC}(203)P_C(0) + 2P_{ABC}(303)P_C(0) + P_{ABC}(003)P_C(0) + P_{ABC}(013)P_C(0)+ \\
P_{ABC}(020)P_C(2) + P_{ABC}(022)P_C(2) + P_{ABC}(023)P_C(0) + P_{ABC}(023)P_C(2)+ \\
P_{ABC}(030)P_C(2) + P_{ABC}(031)P_C(2) + P_{ABC}(032)P_C(2) + P_{ABC}(033)P_C(0)+ \\
P_{ABC}(033)P_C(2) + P_{ABC}(103)P_C(0) + P_{ABC}(113)P_C(0) + P_{ABC}(123)P_C(0)+ \\
P_{ABC}(133)P_C(0) + P_{ABC}(200)P_C(0) + P_{ABC}(200)P_C(1) + P_{ABC}(200)P_C(2)+ \\
P_{ABC}(200)P_C(3) + P_{ABC}(201)P_C(0) + P_{ABC}(201)P_C(1) + P_{ABC}(201)P_C(2)+ \\
P_{ABC}(201)P_C(3) + P_{ABC}(203)P_C(1) + P_{ABC}(203)P_C(2) + P_{ABC}(203)P_C(3)+ \\
P_{ABC}(213)P_C(0) + P_{ABC}(223)P_C(0) + P_{ABC}(300)P_C(0) + P_{ABC}(300)P_C(1)+ \\
P_{ABC}(300)P_C(2) + P_{ABC}(300)P_C(3) + P_{ABC}(301)P_C(0) + P_{ABC}(301)P_C(1)+ \\
P_{ABC}(301)P_C(2) + P_{ABC}(301)P_C(3) + P_{ABC}(302)P_C(1) + P_{ABC}(303)P_C(1)+ \\
P_{ABC}(303)P_C(2) + P_{ABC}(303)P_C(3) + P_{ABC}(313)P_C(0) - P_{ABC}(000)P_C(3) \ge 0
\end{aligned} \tag{24}$$

By construction, every marginal model $\mathcal{P}^{\mathcal{M}} = \{P_{ABC}, P_{AB}, P_{BC}, P_{AC}, P_A, P_B, P_C\}$ compatible with the triangle scenario must satisfy eq. (24). The Fritz distribution however violates eq. (24).

Notice the final term $P_{ABC}(000)P_C(3)$ has a coefficient -1. For the Fritz distribution, $P_{ABC}(000)P_C(3) = \frac{1}{128}\left(2 + \sqrt{2}\right) \simeq 0.0267$ but all of the remaining terms sum to $\simeq 0.0136$ meaning that the Fritz distribution violates this inequality with violation $-0.0129\ldots \not\geq 0$.

It is possible to simplify this inequality by marginalizing over sets of terms such as $P_{ABC}(300)P_C(0) + P_{ABC}(300)P_C(1) + P_{ABC}(300)P_C(2) + P_{ABC}(300)P_C(3) = P_{ABC}(300)$. Upon doing so, one obtains the following,

$$P_{ABC}(000)P_C(3) + P_{ABC}(021)P_C(2) + P_{ABC}(202) + P_{ABC}(302) + P_{ABC}(233)P_C(0) + P_{AC}(33)P_C(0) \leq$$
$$P_{ABC}(303)P_C(0) + P_{ABC}(313)P_C(0) + P_{ABC}(302)P_C(1)+$$
$$+P_{AB}(02)P_C(2) + P_{AB}(03)P_C(2) + P_{AB}(20) + P_{AB}(30) + P_C(3)P_C(0)$$

There are a couple of important concepts to understand. First, the Fritz distribution is known to be incompatible with the Triangle scenario due to the perfect correlation between C and the measurement pseudo-settings. The inequality contained in eq. (24) represents a significant advancement in our understanding of non-classicality in the triangle scenario. For example, eq. (24) is violated by distributions that do not exhibit perfect correlation between C and A_l, B_l. Moreover, although the existence of an inequality capable of witnessing Fritz distribution is guaranteed, there was no guarantee that the marginal model associated with the Wagon-Wheel inflation (eq. (20)) would be capable of producing such an inequality. In fact, smaller inflations of the triangle scenario such as the Spiral inflation (fig. 5b) and Cut inflation (fig. 5a) are *not* able to produce causal compatibility inequalities that witness the Fritz distribution.

Larger inflations such as the Web inflation (fig. 5d) were also used to generate causal compatibility inequalities that are violated by the Fritz distribution. The set of maximal pre-injectable sets for the Web inflation is considerably larger than eq. (19).

Maximal Pre-injectable Sets	Ancestral Independences
$\{A_1, B_1, C_1, A_4, B_4, C_4\}$	$\{A_1, B_1, C_1\} \perp \{A_4, B_4, C_4\}$
$\{A_1, B_2, C_3, A_4, B_3, C_2\}$	$\{A_1, B_2, C_3\} \perp \{A_4, B_3, C_2\}$
$\{A_2, B_3, C_1, A_3, B_2, C_4\}$	$\{A_2, B_3, C_1\} \perp \{A_3, B_2, C_4\}$
$\{A_2, B_4, C_3, A_3, B_1, C_2\}$	$\{A_2, B_4, C_3\} \perp \{A_3, B_1, C_2\}$
$\{A_1, B_3, C_4\}$	$\{A_1\} \perp \{B_3\} \perp \{C_4\}$
$\{A_1, B_4, C_2\}$	$\{A_1\} \perp \{B_4\} \perp \{C_2\}$
$\{A_2, B_1, C_4\}$	$\{A_2\} \perp \{B_1\} \perp \{C_4\}$
$\{A_2, B_2, C_2\}$	$\{A_2\} \perp \{B_2\} \perp \{C_2\}$
$\{A_3, B_3, C_3\}$	$\{A_3\} \perp \{B_3\} \perp \{C_3\}$
$\{A_3, B_4, C_1\}$	$\{A_3\} \perp \{B_4\} \perp \{C_1\}$
$\{A_4, B_1, C_3\}$	$\{A_4\} \perp \{B_1\} \perp \{C_3\}$
$\{A_4, B_2, C_1\}$	$\{A_4\} \perp \{B_2\} \perp \{C_1\}$

(25)

Again, analogously to theorem 6, this marginal scenario defines an incidence matrix M that has $|\mathcal{S}(\mathcal{M})| = 4 \cdot 4^6 + 8 \cdot 4^3 = 16,896$ rows, $|\mathcal{S}(\mathcal{J})| = 4^{12} = 16,777,216$ and $201,326,592$ non-zero entires. Using this incidence matrix, a number of incompatibility certificates were derived. For inspection, one is reproduced in appendix C 1.

In summary, these causal compatibility inequalities represent a completion of objective item **R.1** from section IV. These inequalities are derived without making use of Bell's theorem and instead use the Inflation Technique which is completely independent of Bell's theorems. Additionally, the derivation of causal compatibility inequalities allows one prove incompatibility even when perfect correlations are absent. Moreover, they allow one to probe the amount of noise the Fritz distribution can sustain before becoming compatible (see section VII). Unfortunately, infeasibility certificates are found by linear programing software to certify the infeasibility of a *particular distribution*. This means that inequalities like eq. (24), the one presented in appendix C 1, and others are specifically manufactured in order witness a particular type of quantum non-classicality. In an effort to derive inequalities that could lead to potentially new forms of non-classicality, we searched for inequalities that did not admit biased towards the Fritz distribution.

C. Symmetric Inequalities

Quite surprisingly, it is possible to derive causal compatibility inequalities for the triangle scenario that are symmetric under the exchange of parties. This is done by grouping marginal events $m \in \mathcal{S}(\mathcal{M})$ into orbits under the

action of party exchange. Even more interestingly, these inequalities are violated by the Fritz distribution. This comes as a surprise because the Fritz distribution is not invariant under the exchange of parties; in fact the Fritz distribution is noticeably biased toward C.

In general, exploiting symmetries of the marginal scenario is useful for a few distinct reasons. First, Bancal *et al.* [39] discuss computational advantages in considering symmetric versions of marginal polytopes mentioned in section VI; 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 [21]) 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[12].

To demonstrate what types of symmetries we have in mind, we present the following example:

Example 11. 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: $O_A = O_B = O_C = O_D = \{0, 1\}$. Moreover, let $I_{\mathcal{M}}$ be the following compatibility inequality over $\mathcal{M}$ for some causal structure $\mathcal{G}$,

$$I_{\mathcal{M}} = \{P_{ABC}(010) \leq P_{CD}(00) + P_{AD}(01)\}$$

Intuitively, the compatibility of $\mathcal{P}^{\mathcal{M}}$ is *independent* of the how one labels outcomes. Suppose $\varphi : \mathcal{J} \to \mathcal{J} : \{a, b, c, d\} \mapsto \varphi[\{a, b, c, d\}] = \{c, a, d, b\}$ is a permutation of the variables $\mathcal{J}$. Intuitively φ acts on $I_{\mathcal{M}}$ in the following way,

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

In general (as in this example), permutations over all variables can modify the marginal scenario $\varphi[\mathcal{M}] \neq \mathcal{M}$; in general we write $\varphi[I_{\mathcal{M}}] = \varphi[I]_{\varphi[\mathcal{M}]}$. Although $\varphi[I_{\mathcal{M}}]$ is still a valid causal compatibility inequality, in constrains marginal models of a different marginal scenario $\varphi[\mathcal{M}]$ and are in general of no interest.

Following the observations of example 11, it is always possible to permute the random variables in a marginal scenario to produce new and valid inequalities. The symmetry group over a the complete set of variables $\mathcal{J}$ is simply the permutation group[13] $\mathsf{Perm}(\mathcal{J})$. As was noted in example 11, some permutations $\varphi \in \mathsf{Perm}(\mathcal{J})$ have the undesired possibility of transforming $\mathcal{M}$ into a different marginal scenario.

$$\varphi[\mathcal{M}] \equiv \{\varphi[V_1], \ldots, \varphi[V_k]\} \neq \mathcal{M}$$

This means that there could exist $V \in \mathcal{M}$ such that $\varphi[V] \notin \mathcal{M}$. Permutations φ retaining this property are unable to generate compatibility inequalities over the same marginal scenario. Therefore, the desired set of symmetries is a subgroup that takes $\mathcal{M}$ to $\mathcal{M}$.

Definition 12. The **variable permutation group** Φ 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 [39], we divert our attention to finding *symmetric* inequalities themselves.

$$\forall \varphi \in \Phi(\mathcal{M}) : \varphi[I_{\mathcal{M}}] = I_{\mathcal{M}}$$

---

[12] If the extremal distribution happens to not be symmetric, there is indication that the space of accessible distributions is non-convex.
[13] The permutation group over a set S is the set of all bijective maps $\varphi : S \to S$.

In order to do this, we must first define how elements $\varphi \in \Phi$ act on marginal events. The action of $\varphi \in \Phi$ on any outcome $f \in \mathcal{S}(V)$ (denoted $\omega[f]$) is defined as,

$$\forall v \in \mathcal{D}(\varphi[f]) : (\varphi[f])(v) \equiv f(\varphi^{-1}[v]) \tag{26}$$

Equation (26) is nothing more than a mathematical definition for the intuitive action used in example 11. 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 $\mathcal{S}(\mathcal{J})$.

$$\mathsf{Orb}_\Phi(m) \equiv \{\varphi[m] \mid \varphi \in \Phi\}$$
$$\mathsf{Orb}_\Phi(j) \equiv \{\varphi[j] \mid \varphi \in \Phi\}$$

Using these group orbits, it is possible to contract the incidence matrix M of a marginal scenario into a symmetrized version.

Definition 13. 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 $\mathsf{Orb}_\Phi(m)$,

$$\left(\mathcal{P}_\Phi^{\mathcal{M}}\right)_{\mathsf{Orb}_\Phi(m)} = \sum_{m' \in \mathsf{Orb}_\Phi(m)} \mathcal{P}_m^{\mathcal{M}}$$

Which altogether 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 section VI 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 scenario, 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. 5d), we have obtained $\Phi(\mathsf{PreInj}_\mathcal{G}(\mathcal{G}'))$ which is an order 48 group with the following 4 generators:

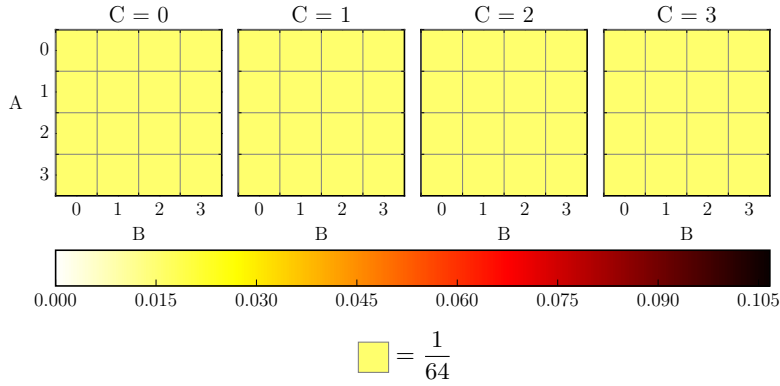| φ_1	φ_2	φ_3	φ_4
$A_1 \to A_4$	$A_1 \to A_1$	$A_1 \to C_1$	$A_1 \to A_1$
$A_2 \to A_3$	$A_2 \to A_3$	$A_2 \to C_2$	$A_2 \to A_2$
$A_3 \to A_2$	$A_3 \to A_2$	$A_3 \to C_3$	$A_3 \to A_3$
$A_4 \to A_1$	$A_4 \to A_4$	$A_4 \to C_4$	$A_4 \to A_4$
$B_1 \to B_4$	$B_1 \to C_1$	$B_1 \to A_1$	$B_1 \to B_2$
$B_2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

FIG. 6: The completely uniform distribution $\mathcal{U}$ visualized using a $4 \times 4 \times 4$ grid.

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 pre-injectable sets of eq. (25). Importantly,

$$\Phi\big(\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')\big) \sim \mathsf{Perm}(A, B, C)$$

To see this, φ_1 and φ_4 become the identity element $\mathbb{1}$ 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. This inequality is reproduced in appendix C 2.

VII.   VIOLATIONS & NOISE

Causal compatibility inequalities for the triangle scenario (such as those derived in section VI) offer an avenue into studying quantum non-classicality. In this section, we illuminate two tasks that can be performed in order to further understand the nature of quantum non-classical by using inequalities.

A.   Noise

In his original work, Fritz [12] was able to find quantum non-classicality in the triangle scenario by embedding Bell non-classicality. As was mentioned in section IV, 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.

In the triangle scenario, we define a completely noisy observation as the uniform probability distribution $\mathcal{U} = \mathcal{U}_{ABC}$ (see fig. 6)

$$\mathcal{U}(abc) = \mathcal{U}_{ABC}(abc) = \frac{1}{64} \quad \forall a, b, c \in \{0, 1, 2, 3\}$$

Since the uniform distribution $\mathcal{U}$ is trivially compatible with the triangle scenario, the objective of this section is to determine how much uniform noise needs to be added to the Fritz distribution P_{F} before it transitions from

(a) Complete range of noise parameters ε.

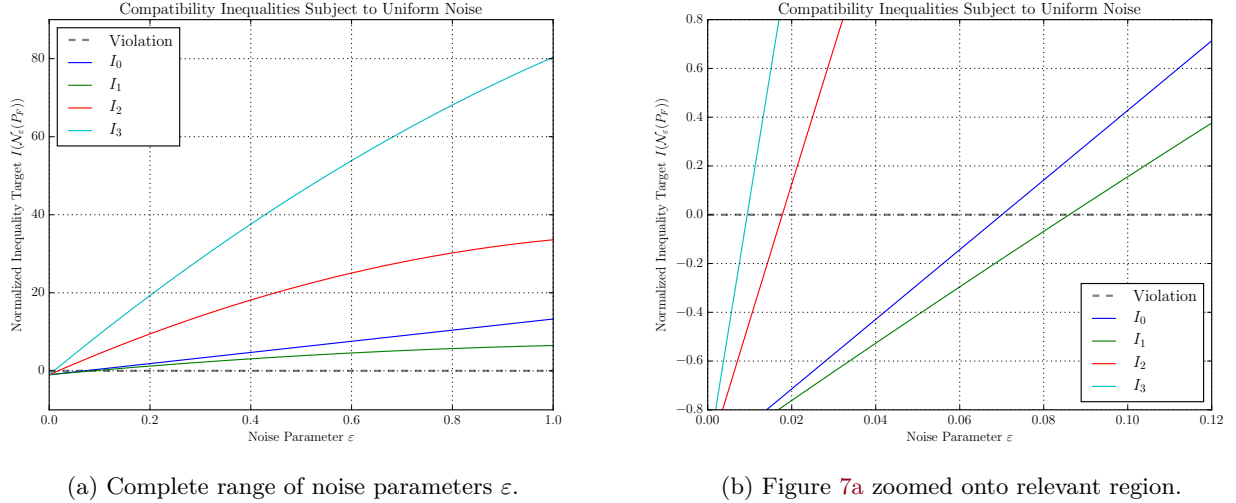(b) Figure 7a zoomed onto relevant region.

FIG. 7: Inequality violations when subjected to uniform noise. Each inequality is normalized so that maximum violation is -1.

incompatible to compatible. We do this by defining an affine noise parameter ε such that $0 \leq \varepsilon \leq 1$ along with the **noisy variant** $\mathcal{N}_\varepsilon$ of the Fritz distribution,

$$\mathcal{N}_\varepsilon = (1 - \varepsilon)P_\mathrm{F} + \varepsilon\mathcal{U}$$

Therefore as ε varies from 0 to 1, more and more noise is added to the Fritz distribution as $\mathcal{N}_\varepsilon$ transitions from an incompatible distribution $\mathcal{N}_0 = P_\mathrm{F}$ to a compatible distribution $\mathcal{N}_1 = \mathcal{U}$. For each inequality I and distribution P, let $I(P)$ denote a scalar value representing the amount of violation seen by P. As a standard, let $I(P) < 0$ represent violations while $I(P) \leq 0$ represents satisfaction of I (with $I(P) = 0$ corresponding to complete saturations of the inequality).

To analyze noise, 4 inequalities where chosen. First, I_0 is the inequality derived using the Wagon-Wheel wheel inflation eq. (24). The other three inequalities I_1, I_2, I_3 were all derived using the Web inflation; I_1 can be found in appendix C 1 while I_3 is the symmetric inequality in appendix C 2. Inequality I_2 is not reproduced anywhere in this document but is similar to appendix C 1.

The results of our experiments are plotted in fig. 7. Figure 7a illustrates that each curve $I(\mathcal{N}_\varepsilon(P_\mathrm{F}))$ as a function of ε is monotonically increasing as expected; the more noise added to the probability distribution, the more classical it becomes. Figure 7b is the same plot as fig. 7a zoomed onto the interval $\varepsilon \in [0, 0.12]$. These tests suggest that inequality I_1 (appendix C 1), when compared against all others is most robust to noise.

Overall, this noise model demonstrates that the Fritz distribution remains incompatible with the triangle scenario up to at least a noise parameter of $\varepsilon \simeq 0.085$. The associated distribution $\mathcal{N}_{0.085}(P_\mathrm{F})$ is plotted in fig. 8. There remains the possibility that another inequality will be able to withstand a larger degree of noise that I_1. Given that I_0, I_1, I_2, I_3 are far from a representative inequalities and are likely not facets of any inflated marginal polytope, it is save to say that this is possible.

B. Numerical Optimizations

Causal compatibility inequalities act as a *measure* of non-classicality. If two non-classical distributions P_1 and P_2 violate the same inequality I with violations $I(P_1)$ and $I(P_2)$ respectively, it is somewhat reasonable to say that P_1 is *more* non-classical than P_2 whenever $I(P_1) < I(P_2)$[14] Nonetheless, this statement comes from several caveats. First,

----

[14] Remember that larger violations are more negative when the inequalities are written in the proposed canonical form.
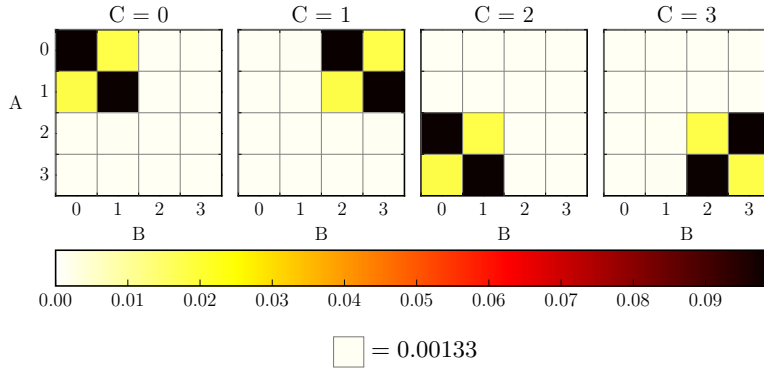
FIG. 8: $\mathcal{N}_{0.085}$: A noisy yet still non-classical variant of the Fritz distribution.

it is possible (and frequently the case) for $I(P_1) < I(P_2)$ to hold for one inequality I but have $\tilde{I}(P_2) < \tilde{I}(P_1)$ hold for another inequality $\tilde{I}$. In this case, all one can claim is that P_1 is more incompatible than P_2 with respective to I but the contrary is true with respect to $\tilde{I}$. Second, not all inequalities I are *good* measures of non-classicality. In the case of the Bell scenario, the linear inequality constraints associated with the facets of the marginal polytope are the best measures of non-classicality. This is because the facets form a necessary and sufficient set with all other inequalities following from positive mixtures of the facets. However, good measures of incompatibility are less understood in the triangle scenario where the space of classical correlations is non-convex. At the very least, the inequalities derived in this report are likely *not* the best measures of incompatibility (especially because they are curated toward the Fritz distribution). Nonetheless, these inequalities act the *first* measures of non-classicality in the triangle scenario.

In order to find quantum distributions that are more non-classical than the Fritz, we performed numerical optimizations against a objective scalar function $f : \mathbb{R}^n \in \mathbb{R}$ using the following pipeline: To begin, a set of real-valued parameters $\lambda = (\lambda_0, \ldots, \lambda_n) \in \mathbb{R}^n$ are used to parameterize a set of quantum states $\rho_{AB}, \rho_{BC}, \rho_{CA}$ and measurements M_A, M_B, M_C in agreement with the quantum version of the triangle scenario in fig. 9. Second, a 4-outcome distribution P_{ABC} is obtained via eq. (27) and supplanted into an inequality I in homogeneous and canonical form $I(P_{ABC}) \leq 0$. The objective value is simply the degree of violation $f = I(P_{ABC})$ and used to make adjustments to λ in order to achieve largest and larger violations.

Finding larger violations to compatibility inequalities is nothing more that a numerical optimization (specifically minimization) of $f(\lambda)$. 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 [40, p. 142] and the Nelder-Mead simplex method [40, p. 238] were used along with a hybrid method called Basin Hopping [41].

C.   Parameterizing Quantum Distributions

In search of new, incompatible, quantum distributions that can be realized on the triangle scenario, numerical optimizations over the the space of quantum-accessible probability distributions are performed. To do so, we are interested in a parameterization all distributions that can be expressed as follows:

$$P_{ABC}(abc) = \text{Tr}[\Pi^{\intercal} \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA} \Pi M_{A,a} \otimes M_{B,b} \otimes M_{C,c}] \tag{27}$$

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[15].

In order to qualify the scope of eq. (27) and associated computational complexity of the parameterization, there are a two restrictions that are made with justification. The states ρ are taken to be bipartite *qubit* states acting on $\mathcal{H}^2 \otimes \mathcal{H}^2$.

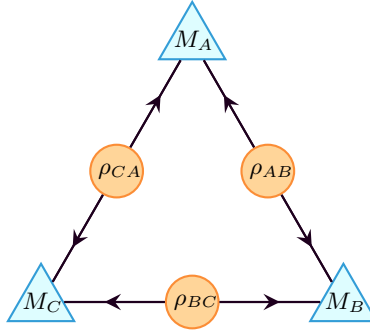[15] Π is discussed more thoroughly in section VII C 4.

FIG. 9: The triangle scenario as modeled by quantum states and measurements.

This is motivated by section IV in which the Fritz distribution only requires qubit states. In generality, one could consider n-dimensional states acting on $\mathcal{H}^n \otimes \mathcal{H}^n$. However, in such cases the joint density matrix $\rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA}$ then acts on $(\mathcal{H}^n)^{\otimes 6}$ requiring a $n^6 \times n^6$ matrix with n^{12} entries. Therefore, $n = 2$ is far more computationally feasible than $n > 2$. Additionally, restrict our focus to projective-valued measures (PVMs) instead of projective-operator valued measures (POVMs) for three reasons. First, section IV demonstrates that PVMs are sufficient for witnessing incompatible quantum distributions in the triangle scenario. Second, generating k-outcome POVM measurements is possible using rejection sampling techniques [42], however a valid, unbiased parameterization was not found for $k > 2$. Finally, PVMs provide considerable computational advantage over POVMs as they permit eq. (27) to be re-written as follows:

$$P_{ABC}(abc) = \langle m_{A,a} m_{B,b} m_{C,c} | \Pi^{\mathsf{T}} \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA} \Pi | m_{A,a} m_{B,b} m_{C,c} \rangle \tag{28}$$

In order to parameterize all such distributions, we elect to parameterize the states and measurements separately. Although there are numerous techniques that can used when parameterizing quantum states and measurements [4, 42–46], a single technique is presented that was found to be most computationally suitable for our purposes.

1. Unitary Group

To facilitate the parameterization of quantum states states and PVMs, a parameterization of the unitary group of dimension d (denoted $\mathcal{U}(d)$) by Spengler, Huber and Hiesmayr [47] is introduced. Discussions of its application to states and measurements are deferred to section VII C 3 and section VII C 2 respectively.

In Ref. [47], Spengler *et al.* proved that all unitaries $U \in \mathcal{U}(d)$ can be parameterized without degeneracy as follows:

$$U = \left[\prod_{m=1}^{d-1} \left(\prod_{n=m+1}^{d} \exp(iP_n \lambda_{n,m}) \exp(i\sigma_{m,n} \lambda_{m,n}) \right) \right] \cdot \left[\prod_{l=1}^{d} \exp(iP_l \lambda_{l,l}) \right] \tag{29}$$

Where $\lambda = \{\lambda_{n,m} \mid n, m \in 1, \ldots, d\}$ form a $d \times d$ matrix of real-valued parameters with periodicities $\lambda_{m,n} \in \left[0, \frac{\pi}{2}\right]$ for $m < n$ and $\lambda_{m,n} \in [0, 2\pi]$ for $m \geq n$.

$$\lambda = \begin{pmatrix} \lambda_{1,1} & \cdots & \lambda_{1,d} \\ \vdots & \ddots & \vdots \\ \lambda_{d,1} & \cdots & \lambda_{d,d} \end{pmatrix}$$

And where P_l are one-dimensional projective operators,

$$P_l = |l\rangle\langle l| \qquad 1 \leq l \leq d \tag{30}$$

and the $\sigma_{m,n}$ are generalized anti-symmetric σ-matrices,

$$\sigma_{m,n} = -i|m\rangle\langle n| + i|n\rangle\langle m| \qquad 1 \leq m < n \leq d$$

The SSH parameterization (eq. (29)) is unique in that each of the real-valued parameters $\lambda_{n,m}$ has physical significance. For the sake of reference, let us label the matrix exponential terms in eq. (29) in a manner that corresponds to their affect on an orthonormal basis $\{|1\rangle, \ldots, |d\rangle\}$.

$$GP_l = \exp(iP_l\lambda_{l,l})$$
$$RP_{n,m} = \exp(iP_n\lambda_{n,m}) \tag{31}$$
$$R_{m,n} = \exp(i\sigma_{m,n}\lambda_{m,n})$$

For example, $R_{m,n} = \exp(i\sigma_{m,n}\lambda_{m,n})$ applies a rotation to the sub-space spanned by $|m\rangle$ and $|n\rangle$ for $m < n$. Analogously, $RP_{n,m}$ generates the relative phase between $|m\rangle$ and $|n\rangle$ for $m > n$ and GP_l fixes the global phase of $|l\rangle$. Possessing a parameterization that maintains this physical interpretation is useful as it allows one to readily eliminate any degeneracies. For the purposes of quantum distributions such as eq. (27), it should be clear that any contributions to global phase are irrelevant. The SSH parameterization becomes especially attractive because it readily permits one to drop the global phase terms by setting $\lambda_{l,l} = 0$ for all $l = 1, \ldots, d$. It is convenient to denote the set of all unitaries up to global phase considerations as $\tilde{\mathcal{U}}(d)$ and express $\tilde{U} \in \tilde{\mathcal{U}}(d)$ using $d(d-1)$ real-valued parameters:

$$\tilde{U} = \prod_{m=1}^{d-1}\left(\prod_{n=m+1}^{d} RP_{n,m}R_{m,n}\right) \tag{32}$$

Another attractive feature of the SSH parameterization not mentioned in [47] is that eq. (29), and thus eq. (32), can be implemented in a computationally efficient manner. This is discussed in appendix B.

2. Measurements

For each measurement set M in eq. (27), we consider a d-element projective-valued measure (PVM) $M = \{M_1, \ldots, M_d\}$ satisfying a number of familiar constraints:

$$\forall|\phi\rangle \in \mathcal{H}^d : \langle\phi|M_i|\phi\rangle \geq 0 \qquad \sum_{i=1}^{d} M_i = \mathbb{1}_d$$
$$M_i = |m_i\rangle\langle m_i| \qquad M_iM_j = \delta_{ij}M_i$$

Therefore, parameterizing M corresponds to parameterizing the set of all orthonormal basis $\{|\psi_1\rangle, \ldots, |\psi_d\rangle\}$ of $\mathcal{H}^d$. First note that any such basis can be transformed into the computational basis $\{|1\rangle, \ldots, |d\rangle\}$ by a unitary denoted $U^{-1} \in \mathcal{U}(d)$.

$$\forall i : U^{-1}|\psi_i\rangle = |i\rangle \tag{33}$$

With this observation, all that is required is to parameterize the set of all unitaries $U \in \mathcal{U}(d)$. Specifically, the projective property each $M_i \in M$ means that the global phase of U is completely arbitrary; one only needs to consider parameterizing unitaries up to global phase using eq. (32).

$$M = \left\{\tilde{U}|i\rangle\langle i|\tilde{U}^\dagger \mid i \in \{1, \ldots, d\}\right\}$$

For the purposes discussed in section VII C, only $d = 4$ outcome measurements are utilized and therefore requiring $4(4-1) = 12$ real-valued parameters. This method was inspired by the measurement seeding method for iterative optimization used by Pál and Vértesi [48] .

3. States

Each state $\rho \in (\rho_{AB}, \rho_{BC}, \rho_{CA})$ of eq. (27) is modeled as a two-qubit density matrices acting on $\mathcal{H}^2 \otimes \mathcal{H}^2$. The space of all such states corresponds to the space of all 4×4 positive semi-definite hermitian matrices with unitary trace.

$$\forall|\phi\rangle \in \mathcal{H}^d : \langle\phi|\rho|\phi\rangle \geq 0 \qquad \rho^\dagger = \rho \qquad \mathrm{Tr}(\rho) = 1 \tag{34}$$

Although is common to parameterize all such matrices using a Cholesky decomposition [46], we make use of the non-degenerate SHH parameterization [47] and the spectral decomposition of ρ. Retaining full generality, consider full-rank density matrices:

$$\rho = \sum_{i=1}^{d} p_i |\psi_i\rangle\langle\psi_i| \qquad \sum_{i=1}^{d} p_i = 1, p_i \geq 0 \tag{35}$$

Where $\{p_i\}$ are the eigenvalues of ρ and $\{|\psi_i\rangle\}$ are its eigenstates. It is here that one recognizes the reapplication of eq. (33): any orthonormal basis $\{|\psi_i\rangle\}$ of $\mathcal{H}^d$ can be transformed into a computational basis $\{|i\rangle\}$ by a unitary transformation $U \in \mathcal{U}(d)$ such that $|\psi_i\rangle = U|i\rangle$. Analogous to the projective measurements considered in section VII C 2, the global phase contributions are redundant.

$$\rho = \sum_{i=1}^{d} p_i \tilde{U} |i\rangle\langle i| \tilde{U}^\dagger$$

Parameterizing the eigenvalues requires $d-1$ real-valued parameters due to the trace constraint of eq. (34). The eigenvalues are parameterized without degeneracy using a tuple of $d-1$ parameters $\lambda = (\lambda_1, \ldots, \lambda_{d-1}), \lambda_i \in [0, 2\pi]$ using hyper-spherical coordinates [43, 47]:

$$p_n = \prod_{i=1}^{n-1} \sin^2 \lambda_i$$

$$p_j = \cos^2 \lambda_j \prod_{i=1}^{j-1} \sin^2 \lambda_i \quad \forall j \in 1, \ldots, n-1 \tag{36}$$

Furthermore due to the periodicity of the parameter space $\{\lambda\}$, eq. (36) can be used for either constrained or unconstrained optimization problems. For continuity reasons, unconstrained optimizations are performed whenever possible. In conclusion, it is possible to parametrize all d-dimensional density matrices satisfying eq. (34) using $d(d-1) + d - 1 = d^2 - 1$ real-valued parameters.

Although non-degenerate, eq. (36) suffers from a lack of uniformity; a randomly sampled vector of parameters $\{\lambda\}$ *does not* translate to a randomly sampled set of eigenvalues $\{p_i\}$. Another, albeit degenerate, parameterization of $\{p_i\}$ involves beginning with d real parameters $\lambda = (\lambda_1, \ldots, \lambda_d)$, then forcing normalization by dividing each parameter λ_i by the their sum.[16].

$$p_j = \frac{|\lambda_j|}{\sum_{i=1}^{n} |\lambda_i|} \quad \forall j \in 1, \ldots, n \tag{37}$$

For various optimization tasks sensitive to initial conditions outlined section VII B, the latter parameterization of eq. (37) generally performed better than the former eq. (36).

4. Permutation Matrix

Finally, we introduce the a permutation matrix Π for the triangle scenario. For bipartite qubit states, Π is the 64×64 bit-wise matrix depicted in fig. 10. Upon examination of eq. (28), we consider Π to be acting on the joint global state $\rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA}$. To illuminate its necessity, consider eq. (27) in the absence of Π.

$$P_{ABC}(abc) \stackrel{?}{=} \mathrm{Tr}\big[(\rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA})(M_A^a \otimes M_B^b \otimes M_C^c)\big]$$
$$= \mathrm{Tr}\big[(\rho_{AB}M_A^a) \otimes (\rho_{BC}M_B^b) \otimes (\rho_{CA}M_C^c)\big]$$
$$= \mathrm{Tr}(\rho_{AB}M_A^a)\,\mathrm{Tr}(\rho_{BC}M_B^b)\,\mathrm{Tr}(\rho_{CA}M_C^c)$$
$$= P_{A|\rho_{AB}}(a)P_{B|\rho_{BC}}(b)P_{C|\rho_{CA}}(c)$$

---

[16] Strictly speaking, eq. (37) *also* suffers from non-uniformity; being biased toward eigenvalues that have similar values to each other.

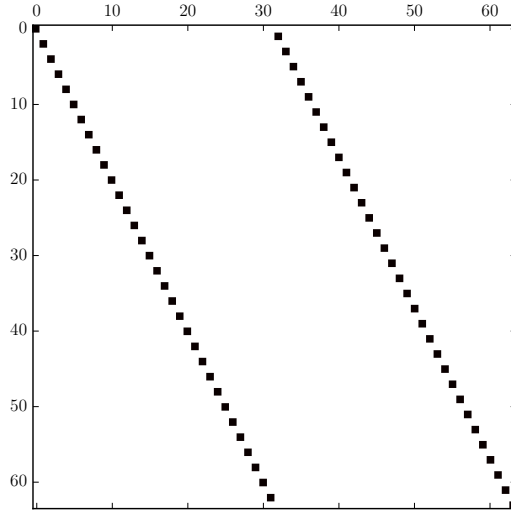FIG. 10: The permutation matrix Π acting on $\left(\mathcal{H}^2\right)^{\otimes 6}$ for the triangle scenario. Black entries represents a value of 1 and white represents 0.

On an operational level, this corresponds to A making a measurement on *both* subsystems of ρ_{AB} and *not* on any component of ρ_{CA}. This is analogously troubling for B and C as well. To resolve this issue, the permutation matrix Π corresponds to *aligning* the underlying 6-qubit joint state ρ with the joint measurement M. To understand its effect, consider its effect on 6-qubit pure state $|q_1\rangle \otimes \cdots \otimes |q_6\rangle = |q_1 q_2 q_3 q_4 q_5 q_6\rangle$ where $\forall i : |q_i\rangle \in \mathcal{H}^2$.

$$\Pi |q_1 q_2 q_3 q_4 q_5 q_6\rangle = |q_2 q_3 q_4 q_5 q_6 q_1\rangle$$

Π acts as a *partial transpose* on $\left(\mathcal{H}^2\right)^{\otimes 6}$ by shifting the underlying tensor structure one subsystem to the "left". It is uniquely defined by its action on all 2^6 orthonormal basis elements of $\left(\mathcal{H}^2\right)^{\otimes 6}$,

$$\Pi \equiv \sum_{|q_i\rangle \in \{|0\rangle, |1\rangle\}} |q_2 q_3 q_4 q_5 q_6 q_1\rangle \langle q_1 q_2 q_3 q_4 q_5 q_6|$$

D.   Numerical Optimization Results

Section VII C expounded on how we elected to parameterized the space of quantum distributions that can be realized on the quantum triangle scenario (fig. 9). Specifically, we make use of fully generic 4×4 density matrices (bipartite qubit) for each shared resource and 4-element projective measurements. For each inequality I, we define the relative violation V_F to be maximum violation relative to the violation achieved by the Fritz distribution.

$$V_F = \frac{\min_P \{I(P)\}}{I(P_F)}$$

The first inequality in consideration is I_1; the non-symmetric Web inflation inequality presented in appendix C 1. At best, numerical optimizations converged to a distribution (fig. 11) with relative violation $V_F \simeq 1.501$. Almost immediately, it is evident that fig. 11 closely resembles the Fritz distribution. In fact, fig. 11 and fig. 4 share the exact same *possibilistic structure* or set of possible events. What does this reveal about I_1 and non-classicality in the triangle scenario? This result is not entirely unexpected. As was mentioned in section VI B, these inequalities are derived specifically to show incompatibility of the triangle scenario. Consequently, it seems that I_1 only has the capacity to demonstrate incompatibility for a specific forms of non-classicality that resemble the Fritz distribution very closely.

An alternative explanation has to do with the sensitively of the initial condition $\lambda_{(0)} = \left(\lambda_{(0),0}, \ldots, \lambda_{(0),n}\right)$ in our optimization. In section VII C, it was shown that our parameter space of choice was 2π-periodic in all entries. Somewhat surprisingly, when using a uniform seed for λ_0, all optimization methods consistently converged to inequality
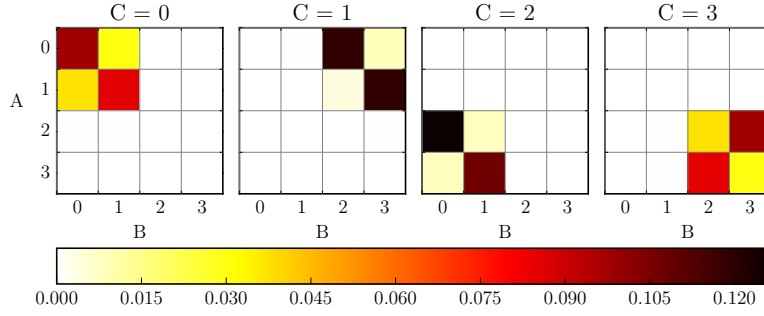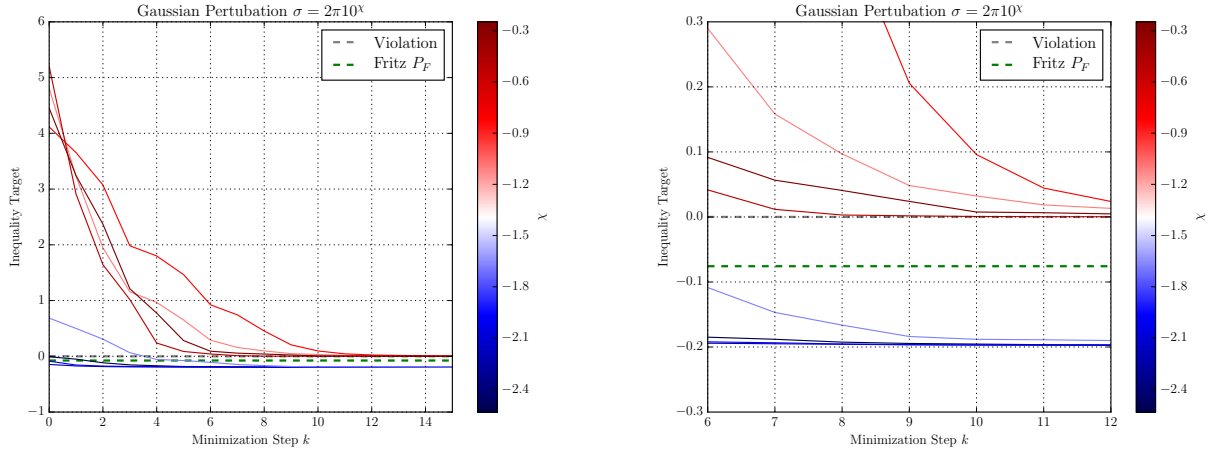
FIG. 11: Probability distribution that maximizes violation of I_1 from appendix C 1 ($V_F \simeq 1.501$).

(a) For each iteration step k, the inequality target decreases.

(b) Figure 7a zoomed to show convergence is sensitive to initial conditions.

FIG. 12: Numerical optimizations aimed at violating causal compatibility inequalities are sensitive to initial conditions.

saturation instead of inequality violation which of course is achievable by the Fritz distribution. Instead, only initial seeds close enough to the *Fritz parameters* $\lambda_{(\mathrm{F})}$ (the parameters that induce the Fritz distribution using the quantum implementation in section IV) where able to achieve violation. To illustrate this graphically, we let the initial seed be a small Gaussian perturbation $\delta\lambda$ away from $\lambda_{(\mathrm{F})}$ in parameter space.

$$\lambda_{(0)} = \lambda_{(\mathrm{F})} + \delta\lambda$$

The Gaussian perturbation was given zero mean $\mu = 0$ and variance $\sigma^2 = (2\pi 10^\chi)^2$ where χ measures an order of magnitude standard deviation. For I_1, deviations of $\chi \gtrsim 1.3$ inhibited convergence to violation. This is depicted in fig. 12. Therefore, it is 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. Most likely, this phenomena is due to a combination of both effects.

The second inequality in consideration is I_3; the *symmetric* Web inflation inequality presented in appendix C 2. This inequality was able to achieve a greater relative violation than I_1 as optimizations converged to a distribution (fig. 13) 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 exchange of parties $\varphi \in \mathsf{Perm}(A, B, C)$, one would naively expect that the maximum violating distribution
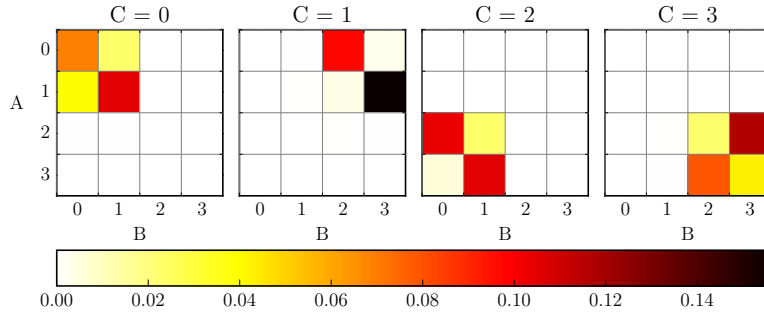
FIG. 13: Probability distribution that maximizes violation of I_3 from appendix C 2 ($V_F \simeq 2.61$).
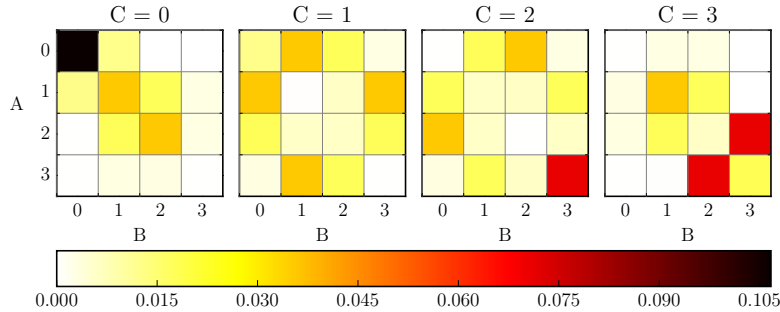

FIG. 14: Symmetrized version of the Fritz distribution ($V_F \simeq 3$).

is be symmetric as well. If it were not, one could construct a distribution that violates I_3 more by symmetrizing P_F by summing over all possible permutations and normalizing. The symmetrized Fritz distribution is plotted in fig. 14. This distribution is also non-classical with relative violation $V_F = 3$. As it turns out, the symmetrized Fritz distribution is *not* quantum-accessible in the sense that is cannot be written as eq. (27)[17] Therefore we conclude that the space of quantum accessible correlations on the triangle scenario is non-convex; a reiteration of a result found in Wolfe *et al.* [18].

Finally for both I_0 and I_2, the maximum relative violation happens to be $V_F = 1$ (up to numerical precision). This means 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 is 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_0 and I_2.

VIII. CONCLUSIONS

In section III, we elucidated that non-classicality in the triangle scenario has been an unsolved problem for nearly a decade. Initially, Branciard *et al.* [16] identified the triangle scenario as a challenging causal structure for studying non-classicality. Afterwards, Henson *et al.* [17] were able to to classify the triangle scenario as *interesting* which partially explains why it remains an enigma. Recently, Fritz [12] discovered the Fritz distribution as the first example of quantum non-classicality in the triangle scenario. Heretofore, researchers have been able to derive causal incompatibility inequalities [14, 17, 18] but unable to show that any are non-trivial in the context of quantum non-classicality. In section IV the Fritz distribution was introduced in detail and refinements where made to a problem originally proposed by Fritz [12]. For this research project, we made heavy use of the Inflation Technique of section V and the Fritz distribution in order to derive the first causal compatibility inequalities capable of having quantum violations.

---

[17] At least numerical optimizations suggest this is the case.

Finally, we used those inequalities in an attempt to discover new forms of non-classicality in the triangle scenario with partial success.

Ultimately, this work accomplishes two important tasks. First and foremost, the derivation of eq. (24) 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 scenario – a task that no other existing techniques succeeded at. Therefore, this work acts as a strong indication that the Inflation technique has the potential to solve a assortment of unsolved causal inference problems including those manifest in quantum foundations and quantum information theory.

Penultimately, the results presented here solve [12, Problem 2.17] insofar as to refine the original problem into three related questions problems: **R.1**, **R.2** and **R.3**. The methods and techniques demonstrated here provide a means of solving **R.1** and gives an opportunity to begin tackling **R.2** and **R.3** in the form of numerical optimizations. Computational and comprehensive limitations have stunted progress towards solving **R.2** and **R.3**, so it still remains unclear if **R.2** and **R.3** have affirmative solutions in the first place.

ACKNOWLEDGMENTS

[1] Peter W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," SIAM Journal on Computing **26**, 1484–1509 (1997).

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

[3] Charles H. Bennett and Gilles Brassard, "Quantum cryptography: Public key distribution and coin tossing," Theoretical Computer Science **560**, 7–11 (2014).

[4] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2011) Chap. The Schmidt decomposition and purifications, p. 109.

[5] M. L. Almeida, J. D. Bancal, N. Brunner, A. Acin, N. Gisin, and S. Pironio, "Guess your neighbour's input: a multipartite non-local game with no quantum advantage," (2010), 10.1103/PhysRevLett.104.230404, arXiv:1003.3844.

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

[7] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?" Phys. Rev. **47**, 777–780 (1935).

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

[9] Judea Pearl, "Causal inference in statistics: An overview," Statistics Surveys **3**, 96–146 (2009).

[10] Christopher J. Wood and Robert W. Spekkens, "The lesson of causal discovery algorithms for quantum correlations: Causal explanations of bell-inequality violations require fine-tuning," (2012), 10.1088/1367-2630/17/3/033002, arXiv:1208.4119.

[11] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner, "Bell nonlocality," (2013), 10.1103/RevModPhys.86.419, arXiv:1303.2849.

[12] Tobias Fritz, "Beyond bell's theorem: Correlation scenarios," (2012), 10.1088/1367-2630/14/10/103001, arXiv:1206.5115.

[13] Tobias Fritz, "Beyond bell's theorem ii: Scenarios with arbitrary causal structure," (2014), 10.1007/s00220-015-2495-5, arXiv:1404.4812.

[14] Bastian Steudel and Nihat Ay, "Information-theoretic inference of common ancestors," (2010), arXiv:1010.5720.

[15] Rafael Chaves, Lukas Luft, and David Gross, "Causal structures from entropic information: geometry and novel scenarios," New Journal of Physics **16**, 043001 (2014).

[16] Cyril Branciard, Denis Rosset, Nicolas Gisin, and Stefano Pironio, "Bilocal versus nonbilocal correlations in entanglement-swapping experiments," Phys. Rev. A **85** (2012), 10.1103/physreva.85.032119.

[17] Joe Henson, Raymond Lal, and Matthew F Pusey, "Theory-independent limits on correlations from generalized bayesian networks," New Journal of Physics **16**, 113043 (2014), arXiv:1405.2572.

[18] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz, "The inflation technique for causal inference with latent variables," (2016), arXiv:1609.00672.

[19] Tobias Fritz and Rafael Chaves, "Entropic inequalities and marginal problems," (2011), 10.1109/TIT.2012.2222863, arXiv:1112.4788.

[20] Barbara M. Terhal and David P. DiVincenzo, "Adaptive quantum computation, constant depth quantum circuits and arthur-merlin games," quant-ph/0205133v6.

[21] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt, "Proposed experiment to test local hidden-variable theories," Phys. Rev. Lett. **23**, 880–884 (1969).

[22] B. S. Cirel'son, "Quantum generalizations of bell's inequality," Lett Math Phys **4**, 93–100 (1980).

[23] N. N. Vorob'ev, "Consistent families of measures and their extensions," Theory of Probability & Its Applications **7**, 147–163 (1962).

[24] Samson Abramsky and Adam Brandenburger, "The sheaf-theoretic structure of non-locality and contextuality," (2011), 10.1088/1367-2630/13/11/113036, arXiv:1102.0264.

[25] Samson Abramsky and Lucien Hardy, "Logical bell inequalities," (2012), 10.1103/PhysRevA.85.062114, arXiv:1203.1352.

[26] Shane Mansfield and Tobias Fritz, "Hardy's non-locality paradox and possibilistic conditions for non-locality," (2011), 10.1007/s10701-012-9640-1, arXiv:1105.1819.

[27] Itamar Pitowsky, "Correlation polytopes: Their geometry and complexity," Mathematical Programming **50**, 395–414 (1991).

[28] George B Dantzig and B Curtis Eaves, "Fourier-motzkin elimination and its dual," Journal of Combinatorial Theory, Series A **14**, 288–297 (1973).

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

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

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

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

[33] Erling D. Andersen, Computational Optimization and Applications **20**, 171–183 (2001).

[34] MOSEK Optimization Suite, "Mosek api," (2016).

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

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

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

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

[39] Jean-Daniel Bancal, Nicolas Gisin, and Stefano Pironio, "Looking for symmetric bell inequalities," Journal of Physics A: Mathematical and Theoretical **43**, 385303 (2010).

[40] Jorge Nocedal and Stephen Wright, *Numerical Optimization (Springer Series in Operations Research and Financial Engineering)* (Springer, 2000).

[41] David J. Wales and Jonathan P. K. Doye, "Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms," J. Phys. Chem. A **101**, 5111–5116 (1997).

[42] Dénes Petz and László Ruppert, "Optimal quantum-state tomography with known parameters," (2015), 10.1088/1751-8113/45/8/085306, arXiv:1511.06666.

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

[44] Kazuyuki Fujii, Kunio Funahashi, and Takayuki Kobayashi, "Jarlskog's parametrization of unitary matrices and qudit theory," (2005), arXiv:quant-ph/0508006.

[45] Daniel F. V. James, Paul G. Kwiat, William J. Munro, and Andrew G. White, "Measurement of qubits," Phys. Rev. A **64** (2001), 10.1103/physreva.64.052312.

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

[47] Christoph Spengler, Marcus Huber, and Beatrix C. Hiesmayr, "A composite parameterization of unitary groups, density matrices and subspaces," (2010), 10.1088/1751-8113/43/38/385306, arXiv:1004.5252.

[48] Károly F. Pál and Tamás Vértesi, "Maximal violation of a bipartite three-setting, two-outcome bell inequality using infinite-dimensional quantum systems," Phys. Rev. A **82** (2010), 10.1103/physreva.82.022116.

[49] Cleve Moler and Charles Van Loan, "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later," SIAM Rev. **45**, 3–49 (2003).

Appendix A: Common Terminology & Notation

1. Graph Theory

Definition 14. A **graph** $\mathcal{G}$ is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *edges* respectively where the nodes have the capacity to represent any object while the edges are defined as *pairs* of nodes $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$. A **directed graph** $\mathcal{G}$ is a graph where each edge $e \in \mathcal{E}$ is assigned an order or **direction** $e = \{n \to n'\}$.

Definition 15. The following definitions are common language in directed graph theory. Let $n, m \in \mathcal{N}$ be example nodes of the graph $\mathcal{G}$.

- The **parents of a node**: $\mathsf{Pa}_{\mathcal{G}}(n) \equiv \{m \mid m \to n\}$

- The **children of a node**: $\mathsf{Ch}_{\mathcal{G}}(n) \equiv \{m \mid n \to m\}$

- The **ancestry of a node**: $\mathsf{An}_{\mathcal{G}}(n) \equiv \bigcup_{i \in \mathbb{W}} \mathsf{Pa}_{\mathcal{G}}^{i}(n)$ where $\mathsf{Pa}_{\mathcal{G}}^{i}(n) \equiv \mathsf{Pa}_{\mathcal{G}}\left(\mathsf{Pa}_{\mathcal{G}}^{i-1}(n)\right)$ and $\mathsf{Pa}_{\mathcal{G}}^{0}(n) = n$

- The **descendants of a node**: $\mathsf{De}_{\mathcal{G}}(n) \equiv \bigcup_{i \in \mathbb{N}} \mathsf{Ch}_{\mathcal{G}}^{i}(n)$ where $\mathsf{Ch}_{\mathcal{G}}^{i}(n) \equiv \mathsf{Ch}_{\mathcal{G}}\left(\mathsf{Ch}_{\mathcal{G}}^{i-1}(n)\right)$ and $\mathsf{Ch}_{\mathcal{G}}^{1}(n) = \mathsf{Ch}_{\mathcal{G}}(n)$

All of these terms can be generalized to sets of nodes $N \subseteq \mathcal{N}$ through union over the elements,

- The **parents of a node set**: $\mathsf{Pa}_{\mathcal{G}}(N) \equiv \bigcup_{n \in N} \mathsf{Pa}_{\mathcal{G}}(n)$

- The **children of a node set**: $\mathsf{Ch}_{\mathcal{G}}(N) \equiv \bigcup_{n \in N} \mathsf{Ch}_{\mathcal{G}}(n)$

- The **ancestry of a node set**: $\mathsf{An}_{\mathcal{G}}(N) \equiv \bigcup_{n \in N} \mathsf{An}_{\mathcal{G}}(n)$

Definition 16. 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.

$$\mathsf{Sub}_{\mathcal{G}}(N) \equiv (N, \{e = \{n \to n'\} \mid n, n' \in N, e \in \mathcal{E}\})$$

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.

$$\mathsf{AnSub}_{\mathcal{G}}(N) \equiv \mathsf{Sub}_{\mathcal{G}}\left(\mathsf{An}_{\mathcal{G}}(N)\right)$$

Definition 17. A **directed acyclic graph** or **DAG** $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ is an directed graph with the additional property that no node n is in its set of **ancestors**.

$$\forall n \in \mathcal{N} : n \notin \bigcup_{i \in \mathbb{N}} \mathsf{Pa}_{\mathcal{G}}^{i}(n)$$

It is critical to note the difference in language between *ancestors* from *ancestry*. The ancestry $\mathsf{An}_{\mathcal{G}}(n)$ of a node n is defined using *whole* numbers $\mathbb{W}$. In contrast, the ancestors of a node are defined similarly with *natural* numbers $\mathbb{N}$.

2. Probability Theory

Definition 18. The **product distribution** of two disjoint probability distributions P_V and P_W (where $V \cap W = \emptyset$) is denoted as $P_V \times P_W$.

$$\forall s \in \mathcal{S}(V \cup W) : (P_V \times P_W)(s) \equiv P_V(s|_V) \cdot P_W(s|_W)$$

A product distribution of k mutually disjoint distributions is defined as,

$$\prod_{i \in [k]} P_{V_i} \equiv (P_{V_1} \times \cdots \times P_{V_k})$$

Definition 19. The **marginalization** of a distribution P_V to a distribution over $W \subseteq V$ is denoted $\sum_{V \setminus W} P_V$ and is defined such that,

$$\forall s \in \mathcal{S}(W) : \left(\sum_{V \setminus W} P_V\right)(s) \equiv \sum_{j \in \mathsf{Ext}_V(s)} P_V(j)$$

Where $\mathsf{Ext}_V(s)$ denotes the set of all events $j \in \mathcal{S}(V)$ that can be restricted to s,

$$\mathsf{Ext}_V(s) = \left\{j \in \mathcal{S}(V) \mid s = j|_{\mathcal{D}(s)}\right\}$$

Appendix B: Computationally Efficient Parameterization of the Unitary Group

If section VII C 1, the SSH parameterization was introduced as a non-degenerate parameterization of the unitary group $\mathcal{U}(d)$. As presented, the SSH parameterization suffers due to the presence of computationally expensive matrix exponential terms [49]. Although not explicitly mentioned in [47], it is possible to remove the reliance on matrix exponential operations in eq. (29) by utilizing the explicit form of the exponential terms in eq. (31). As a first step, recognize the defining property of the projective operators eq. (30),

$$P_l^k = (|l\rangle\langle l|)^k = |l\rangle\langle l| = P_l$$

This greatly simplifies the global phase terms GP_l,

$$GP_l = \exp(iP_l\lambda_{l,l}) = \sum_{k=0}^{\infty} \frac{(iP_l\lambda_{l,l})^k}{k!} = \mathbb{1} + \sum_{k=1}^{\infty} \frac{(i\lambda_{l,l})^k}{k!}P_l^k = \mathbb{1} + P_l\left[\sum_{k=1}^{\infty} \frac{(i\lambda_{l,l})^k}{k!}\right] = \mathbb{1} + P_l(e^{i\lambda_{l,l}} - 1) \tag{B1}$$

Analogously for the relative phase terms $RP_{n,m}$,

$$RP_{n,m} = \cdots = \mathbb{1} + P_n(e^{i\lambda_{n,m}} - 1) \tag{B2}$$

Finally, the rotation terms $R_{m,n}$ can also be simplified by examining powers of $i\sigma_{n,m}$,

$$R_{m,n} = \exp(i\sigma_{m,n}\lambda_{m,n}) = \sum_{k=0}^{\infty} \frac{(|m\rangle\langle n| - |n\rangle\langle m|)^k \lambda_{m,n}^k}{k!}$$

One can verify that the following properties hold,

$$(|m\rangle\langle n| - |n\rangle\langle m|)^0 = \mathbb{1}$$
$$\forall k \in \mathbb{N}, k \neq 0 : (|m\rangle\langle n| - |n\rangle\langle m|)^{2k} = (-1)^k(|m\rangle\langle m| + |n\rangle\langle n|)$$
$$\forall k \in \mathbb{N} : (|m\rangle\langle n| - |n\rangle\langle m|)^{2k+1} = (-1)^k(|m\rangle\langle n| - |n\rangle\langle m|)$$

Revealing the simplified form of $R_{m,n}$,

$$R_{m,n} = \mathbb{1} + (|m\rangle\langle m| + |n\rangle\langle n|)\sum_{j=1}^{\infty}(-1)^j \frac{\lambda_{n,m}^{2j}}{(2j)!} + (|m\rangle\langle n| - |n\rangle\langle m|)\sum_{j=0}^{\infty}(-1)^j \frac{\lambda_{n,m}^{2j+1}}{(2j+1)!}$$

$$R_{m,n} = \mathbb{1} + (|m\rangle\langle m| + |n\rangle\langle n|)(\cos\lambda_{n,m} - 1) + (|m\rangle\langle n| - |n\rangle\langle m|)\sin\lambda_{n,m} \tag{B3}$$

By combining the optimizations of eqs. (B1) to (B3) together we arrive at an equivalent form for eq. (29) that is computational more efficient.

$$U = \left[\prod_{m=1}^{d-1}\left(\prod_{n=m+1}^{d} R_{m,n}RP_{n,m}\right)\right] \cdot \left[\prod_{l=1}^{d} GP_l\right] \tag{B4}$$

Appendix C: Large Triangle Scenario Inequalities

1. Web Inflation Inequality

The following is a causal compatibility inequality for the triangle scenario that is violated by the Fritz distribution outlined in section IV and was derived using the Web inflation of fig. 5d.

$$P(110)P(223) + P(110)P(233) + P(110)P(323) + P(110)P(333)$$

$$\leq$$

$$
\begin{aligned}
&2P(020)P(213) + 2P(023)P(210) + 2P(023)P(310) + 2P(030)P(213) + 2P(033)P(210) + 2P(033)P(310) + 2P(120)P(213) + 2P(123)P(210) + 2P(123)P(310) + 2P(130)P(213)+ \\
&2P(132)P(311) + 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(000)P(223) + 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(023) + P(010)P(033) + 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(010)P(223) + 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(210)+ \\
&P(021)P(211) + P(021)P(212) + P(021)P(213) + P(021)P(310) + P(021)P(311) + P(021)P(312) + P(021)P(313) + P(022)P(210) + 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(212) + P(031)P(213)+ \\
&P(031)P(310) + P(031)P(311) + P(031)P(312) + P(031)P(313) + P(032)P(210) + 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(100)P(223) + 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(110)P(223) + 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(210) + P(121)P(211) + P(121)P(212) + 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(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(212) + P(131)P(213) + P(131)P(310) + P(131)P(311) + P(131)P(312) + P(131)P(313)+ \\
&P(132)P(201) + P(132)P(210) + P(132)P(211) + P(132)P(212) + P(132)P(213) + P(132)P(301) + P(132)P(310) + 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(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(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)
\end{aligned}
$$

Note: $P(abc)$ is shorthand for $P_{ABC}(abc)$.

2. Symmetric Web Inflation Inequality

The following is a symmetric causal compatibility inequality for the triangle scenario that is violated by the Fritz distribution outlined in section IV and was derived using the Web inflation of fig. 5d.

$$2[P(001)P(333)]_3 + 2[P(010)P(323)]_3 + 6[P(000)P(323)]_3 + 6[P(000)P(333)]_1$$

$$\leq$$

$$12[P(031)P(302)]_6 + 12[P(033)P(303)]_6 + 12[P(103)P(130)]_6 + 12[P(203)P(230)]_6 + 12[P(203)P(330)]_6 + 2[P(001)P(320)]_6 + 2[P(002)P(221)]_3 + 2[P(003)P(211)]_6+$$
$$2[P(003)P(331)]_3 + 2[P(011)P(211)]_3 + 2[P(012)P(322)]_6 + 2[P(013)P(313)]_6 + 2[P(013)P(332)]_6 + 2[P(020)P(111)]_3 + 2[P(020)P(211)]_6 + 2[P(021)P(212)]_6+$$
$$2[P(022)P(211)]_3 + 2[P(022)P(212)]_6 + 2[P(022)P(322)]_3 + 2[P(023)P(232)]_6 + 2[P(030)P(212)]_3 + 2[P(031)P(231)]_6 + 2[P(032)P(331)]_6 + 2[P(033)P(333)]_3+$$
$$2[P(101)P(131)]_3 + 2[P(101)P(132)]_6 + 2[P(102)P(131)]_6 + 2[P(102)P(132)]_6 + 2[P(102)P(133)]_6 + 2[P(110)P(133)]_6 + 2[P(110)P(212)]_6 + 2[P(110)P(222)]_3+$$
$$2[P(110)P(223)]_3 + 2[P(112)P(331)]_3 + 2[P(120)P(122)]_6 + 2[P(121)P(201)]_6 + 2[P(122)P(200)]_3 + 2[P(122)P(202)]_6 + 2[P(122)P(210)]_6 + 2[P(122)P(300)]_3+$$
$$2[P(130)P(232)]_6 + 2[P(130)P(233)]_6 + 2[P(131)P(201)]_6 + 2[P(131)P(202)]_3 + 2[P(131)P(313)]_3 + 2[P(133)P(200)]_3 + 2[P(133)P(201)]_6 + 2[P(133)P(211)]_3+$$
$$2[P(133)P(212)]_6 + 2[P(133)P(300)]_3 + 2[P(202)P(231)]_6 + 2[P(210)P(222)]_6 + 2[P(220)P(222)]_3 + 2[P(220)P(313)]_6 + 2[P(221)P(313)]_6 + 2[P(222)P(331)]_3+$$
$$2[P(223)P(331)]_3 + 2[P(230)P(312)]_6 + 2[P(231)P(313)]_6 + 2[P(232)P(320)]_6 + 2[P(302)P(322)]_6 + 2[P(320)P(323)]_6 + 2[P(330)P(332)]_3 + 3[P(000)P(003)]_3+$$
$$3[P(010)P(301)]_6 + 4[P(001)P(131)]_6 + 4[P(002)P(020)]_6 + 4[P(002)P(133)]_6 + 4[P(002)P(323)]_6 + 4[P(010)P(123)]_6 + 4[P(013)P(212)]_6 + 4[P(013)P(312)]_6+$$
$$4[P(023)P(221)]_6 + 4[P(023)P(222)]_6 + 4[P(023)P(322)]_6 + 4[P(031)P(211)]_6 + 4[P(032)P(321)]_6 + 4[P(100)P(123)]_6 + 4[P(100)P(232)]_6 + 4[P(100)P(313)]_6+$$
$$4[P(112)P(310)]_6 + 4[P(122)P(203)]_6 + 4[P(122)P(302)]_6 + 4[P(130)P(222)]_6 + 4[P(130)P(223)]_6 + 4[P(222)P(310)]_6 + 4[P(223)P(320)]_6 + 4[P(231)P(301)]_6+$$
$$4[P(312)P(330)]_6 + 6[P(001)P(031)]_6 + 6[P(001)P(033)]_6 + 6[P(002)P(300)]_6 + 6[P(002)P(330)]_3 + 6[P(003)P(032)]_6 + 6[P(003)P(131)]_6 + 6[P(003)P(132)]_6+$$
$$6[P(011)P(300)]_3 + 6[P(011)P(320)]_6 + 6[P(012)P(200)]_6 + 6[P(012)P(301)]_6 + 6[P(013)P(030)]_6 + 6[P(013)P(110)]_6 + 6[P(013)P(120)]_6 + 6[P(013)P(303)]_6+$$
$$6[P(020)P(102)]_6 + 6[P(020)P(103)]_6 + 6[P(020)P(123)]_6 + 6[P(020)P(202)]_3 + 6[P(020)P(203)]_6 + 6[P(020)P(311)]_6 + 6[P(020)P(322)]_6 + 6[P(020)P(330)]_6+$$
$$6[P(022)P(303)]_6 + 6[P(030)P(033)]_6 + 6[P(030)P(101)]_3 + 6[P(030)P(133)]_6 + 6[P(030)P(202)]_3 + 6[P(030)P(303)]_3 + 6[P(030)P(332)]_6 + 6[P(031)P(203)]_6+$$
$$6[P(032)P(310)]_6 + 6[P(033)P(101)]_6 + 6[P(033)P(130)]_6 + 6[P(033)P(200)]_3 + 6[P(033)P(212)]_6 + 6[P(033)P(220)]_6 + 6[P(033)P(222)]_3 + 6[P(033)P(230)]_6+$$
$$6[P(033)P(322)]_3 + 6[P(100)P(203)]_6 + 6[P(101)P(130)]_6 + 6[P(103)P(310)]_6 + 6[P(113)P(130)]_6 + 6[P(113)P(230)]_6 + 6[P(113)P(330)]_3 + 6[P(122)P(330)]_6+$$
$$6[P(130)P(313)]_6 + 6[P(132)P(303)]_6 + 6[P(133)P(303)]_6 + 6[P(133)P(320)]_6 + 6[P(200)P(203)]_6 + 6[P(201)P(230)]_6 + 6[P(203)P(231)]_6 + 6[P(223)P(300)]_6+$$
$$8[P(003)P(320)]_6 + 8[P(032)P(300)]_6$$

Note: $P(abc)$ is shorthand for $P_{ABC}(abc)$ and $[P(113)P(330)]_3$ is shorthand for the sum over permutations:

$$[P(113)P(330)]_3 = P(113)P(330) + P(131)P(303) + P(311)P(033)$$