

# Causal Compatibility Inequalities Admitting of Quantum Violations in the Triangle Structure

Thomas C. Fraser\*

*Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada  
University of Waterloo, Waterloo, Ontario, Canada*

(Dated: March 8, 2017)

It has long been recognized that certain quantum correlations are often incompatible with particular assumption about classical causal structure. Such correlations are sometimes termed nonlocal, although it would be more accurate to say simply nonclassical, as the presence of correlations certifies the genuinely nonclassical nature of a causal setup in a device-independent fashion. Whenever a quantum correlation violates a causal compatibility inequality (e.g. a Bell inequality) then the correlation is witnessed as nonclassical. Any constraint which is satisfied by all correlations which *are* classically compatible with a given causal scenario defined a causal compatibility criterion. Such criteria were recently derived for the Triangle structure [[arXiv:1609.00672](https://arxiv.org/abs/1609.00672)] in the form of polynomial inequalities, begging the question: do any of those inequalities admit violation by quantum correlations? Numerical investigation suggests not, and we further conjecture that the set of correlations admitted by the classical triangle structure is equivalent to the set of correlations admitted by its quantum generalization whenever the three observable variables are binary. Our main contribution in this work, however, is the derivation of new causal compatibility inequalities for the Triangle structure which *do* admit quantum violation. We conclude by considering the possibility of quantum resources potentially qualitatively different from those known previously.

---

\* [tcfraser@tcfraser.com](mailto:tcfraser@tcfraser.com)

# CONTENTS

|  |    |
|--|----|
| I. Introduction  | 3  |
| II. Causal Compatibility   | 4  |
| A. Marginal Models   | 4  |
| B. Causal Structure  | 4  |
| C. Compatibility   | 5  |
| D. Inequalities  | 6  |
| III. Triangle Structure  | 6  |
| IV. Fritz Distribution   | 7  |
| V. Inflation Technique   | 10 |
| VI. Deriving Inequalities  | 12 |
| A. The Marginal Problem as a Linear Program                        | 14 |
| 1. Events & Outcomes   | 14 |
| 2. Linear Programs   | 15 |
| B. Triangle Structure Inequalities                                 | 18 |
| C. Symmetric Inequalities  | 20 |
| VII. Violations & Noise  | 23 |
| A. Noise   | 23 |
| B. Numerical Optimizations   | 24 |
| C. Parameterizing Quantum Distributions                            | 25 |
| 1. Unitary Group   | 26 |
| 2. Measurements  | 27 |
| 3. States  | 27 |
| 4. Permutation Matrix  | 28 |
| D. Numerical Optimization Results                                  | 29 |
| VIII. Conclusions  | 31 |
| Acknowledgments  | 32 |
| References   | 33 |
| A. Common Terminology & Notation                                   | 35 |
| 1. Graph Theory  | 35 |
| 2. Probability Theory  | 35 |
| B. Computationally Efficient Parameterization of the Unitary Group | 36 |
| C. Large Triangle Structure Inequalities                           | 37 |
| 1. Web Inflation Inequality  | 37 |
| 2. Symmetric Web Inflation Inequality                              | 38 |

## I. INTRODUCTION

In recent decades, the technical utility of quantum mechanics has become abundantly clear. In the realm of computation, quantum algorithms – such as Shor’s algorithm [1] and numerous others [2] – scale exponentially better than their classical counterparts. In the realm of secure communication, quantum protocols — a popular example being Quantum Key Distribution [3] – are able to provide privacy even against hypothetical adversaries with unlimited computational power, a desideratum which classical protocols are unable to fulfill. History, every opportunity of quantum phenomena which fails to be emulated by classical resources has led to a practical exploitation of the nonclassical feature in order to solve a some computational of communicational problem [4]. Motivated by past successes, a primal objective of modern quantum information theory is the discovery of new situations wherein quantum mechanics offers an advantage, and certification that the quantum advantage is genuine.

From a foundational perspective, the most robust demonstrations of quantum phenomena with no classical emulation have involved Bell inequalities [5]. Originally, Bell inequalities were derived as a way to show that no hidden variable theory could ever account for quantum mechanics; in this sense Bell inequalities are a response to the famous EPR paradox [6]. The enumeration of Bell inequalities has since become a widespread systematic way of demonstration the *nonclassicality* of a given observation. More recently it has been appreciated that Bell inequalities can be understood as consequences of **causal inference** [7]. Causal inference is concerned with classifying observations into those which can and cannot be explained by a hypothesized causal structure. The abstract nature of causal inference is responsible for its presence in numerous scientific fields including machine learning and biology [8, 9]. Causal compatibility inequalities, such as Bell of Instrumental inequalities, characterize the space of observations that are compatible with a hypothesized causal structure, albeit the characterization offered by practically derivable inequalities often only an approximation. To derive the traditional Bell inequalities from causal inference one starts with a (classical) causal structure known as the Bell structure, as depicted in Fig. 1. The fundamental Bell structure involves noncommunicating parties making measurements on some hidden shared resource  $\lambda$ , where the measurement outcomes  $(A, B)$  are presumed to be stochastic functions of the the local choices of measurement settings  $(S_A, S_B)$  and the shared resource  $\lambda$ . Quantum nonclassicality in the Bell structure has been thoroughly studied since Bell’s original work [10]. More complex scenarios, however, such as the correlation scenarios proposed by Fritz [11, 12], are much less understood. Here we investigate one particular causal scenario: the Triangle structure (Fig. 2).

The Triangle structure (Fig. 2) is a causal structure composed of 3 parties labeled  $A, B, C$  arranged in a triangular configuration while pair-wise sharing hidden/latent variables  $X, Y, Z$ . It has been extensively studied previously (see [13, Fig. 1], [14, Fig. 6], [15, Fig. 8], [16, Fig. 8, App. E], [11, Fig. 3], [17, Fig. 1], ...). An overview of some milestone results is provided in Section III. Identifying causal compatibility inequalities for this configuration has been seen as particularly challenging Branciard *et al.* [15]. Further identifying causal compatibility inequalities of such high resolution that such inequalities can be violated by quantum-accessible distributions has remained out-of-reach for the Triangle structure.

This work is the first, therefore, to find causal compatibility inequalities for the Triangle structure that are able to be violated by quantum-accessible distributions. This accomplishment was made possible through the combination of two previous developments. The first, an insight made by Fritz [11] regarding the ability to re-interpret the Bell structure as a portion of the Triangle structure. The second, a new framework for solving causal inference problems developed by Wolfe *et al.* [17] called **The Inflation Technique**. Ultimately, this work serves as a validation that the Inflation Technique is sensitive enough to offer new insights into quantum nonclassicality. Moreover, these new inequalities offer an avenue for recognizing previously unknown forms of nonclassicality. The authors attempts to find such novel resources were met with only partial success, suggesting clear refinements for future exploration.

The manuscript is organized into the following sections: Section II recalls important notions from causal inference theory and sets up the notation to be used. Section III discusses the Triangle structure and provides an overview of existing research; identifying its stark differences from the Bell structure and motivating why the Triangle structure was chosen as a case study. Section IV defines and discusses a singularly-quantum correlation first conceived of by Fritz [11], which we term the **Fritz distribution**. The Fritz distribution was proven to be non-classical without the use of inequalities — this report fills in the desired missing inequalities. Section V briefly summarizes the Inflation Technique in the specific context of this work. Although the summary presented here is designed to be self-standing, a much more pedagogical introduction is offered by the original work [17]. Section VI demonstrates how the inflation technique can be used to derive causal compatibility inequalities for the Triangle structure that are violated by quantum-accessible distributions. A sample of these inequalities are presented at the end of Section VI as well as in Appendices C1

and C 2. Finally, Section VII introduces the method for parameterizing the space of quantum-accessible distributions which we used to numerically optimize against our derived inequalities in search of stronger types of non-classicality. Additionally, the Fritz distribution is subjected to noise in order to measure the robustness of the derived inequalities.

## II. CAUSAL COMPATIBILITY

The task of causal inference aims to determine the set of potentially observable probability distribution compatible with some hypothesis about causal relationships. 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  $\mathcal{J} = \{v_1, \dots, v_k\}$  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  $P_{\mathcal{J}} = P_{v_1, \dots, v_n}$ . More generally however, and depending on the context, it is possible to make a measurement on a set of the system  $V \subset \mathcal{J}$  in order to obtain a marginal distribution  $P_V$ . With complete generality however, one could in principle be limited so as to only be able to record observations on portions of the system  $\mathcal{M} = \{V_1, \dots, V_m \mid \forall i : V_i \subseteq \mathcal{J}\}$  obtaining a collection of marginal probability distributions  $\{P_{V_1}, \dots, P_{V_m} \mid \forall i : V_i \in \mathcal{M}\}$ .

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 [18], 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 exclusively consider (without loss of generality) the *maximal* marginal scenario; restricting our focus to the largest contexts in the marginal scenario. Additionally, all marginal scenarios are taken to be complete in the sense that the marginal scenario covers the complete set of observable variables  $\mathcal{J}$ .

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

### B. Causal Structure

A hypothesis of causal mechanism is formally referred to as a **causal structure**. Mathematically, a causal structure can be represented as 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 structure has 3 observable nodes and 3 latent nodes,

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

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

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

### 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 absence of certain causal connections within  $\mathcal{G}$  should restrict and limit the set of possible observations that can be made on  $\mathcal{G}$ ; equivalently, the causal connections within  $\mathcal{G}$  should explain  $P^\mathcal{M}$ . If each variable  $n \in \mathcal{N}$  is caused by its parents  $\text{Pa}_\mathcal{G}(n)$ , then  $n$  should be independent of all other variables that are not its descendants. This is known as the **local Markov property**:

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

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:

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

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 causal parameters:<sup>1</sup>

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

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  $\mathcal{N}_O$ :

$$P_{\mathcal{N}_O} = \sum_{\mathcal{N}_L} P_\mathcal{N} \quad (4)$$

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

$$P_V = \sum_{\mathcal{N}_O \setminus V} P_{\mathcal{N}_O} \quad (5)$$

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**.

---

<sup>1</sup> Here  $\prod$  denotes the product distribution  $\prod_{n \in \mathcal{N}} P_{n|\text{Pa}_\mathcal{G}(n)} = P_{n_1|\text{Pa}_\mathcal{G}(n_1)} \times \cdots \times P_{n_k|\text{Pa}_\mathcal{G}(n_k)}$ .

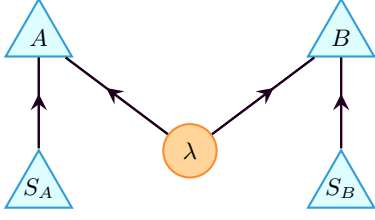


FIG. 1: The Bell structure consisting of two observers  $A, B$  together with measurement settings  $S_A$  and  $S_B$  respectively. The shared hidden variable is labeled  $\lambda$ .

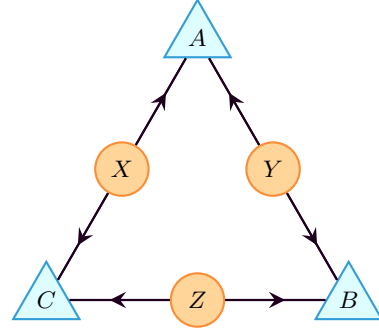


FIG. 2: The casual structure of the Triangle structure. Three variables  $A, B, C$  are observable, while  $X, Y, Z$  are latent variables.

#### 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**<sup>2</sup> 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 structure.

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.

### III. TRIANGLE STRUCTURE

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

Following the definition of causal compatibility from Section II, a distribution  $P_{N_O} = P_{ABC}$  is compatible with the Triangle structure if and only if there exists a choice of causal parameters  $\{P_{A|X,Y}, P_{B|Y,Z}, P_{C|Z,X}, P_X, P_Y, P_Z\}$  such that  $P_{ABC}$  is a marginalization of  $P_{ABCXYZ}$  over  $X, Y, Z$ :

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

<sup>2</sup> We refer to these inequalities as “causal compatibility inequalities” instead of simply Bell inequalities for two reasons. Firstly, “Bell inequalities” usually are associated specifically with the Bell structure/Bell causal structure. Secondly, the inequalities derived in this work are fundamentally distinct from a typical Bell inequality in that these inequalities are *polynomial* over  $P^{\mathcal{M}}$  instead of *linear*.

For certain causal structures, the conditional independence relations provided by  $d$ -separation relations are a sufficient characterization of compatibility. Recently, Henson *et al.* [16], classified the Triangle structure as an *interesting* causal structure in the sense that conditional independence relations *do not* provide a sufficient characterization of non-classicality.<sup>3</sup> Moreover, Wolfe *et al.* [17] concretely demonstrate that the space of classical distributions on the Triangle structure is non-convex, unlike the Bell structure. It can be argued that convexity of the Bell structure distributions is partially responsible for the wealth of knowledge that is known about the Bell structure. It is also reasonable to speculate that quantum non-classicality in the triangle can be translated directly into a computational advantage for certain computational circuits [19]. Characterizing quantum non-classicality for interesting causal structures remains an unsolved problem, and the Triangle structure is a small, non-trivial case study.

Additionally, Fritz [11] demonstrated that the Triangle structure is the *smallest* correlation scenario<sup>4</sup> in which there 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-classicality in the Triangle structure. **Comment (Elie Wolfe):** Revisions processed up to here thusfar.

Finally Wolfe *et al.* [17] 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 structure and the  $W$ -type distribution; a quantum non-accessible distribution whose incompatibility is not witnessable by *any* entropic inequalities or any other known constraints. Therefore as an algorithmic framework, the Inflation Technique comes closer to characterizing non-classicality in the Triangle structure than all others.

It is worth noting that a number of causal compatibility inequalities for the Triangle structure have been derived in previous works. For example, Steudel and Ay [13] derived an inequality for a series of causal structures with  $n \in \mathbb{N}$  variables where at most  $c \in \mathbb{N}$  share a common ancestor. The Triangle structure is a special case with  $c = 2$ ,  $n = 3$ . Moreover, Henson *et al.* [16] derive a family of entropic inequalities constraining the set of compatible correlations. Most impressively, Wolfe *et al.* [17] derive a family of polynomial, 2-outcome, causal compatibility inequalities; 37 non-trivial representatives are printed in [17].

As a preliminary search for quantum incompatibility in the Triangle structure, we performed numerical optimizations using bipartite qubit density matrices and 2-outcome POVM measurements against the compatibility inequalities printed in [17] as well as the entropic inequalities presented by Henson *et al.* [16]. Unfortunately, none of these inequalities could be violated. These early results suggest, although do not prove, that quantum non-classicality does not exist in the Triangle structure for *two-outcome* measurements. We make *no claim* that this is a universal truth and simply take it as motivation to explore incompatibility in Triangle structure for a degree of outcomes.

In conclusion, the Triangle structure is a causal structure whose non-classicality remains an open problem. The Triangle structure 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 structure 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 [11], it is possible to construct quantum distributions incompatible with the Triangle structure by utilizing quantum distributions incompatible with the familiar Bell structure. To explain, imagine rearranging the triangle into the configuration depicted in Fig. 3 and contrast it with the Bell structure of Fig. 1. Evidently, under the correct relabeling, large portions of the Triangle structure resemble the Bell structure. The crucial distinction is that the measurement settings  $S_A, S_B$  are *observable variables* in the Bell structure but become *latent variables* in the Triangle structure. In order to embed distributions that are incompatible with the Bell structure into the Triangle structure, the latent nodes  $S_A, S_B$  need to be observable and independent of the shared

<sup>3</sup> Note that the Triangle structure possesses *zero* conditional independence relations over the observable nodes  $A, B, C$ .

<sup>4</sup> Correlation scenarios differ from the Bell structure in that measure setting variables are non-existent in correlation scenarios.

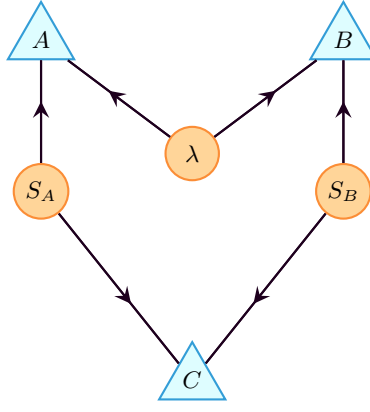


FIG. 3: The Triangle structure re-imagined to mimic the Bell structure. The measurement settings  $S_A, S_B$  are latent nodes unlike the Bell structure (Fig. 1).

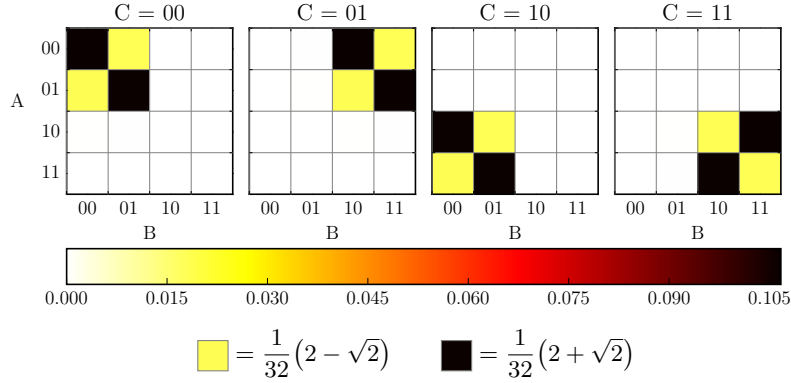


FIG. 4: The Fritz distribution visualized using a  $4 \times 4 \times 4$  grid. The 4 outcomes of  $A, B, C$  are written in binary as a doublet of bits to illustrate that certain bits act as measurement pseudo-settings.

latent node  $\lambda$ . This can be accomplished by having  $C$  “measure” the measurement settings  $S_A, S_B$  and announce them as an outcome. Consequently, any distribution over  $A, B, S_A, S_B$  that is incompatible with Bell structure is also incompatible with the Triangle structure provided that  $C$  is *perfectly correlated* with  $S_A, S_B$ . 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 Triangle structure [11]. In the Fritz distribution, each of the variables  $A, B, C$  have 4 possible outcomes. Explicitly,  $P_F$  can be written as:

$$\begin{aligned} P_F(000) = P_F(110) = P_F(021) = P_F(131) = P_F(202) = P_F(312) = P_F(233) = P_F(323) &= \frac{1}{32}(2 + \sqrt{2}) \\ P_F(010) = P_F(100) = P_F(031) = P_F(121) = P_F(212) = P_F(302) = P_F(223) = P_F(333) &= \frac{1}{32}(2 - \sqrt{2}) \end{aligned} \quad (7)$$

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  $2 \times 2$  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 structure defined over  $A, B, S_A, S_B$  can be directly converted to an inequality for the Triangle structure 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 [20],

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

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

$$\langle A_r B_r | C = 00 \rangle + \langle A_r B_r | C = 01 \rangle + \langle A_r B_r | C = 10 \rangle - \langle A_r B_r | C = 11 \rangle \leq 2 \quad (8)$$

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 [21],

$$3\left(\frac{1}{\sqrt{2}}\right) - \left(-\frac{1}{\sqrt{2}}\right) = 2\sqrt{2} \not\leq 2$$

Therefore, the Fritz distribution is incompatible with the Triangle structure 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  $\lambda$  in Fig. 3) is the maximally entangled Bell state:

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

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

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

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

$$\begin{aligned} M_A &= \{|0\psi_1\rangle\langle 0\psi_1|, |0\psi_5\rangle\langle 0\psi_5|, |1\psi_3\rangle\langle 1\psi_3|, |1\psi_7\rangle\langle 1\psi_7|\} \\ M_B &= \{|\psi_6 0\rangle\langle \psi_6 0|, |\psi_2 0\rangle\langle \psi_2 0|, |\psi_0 1\rangle\langle \psi_0 1|, |\psi_4 1\rangle\langle \psi_4 1|\} \\ M_C &= \{|00\rangle\langle 00|, |10\rangle\langle 10|, |01\rangle\langle 01|, |11\rangle\langle 11|\} \end{aligned}$$

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

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

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

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

Where  $\Pi$  is a  $64 \times 64$  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  $\rho_{AB}$  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 structure 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 [11]. 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 structure 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 structure. Fritz [11] presented the following problem: *Find an example of non-classical quantum correlations in the Triangle structure together with a proof of its non-classicality which does not hinge on Bell’s Theorem.*<sup>5</sup> 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 structure that are *not* reliant on perfect correlations.
- R.2** Find incompatible quantum distributions on the Triangle structure that satisfy all Bell structure 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.* [17] and inspired by the *do calculus* and *twin networks* of Pearl [8], is a family of causal inference techniques that can be used to determine if a 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 an example, inflations of the Triangle structure 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:

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

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.

---

<sup>5</sup> In the way Fritz defines, *non-classical* correlations are the class of *incompatible* correlations.

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  $P_{\mathcal{N}}$ ,

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

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

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

Which in turn uniquely defines a compatible joint distribution  $P_{\mathcal{N}'}$  on  $\mathcal{G}'$ <sup>6</sup>.

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

Together, Eqs. (10-12) 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  $\text{AnSub}_{\mathcal{G}}(N) \sim \text{AnSub}_{\mathcal{G}'}(N')$  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  $\mathcal{G}$ .

$$\begin{aligned} \text{Inj}_{\mathcal{G}'}(\mathcal{G}') &\equiv \{N' \subseteq \mathcal{N}' \mid \exists N \subseteq \mathcal{N} : N \sim N'\} \\ \text{ImInj}_{\mathcal{G}}(\mathcal{G}') &\equiv \{N \subseteq \mathcal{N} \mid \exists N' \subseteq \mathcal{N}' : N \sim N'\} \end{aligned}$$

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** [17, Lemma 3] *Given a particular inflation  $\mathcal{G}'$  of  $\mathcal{G}$  and a marginal model  $\{P_N \mid N \in \text{ImInj}_{\mathcal{G}}(\mathcal{G}')\}$  compatible with  $\mathcal{G}$  then all marginal models  $\{P_{N'} \mid N' \in \text{Inj}_{\mathcal{G}'}(\mathcal{G}')\}$  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:

$$\begin{array}{ccc} \underbrace{\{P_N \mid N \in \text{ImInj}_{\mathcal{G}}(\mathcal{G}')\}}_{\text{compatible with } \mathcal{G}} & \longrightarrow & \{P_{n|\text{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N}\} \\ & & \downarrow \text{define} \\ \underbrace{\{P_{N'} \mid N' \in \text{Inj}_{\mathcal{G}'}(\mathcal{G}')\}}_{\text{compatible with } \mathcal{G}'} & \longleftarrow & \{P_{n'|\text{Pa}_{\mathcal{G}'}(n')} \mid n' \in \mathcal{N}'\} \end{array}$$

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  $\{P_{N'} \mid N' \in \text{Inj}_{\mathcal{G}'}(\mathcal{G}')\}$  can be *deflated* into a valid compatibility constraint on marginal models  $\{P_N \mid N \in \text{ImInj}_{\mathcal{G}}(\mathcal{G}')\}$  of the original causal structure by

---

<sup>6</sup> 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.

dropping all references to the copy-indices. Additionally, the inflation lemma holds when considering any subset of  $\text{Inj}_{\mathcal{G}}(\mathcal{G}')$  (analogously  $\text{ImInj}_{\mathcal{G}}(\mathcal{G}')$ ). 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  $N' \notin \text{Inj}_{\mathcal{G}}(\mathcal{G}')$ . 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,  $N' \notin \text{Inj}_{\mathcal{G}}(\mathcal{G}')$  could be partitioned into two injectable subsets  $N'_1 \in \text{Inj}_{\mathcal{G}}(\mathcal{G}')$  and  $N'_2 \in \text{Inj}_{\mathcal{G}}(\mathcal{G}')$  that where **ancestrally independent** in  $\mathcal{G}'$ ,

$$N'_1 \perp N'_2 \iff \text{An}_{\mathcal{G}'}(N'_1) \cap \text{An}_{\mathcal{G}'}(N'_2) = \emptyset$$

Then the compatible marginal context  $P_{N'}$  must be probabilistically separable  $P_{N'} = P_{N'_1}P_{N'_2}$  due to Eq. (1). Since  $N'_1$  and  $N'_2$  are both injectable, the inflation lemma now applies and the original inequality can be deflated.

$$\begin{array}{ccc} P_{N'} & \xrightarrow[N'_1 \perp N'_2]{N'_1 \cup N'_2 = N} & P_{N'_1}P_{N'_2} \\ | & & | \\ N' \notin \text{Inj}_{\mathcal{G}}(\mathcal{G}') & & N'_1, N'_2 \in \text{Inj}_{\mathcal{G}}(\mathcal{G}') \\ \text{\textcolor{red}{X}} & & \downarrow \\ ?? & & P_{N'_1}P_{N'_2} \end{array}$$

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 sets  $N' = \coprod_i N'_i \mid N'_i \in \text{Inj}_{\mathcal{G}}(\mathcal{G}')$  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  $\text{PreInj}_{\mathcal{G}}(\mathcal{G}')$  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<sup>7</sup>.

## 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 IIC, 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

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

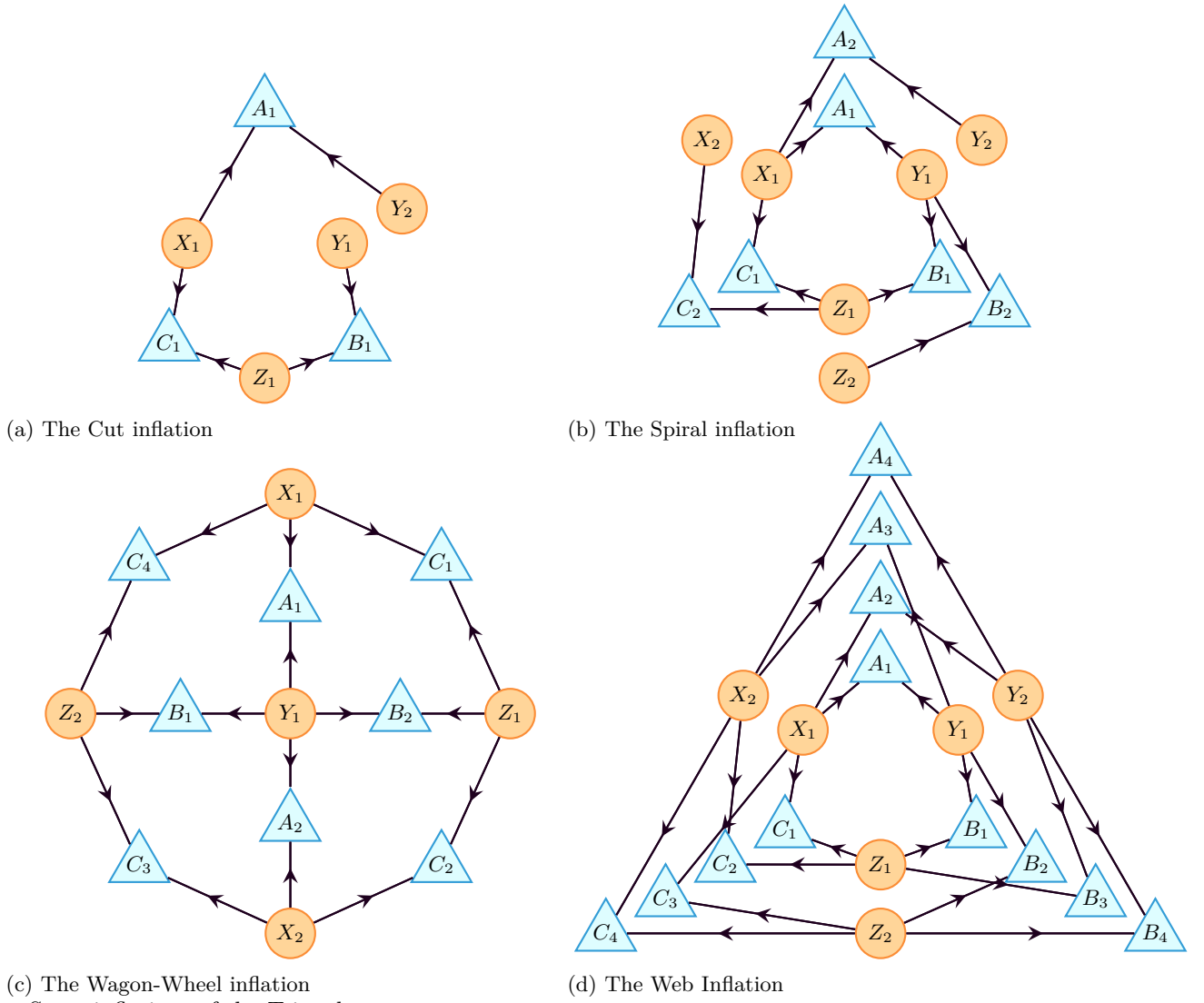


FIG. 5: Some inflations of the Triangle structure.

of operations. First choose a set of causal parameters  $\{P_{n|\text{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N}\}$  and then construct a joint distribution  $P_{\mathcal{N}} = \prod_{n \in \mathcal{N}} P_{n|\text{Pa}_{\mathcal{G}}(n)}$  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. (4,5)). 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. (4,5)).
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  $P_{\mathcal{J}}$ ?

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

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 [22], database theory, knowledge integration of expert system, and of course, quantum information theory [18].

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 VIA). 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  $v \in \mathcal{J}$  [18]. Entropic inequalities however have been shown to be less constraining than inequalities defined over variables with finite cardinality [17]. 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**  $v$   $\mathcal{S}(v)$ <sup>8</sup> to be the set of all functions  $s : \{v\} \rightarrow O_v$  each representing the event that a measurement on  $v$  was made where  $s(v) \in O_v$  was observed.

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

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 [23] demands the distinction. Specifically for this work, the distinction becomes essential for our discussion and exploit

---

<sup>8</sup> In the language of sheaf-theory,  $\mathcal{S}(v)$  is the *sheaf of events* [23].

of marginal symmetries in Section **VIC**. As a natural generalization we define the event over a collection of random variables  $V = \{v_1, \dots, v_n\}$  in a parallel manner,

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

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

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

The domain  $\mathcal{D}(s)$  of an event  $s$  is the set of random variables it values, 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  $\mathcal{M} = \{V_1, \dots, V_k\}$ , 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  $j \in \mathcal{S}(\mathcal{J})$ .

## 2. Linear Programs

The marginalization operation (Eq. (13)) of the marginal problem is a linear operator. This linear operator maps a joint probability distribution  $P_{\mathcal{J}} : \mathcal{S}(\mathcal{J}) \rightarrow [0, 1]$  into a marginal model  $P^{\mathcal{M}} = \{P_V : \mathcal{S}(V) \rightarrow [0, 1] \mid V \in \mathcal{M}\}$ . 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  $\mathcal{M} = \{V_1, \dots, V_k\}$  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:

$$M = \begin{array}{c} \begin{array}{l} (A, B, C) \mapsto \\ (A \mapsto 0, B \mapsto 0) \\ (A \mapsto 0, B \mapsto 1) \\ (A \mapsto 1, B \mapsto 0) \\ (A \mapsto 1, B \mapsto 1) \\ (B \mapsto 0, C \mapsto 0) \\ (B \mapsto 0, C \mapsto 1) \\ (B \mapsto 1, C \mapsto 0) \\ (B \mapsto 1, C \mapsto 1) \\ (A \mapsto 0, C \mapsto 0) \\ (A \mapsto 0, C \mapsto 1) \\ (A \mapsto 1, C \mapsto 0) \\ (A \mapsto 1, C \mapsto 1) \end{array} \mapsto \begin{pmatrix} (0,0,0) & (0,0,1) & (0,1,0) & (0,1,1) & (1,0,0) & (1,0,1) & (1,1,0) & (1,1,1) \\ \mathbf{1} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{1} \\ \mathbf{1} & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & \mathbf{1} \end{pmatrix} \end{array} \quad (15)$$

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  $P^{\mathcal{M}} = \{P_{V_1}, \dots, P_{V_k}\}$  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  $\mathcal{P}^{\mathcal{J}}$ .

$$\mathcal{P}^{\mathcal{M}} = M \cdot \mathcal{P}^{\mathcal{J}} \quad (16)$$

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 [18, 24, 25]: *Given a marginal scenario  $\mathcal{M}$ , what is the set of all non-contextual marginal models?* Pitowsky [26] 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 [10]. 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** [17, 24, 27, 28]. Applying the Fourier-Motzkin procedure to completely solve the marginal problem is discussed in more detail in Fritz and Chaves [18]. An excellent survey of existing techniques for solving the marginal problem including Equality Set Projection [28] and Hardy-type hypergraph transversals can be found in Wolfe *et al.* [17]. In conclusion, there are a number of computational tools available to solve the marginal problem completely whenever no marginal model is provided.

Each of the above mentioned techniques suffers from computational complexity limitations. For example, the Fourier-Motzkin procedure is in the worst case doubly exponential in the number of initial inequalities [27]. For the purposes of this research, solving the marginal problem without reference to a marginal model was intractable. This will become apparent in Section VIB when the inflation technique is applied to the Triangle structure, 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 structure; 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”<sup>9</sup> 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  $\mathcal{P}^{\mathcal{J}}$ .

To every linear program, there exists a dual linear program that characterizes the feasibility of the original [29]. Constructing the dual linear program is straightforward [30].

**Definition 10. The Dual Marginal Linear Program:**

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

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

---

<sup>9</sup> “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*.

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:

$$y \cdot \mathcal{P}^{\mathcal{M}} = y \cdot M \cdot x \geq 0 \quad (18)$$

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.<sup>10</sup> This is manifestation of Farkas's lemma [29].

An **infeasibility certificate** [32] is any vector  $y$  that satisfies the constraint of Theorem 10 and also permits violation of Eq. (18) for some marginal distribution vector  $\mathcal{P}^{\mathcal{M}}$ .

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

Most linear program software packages such as Mosek [33], Gurobi [34], CPLEX [35] and CVX/CVX OPT [36, 37] 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 Structure 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 structure  $\mathcal{G}$  and identify the maximal pre-injectable sets  $\text{PreInj}_{\mathcal{G}}(\mathcal{G}')$ . 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 programming 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<sup>11</sup>, the maximal pre-injectable sets of the Wagon-Wheel inflation are as follows:

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

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

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

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

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

<sup>10</sup> 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 [31]. 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 [31]; if the dual is unbounded, then the primal is infeasible.

<sup>11</sup> Computing the injectable and pre-injectable sets computationally is possible and a technique is discussed in Wolfe *et al.* [17].

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,

$$\mathcal{P}^{\mathcal{M}\top} = \underbrace{(P_{A_2B_1C_3C_1}(0000), \dots, P_{A_1B_1C_4C_2}(3232), \dots, P_{A_2B_2C_2C_4}(3333))}_{1024 \text{ entries}} \quad (21)$$

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

$$\mathcal{P}^{\mathcal{M}\top} = (P_{A_2B_1C_3}(000)P_{C_1}(0), \dots, P_{A_1B_1C_4}(323)P_{C_2}(2), \dots, P_{A_2B_2C_2}(333)P_{C_4}(3)) \quad (22)$$

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

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

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.

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

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 programming software and an infeasibility certificate  $y$  was obtained. Below is an example of an infeasibility inequality obtained over the entries of Eq. (23).

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

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 structure 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,  $P_{ABC}(000)P_C(3) = \frac{1}{128} (2 + \sqrt{2}) \simeq 0.0267$  but all of the remaining terms sum to  $\simeq 0.0136$  meaning that the Fritz distribution violates this inequality with violation  $-0.0129 \dots \not\geq 0$ .

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,

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

There are a couple of important concepts to understand. First, the Fritz distribution is known to be incompatible with the Triangle structure 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 structure. 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 structure 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 entries. Using this incidence matrix, a number of incompatibility certificates were derived. For inspection, one is reproduced in Appendix C1.

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 programming software to certify the infeasibility of a *particular distribution*. This means that inequalities like Eq. (24), the one presented in Appendix C1, and others are specifically manufactured in order to 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 structure 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.* [38] 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 [20]) 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<sup>12</sup>.

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  $\varphi : \mathcal{J} \rightarrow \mathcal{J} : \{a, b, c, d\} \mapsto \varphi[\{a, b, c, d\}] = \{c, a, d, b\}$  is a permutation of the variables  $\mathcal{J}$ . Intuitively  $\varphi$  acts on  $I_{\mathcal{M}}$  in the following way,

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

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, it 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<sup>13</sup>  $\text{Perm}(\mathcal{J})$ . As was noted in Example 11, some permutations  $\varphi \in \text{Perm}(\mathcal{J})$  have the undesired possibility of transforming  $\mathcal{M}$  into a different marginal scenario.

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

This means that there could exist  $V \in \mathcal{M}$  such that  $\varphi[V] \notin \mathcal{M}$ . Permutations  $\varphi$  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**  $\Phi$  for a marginal scenario  $\mathcal{M}$  is the joint permutation subgroup that stabilizes the marginal scenario.

$$\Phi(\mathcal{M}) \equiv \{\varphi \in \text{Perm}(\mathcal{J}) \mid \forall V \in \mathcal{M} : \varphi[V] \in \mathcal{M}\}$$

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 [38], we divert our attention to finding *symmetric* inequalities themselves.

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

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,

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

<sup>12</sup> If the extremal distribution happens to not be symmetric, there is indication that the space of accessible distributions is non-convex.

<sup>13</sup> The permutation group over a set  $S$  is the set of all bijective maps  $\varphi : S \rightarrow S$ .

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  $\Phi$  in  $\mathcal{S}(\mathcal{M})$  and  $\mathcal{S}(\mathcal{J})$ .

$$\begin{aligned}\text{Orb}_\Phi(m) &\equiv \{\varphi[m] \mid \varphi \in \Phi\} \\ \text{Orb}_\Phi(j) &\equiv \{\varphi[j] \mid \varphi \in \Phi\}\end{aligned}$$

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**  $M_\Phi$  for a marginal scenario  $\mathcal{M}$  and the variable permutation group  $\Phi$  is a contracted version of the incidence matrix  $M$  for  $\mathcal{M}$ . Each row of  $M_\Phi$  corresponds to a marginal orbit  $\text{Orb}_\Phi(m)$ . Analogously each column of  $M_\Phi$  corresponds to a joint orbit  $\text{Orb}_\Phi(j)$ . The entries of  $M_\Phi$  are integers and correspond to summing over the rows and columns of  $M$  that belong to each orbit.

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

It is possible to analogously define a **symmetric joint distribution vector**  $\mathcal{P}_\Phi^\mathcal{J}$  indexed by  $\text{Orb}_\Phi(j)$ ,

$$(\mathcal{P}_\Phi^\mathcal{J})_{\text{Orb}_\Phi(j)} = \sum_{j' \in \text{Orb}_\Phi(j)} \mathcal{P}_{j'}^\mathcal{J}$$

And a **symmetric marginal distribution vector**  $\mathcal{P}_\Phi^\mathcal{M}$  indexed by  $\text{Orb}_\Phi(m)$ ,

$$(\mathcal{P}_\Phi^\mathcal{M})_{\text{Orb}_\Phi(m)} = \sum_{m' \in \text{Orb}_\Phi(m)} \mathcal{P}_{m'}^\mathcal{M}$$

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 structure, the variable permutation group is simply the set of permutations on  $A, B, C$ :  $\Phi(\mathcal{M}) = \text{Perm}(A, B, C)$ . For the Web inflation (Fig. 5d), we have obtained  $\Phi(\text{Prelnj}_G(\mathcal{G}'))$  which is an order 48 group with the following 4 generators:

| $\varphi_1$           | $\varphi_2$           | $\varphi_3$           | $\varphi_4$           |
|-----------------------|-----------------------|-----------------------|-----------------------|
| $A_1 \rightarrow A_4$ | $A_1 \rightarrow A_1$ | $A_1 \rightarrow C_1$ | $A_1 \rightarrow A_1$ |
| $A_2 \rightarrow A_3$ | $A_2 \rightarrow A_3$ | $A_2 \rightarrow C_2$ | $A_2 \rightarrow A_2$ |
| $A_3 \rightarrow A_2$ | $A_3 \rightarrow A_2$ | $A_3 \rightarrow C_3$ | $A_3 \rightarrow A_3$ |
| $A_4 \rightarrow A_1$ | $A_4 \rightarrow A_4$ | $A_4 \rightarrow C_4$ | $A_4 \rightarrow A_4$ |
| $B_1 \rightarrow B_4$ | $B_1 \rightarrow C_1$ | $B_1 \rightarrow A_1$ | $B_1 \rightarrow B_2$ |
| $B_2 \rightarrow B_3$ | $B_2 \rightarrow C_3$ | $B_2 \rightarrow A_2$ | $B_2 \rightarrow B_1$ |
| $B_3 \rightarrow B_2$ | $B_3 \rightarrow C_2$ | $B_3 \rightarrow A_3$ | $B_3 \rightarrow B_4$ |
| $B_4 \rightarrow B_1$ | $B_4 \rightarrow C_4$ | $B_4 \rightarrow A_4$ | $B_4 \rightarrow B_3$ |
| $C_1 \rightarrow C_4$ | $C_1 \rightarrow B_1$ | $C_1 \rightarrow B_1$ | $C_1 \rightarrow C_3$ |
| $C_2 \rightarrow C_3$ | $C_2 \rightarrow B_3$ | $C_2 \rightarrow B_2$ | $C_2 \rightarrow C_4$ |
| $C_3 \rightarrow C_2$ | $C_3 \rightarrow B_2$ | $C_3 \rightarrow B_3$ | $C_3 \rightarrow C_1$ |
| $C_4 \rightarrow C_1$ | $C_4 \rightarrow B_4$ | $C_4 \rightarrow B_4$ | $C_4 \rightarrow C_2$ |

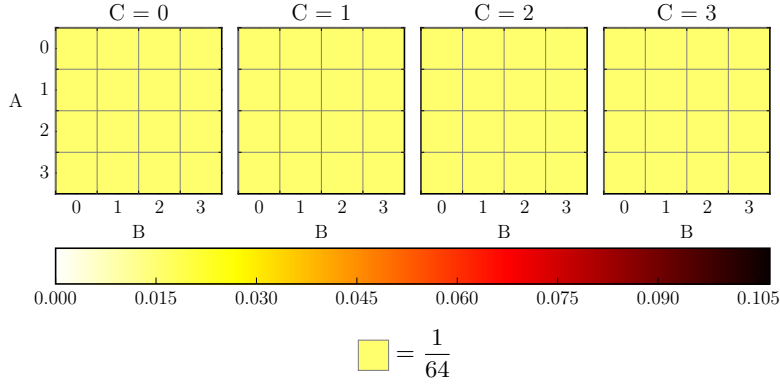


FIG. 6: The completely uniform distribution  $\mathcal{U}$  visualized using a  $4 \times 4 \times 4$  grid.

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,

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

To see this,  $\varphi_1$  and  $\varphi_4$  become the identity element  $\mathbb{1}$  in  $\text{Perm}(A, B, C)$  upon removal of the copy-index, leaving  $\varphi_2$  to generate reflections and  $\varphi_3$  to generate rotations.

The symmetric incidence matrix  $M_{\Phi}$  for the Web inflation is considerably smaller than  $M$ . The number of rows of  $M_{\Phi}$  is a number of distinct orbits  $\text{Orb}_{\Phi}(m)$  in  $\mathcal{S}(\mathcal{M})$ . Likewise the number of columns is the number of distinct orbits  $\text{Orb}_{\Phi}(j)$  in  $\mathcal{S}(\mathcal{J})$ . For the Web inflation in particular,  $M_{\Phi}$  is  $450 \times 358,120$ . Using the symmetric incidence matrix and linear programming methods, an infeasibility certificate was found that is capable of witnessing the Fritz distribution. This inequality is reproduced in Appendix C 2.

## VII. VIOLATIONS & NOISE

Causal compatibility inequalities for the Triangle structure (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

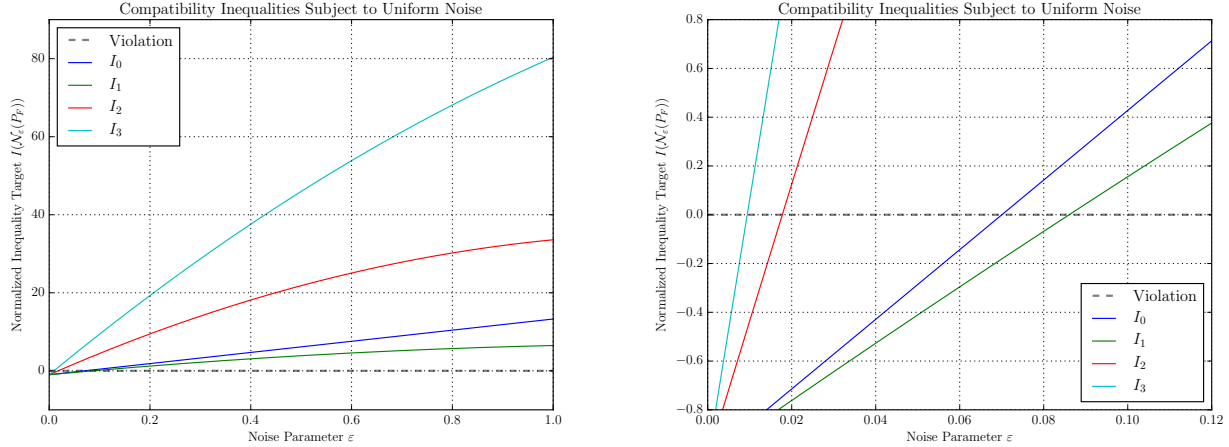
In his original work, Fritz [11] was able to find quantum non-classicality in the Triangle structure by embedding Bell non-classicality. As was mentioned in Section IV, this embedding required a very idealistic condition to hold: perfect correlations between  $C$  and  $A_i, B_i$ . From an experimental 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 structure, 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 structure, 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  $\varepsilon$ .

(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  $\varepsilon$  such that  $0 \leq \varepsilon \leq 1$  along with the **noisy variant**  $\mathcal{N}_\varepsilon$  of the Fritz distribution,

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

Therefore as  $\varepsilon$  varies from 0 to 1, more and more noise is added to the Fritz distribution as  $\mathcal{N}_\varepsilon$  transitions from an incompatible distribution  $\mathcal{N}_0 = P_F$  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 were 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 C1 while  $I_3$  is the symmetric inequality in Appendix C2. Inequality  $I_2$  is not reproduced anywhere in this document but is similar to Appendix C1.

The results of our experiments are plotted in Fig. 7. Figure 7a illustrates that each curve  $I(\mathcal{N}_\varepsilon(P_F))$  as a function of  $\varepsilon$  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 C1), when compared against all others is most robust to noise.

Overall, this noise model demonstrates that the Fritz distribution remains incompatible with the Triangle structure up to at least a noise parameter of  $\varepsilon \simeq 0.085$ . The associated distribution  $\mathcal{N}_{0.085}(P_F)$  is plotted in Fig. 8. There remains the possibility that another inequality will be able to withstand a larger degree of noise than  $I_1$ . 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



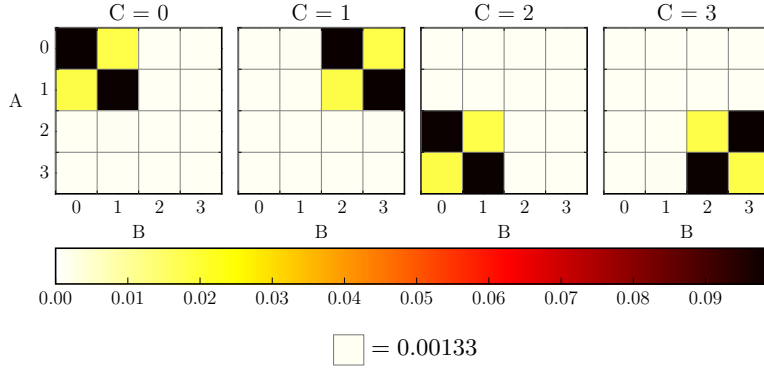


FIG. 8:  $\mathcal{N}_{0.085}$ : A noisy yet still non-classical variant of the Fritz distribution.

more non-classical than  $P_2$  whenever  $I(P_1) < I(P_2)$ <sup>14</sup>. Nonetheless, this statement comes from several caveats. First, 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 respect 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 structure, 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 structure 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 structure.

In order to find quantum distributions that are more non-classical than the Fritz, we performed numerical optimizations against a objective scalar function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  using the following pipeline: To begin, a set of real-valued parameters  $\lambda = (\lambda_0, \dots, \lambda_n) \in \mathbb{R}^n$  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 structure 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  $\lambda$  in order to achieve largest and larger violations.

Finding larger violations to compatibility inequalities is nothing more than 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 [39, p. 142] and the Nelder-Mead simplex method [39, p. 238] were used along with a hybrid method called Basin Hopping [40].

### C. Parameterizing Quantum Distributions

In search of new, incompatible, quantum distributions that can be realized on the Triangle structure, numerical optimizations over 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:

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

Where  $\rho_{AB}, \rho_{BC}, \rho_{CA}$  are bipartite density matrices,  $M_A, M_B, M_C$  are generic measurements sets and  $\Pi$  is a permutation matrix that aligns the states and measurements appropriately<sup>15</sup>.

<sup>14</sup> Remember that larger violations are more negative when the inequalities are written in the proposed canonical form.

<sup>15</sup>  $\Pi$  is discussed more thoroughly in Section VII C 4.

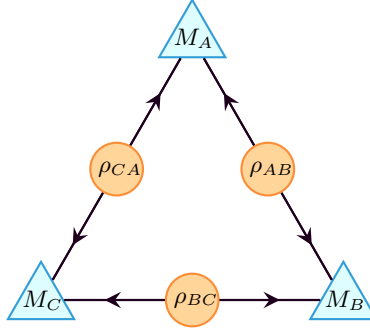


FIG. 9: The Triangle structure as modeled by quantum states and measurements.

In order to qualify the scope of Eq. (27) and associated computational complexity of the parameterization, there are two restrictions that are made with justification. The states  $\rho$  are taken to be bipartite *qubit* states acting on  $\mathcal{H}^2 \otimes \mathcal{H}^2$ . 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 structure. Second, generating  $k$ -outcome POVM measurements is possible using rejection sampling techniques [41], 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:

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

In order to parameterize all such distributions, we elect to parameterize the states and measurements separately. Although there are numerous techniques that can be used when parameterizing quantum states and measurements [4, 41–45], 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 and PVMs, a parameterization of the unitary group of dimension  $d$  (denoted  $\mathcal{U}(d)$ ) by Spengler, Huber and Hiesmayr [46] 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. [46], Spengler *et al.* proved that all unitaries  $U \in \mathcal{U}(d)$  can be parameterized without degeneracy as follows:

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

Where  $\lambda = \{\lambda_{n,m} \mid n, m \in 1, \dots, d\}$  form a  $d \times d$  matrix of real-valued parameters with periodicities  $\lambda_{m,n} \in [0, \frac{\pi}{2}]$  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,

$$P_l = |l\rangle\langle l| \quad 1 \leq l \leq d \quad (30)$$

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

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

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  $\{|1\rangle, \dots, |d\rangle\}$ .

$$\begin{aligned} GP_l &= \exp(iP_l \lambda_{l,l}) \\ RP_{n,m} &= \exp(iP_n \lambda_{n,m}) \\ R_{m,n} &= \exp(i\sigma_{m,n} \lambda_{m,n}) \end{aligned} \quad (31)$$

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  $l = 1, \dots, d$ . 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:

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

Another attractive feature of the SSH parameterization not mentioned in [46] 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)  $M = \{M_1, \dots, M_d\}$  satisfying a number of familiar constraints:

$$\begin{aligned} \forall |\phi\rangle \in \mathcal{H}^d : \langle \phi | M_i | \phi \rangle &\geq 0 & \sum_{i=1}^d M_i &= \mathbb{1}_d \\ M_i &= |m_i\rangle\langle m_i| & M_i M_j &= \delta_{ij} M_i \end{aligned}$$

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

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

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).

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

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 Vértési [47].

## 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  $4 \times 4$  positive semi-definite hermitian matrices with unitary

trace.

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

Although is common to parameterize all such matrices using a Cholesky decomposition [45], we make use of the non-degenerate SHH parameterization [46] and the spectral decomposition of  $\rho$ . Retaining full generality, consider full-rank density matrices:

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

Where  $\{p_i\}$  are the eigenvalues of  $\rho$  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  $\lambda = (\lambda_1, \dots, \lambda_{d-1})$ ,  $\lambda_i \in [0, 2\pi]$  using hyper-spherical coordinates [42, 46]:

$$\begin{aligned} p_n &= \prod_{i=1}^{n-1} \sin^2 \lambda_i \\ p_j &= \cos^2 \lambda_j \prod_{i=1}^{j-1} \sin^2 \lambda_i \quad \forall j \in 1, \dots, n-1 \end{aligned} \quad (36)$$

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 parameterize 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  $\lambda = (\lambda_1, \dots, \lambda_d)$ , then forcing normalization by dividing each parameter  $\lambda_i$  by the their sum.<sup>16</sup>

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

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  $\Pi$  for the Triangle structure. For bipartite qubit states,  $\Pi$  is the  $64 \times 64$  bit-wise matrix depicted in Fig. 10. Upon examination of Eq. (28), we consider  $\Pi$  to be acting on the joint

---

<sup>16</sup> Strictly speaking, Eq. (37) *also* suffers from non-uniformity; being biased toward eigenvalues that have similar values to each other.

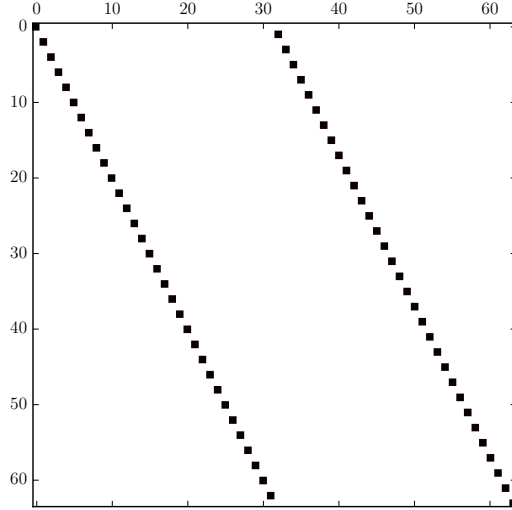


FIG. 10: The permutation matrix  $\Pi$  acting on  $(\mathcal{H}^2)^{\otimes 6}$  for the Triangle structure. Black entries represents a value of 1 and white represents 0.

global state  $\rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA}$ . To illuminate its necessity, consider Eq. (27) in the absence of  $\Pi$ .

$$\begin{aligned}
 P_{ABC}(abc) &\stackrel{?}{=} \text{Tr}[(\rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA})(M_A^a \otimes M_B^b \otimes M_C^c)] \\
 &= \text{Tr}[(\rho_{AB} M_A^a) \otimes (\rho_{BC} M_B^b) \otimes (\rho_{CA} M_C^c)] \\
 &= \text{Tr}(\rho_{AB} M_A^a) \text{Tr}(\rho_{BC} M_B^b) \text{Tr}(\rho_{CA} M_C^c) \\
 &= P_{A|\rho_{AB}}(a) P_{B|\rho_{BC}}(b) P_{C|\rho_{CA}}(c)
 \end{aligned}$$

On an operational level, this corresponds to  $A$  making a measurement on *both* subsystems of  $\rho_{AB}$  and *not* on any component of  $\rho_{CA}$ . This is analogously troubling for  $B$  and  $C$  as well. To resolve this issue, the permutation matrix  $\Pi$  corresponds to *aligning* the underlying 6-qubit joint state  $\rho$  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  $\forall i : |q_i\rangle \in \mathcal{H}^2$ .

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

$\Pi$  acts as a *partial transpose* on  $(\mathcal{H}^2)^{\otimes 6}$  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  $(\mathcal{H}^2)^{\otimes 6}$ ,

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

#### 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 structure (Fig. 9). Specifically, we make use of fully generic  $4 \times 4$  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

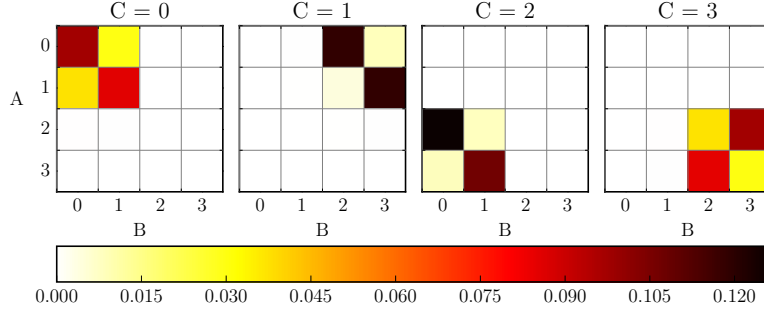
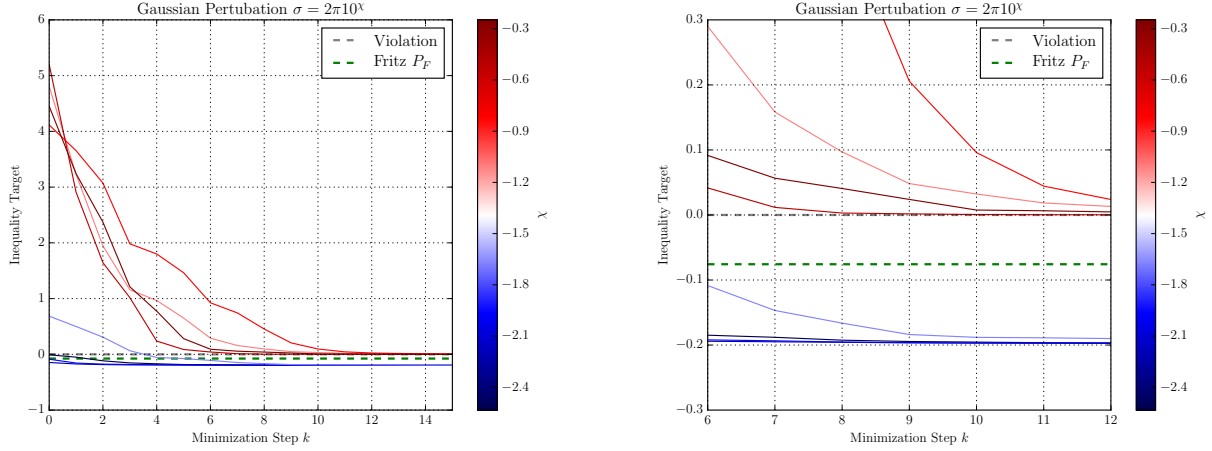


FIG. 11: Probability distribution that maximizes violation of  $I_1$  from Appendix C 1 ( $V_F \simeq 1.501$ ).



(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.

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 structure? This result is not entirely unexpected. As was mentioned in Section VIB, these inequalities are derived specifically to show incompatibility of the Triangle structure. 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 sensitivity of the initial condition  $\lambda_{(0)} = (\lambda_{(0),0}, \dots, \lambda_{(0),n})$  in our optimization. In Section VII C, it was shown that our parameter space of choice was  $2\pi$ -periodic in all entries. Somewhat surprisingly, when using a uniform seed for  $\lambda_0$ , all optimization methods consistently converged to inequality saturation instead of inequality violation which of course is achievable by the Fritz distribution. Instead, only initial seeds close enough to the *Fritz parameters*  $\lambda_{(F)}$  (the parameters that induce the Fritz distribution using the quantum implementation in Section IV) were able to achieve violation. To illustrate this graphically, we let the initial seed be a small Gaussian perturbation  $\delta\lambda$  away from  $\lambda_{(F)}$  in parameter space.

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

The Gaussian perturbation was given zero mean  $\mu = 0$  and variance  $\sigma^2 = (2\pi 10^\chi)^2$  where  $\chi$  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.

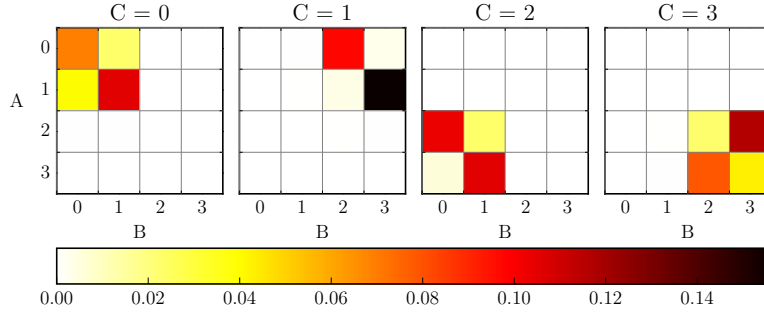


FIG. 13: Probability distribution that maximizes violation of  $I_3$  from Appendix C 2 ( $V_F \simeq 2.61$ ).

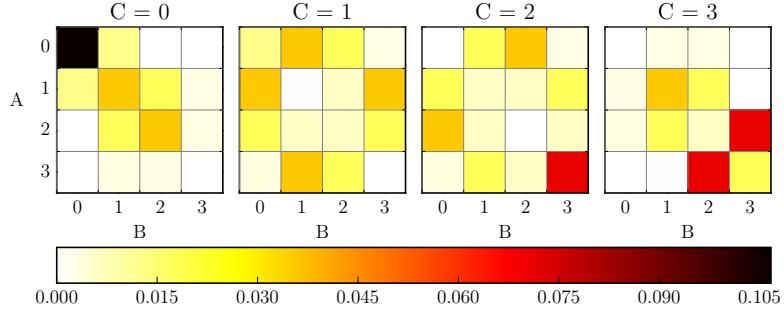


FIG. 14: Symmetrized version of the Fritz distribution ( $V_F \simeq 3$ ).

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  $\varphi \in \text{Perm}(A, B, C)$ , one would naively expect that the maximum violating distribution 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 it cannot be written as Eq. (27)<sup>17</sup>. Therefore we conclude that the space of quantum accessible correlations on the Triangle structure is non-convex; a reiteration of a result found in Wolfe *et al.* [17].

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 it is 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 structure has been an unsolved problem for nearly a decade. Initially, Branciard *et al.* [15] identified the Triangle structure as a challenging causal structure for studying non-classicality. Afterwards, Henson *et al.* [16] were able to classify the Triangle structure as *interesting*

<sup>17</sup> At least numerical optimizations suggest this is the case.

which partially explains why it remains an enigma. Recently, Fritz [11] discovered the Fritz distribution as the first example of quantum non-classicality in the Triangle structure. Heretofore, researchers have been able to derive causal incompatibility inequalities [13, 16, 17] but unable to show that any are non-trivial in the context of quantum non-classicality. In Section IV the Fritz distribution was introduced in detail and refinements were made to a problem originally proposed by Fritz [11]. 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. Finally, we used those inequalities in an attempt to discover new forms of non-classicality in the Triangle structure 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 structure – 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 [11, 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

T.C. Fraser would like to thank E. Wolfe for his excellent guidance and mentor-ship throughout the duration of this research project. Moreover, T.C. Fraser would like to thank the Perimeter Institute for financing this research and providing access to much needed computational resources.



- 
- [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] John S Bell, “On the Einstein-Podolsky-Rosen paradox,” *Physics* **1**, 195–200 (1964).
  - [6] A. Einstein, B. Podolsky, and N. Rosen, “Can quantum-mechanical description of physical reality be considered complete?” *Phys. Rev.* **47**, 777–780 (1935).
  - [7] 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,” *New Journal of Physics* **17**, 033002 (2012), [arXiv:1208.4119](#).
  - [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] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner, “Bell nonlocality,” *Reviews of Modern Physics* **86**, 419–478 (2013), [arXiv:1303.2849](#).
  - [11] Tobias Fritz, “Beyond Bell’s Theorem: Correlation Scenarios,” *New Journal of Physics* **14**, 103001 (2012), [arXiv:1206.5115](#).
  - [12] Tobias Fritz, “Beyond Bell’s Theorem II: Scenarios with arbitrary causal structure,” *Communications in Mathematical Physics* **341**, 391–434 (2014), [arXiv:1404.4812](#).
  - [13] Bastian Steudel and Nihat Ay, “Information-theoretic inference of common ancestors,” (2010), [arXiv:1010.5720](#).
  - [14] Rafael Chaves, Lukas Luft, and David Gross, “Causal structures from entropic information: geometry and novel scenarios,” *New Journal of Physics* **16**, 043001 (2014).
  - [15] 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](#).
  - [16] 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](#).
  - [17] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz, “The inflation technique for causal inference with latent variables,” (2016), [arXiv:1609.00672](#).
  - [18] Tobias Fritz and Rafael Chaves, “Entropic inequalities and marginal problems,” *IEEE Transactions on Information Theory* **59**, 803–817 (2011), [arXiv:1112.4788](#).
  - [19] Barbara M. Terhal and David P. DiVincenzo, “Adaptive quantum computation, constant depth quantum circuits and arthur-merlin games,” [quant-ph/0205133v6](#).
  - [20] 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).
  - [21] B. S. Cirel’son, “Quantum generalizations of bell’s inequality,” *Lett Math Phys* **4**, 93–100 (1980).
  - [22] N. N. Vorob’ev, “Consistent families of measures and their extensions,” *Theory of Probability & Its Applications* **7**, 147–163 (1962).
  - [23] Samson Abramsky and Adam Brandenburger, “The sheaf-theoretic structure of non-locality and contextuality,” *New Journal of Physics* **13**, 113036 (2011), [arXiv:1102.0264](#).
  - [24] Samson Abramsky and Lucien Hardy, “Logical bell inequalities,” (2012), [10.1103/PhysRevA.85.062114](#), [arXiv:1203.1352](#).
  - [25] Shane Mansfield and Tobias Fritz, “Hardy’s non-locality paradox and possibilistic conditions for non-locality,” *Foundations of Physics* **42**, 709–719 (2011), [arXiv:1105.1819](#).
  - [26] Itamar Pitowsky, “Correlation polytopes: Their geometry and complexity,” *Mathematical Programming* **50**, 395–414 (1991).
  - [27] George B Dantzig and B Curtis Eaves, “Fourier-motzkin elimination and its dual,” *Journal of Combinatorial Theory, Series A* **14**, 288–297 (1973).
  - [28] 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).
  - [29] Alexander Schrijver, *Theory of Linear and Integer Programming* (Wiley, 1998).
  - [30] Sébastien Lahaie, “How to take the dual of a linear program,” (2008).
  - [31] Bradley Hax, and Magnanti, *Applied Mathematical Programming* (Addison-Wesley, 1977) Chap. 4, pp. 143–144.
  - [32] Erling D. Andersen, *Computational Optimization and Applications* **20**, 171–183 (2001).
  - [33] MOSEK Optimization Suite, “Mosek api,” (2016).
  - [34] Gurobi Optimizer, “Farkasdual,” (2016).
  - [35] IBM - CPLEX Optimization Studio, “Cplex,” (2016).
  - [36] CVX - Matlab Software for Disciplined Convex Programming, “Dual variables,” (2016).
  - [37] Martin Andersen, Joachim Dahl, and Lieven Vandenberghe, “Linear cone programs,” (2016).
  - [38] Jean-Daniel Bancal, Nicolas Gisin, and Stefano Pironio, “Looking for symmetric bell inequalities,” *Journal of Physics A: Mathematical and Theoretical* **43**, 385303 (2010).

- [39] Jorge Nocedal and Stephen Wright, *Numerical Optimization (Springer Series in Operations Research and Financial Engineering)* (Springer, 2000).
- [40] 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).
- [41] Dénes Petz and László Ruppert, “Optimal quantum-state tomography with known parameters,” *Journal of Physics A: Mathematical and Theoretical* **45**, 085306 (2015), [arXiv:1511.06666](#).
- [42] S. R. Hedemann, “Hyperspherical parameterization of unitary matrices,” (2013), [arXiv:1303.5904](#).
- [43] Kazuyuki Fujii, Kunio Funahashi, and Takayuki Kobayashi, “Jarlskog’s parametrization of unitary matrices and qudit theory,” (2005), [arXiv:quant-ph/0508006](#).
- [44] 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](#).
- [45] Markus Grasmair, “On the existence of a cholesky factorization,” (2014).
- [46] Christoph Spengler, Marcus Huber, and Beatrix C. Hiesmayr, “A composite parameterization of unitary groups, density matrices and subspaces,” *Journal of Physics A: Mathematical and Theoretical* **43**, 385306 (2010), [arXiv:1004.5252](#).
- [47] 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](#).
- [48] 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**  $e = \{n \rightarrow n'\}$ .

**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**:  $\text{Pa}_{\mathcal{G}}(n) \equiv \{m \mid m \rightarrow n\}$
- The **children of a node**:  $\text{Ch}_{\mathcal{G}}(n) \equiv \{m \mid n \rightarrow m\}$
- The **ancestry of a node**:  $\text{An}_{\mathcal{G}}(n) \equiv \bigcup_{i \in \mathbb{W}} \text{Pa}_{\mathcal{G}}^i(n)$  where  $\text{Pa}_{\mathcal{G}}^i(n) \equiv \text{Pa}_{\mathcal{G}}(\text{Pa}_{\mathcal{G}}^{i-1}(n))$  and  $\text{Pa}_{\mathcal{G}}^0(n) = n$
- The **descendants of a node**:  $\text{De}_{\mathcal{G}}(n) \equiv \bigcup_{i \in \mathbb{N}} \text{Ch}_{\mathcal{G}}^i(n)$  where  $\text{Ch}_{\mathcal{G}}^i(n) \equiv \text{Ch}_{\mathcal{G}}(\text{Ch}_{\mathcal{G}}^{i-1}(n))$  and  $\text{Ch}_{\mathcal{G}}^1(n) = \text{Ch}_{\mathcal{G}}(n)$

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

**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  $N$ .

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

An **ancestral subgraph** of  $\mathcal{G}$  for a subset of nodes  $N \subseteq \mathcal{N}$  is the induced subgraph due to the ancestry of  $N$ .

$$\text{AnSub}_{\mathcal{G}}(N) \equiv \text{Sub}_{\mathcal{G}}(\text{An}_{\mathcal{G}}(N))$$

**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**.

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

It is critical to note the difference in language between *ancestors* from *ancestry*. The ancestry  $\text{An}_{\mathcal{G}}(n)$  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,

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

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

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

## Appendix B: Computationally Efficient Parameterization of the Unitary Group

If Section **VIII 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 [48]. Although not explicitly mentioned in [46], 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  $GP_l$ ,

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

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

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

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

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

One can verify that the following properties hold,

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

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

$$\begin{aligned} R_{m,n} &= \mathbb{1} + (|m\rangle\langle m| + |n\rangle\langle n|) \sum_{j=1}^{\infty} (-1)^j \frac{\lambda_{n,m}^{2j}}{(2j)!} + (|m\rangle\langle n| - |n\rangle\langle m|) \sum_{j=0}^{\infty} (-1)^j \frac{\lambda_{n,m}^{2j+1}}{(2j+1)!} \\ R_{m,n} &= \mathbb{1} + (|m\rangle\langle m| + |n\rangle\langle n|)(\cos \lambda_{n,m} - 1) + (|m\rangle\langle n| - |n\rangle\langle m|) \sin \lambda_{n,m} \end{aligned} \quad (\text{B3})$$

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

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

## Appendix C: Large Triangle Structure Inequalities

### 1. Web Inflation Inequality

The following is a causal compatibility inequality for the Triangle structure that is violated by the Fritz distribution outlined in Section IV and was derived using the Web inflation of Fig. 5d.

$$\begin{aligned}
& P(000)P(202) + P(000)P(212) + P(202)P(233) + P(302)P(312) \\
& \leq \\
& 2P(123)P(210) + 2P(123)P(310) + 2P(130)P(213) + 2P(133)P(210) + 2P(133)P(310) + P(000)P(003) + P(000)P(013) + P(000)P(023) + \\
& P(000)P(033) + P(000)P(103) + P(000)P(113) + P(000)P(123) + P(000)P(133) + P(000)P(203) + P(000)P(213) + P(003)P(010) + \\
& P(003)P(020) + P(003)P(030) + P(003)P(100) + P(003)P(110) + P(003)P(120) + P(003)P(130) + P(003)P(200) + P(003)P(210) + \\
& P(003)P(220) + P(003)P(230) + P(003)P(300) + P(003)P(310) + P(003)P(320) + P(003)P(330) + P(010)P(013) + P(010)P(021) + \\
& P(010)P(023) + P(010)P(033) + P(010)P(100) + P(010)P(103) + P(010)P(113) + P(010)P(123) + P(010)P(133) + P(010)P(203) + \\
& P(010)P(213) + P(013)P(020) + P(013)P(030) + P(013)P(100) + P(013)P(110) + P(013)P(120) + P(013)P(130) + P(013)P(200) + \\
& P(013)P(210) + P(013)P(220) + P(013)P(230) + P(013)P(300) + P(013)P(310) + P(013)P(320) + P(013)P(330) + P(020)P(023) + \\
& P(020)P(033) + P(020)P(103) + P(020)P(113) + P(020)P(123) + P(020)P(133) + P(020)P(203) + P(020)P(210) + P(020)P(211) + \\
& P(020)P(212) + P(020)P(223) + P(020)P(233) + P(020)P(310) + P(020)P(311) + P(020)P(312) + P(020)P(313) + P(021)P(100) + \\
& P(021)P(121) + P(021)P(210) + P(021)P(211) + P(021)P(213) + P(021)P(310) + P(021)P(311) + P(021)P(313) + P(022)P(210) + \\
& P(022)P(211) + P(022)P(212) + P(022)P(213) + P(022)P(310) + P(022)P(311) + P(022)P(312) + P(022)P(313) + P(023)P(030) + \\
& P(023)P(100) + P(023)P(110) + P(023)P(120) + P(023)P(130) + P(023)P(200) + P(023)P(211) + P(023)P(212) + P(023)P(213) + \\
& P(023)P(220) + P(023)P(230) + P(023)P(300) + P(023)P(311) + P(023)P(312) + P(023)P(313) + P(023)P(320) + P(023)P(330) + \\
& P(030)P(033) + P(030)P(103) + P(030)P(113) + P(030)P(123) + P(030)P(133) + P(030)P(203) + P(030)P(210) + P(030)P(211) + \\
& P(030)P(212) + P(030)P(223) + P(030)P(233) + P(030)P(310) + P(030)P(311) + P(030)P(312) + P(030)P(313) + P(031)P(210) + \\
& P(031)P(211) + P(031)P(213) + P(031)P(223) + P(031)P(310) + P(031)P(311) + P(031)P(313) + P(032)P(210) + P(032)P(211) + \\
& P(032)P(212) + P(032)P(213) + P(032)P(310) + P(032)P(311) + P(032)P(312) + P(032)P(313) + P(033)P(100) + P(033)P(110) + \\
& P(033)P(120) + P(033)P(130) + P(033)P(200) + P(033)P(211) + P(033)P(212) + P(033)P(213) + P(033)P(220) + P(033)P(230) + \\
& P(033)P(300) + P(033)P(311) + P(033)P(312) + P(033)P(313) + P(033)P(320) + P(033)P(330) + P(100)P(103) + P(100)P(113) + \\
& P(100)P(123) + P(100)P(133) + P(100)P(203) + P(100)P(213) + P(103)P(110) + P(103)P(120) + P(103)P(130) + P(103)P(200) + \\
& P(103)P(210) + P(103)P(220) + P(103)P(230) + P(103)P(300) + P(103)P(310) + P(103)P(320) + P(103)P(330) + P(110)P(113) + \\
& P(110)P(123) + P(110)P(133) + P(110)P(203) + P(110)P(213) + P(113)P(120) + P(113)P(130) + P(113)P(200) + P(113)P(210) + \\
& P(113)P(220) + P(113)P(230) + P(113)P(300) + P(113)P(310) + P(113)P(320) + P(113)P(330) + P(120)P(123) + P(120)P(133) + \\
& P(120)P(203) + P(120)P(210) + P(120)P(211) + P(120)P(212) + P(120)P(223) + P(120)P(233) + P(120)P(310) + P(120)P(311) + \\
& P(120)P(312) + P(120)P(313) + P(121)P(131) + P(121)P(210) + P(121)P(211) + P(121)P(213) + P(121)P(310) + P(121)P(311) + \\
& P(121)P(312) + P(121)P(313) + P(122)P(210) + P(122)P(211) + P(122)P(212) + P(122)P(213) + P(122)P(310) + P(122)P(311) + \\
& P(122)P(312) + P(122)P(313) + P(123)P(130) + P(123)P(200) + P(123)P(211) + P(123)P(212) + P(123)P(213) + P(123)P(220) + \\
& P(123)P(230) + P(123)P(300) + P(123)P(311) + P(123)P(312) + P(123)P(313) + P(123)P(320) + P(123)P(330) + P(130)P(133) + \\
& P(130)P(203) + P(130)P(210) + P(130)P(211) + P(130)P(212) + P(130)P(223) + P(130)P(233) + P(130)P(310) + P(130)P(311) + \\
& P(130)P(312) + P(130)P(313) + P(131)P(210) + P(131)P(211) + P(131)P(213) + P(131)P(310) + P(131)P(311) + P(131)P(313) + \\
& P(132)P(210) + P(132)P(211) + P(132)P(212) + P(132)P(213) + P(132)P(310) + P(132)P(311) + P(132)P(312) + P(132)P(313) + \\
& P(133)P(200) + P(133)P(211) + P(133)P(212) + P(133)P(213) + P(133)P(220) + P(133)P(230) + P(133)P(300) + P(133)P(311) + \\
& P(133)P(312) + P(133)P(313) + P(133)P(320) + P(133)P(330) + P(200)P(203) + P(200)P(213) + P(200)P(223) + P(200)P(233) + \\
& P(200)P(303) + P(200)P(313) + P(200)P(323) + P(200)P(333) + P(203)P(210) + P(203)P(220) + P(203)P(230) + P(203)P(300) + \\
& P(203)P(310) + P(203)P(320) + P(203)P(330) + P(210)P(213) + P(210)P(223) + P(210)P(233) + P(210)P(303) + P(210)P(313) + \\
& P(210)P(323) + P(210)P(333) + P(212)P(212) + P(212)P(333) + P(213)P(220) + P(213)P(230) + P(213)P(300) + P(213)P(310) + \\
& P(213)P(320) + P(213)P(330) + P(220)P(223) + P(220)P(233) + P(220)P(303) + P(220)P(313) + P(220)P(323) + P(220)P(333) + \\
& P(223)P(230) + P(223)P(300) + P(223)P(310) + P(223)P(320) + P(223)P(330) + P(230)P(233) + P(230)P(303) + P(230)P(313) + \\
& P(230)P(323) + P(230)P(333) + P(233)P(300) + P(233)P(310) + P(233)P(320) + P(233)P(330) + P(233)P(333) + P(300)P(303) + \\
& P(300)P(313) + P(300)P(323) + P(300)P(333) + P(303)P(310) + P(303)P(320) + P(303)P(330) + P(310)P(313) + P(310)P(323) + \\
& P(310)P(333) + P(313)P(320) + P(313)P(330) + P(320)P(323) + P(320)P(333) + P(323)P(330) + P(330)P(333)
\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 structure that is violated by the Fritz distribution outlined in Section IV and was derived using the Web inflation of Fig. 5d.

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

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

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