Inequalities Witnessing Quantum Incompatibility in The Triangle Scenario

Thomas C. Fraser*
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada
University of Waterloo, Waterloo, Ontario, Canada
(Dated: September 14, 2016)

Quantum correlations are often incompatible with a classical assumption of causal structure. This nonclassicality is often known as quantum nonlocality, and it is witnessed through the violation of causal compatability inequalities, such as Bell inequalities. Such inequalities were recently derived for the Triangle scenario [arXiv:1609.00672], begging the question: can these inequalities be violated by quantum correlations? Here we answer this affirmatively, and discuss specific Triangle scenario inequalities and quantum configurations which manifest nonclassical correlations. We also report the development of novel computational techniques suitable for large causal structures with numerous outcomes. Numerical optimzations reveal quantum resources potentially qualitatively different from those known previously.

* tcfraser@tcfraser.com

CONTENTS

I. INTRODUCTION

Todo (TC Fraser): Discuss why the marginal problem is important.

II. DEFINITIONS & NOTATION

Comment (TC Fraser): This entire section should be integrated into the body when logical to do so.

Definition 1. Following common notation used in [1], we use the notation $[k] = \{1, \ldots, k\}$ to be a finite index set. Moreover, when referencing a finite set with k elements, we can write,

$$A_{[k]} \equiv \{A_i\}_{i\in[k]} = \{A_1, \ldots, A_k\}$$

Definition 2. Each random variable v has a set of possible outcomes which are labeled by **outcome labels** O_v. The set of all outcomes themselves is its **outcome space** and is represented by the set of all functions from v to O_v[1],

$$\mathsf{Out}(v) \equiv \{f : \{v\} \to O_v\}$$

Each of these functions has a singular domain (v) and thus can be concisely written in a compact form $\{v \mapsto o\}$ where $o \in O_v$ represents some generic outcome label. Each function f should be interpreted as the *event* that v obtains the outcome o, as well as the function that takes v to o $\{v \mapsto f(v) = o\}$. Evidently $\mathsf{Out}(v)$ is isomorphic to O_v, and in many cases the two can be considered equivalent; however we make this particular semantic distinction now in preparation of section IX.

Definition 3. As a natural generalization, when considering a larger set of random variables $V = \{v_i\}_{i\in[k]}$ we define the joint outcome space in an analogous way,

$$\mathsf{Out}(V) \equiv \left\{f : \{v_i\}_{i\in[k]} \to \{O_{v_i}\}_{i\in[k]} \mid \forall i \in [k] : f(v_i) \in O_{v_i}\right\} \tag{1}$$

Each outcome f can be compactly represented as a set of mappings over each element of V.

$$f = \{v_1 \mapsto f(v_1), \ldots, v_k \mapsto f(v_k)\} = \{v_i \mapsto f(v_i)\}_{i\in[k]}$$

The domain of f refers to the set of random variables it valuates and is denoted as $\mathsf{Dom}(f)$. If $f \in \mathsf{Out}(V)$, then $\mathsf{Dom}(f) = V$.

Example 4. Interpreting $f \in \mathsf{Out}(V)$ as an *event* or *outcome* on V, a probability distribution P_V over V will assign a non-negative probability $P_V(f)$ to f. Let $V = \{v_1, v_2\}$ with each $v \in V$ having 3 distinct outcomes $O_v = \{0, 1, 2\}$. Furthermore let $f \in \mathsf{Out}(V)$ be the outcome $(v_1 \mapsto 2, v_2 \mapsto 1)$. The following class of notations will be considered interchangeable,

$$P_V(f) = P_{v_1,v_2}(f) = P(v_1 \mapsto f_1(v_1), v_2 \mapsto f_2(v_2)) = P(v_1 \mapsto 2, v_2 \mapsto 1) = P_{v_1v_2}(21)$$

Each of which should be read as "the probability that v_1 gets outcome 2 AND v_2 gets outcome 1".

Remark 5. In each of the above notations, P_V always represents a probability distribution over the *domain* of $f \in \mathsf{Out}(V)$. In cases where the subscript is omitted, it should be assumed to be over the domain of f.

$$P_V(f) = P_{\mathsf{Dom}(f)}(f) = P(f)$$

Definition 6. Let $V = \{v_1, \ldots, v_k\}$ be a set of random variables and $W \subseteq V$ be a subset of V. Moreover, let $f \in \mathsf{Out}(V)$ be a joint outcome over V; assigning a unique label to each variable in V. The **restriction of f onto W** (denoted $f|_W \in \mathsf{Out}(W)$) is the outcome of $\mathsf{Out}(W)$ that *agrees* with each of f's assignments for variables in W.

$$\forall v \in W : f(v) = f|_W(v)$$

Definition 7. Consider two sets of random variables V and W and two specific outcomes $f_V \in \mathsf{Out}(V)$ and $f_W \in \mathsf{Out}(W)$. The outcomes f_V and f_W are said to be **compatible** if they agree on their valuations of $V \cap W$. Compatibility will be denoted with '$\rightarrowtail\!\!\!\!\prec$'.

$$f_V \rightarrowtail\!\!\!\!\prec f_W \iff f_V|_{V\cap W} = f_W|_{V\cap W}$$

Compatibility is a reflexive and symmetric, but *not* transitive. There are two cases when compatibility is transitive; whenever $V \cap W = \emptyset$ or $V = W$. Whenever $V \cap W = \emptyset$, f_V and f_W can be said to be *trivially* compatible. Moreover if $V = W$, compatibility becomes ordinary equivalence,

$$V = W \quad \text{and} \quad f_V \rightarrowtail\!\!\!\!\prec f_W \iff f_V = f_W$$

---

[1] Using functional language for outcome events is necessary for a sheaf-theoretic treatment of contextuality [2].

Definition 8. Let $f_V \in \mathsf{Out}(V)$ and $f_W \in \mathsf{Out}(W)$ be outcomes of V and W respectively as in definition 7. If f_V and f_W are compatible, and $W \subseteq V$ then f_W is said to be **extendable** to f_V and is denoted '$\prec$'.

$$f_W \prec f_V \iff f_W \rightarrowtail\!\!\!\prec f_V \quad \text{and} \quad W \subseteq V \iff f_W = f_V|_W$$

The dual language is also appropriate: f_W is a **restriction** of f_V (denoted $f_V \succ f_W$) if and only if f_W is extendable to f_V. If $f_W \prec f_V$ then there exists a unique element $f_{V\setminus W} \in \mathsf{Out}(V \setminus W)$ such that f_V can be reconstructed by gluing together $f_{V\setminus W}$ and f_W.

$$\forall v \in V : f_V(v) = \begin{cases} f_W(v) & v \in W \\ f_{V\setminus W}(v) & v \in V \setminus W \end{cases}$$

Remark 9. The notation '$g \rightarrowtail\!\!\!\prec h$' is used to indicate compatibility between $g \in \mathsf{Out}(V)$ and $h \in \mathsf{Out}(W)$ because of the following observation; g and h are compatible if and only if extensions of some outcome $f \in \mathsf{Out}(V \cap W)$.

$$g \rightarrowtail\!\!\!\prec h \iff \exists f \in \mathsf{Out}(V \cap W) \mid g \succ f \prec h$$

In fact, such an f is uniquely defined,

$$\forall v \in V \cap W : f(v) = g(v) = g(v)$$

Definition 10. The set of all extendable outcomes of $f \in \mathsf{Out}(W)$ into V is called the **extendable set of f into V** and can be written as,

$$\mathsf{Ext}_V(f) \equiv \{g \in \mathsf{Out}(V) \mid f \prec g\} \subseteq \mathsf{Out}(V)$$

The extendable set of f into V is the subset of outcomes in $\mathsf{Out}(V)$ that *agree* with f about valuations for variables in W.

Example 11. Let $W = \{a, b\}$ and $V = \{a, b, c\}$ be two sets of random variables. Clearly $W \subseteq V$; a prerequisite for extendability. Further impose that the outcome labels for each variable are the same: $O_a = O_b = O_c = \{0, 1, 2\}$. As an example, let $f = \{a \mapsto f(a) = 1, b \mapsto f(b) = 2\} \in \mathsf{Out}(W)$; f is extendable to the outcome $g = \{a \mapsto g(a) = 1, b \mapsto g(b) = 2, c \mapsto g(c) = 1\} \in \mathsf{Out}(V)$, and the extendable set of f into V is,

$$\mathsf{Ext}_V(f) = \mathsf{Ext}_{a,b,c}(a \mapsto 1, b \mapsto 2) = \{\{a \mapsto 1, b \mapsto 2, c \mapsto 0\}, \{a \mapsto 1, b \mapsto 2, c \mapsto 1\}, \{a \mapsto 1, b \mapsto 2, c \mapsto 2\}\}$$

Definition 12. A **graph** is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *edges* respectively where the nodes can represent any object and the edges are pairs of nodes. For convenience of notation, one defines an index set over the nodes denoted $\mathcal{I}_\mathcal{N}$.

$$\mathcal{N} = \{n_i \mid i \in \mathcal{I}_\mathcal{N}\} \quad \mathcal{E} = \{\{n_j, n_k\} \mid j, k \in \mathcal{I}_\mathcal{N}\}$$

Definition 13. A **directed graph** $\mathcal{G}$ is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *edges* respectively where the nodes can represent any object and the edges are *ordered* pairs of nodes. For convenience of notation, one defines an index set over the nodes denoted $\mathcal{I}_\mathcal{N}$.

$$\mathcal{N} = \{n_i \mid i \in \mathcal{I}_\mathcal{N}\} \quad \mathcal{E} = \{n_j \to n_k \mid j, k \in \mathcal{I}_\mathcal{N}\}$$

Definition 14. The following definitions are common language in directed graph theory. Let $n, m \in \mathcal{N}$ be example nodes of the graph $\mathcal{G}$.

- The **parents of a node**: $\mathsf{Pa}_\mathcal{G}(n) \equiv \{m \mid m \to n\}$
- The **children of a node**: $\mathsf{Ch}_\mathcal{G}(n) \equiv \{m \mid n \to m\}$
- The **ancestry of a node**: $\mathsf{An}_\mathcal{G}(n) \equiv \bigcup_{i \in \mathbb{W}} \mathsf{Pa}^i_\mathcal{G}(n)$ where $\mathsf{Pa}^i_\mathcal{G}(n) \equiv \mathsf{Pa}_\mathcal{G}\big(\mathsf{Pa}^{i-1}_\mathcal{G}(n)\big)$ and $\mathsf{Pa}^0_\mathcal{G}(n) = n$

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

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

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

Moreover, an **induced subgraph** of $\mathcal{G}$ due to a set of nodes $N \subseteq \mathcal{N}$ is the graph composed of N and all edges $e \in \mathcal{E}$ of the original graph that are contained in N.

$$\mathsf{Sub}_{\mathcal{G}}(N) \equiv (N, \{e_i \mid i \in \mathcal{I}_{\mathcal{E}}, e_i \subseteq N\})$$

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

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

Definition 15. A **directed acyclic graph** or **DAG** $\mathcal{G}$ is an directed graph definition 13 with the additional property that no node n is in its set of **ancestors**.

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

Notice the difference between using the natural numbers $\mathbb{N}$ to distinguish *ancestors* from *ancestry*.

Definition 16. A **causal structure** is simply a DAG with the extra classification of each node into one of two categories; the **latent nodes** and **observed nodes** denoted $\mathcal{N}_L$ and $\mathcal{N}_O$. The latent nodes correspond to random variables that are either hidden through some fundamental process or cannot/will not be measured. The observed nodes are random variables that are measurable. Every node is either latent or observed and no node is both:

$$\mathcal{N}_L \cap \mathcal{N}_O = \emptyset \qquad \mathcal{N}_L \cup \mathcal{N}_O = \mathcal{N}$$

Definition 17. The **product distribution** of two disjoint distributions P_V and P_W (where $V \cap W = \emptyset$) is denoted as usual with $P_V \times P_W$ and is defined as,

$$\forall f \in \mathsf{Out}(V \cup W) : (P_V \times P_W)(f) \equiv P_V(f|_V) \cdot P_W(f|_W)$$

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 18. 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,

$$\forall f \in \mathsf{Out}(W) : \left(\sum_{V\setminus W} P_V\right)(f) \equiv \sum_{g\in \mathsf{Ext}_V(f)} P_V(g)$$

Comment (TC Fraser): How many definitions do I need to write??

III. COMPATIBILITY, CONTEXTUALITY AND THE MARGINAL PROBLEM

In order to determine if a given marginal distribution P_V or set of marginal distributions $\{P_{V_i}\}_{i\in[k]} \equiv \{P_{V_1}, \dots, P_{V_k}\}$ is compatible with a causal structure $\mathcal{G}$, one should first formalize what is meant by *compatible*.

Definition 19. A set of **causal parameters** for a particular causal structure $\mathcal{G}$ is the specification of a conditional distribution for every node $n \in \mathcal{N}$ given it's parents in $\mathcal{G}$.

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

Definition 20. A marginal distribution P_V is **compatible** with a causal structure $\mathcal{G}$ (where it is assumed that $V \subseteq \mathcal{N}_O$) if there exists a *choice* of causal parameters $\left\{P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N}\right\}$ such that P_V can be *recovered* from the following series of operations:

1. First obtain a joint distribution over *all* nodes of of the causal structure,

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

2. Then marginalize over the latent nodes of $\mathcal{G}$,

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

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

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

A set of marginal distributions $\{P_{V_i}\}_{i\in[k]}$ is compatible with $\mathcal{G}$ if each distribution P_{V_i} can be made compatible by the *same* choice of causal parameters. A distribution P_V or set of distributions $\{P_{V_i}\}_{i\in[k]}$ is said to be **incompatible** with a causal structure if there *does not exist* a set of causal parameters with the above mentioned property.

Todo (TC Fraser): Source this?

Operations 2 and 3 of definition 20 are related to the *marginal problem*.

Definition 21. **The Marginal Problem:** Given a set of distributions $\{P_{V_i}\}_{i\in[k]} \equiv \{P_{V_1}, \dots, P_{V_k}\}$ where $V_i \subseteq \mathcal{J}$ for some set of random variables $\mathcal{J}$, does there exist a joint distribution $P_{\mathcal{J}}$ such that each given distribution P_{V_i} can be obtained from marginalizing $P_{\mathcal{J}}$?

$$\forall i \in [k] : P_{V_i} = \sum_{\mathcal{J}\setminus V_i} P_{\mathcal{J}}$$

Typically (although not strictly necessary[2]), $\mathcal{J}$ is taken to mean the union of all V_i's.

$$\mathcal{J} = V_1 \cup \dots \cup V_k = \bigcup_{i\in[k]} V_i \tag{2}$$

Todo (TC Fraser): Mention polytope perspective and existing methods

Definition 22. A reoccurring motif of these discussions will be the set of distributions $\{P_{V_i}\}_{i\in[k]} \equiv \{P_{V_1}, \dots, P_{V_k}\}$ mentioned in definition 21. In agreement with [1] we call the set of subsets $\mathcal{M} \equiv \{V_1, \dots, V_k\}$ the marginal contexts or the **maximal marginal scenario**[3]. An individual $V_i \in \mathcal{M}$ is a **marginal context**. The union of all contexts is denoted $\mathcal{J}$ and is defined exactly as in eq. (2). Moreover, we will call the set of distributions a **marginal model** and denote it $P^{\mathcal{M}}$ provided that they are *compatible* Todo (TC Fraser): Mention quantum non-signalling and maybe $k \geq 3$:

$$\forall i \neq j \text{ if } V_i \cap V_j \neq \emptyset \text{ then } \sum_{V_i\setminus V_j} P_{V_i} = \sum_{V_j\setminus V_i} P_{V_j}$$

Following definition 2.3 in [1], a marginal model is said to be **contextual** if it *does not* admit a joint distribution $P_{\mathcal{J}}$ and **non-contextual** otherwise.

In addition, we elect to define the **marginal outcomes** $\mathsf{Out}(\mathcal{M})$ to be the set of all outcomes belonging to outcomes $\mathsf{Out}(V_i)$ of the marginal context[4].

$$\mathsf{Out}(\mathcal{M}) \equiv \coprod_{i\in[k]} \mathsf{Out}(V_i)$$

Each marginal outcome $m \in \mathsf{Out}(\mathcal{M})$ belongs to a unique marginal outcome space $\mathsf{Out}(V_i)$. To make reference to the marginal context V_i associated with m, recall remark 5: the context for m is precisely the domain of m ($V_i = \mathsf{Dom}(m)$). Note that $\mathsf{Out}(\mathcal{M})$ is a slight abuse of notation; $\mathsf{Out}(\mathcal{M})$ is *not* a valid outcome space, analogous to the fact that $P^{\mathcal{M}}$ is *not* a probability distribution. Instead $\mathsf{Out}(\mathcal{M})$ is a *flattened* collection of outcome spaces just as $P^{\mathcal{M}}$ is a collection of distributions. The **joint outcomes** $\mathsf{Out}(\mathcal{J})$ are defined in the usual fashion of definition 2.

---

[2] Appending any additional variables to $\mathcal{J}$ will not affect the existence of a joint distribution.

[3] Rigorously speaking as defined by [1], a marginal scenario forms an *abstract simplicial complex* where it is required that $V' \in \mathcal{M}$ whenever $V' \subseteq V_i$ for some $i \in [k]$. "Maximal" refers to the restriction that $\forall i, j : V_i \not\subseteq V_j$.

[4] Here '$\coprod$' refers to the *disjoint union* of outcome contexts; when $i \neq j$, $\mathsf{Out}(V_i) \cap \mathsf{Out}(V_j) = \emptyset$ since $V_i \neq V_j$.

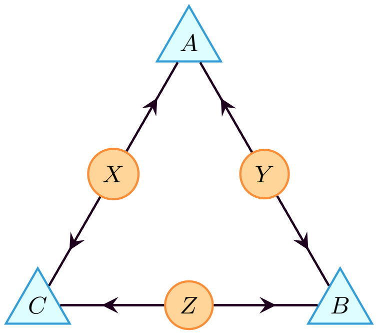

FIG. 1. The casual structure of the Triangle Scenario. Three variables A, B, C are observable and illustrated as triangles, while X, Y, Z are latent variables illustrated as circles.

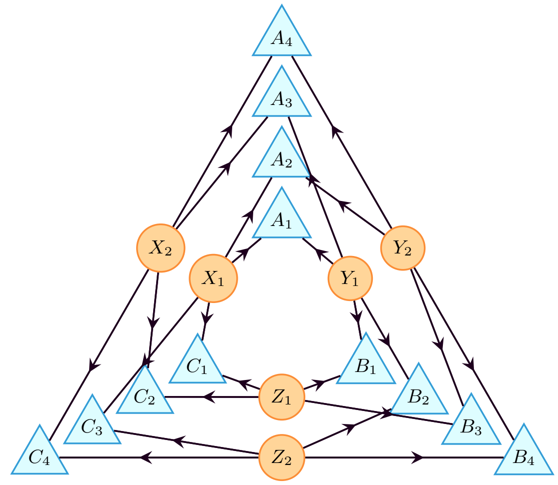

FIG. 2. An inflated causal structure of the Triangle Scenario fig. 1.

Definition 23. A **causal compatibility inequality** $I_{\mathcal{M}}$ for a marginal scenario $\mathcal{M}$ is a probabilistic inequality that is obeyed for every compatible marginal model $P^{\mathcal{M}}$. Whenever the marginal scenario $\mathcal{M}$ in question is evident by context, the subscript of $I_{\mathcal{M}}$ will be dropped leaving just I.

Todo (TC Fraser): Discuss Compatibility, connection to cooperative games/resources, bell incompatibility? Todo (TC Fraser): Connection between contextuality and Compatibility via the marginal problem for causal parameters Todo (TC Fraser): Discuss what is meant by a 'complete' solution to the marginal problem Todo (TC Fraser): Maybe define the possibilistic marginal problem for later

IV. THE TRIANGLE SCENARIO

The triangle scenario $\mathcal{G}$ is a causal structure composed of 3 parties A, B, C arranged in a triangular configuration while pair-wise sharing latent variables X, Y, Z. It has that has been studied extensively in existing literature (see [3, Fig. 1], [4, Fig. 6], [5, Fig. 8], [6, Fig. 8, Appendix E], [7, Fig. 3], [8, Fig. 1], ...) and is reproduced here in fig. 1 for convenience. At the time of publication for [5] it was noted that characterizing locality in $\mathcal{G}$ remained an open problem and that identifying locality constraints in this configuration seemed challenging. In [6], $\mathcal{G}$ was classified as an "interesting" causal structure.[5] Additionally $\mathcal{G}$ is presented along with a family of entropic inequalities that constrain the set of compatible distributions [6]. Furthermore in [7], Fritz demonstrated that $\mathcal{G}$ is the *smallest* correlation scenario in which their exists quantum incompatible distributions by explicitly mentioning one (which we elect to call the *Fritz distribution* in section V) and poses an important question asking if other incompatible quantum distributions exist. Finally in [8], Wolfe et. al. make use of $\mathcal{G}$ along with the Bell Scenario to demonstrate a novel technique for causal inference called the *inflation technique* (summarized in section VI). Wolfe et. al. validate the power and applicability of the inflation technique by proving incompatibility between $\mathcal{G}$ and the W-type distribution; a quantum non-accessible distribution whose incompatibility is not witness-able by *any* entropic inequalities or any other known constraints [8]. Moreover, a family of polynomial, 2-outcome, compatibility inequalities are explicitly derived for an exemplary inflation of $\mathcal{G}$ and 37 non-trivial representatives are printed in [8].

Identifying quantum-accessible distributions that are incompatible with the triangle scenario is of particular importance as it Todo (TC Fraser): Discuss why we are doing this.

As a preliminary search for quantum incompatibility in $\mathcal{G}$, we performed numerical optimizations using bipartite qubit density matrices and 2-outcome POVM measurements (see appendix D for details) against the 37 compatibility inequalities printed in [8] as well as the entropic inequalities presented by [6]. Unfortunately, none of these inequalities could be violated. These early results suggest, although do not prove, that quantum non-locality is not present in $\mathcal{G}$

[5] Indicating that conditional independence relations are not a sufficient characterization of compatibility for $\mathcal{G}$ [6]. Note that in $\mathcal{G}$ there are *zero* conditional independence relations over the observable nodes A, B, C.

for *two-outcome* measurements. We make *no claim* that this is a universal truth and simply take it as motivation to explore incompatibility in $\mathcal{G}$ for a larger number of outcomes.

Specifically, we make use of the inflation technique for an inflated triangle scenario $\mathcal{G}'$ depicted in fig. 2 and cast the *causal compatibility* problem as a *marginal problem.* We develop a new technique which we call the *extension covering procedure (ECP)* to find a complete solution to the marginal problem that is suitable for large causal structures where the observable nodes are equipped with numerous outcomes. We use this technique to find numerous compatibility inequalities, some of which witness the incompatibility of the Fritz distribution. Before discussing the ECP, the Fritz distribution will be re-defined explicitly here, followed by a quick summary of the inflation technique of [8]. Following these summaries of existing work, we present the ECP and its connection to Hardy-type paradoxes [8–10], certificate inequalities and logical Bell inequalities [2]. Moreover, the ECP can be adapted to derive symmetric inequalities or inequalities bounding any subspace of marginal models. To conclude, we subject our found inequalities to non-linear optimizations and discuss the implications of our results.

V. THE FRITZ DISTRIBUTION

The **Fritz distribution** P_F is a quantum-accessible distribution known to be incompatible with the Triangle Scenario [7]. Explicitly, P_F is a three-party (A, B, C), four-outcome $(1, 2, 3, 4)$ distribution that has form as follows:

$$\begin{aligned} P_F(111) = P_F(221) = P_F(412) = P_F(322) = P_F(233) = P_F(143) = P_F(344) = P_F(434) &= \frac{1}{32}\left(2+\sqrt{2}\right) \\ P_F(121) = P_F(211) = P_F(422) = P_F(312) = P_F(243) = P_F(133) = P_F(334) = P_F(444) &= \frac{1}{32}\left(2-\sqrt{2}\right) \end{aligned} \tag{3}$$

Here the notation $P_F(abc) = P_{ABC}(abc) = P(A = a, B = b, C = c)$ is used. The Fritz distribution P_F can be realized with the following quantum configuration:

$$\begin{gathered} \rho_{AB} = \left|\Psi^+\right\rangle\left\langle\Psi^+\right| \quad \rho_{BC} = \rho_{CA} = \left|\Phi^+\right\rangle\left\langle\Phi^+\right| \\ M_A = \left\{|0\psi_0\rangle\langle 0\psi_0|, |0\psi_\pi\rangle\langle 0\psi_\pi|, \left|1\psi_{-\pi/2}\right\rangle\left\langle 1\psi_{-\pi/2}\right|, \left|1\psi_{\pi/2}\right\rangle\left\langle 1\psi_{\pi/2}\right|\right\} \\ M_B = \left\{\left|\psi_{\pi/4}0\right\rangle\left\langle\psi_{\pi/4}0\right|, \left|\psi_{5\pi/4}0\right\rangle\left\langle\psi_{5\pi/4}0\right|, \left|\psi_{3\pi/4}1\right\rangle\left\langle\psi_{3\pi/4}1\right|, \left|\psi_{-\pi/4}1\right\rangle\left\langle\psi_{-\pi/4}1\right|\right\} \\ M_C = \{|00\rangle\langle 00|, |01\rangle\langle 01|, |10\rangle\langle 10|, |11\rangle\langle 11|\} \end{gathered} \tag{4}$$

Where for convenience of notation ψ_x is used to denote the superposition,

$$|\psi_x\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{ix}|1\rangle\right)$$

Additionally $|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ and $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ are two maximally entangled Bell states.

Fritz first proved it's incompatibility [7] by showing C acts a moderator to ensure measurement pseudo-settings for A and B are independent, satisfying non-broadcasting requirements for the standard Bell scenario. In fact, by coarse-graining outcomes for A and B and treating C as a measurement-setting moderator, P_F maximally violates the CHSH inequality. To illustrate this, begin with the CHSH inequality [11],

$$\langle AB|S_A = 1, S_B = 1\rangle + \langle AB|S_A = 1, S_B = 2\rangle + \langle AB|S_A = 2, S_B = 1\rangle - \langle AB|S_A = 2, S_B = 2\rangle \le 2 \tag{5}$$

Where $\langle AB|S_A = i, S_B = j\rangle$ is the correlation between A and B given the measurement settings for A (B) is i (j) respectively. Next each of C's outcomes become the conditioned settings in eq. (5),

$$\langle AB|C = 2\rangle + \langle AB|C = 3\rangle + \langle AB|C = 4\rangle - \langle AB|C = 1\rangle \le 2 \tag{6}$$

Finally, specifying the two outcome coarse-graining as $\{1, 2, 3, 4\} \to \{(1, 4), (2, 3)\}$ gives a definition of correlation,

$$\langle AB\rangle \equiv 2\{P(11) + P(14) + P(41) + P(44) + P(22) + P(23) + P(32) + P(33)\} - 1$$

Which when applied to the Fritz distribution eq. (3) gives the following correlations:

$$\langle AB|C = 2\rangle = 2\left\{\frac{1}{32}\left[2\left(2+\sqrt{2}\right) + 2\left(2-\sqrt{2}\right)\right]\right\} - 1 = \frac{\sqrt{2}}{2}$$

$$\langle AB|C = 3\rangle = \langle AB|C = 4\rangle = \frac{\sqrt{2}}{2} \quad \langle AB|C = 1\rangle = -\frac{\sqrt{2}}{2}$$

Which when applied to eq. (6) gives the familiar Tsirelson violation of $2\sqrt{2} \nleq 2$ [12].

Before continuing it is worth noting that eq. (3) is non-unique. Any distribution that is equal to eq. (3) via a permutation of outcomes or exchange of parties is also referred to as a Fritz distribution. Moreover, the quantum realization of eq. (4) is non-unique. In fact, eq. (4) is not the realization that Fritz originally had in mind. Nonetheless eq. (3) is taken as *the* Fritz distribution for concreteness throughout this paper.

In [7], Fritz posits the following question: *Find an example of non-classical quantum correlations in $\mathcal{G}$ together with a proof of its non-classicality which does not hinge on Bell's Theorem.*

Todo (TC Fraser): Summarize Problem 2.17 in fritz BBT, make it more formal

VI. SUMMARY OF THE INFLATION TECHNIQUE

The causal inflation technique, first pioneered by Wolfe, Spekkens, and Fritz [8] and inspired by the *do calculus* and *twin networks* of Ref. [13], is a family of causal inference techniques that can be used to determine if a probability distribution is compatible or incompatible with a given causal structure. 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. Copies of the original nodes are distinguished by an additional subscript called the **copy-index**. For example node A of fig. 1 has copies A_1, A_2, A_3, A_4 in the inflation of fig. 2. Following the notions of Ref. [8], all such copies are deemed equivalent via a **copy-index equivalence** relation denoted '$\sim$'. A copy-index is effectively arbitrary, so A' will refer to an arbitrary inflated copy of A.

$$A \sim A_1 \sim A' \nsim B \sim B_1 \sim B'$$

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 $X = Y$ upon removal of the copy-index. Equipped with the common graph-theoretic terminology and notation of definition 14, an inflation can be formally defined as follows:

Definition 24. An **inflation** of a causal structure $\mathcal{G}$ is another causal structure $\mathcal{G}'$ such that:

$$\forall n' \in \mathcal{N}' : \mathsf{AnSub}_{\mathcal{G}'}(n') \sim \mathsf{AnSub}_{\mathcal{G}}(n) \tag{7}$$

The motivation being that since 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 $\mathcal{G}$.

$$\forall n \in \mathcal{N} : P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} = \sum_{\mathcal{N}\setminus n} P_{\mathcal{N}} \tag{8}$$

Now since eq. (7) enforces that $\mathsf{Pa}_{\mathcal{G}'}(n') \sim \mathsf{Pa}_{\mathcal{G}}(n)$, one can create a set of **inflated causal parameters** using eq. (8),

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

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

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

This is known as the **weak inflation lemma**; compatible joint distributions over $\mathcal{N}$ induce compatible joint distributions over $\mathcal{N}'$. Before generalizing to the inflation lemma, a careful observation needs to be made. For any pair of subsets of nodes $N \subseteq \mathcal{N}$ and $N' \subseteq \mathcal{N}'$ that are equivalent up to copy-index $N \sim N'$, if $\mathsf{AnSub}_{\mathcal{G}}(N) \sim \mathsf{AnSub}_{\mathcal{G}'}(N')$ then any compatible *marginal* distribution P_N over N induces a compatible marginal distribution $P_{N'}$ over N'. In fact since N' contains no duplicate nodes up to copy index (simply because $N \sim N'$ and N cannot contain duplicate nodes), $P_N = P_{N'}$. We refer such sets N' as the **injectable sets** of $\mathcal{G}'$ and N the **images of the injectable sets** of $\mathcal{G}$.

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

[6] Of course not all compatible joint distributions on $\mathcal{N}'$ are constructed in this way; all that is demonstrated is that all joint distributions constructed in this way are compatible.

Lemma 25. ***The Inflation Lemma*** *Given a particular inflation $\mathcal{G}'$ of $\mathcal{G}$, if a marginal model $\left\{P_N \mid N \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\right\}$ is compatible with $\mathcal{G}$ then all marginal models $\left\{P_{N'} \mid N' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\right\}$ are compatible with $\mathcal{G}'$ provided that $P_N = P_{N'}$ for all instances where $N \sim N'$. This is lemma 3 of [8].*

The inflation lemma is the most important result of the causal inflation technique [8]. The contrapositive version of lemma 25 is a powerful tool for determining compatibility. Any compatibility constraint on marginal models $\left\{P_{N'} \mid N' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\right\}$ of the inflated causal structure $\mathcal{G}'$ correspond to valid compatibility constraints on marginal models $\left\{P_N \mid N \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\right\}$ of the original causal structure. Corollary 5 of [8] proves this explicitly for incompatibility inequalities I; which the remainder of this work focuses on. Additionally, the inflation lemma holds when considering any subset of $\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ (analogously $\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')$). Therefore, in situations where latent nodes are present in $\mathcal{G}$, one only needs to consider injectable sets that are composed of *observable nodes*.

In this work, we obtain inequalities that constrain the set of injectable marginal models $\left\{P_{N'} \mid N' \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\right\}$ for the inflated Triangle Scenario of fig. 1 by considering the marginal problem (see section III) for the set of **maximally pre-injectable sets**[7]. A pre-injectable set V is a subset of $\mathcal{N}'$ that can be decomposed into the disjoint union of injectable sets $V = \coprod_i N'_i \mid N'_i \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ that are mutually **ancestrally independent**,

$$\forall i, j : N'_i \perp N'_j \iff \mathsf{An}_{\mathcal{G}'}(N'_i) \cap \mathsf{An}_{\mathcal{G}'}(N'_j) = \emptyset$$

In doing so, any distribution over all nodes of a pre-injectable set $P_{\cup_i N'_i}$ will factorize according to graphical d-separation conditions [13],

$$P_V = P_{\cup_i N'_i} = \prod_i P_{N'_i}$$

This turns linear inequalities over the pre-injectable distributions into polynomial inequalities over the injectable distributions, allowing one to replace all such distributions with equivalent distributions over the original random variables of $\mathcal{N}$. $\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')$ will denote the set of all pre-injectable sets.

Focusing on the inflation $\mathcal{G}'$ depicted in fig. 2, we obtained the injectable sets $\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ along with the maximally pre-injectable sets through a simple graphical procedure outlined in [8].

$$\begin{array}{cc}
\textbf{Maximal Pre-injectable Sets} & \textbf{Ancestral Independences} \\
\{A_1, B_1, C_1, A_4, B_4, C_4\} & \{A_1, B_1, C_1\} \perp \{A_4, B_4, C_4\} \\
\{A_1, B_2, C_3, A_4, B_3, C_2\} & \{A_1, B_2, C_3\} \perp \{A_4, B_3, C_2\} \\
\{A_2, B_3, C_1, A_3, B_2, C_4\} & \{A_2, B_3, C_1\} \perp \{A_3, B_2, C_4\} \\
\{A_2, B_4, C_3, A_3, B_1, C_2\} & \{A_2, B_4, C_3\} \perp \{A_3, B_1, C_2\} \\
\{A_1, B_3, C_4\} & \{A_1\} \perp \{B_3\} \perp \{C_4\} \\
\{A_1, B_4, C_2\} & \{A_1\} \perp \{B_4\} \perp \{C_2\} \\
\{A_2, B_1, C_4\} & \{A_2\} \perp \{B_1\} \perp \{C_4\} \\
\{A_2, B_2, C_2\} & \{A_2\} \perp \{B_2\} \perp \{C_2\} \\
\{A_3, B_3, C_3\} & \{A_3\} \perp \{B_3\} \perp \{C_3\} \\
\{A_3, B_4, C_1\} & \{A_3\} \perp \{B_4\} \perp \{C_1\} \\
\{A_4, B_1, C_3\} & \{A_4\} \perp \{B_1\} \perp \{C_3\} \\
\{A_4, B_2, C_1\} & \{A_4\} \perp \{B_2\} \perp \{C_1\}
\end{array} \tag{9}$$

As can be counted, there are 12 maximally pre-injectable sets $\mathcal{G}'$ which will be indexed 1 through 12 in the order seen above ($\{V_1, \ldots, V_{12}\}$). The maximally pre-injectable sets $\{V_i\}_{i\in[12]}$ form a maximal marginal scenario $\mathcal{M}$ over the observable nodes of $\mathcal{G}'$ when equipped with outcome labels. Using lemma 25 inequalities $I_{\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')}$ constraining the set of contextual marginal models $P^{\mathcal{M}}$ can be deflated into polynomial inequalities over $I_{\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')}$ bounding compatible marginal models over the images of the injectable sets in $\mathcal{G}$.

$$\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}') = \{\{A, B, C\}, \{A\}, \{B\}, \{C\}\}$$

Comment (TC Fraser): This is not all the images. I should clarify this.

[7] The set of all pre-injectable sets forms a topological *simplicial conplex*. ‘Maximal’ refers to the pre-injectable sets that are no proper subset of another.

VII. CERTIFICATE INEQUALITIES

A. Casting the Marginal Problem as a Linear Program

After obtaining the maximal pre-injectable sets associated with a particular inflation, one can write the marginal problem of definition 21 as a linear program. The key observation is that marginalization is a *linear* operator that can be performed via a matrix multiplication. To do this, we will define the *incidence matrix*.

Definition 26. 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* outcomes $j \in \mathsf{Out}(\mathcal{J})$ and the rows are indexed by *marginal* outcomes $m \in \mathsf{Out}(\mathcal{M})$. The entries of M are populated whenever a row index is extendable to a column index.

$$M[m, j] = \delta_{m \prec j} \equiv \begin{cases} 1 & m \prec j \\ 0 & \text{otherwise} \end{cases}$$

Where $\delta_{m \prec j}$ is an *extendability indicator*. The incidence matrix has $|\mathsf{Out}(\mathcal{J})|$ columns and $|\mathsf{Out}(\mathcal{M})| = \sum_{i \in [k]} |\mathsf{Out}(V_i)|$ rows. The number of non-zero entries of M is a simple expression,

$$\sum_{i \in [k]} |\mathsf{Out}(V_i)| |\mathsf{Out}(\mathcal{J} \setminus V_i)| = \sum_{i \in [k]} |\mathsf{Out}(\mathcal{J})| = k |\mathsf{Out}(\mathcal{J})|$$

This is due to the fact that each marginal context contributes a single non-zero entry to each column of M, resulting in $k|\mathsf{Out}(\mathcal{J})|$ total non-zero entries.

Todo (TC Fraser): Computationally Efficient generation?

To illustrate this concretely, consider the following example:

Example 27. Let $\mathcal{J}$ be 3 binary variables $\mathcal{J} = \{a, b, c\}$ and consider the marginal scenario $\mathcal{M} = \{\{a, c\}, \{b\}\}$. The incidence matrix becomes:

$$M = \begin{array}{c} (a,b,c) \mapsto \\ (a\mapsto 0, c\mapsto 0) \\ (a\mapsto 0, c\mapsto 1) \\ (a\mapsto 1, c\mapsto 0) \\ (a\mapsto 1, c\mapsto 1) \\ (b\mapsto 0) \\ (b\mapsto 1) \end{array} \begin{array}{c} \begin{array}{cccccccc} (0,0,0) & (0,0,1) & (0,1,0) & (0,1,1) & (1,0,0) & (1,0,1) & (1,1,0) & (1,1,1) \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

Example 28. The marginal problem for the inflated causal structure $\mathcal{G}'$ depicted in fig. 2 concerns itself with the pre-injectable marginal scenario $\mathcal{M} = \mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}') = \{V_1, \dots, V_{12}\}$. Equipping the observed nodes of $\mathcal{G}'$ with 4-element outcome labels, the incidence matrix M has $16,896$ rows and $16,777,216$ columns.

$$\# \text{ Rows} = |\mathsf{Out}(\mathcal{M})| = \sum_{i=1}^{12} |\mathsf{Out}(V_i)| = \sum_{i=1}^{12} 4^{|V_i|} = 4 \cdot 4^6 + 8 \cdot 4^3 = 16,896$$
$$\# \text{ Columns} = |\mathsf{Out}(\mathcal{J})| = 4^{|\mathcal{J}|} = 4^{12} = 16,777,216$$

With $12 \cdot 4^{12} = 201,326,592$ non-zero entries. For obvious reasons, no attempt is made to reproduce it here.

In order to describe how marginalization can be written as matrix multiplication $M \cdot x = b$, two more quantities need to be described:

Definition 29. 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 outcome of $j \in \mathsf{Out}(\mathcal{J})$. $\mathcal{P}_{\mathcal{J}}$ shares the same indices as the *column* indices of M.

$$\forall j \in \mathsf{Out}(\mathcal{J}) : \mathcal{P}_{\mathcal{J}}^{\intercal}[j] = P_{\mathcal{J}}(j)$$

Definition 30. The **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 $\mathsf{Out}(\mathcal{M})$. $\mathcal{P}^{\mathcal{M}}$ shares the same indices as the *row* indices of M.

$$\forall m \in \mathsf{Out}(\mathcal{M}) : \mathcal{P}^{\mathcal{M}\intercal}[m] = P_{\mathsf{Dom}(m)}(m)$$

Where $\mathsf{Dom}(m) \in \mathcal{M}$ and $P_{\mathsf{Dom}(m)} \in P^{\mathcal{M}}$ as per remark 5.

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

$$\mathcal{P}^{\mathcal{M}} = M \cdot \mathcal{P}_{\mathcal{J}} \tag{10}$$

Remark 31. 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 $\mathcal{P}^{\mathcal{M}}$. Throughout this work, a *canonical ordering* is used: the marginal contexts are written in the same order as listed in the marginal scenario $\mathcal{M}$ and outcomes within a marginal context are enumerated in a *raveling* order; exemplified in example 27.

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. (10) holds?*

Definition 32. 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"[8] is *feasible*, then there exists a vector x than can satisfy eq. (10) and is a valid joint distribution vector. Therefore feasibility implies that $P_{\mathcal{J}} = x$, solving the marginal problem with positive result. Moreover if the marginal linear program is *infeasible*, then there *does not* exist a joint distribution $P_{\mathcal{J}}$ over all random variables.

Definition 33. The Dual Marginal Linear Program is the dual of definition 32 formulated via a procedure similar to [14]:

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

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

B. Infeasibility Certificates

The dual marginal linear program also provides an answer to the marginal problem. To prove this, first notice that the dual problem is *never infeasible*; by choosing y to be trivially the null vector $\emptyset$ of appropriate size, all constraints are satisfied. Secondly if $y \cdot M \succeq 0$ and $x \succeq 0$, then the following must hold if the primal is feasible:

$$y \cdot \mathcal{P}^{\mathcal{M}} = y \cdot M \cdot x \geq 0 \tag{11}$$

Therefore the *sign* of the dual value $d \equiv \min\left(y \cdot \mathcal{P}^{\mathcal{M}}\right)$ solves the marginal problem. If $d < 0$ then eq. (11) is violated and therefore the marginal problem has negative result. Likewise if d satisfies eq. (11), then a joint distribution $P_{\mathcal{J}}$ exists. Before continuing, an important observation needs to be made. 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 [15]. Comment (TC Fraser): Is this really the CSP? Moreover, if $d < 0$, then it is unbounded $d = -\infty$. This latter point becomes clear upon recognizing that for any y such that $d < 0$, another y' can be constructed by multiplying y by a real constant α greater than one such that,

$$y' = \alpha y \mid \alpha > 1 \implies d' = \alpha d < d$$

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

Comment (TC Fraser): Farkas's lemma here?

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

Definition 34. An **infeasibility certificate** [16] is any vector y that satisfies the constraints of definition 33 and also permits violation of eq. (11) for some marginal distribution vector $\mathcal{P}^{\mathcal{M}}$.

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

Furthermore, any y satisfying $y \cdot M \succeq 0$ induces a **certificate inequality** that constraints the space of marginal distribution vectors which takes the symbolic form of eq. (11),

$$y \cdot \mathcal{P}^{\mathcal{M}} \geq 0$$

Where the entries of the certificate y act as coefficients for the entries of $\mathcal{P}^{\mathcal{M}}$.

Todo (TC Fraser): Discuss Infeasibility Certificates basis Todo (TC Fraser): Discuss the certificate inequalities we found.

$$I_{\text{Mosek-Cert}}, I_{\text{Cert}}$$

VIII. COVERING INEQUALITIES

Section VII discusses how one can obtain a valid compatibility inequality that witnesses incompatibility for a particular marginal model $P^{\mathcal{M}}$ by writing the marginal problem as a linear program. It was demonstrated that every incompatible distribution is witness-able by some certificate inequality. However, when a particular marginal model $P^{\mathcal{M}}$ is not known, a linear optimization offers no utility. Todo (TC Fraser): Motivate why certificates are not enough, want many solutions, want logical foundation

In this section, we will first summarize a method that can be used to obtain a complete solution to the *possibilistic* marginal problem and indirectly offer a *partial* solution to the *probabilistic* marginal problem. Doing so corresponds to enumerating all

A. Logical Covers

In this section, we summarize the familiar notion of logical implications as witnesses for contextuality. Let $a \in \mathsf{Out}(\mathcal{M})$ be a marginal outcome and let $C = \{c_1, \ldots, c_{|C|}\} \subseteq \mathsf{Out}(\mathcal{M})$ be a subset of marginal outcomes. The subset C is a **logical cover** of a if the following possibilistic statement holds for *all* marginal models $P^{\mathcal{M}}$.

$$a \implies c_1 \vee \ldots \vee c_{|C|} = \bigvee_{c \in C} c \tag{12}$$

The marginal outcome a will be referred to as the *antecedent* and C as the *consequent set*. A marginal model $P^{\mathcal{M}}$ satisfies eq. (12) if there always exists at least one $c \in C$ that is possible ($P(c) > 0$) whenever a is possible. A marginal model $P^{\mathcal{M}}$ violates eq. (12) whenever *none* of the outcomes in C are possible while a is nonetheless still possible. Marginal models that violate logical statements such as eq. (12) are known as **Hardy Paradoxes** [8, 10, 17]. The concept of witnessing quantum incompatibility on a logical level has been analyzed thoroughly for decades [18, 19]; being motivated by a false sense of robustness compared to probabilistic constraints. Logical covers such as eq. (12) are in one-to-one correspondence with a subset of probabilistic inequalties.

Lemma 35. *Let $a \in \mathsf{Out}(\mathcal{M})$ be some marginal outcome of the marginal scenario $\mathcal{M}$. If any marginal model $P^{\mathcal{M}}$ is to be non-contextual, the following condition must hold.*

$$a \iff \bigvee_{j \in \mathsf{Ext}_{\mathcal{J}}(a)} j \tag{13}$$

The idea being that if a joint distribution exists, then the event a represents partial knowledge of the entire system $\mathcal{J}$; whenever a occurs, exactly one of the extendable joint outcomes has to have occurred. Likewise lemma 35 applies to a collection of marginal outcomes $C \in \mathsf{Out}(\mathcal{M})$,

$$\bigvee_{c \in C} c \iff \bigvee_{c \in C} \bigvee_{j \in \mathsf{Ext}_{\mathcal{J}}(c)} j \tag{14}$$

If a subset C can be constructed such that its extension *covers* the extendable set of a,

$$\mathsf{Ext}_{\mathcal{J}}(a) \subseteq \bigcup_{c \in C} \mathsf{Ext}_{\mathcal{J}}(c) \tag{15}$$

Then the combination of eqs. (13) to (15) yields eq. (12). Hence the motivation for calling eq. (12) a logical cover.

The principle statement of lemma 35 can be taken in a probabilistic form. If the marginal model $P^{\mathcal{M}}$ is to admit a joint distribution $P_{\mathcal{J}}$, then it must be that the probability $P(a)$ of any marginal outcome $a \in \mathsf{Out}(\mathcal{M})$ is equal to a marginalization over its extendable set into $\mathcal{J}$.

$$P(a) = \sum_{j \in \mathsf{Ext}_{\mathcal{J}}(a)} P_{\mathcal{J}}(j) \tag{16}$$

A logical cover $C \subseteq \mathsf{Out}(\mathcal{M})$ of a has probabilistic extension,

$$\sum_{c \in C} P(c) = \sum_{c \in C} \sum_{j \in \mathsf{Ext}_{\mathcal{J}}(c)} P_{\mathcal{J}}(j) \tag{17}$$

Since C is a logical cover of a, the right-hand side of eq. (16) algebraically covers eq. (17)[9]. Therefore one recovers a probabilistic version of eq. (12),

$$P(a) \leq \sum_{c \in C} P(c) \tag{18}$$

Remark 36. It is also possible to recover eq. (18) from eq. (12) directly, by putting eq. (12) into disjunction normal form and assigning it a certain probability,

$$P\left(\neg a \vee \bigvee_{c \in C} c\right) = 1$$

Then applying Boole's indentity and the negative rule $P(\neg a) = 1 - P(a)$ one recovers eq. (18).

Remark 37. The probabilistic logical cover eq. (18) is an *upgraded* version of eq. (12) in the following sense.

- Every non-contextual marginal model $P^{\mathcal{M}}$ satisfies both eq. (18) and eq. (12).
- Every contextual marginal model $P^{\mathcal{M}}$ that violates eq. (12) will also violate eq. (18); but the converse is not necessarily true.

Computationally enumerating all logical covers for a particular antecedent a corresponds to a hypergraph transversal problem [8], an idea we return to in the more general picture to follow.

B. Extension Covers

We will now generalize the notion of a logical cover by considering more exotic antecedents corresponding to multiple marginal outcomes along with weights γ^{A} for each outcome. Afterwards we demonstrate that these generalized covers also admit non-trivial probabilistic compatibility inequalities.

Definition 38. We first define the marginal indicator vector δ_m to be the unit vector indexed by marginal outcomes $\mathsf{Out}(\mathcal{M})$ that is 1 whenever the index is m, and 0 otherwise.

$$\delta_m(m) = 1 \quad \forall m' \neq m : \delta_m(m') = 0$$

Analogously defined is the joint indicator vector δ_j.

$$\delta_j(j) = 1 \quad \forall j' \neq j : \delta_j(j') = 0$$

[9] If there were a joint outcome j^* present on the right-hand side of eq. (16) but not the right-hand side of eq. (17), then eq. (18) would be violated by the deterministic distribution $P_{\mathcal{J}}(j^*) = 1$

We first recall that every non-contextual marginal model $P^{\mathcal{M}}$ can be written as the convex combination of columns of the incidence matrix M.

$$\mathcal{P}^{\mathcal{M}} = M \cdot \mathcal{P}_{\mathcal{J}} = \sum_{j \in \mathsf{Out}(\mathcal{J})} \delta_{m \prec j} P_{\mathcal{J}}(j)$$

Thus we define the set of all non-contextual marginal models as $\mathcal{NC}$,

$$\mathcal{NC} \equiv \left\{ \mathcal{P}^{\mathcal{M}} \mid \mathcal{P}^{\mathcal{M}} = \sum_{j} \mu_j \delta_{j \succ \mathcal{M}}, \mu_j \geq 0, \sum_{j} \mu_j = 1 \right\}$$

Definition 39. Let $\mathcal{M}$ be a marginal scenario and let $m \in \mathsf{Out}(\mathcal{M})$ be a marginal outcome. The **extension of m onto $\mathcal{J}$** is the function $\delta_{m \prec \mathcal{J}} : \mathsf{Out}(\mathcal{J}) \to \{0, 1\}$ that indicates membership of $\mathsf{Ext}_{\mathcal{J}}(m)$.

$$\delta_{m \prec \mathcal{J}}(j) \equiv \begin{cases} 1 & j \prec m \\ 0 & \text{otherwise} \end{cases}$$

As a vector indexed by joint outcomes $j \in \mathsf{Out}(\mathcal{J})$, $\delta_{m \prec \mathcal{J}}$ is precisely the m-th row of the incidence matrix.

Definition 40. Let $\mathcal{M}$ be a marginal scenario and let $j \in \mathsf{Out}(\mathcal{J})$ be a joint outcome. The **restriction of j onto $\mathcal{M}$** is deterministic marginal model $\delta_{j \succ \mathcal{M}} : \mathsf{Out}(\mathcal{M}) \to \{0, 1\}$.

$$\delta_{j \succ \mathcal{M}}(m) \equiv \begin{cases} 1 & j \prec m \\ 0 & \text{otherwise} \end{cases}$$

As a vector indexed by marginal outcomes $m \in \mathsf{Out}(\mathcal{M})$, $\delta_{j \prec \mathcal{M}}$ is precisely the j-th column of the incidence matrix.

Definition 41. A **multiplicity set** or **m-set** γ^S for a marginal scenario $\mathcal{M}$ is a vector indexed by marginal outcomes $m \in \mathsf{Out}(\mathcal{M})$ whose entries are non-negative integers $\gamma^S_m \in \mathbb{Z}_{\geq 0}$. The **support** $\sigma\big(\gamma^S\big)$ of an m-set is the set of marginal outcomes that are assigned strictly positive multiplicity.

$$\sigma\big(\gamma^S\big) \equiv \big\{ m \in \mathsf{Out}(\mathcal{M}) \mid \gamma^S_m \in \mathbb{N} \big\}$$

Definition 42. The **extension of an m-set** γ^S into $\mathcal{J}$ denoted Γ^S is a vector indexed by *joint* outcomes $j \in \mathsf{Out}(\mathcal{J})$ and is defined as follows,

$$\Gamma^S_j \equiv \sum_{s \in \sigma(\gamma^S)} \gamma^S_s \delta_{s \prec j} \tag{19}$$

As an immediate consequence, the support of Γ^S is immediately known,

$$\sigma\big(\Gamma^S\big) = \bigcup_{s \in \sigma(\gamma^S)} \mathsf{Ext}_{\mathcal{J}}(s)$$

Definition 43. An m-set γ^C is an **extension cover** of an m-set γ^A for a marginal scenario $\mathcal{M}$ if and only if the following two conditions hold[10]:

$$\sigma\big(\Gamma^A\big) \subseteq \sigma\big(\Gamma^C\big) \tag{20}$$

$$\forall j \in \sigma\big(\Gamma^A\big) : \Gamma^A_j \leq \Gamma^C_j \tag{21}$$

Identifying extension covers is of great importance because of the following theorem.

Theorem 44. *If the m-set γ^C is an extension cover of the m-set γ^A for a marginal scenario $\mathcal{M}$, then the following is a valid compatibility $I_{\mathcal{M}}$ for the marginal scenario $\mathcal{M}$.*

$$\sum_{a \in \sigma(\gamma^A)} \gamma^A_a P(a) \leq \sum_{c \in \sigma(\gamma^C)} \gamma^C_c P(c) \tag{22}$$

---

[10] In fact eq. (21) implies eq. (20) since $\forall j \notin \sigma\big(\Gamma^A\big) : \Gamma^A_j = 0$.

Proof. To prove that eq. (22) is a valid compatibility inequality $I_{\mathcal{M}}$ for all marginal models $P^{\mathcal{M}}$, we will demonstrate that eq. (22) follows from the assumption that a joint distribution $P_{\mathcal{J}}$ exists.

For each m-set γ^S write the weighted sum over the probabilities $\{P(s)\}_{s\in\sigma(\gamma^S)}$ in terms of the joint distribution $P_{\mathcal{J}}$ using eq. (19).

$$\begin{aligned}\sum_{s\in\sigma(\gamma^S)} \gamma^S_s P(s) &= \sum_{s\in\sigma(\gamma^S)} \gamma^S_s \sum_{j\in\mathsf{Ext}_{\mathcal{J}}(s)} P_{\mathcal{J}}(j)\\ &= \sum_{s\in\sigma(\gamma^S)} \gamma^S_s \sum_{j\in\mathsf{Out}(\mathcal{J})} \delta_{s\prec j} P_{\mathcal{J}}(j)\\ &= \sum_{j\in\mathsf{Out}(\mathcal{J})} \left[\sum_{s\in\sigma(\gamma^S)} \gamma^S_s \delta_{s\prec j}\right] P_{\mathcal{J}}(j)\\ &= \sum_{j\in\sigma(\Gamma^S)} \Gamma^S_j P_{\mathcal{J}}(j)\end{aligned} \tag{23}$$

First examine the substitution $S \to C$,

$$\sum_{c\in\sigma(\gamma^C)} \gamma^C_c P(c) = \sum_{j\in\sigma(\Gamma^C)} \Gamma^C_j P_{\mathcal{J}}(j)$$

By eq. (20) $\sigma(\Gamma^A) \subseteq \sigma(\Gamma^C)$,

$$\sum_{c\in\sigma(\gamma^C)} \gamma^C_c P(c) = \sum_{j\in\sigma(\Gamma^C)} \Gamma^C_j P_{\mathcal{J}}(j) + \sum_{j\in\sigma(\Gamma^C)\setminus\sigma(\Gamma^A)} \Gamma^C_j P_{\mathcal{J}}(j)$$

Notice that the sum over $j \in \sigma(\Gamma^C) \setminus \sigma(\Gamma^A)$ is entirely positive since $\Gamma^C_j \in \mathbb{N}$ and $P_{\mathcal{J}}(j) \geq 0$. Moreover eq. (21) gives $\Gamma^C_j \geq \Gamma^A_j$. Combining these observations along with eq. (23) for $S \to A$, one arrives at the inequality,

$$\sum_{c\in\sigma(\gamma^C)} \gamma^C_c P(c) \geq \sum_{j\in\sigma(\Gamma^A)} \Gamma^A_j P_{\mathcal{J}}(j) = \sum_{a\in\sigma(\gamma^A)} \gamma^A_a P(a)$$

Which is equivalent to eq. (22); every non-contextual marginal model $P^{\mathcal{M}}$ must satisfy eq. (22). ☐

Corollary 45. *Every linear, integer coefficient compatibility inequality that is homogeneous is an extension covering inequality.*

Proof. A proof is easy to see by contradiction. Let $I_{\mathcal{M}} = \left\{\gamma \cdot \mathcal{P}^{\mathcal{M}} \geq 0\right\}$ where $\gamma_m \in \mathbb{Z} | \forall m \in \mathsf{Out}(\mathcal{M})$ be a homogeneous, linear, integer coefficient, compatibility inequality such that it is *not* an extension covering inequality. Next seperate γ into its positive and negative components,

$$\gamma^+_m = \begin{cases}\gamma_m & \gamma_m > 0\\ 0 & \gamma_m \leq 0\end{cases} \qquad \gamma^-_m = \begin{cases}-\gamma_m & \gamma_m < 0\\ 0 & \gamma_m \geq 0\end{cases}$$

Such that both γ^+ and γ^- are valid m-sets. Therefore $I_{\mathcal{M}} = \left\{(\gamma^+ - \gamma^-) \cdot \mathcal{P}^{\mathcal{M}} \geq 0\right\}$. Since this inequality is *not* an extension cover, it must be that there exists at least one joint outcome $j' \in \mathsf{Out}(\mathcal{J})$ such that $\Gamma^-_{j'} > \Gamma^+_{j'}$. Consequently by eq. (23), the deterministic distribution $\mathcal{P}^{\mathcal{M}} = \delta_{j'\succ\mathcal{M}}$ *violates* $I_{\mathcal{M}}$. However since $\delta_{j'\succ\mathcal{M}} \in \mathcal{NC}$ and violates $I_{\mathcal{M}}$, $I_{\mathcal{M}}$ can not be a compatibility inequality; disagreeing with the assumption that it was. ☐

C. Completeness of Extension Covers

Corollary 45 alludes to completeness of extension covers. It vaguely suggests that every non-contextuality inequality needs to be an extension cover, otherwise there would exist a compatible and deterministic distribution that violates it. We will now demonstrate this completeness explicitly by first discussing what is required of a complete solution to the marginal problem.

Definition 46. A **complete solution to the marginal problem** is a finite set of q vectors $\gamma_1, \dots, \gamma_q$ together with q real numbers $k \in \mathbb{R}$

Theorem 47. *The set of all logical covering inequalities forms a sufficient solution to the possibilistic marginal problem.*

Theorem 48. *The set of all extension covering inequalities forms a sufficient solution to the probabilistic marginal problem.*

Proof. □

Definition 49. An extension cover γ^C of γ^A is said to be **minimal** if there does not exist an extension cover $\gamma^{C'}$ different from γ^C such that,

$$\forall j \in \mathsf{Out}(\mathcal{J}) : \gamma_j^{C'} \leq \gamma_j^C$$

If a more minimal extension cover $\gamma^{C'}$ does exists, then the inequality it defines (as per theorem 44 directly implies the inequality defined by γ^C.

Definition 50. An extension cover γ^C of γ^A is said to be **reducible** if there exists a *partition* of γ^C and γ^A,

$$\gamma^C = \gamma^{C_1} + \gamma^{C_2} \qquad \gamma^A = \gamma^{A_1} + \gamma^{A_2}$$

Such that γ^{C_1} is an extension cover for γ^{A_1} and γ^{C_2} is an extension cover for γ^{A_2}. If $\left(\gamma^C, \gamma^A\right)$ does *not* admit a partition, then it is said to be **irreducible**.

Todo (TC Fraser): Basis certificates

D. Extension Cover Procedure

The **extension cover procedure** begins with an antecedent m-set γ^A and iteratively produces consequent m-sets γ^C that are extension covers of γ^A. To facilitate this, we define a **partial consequent m-set** to be any marginal m-set γ^{*C} that *has yet to* become an extension cover of γ^A. Also define the null m-set $\gamma^{\emptyset}$ to be the m-set assigning zero multiplicity to each marginal outcome. As a preliminary insight, it is realatively easy to construct a valid extension cover of γ^A by first beginning with $\gamma^{*C} = \gamma^{\emptyset}$ and iteratively increasing entries in γ^{*C} until eq. (21) is satisfied. To ensure efficient construction, it is wise to first identity a particular joint outcome j' that violates eq. (21) $\Gamma_{j'}^{*C} < \Gamma_{j'}^{A}$, and then append to Γ^{*C} a marginal outcome m ($\Gamma^{*C} \mapsto \Gamma^{*C} + \delta_m$) that is extendable to j'. Eventually, enough marginal outcomes will be added to the partial extension cover Γ^{*C} for eq. (21) to hold.

This procedure can be generalized to find *all* minimal extension covers for a chosen antecdent m-set γ^A and is best formalized as a *weighted hypergraph transversal* problem.

1. *Hypergraphs*

Definition 51. A **hypergraph** denoted $\mathcal{H}$ is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *hyperedges* respectively where the nodes can represent any object and the hyperedges are *subsets* of nodes. For convenience of notation, one defines an index set over the nodes and hyperedges of a hypergraph $\mathcal{H}$ denoted $\mathcal{I}_{\mathcal{N}}$ and $\mathcal{I}_{\mathcal{E}}$ respectively.

$$\mathcal{H} = (\mathcal{N}, \mathcal{E}) \quad \mathcal{N} = \{n_i \mid i \in \mathcal{I}_{\mathcal{N}}\} \quad \mathcal{E} = \{e_i \mid i \in \mathcal{I}_{\mathcal{E}}, e_i \subseteq \mathcal{N}\}$$

Note that whenever the hyperedge or node index is arbitrary, it will be omitted. There is a dual correspondence between hyperedges $e \in \mathcal{E}$ and nodes $n \in \mathcal{N}$ in a Hypergraph. A hyperedge e is viewed as a set of nodes $\{n_i\}$, and a node n can be viewed as the set of hyperedges $\{e_i\}$ that contain it.

Definition 52. A **hypergraph transversal** (or edge hitting set) $\mathcal{T}$ of a hypergraph $\mathcal{H}$ is a set of nodes $\mathcal{T} \subseteq \mathcal{N}$ that have non-empty intersections with every hyperedge in $\mathcal{E}$.

$$\mathcal{T} = \{n_i \in \mathcal{N} \mid i \in \mathcal{I}_{\mathcal{T}}\} \quad \forall e \in \mathcal{E} : \mathcal{T} \cap e \neq \emptyset$$

Definition 53. A **necessary node** of a transversal $\mathcal{T}$ is a node n such that $\mathcal{T} \setminus n$ is no longer a valid transversal. The set of all necessary nodes is denoted $\mathsf{Nec}(\mathcal{T})$,

$$\mathsf{Nec}(\mathcal{T}) = \{n \in \mathcal{T} \mid \exists e \in \mathcal{E} : (\mathcal{T} \setminus n) \cap e = \emptyset\}$$

An **unnecessary node** of a transversal $\mathcal{T}$ is any node that is *not* necessary. The set of all unnecessary nodes is denoted $\mathsf{UnNec}(\mathcal{T})$,

$$\mathsf{UnNec}(\mathcal{T}) = \mathcal{T} \setminus \mathsf{Nec}(\mathcal{T})$$

A **minimal hypergraph transversal** $\mathcal{T}$ is any valid transversal of $\mathcal{H}$ where every node n is *necessary*.

$$\mathcal{T} = \mathsf{Nec}(\mathcal{T})$$

Definition 54. A **weighted hypergraph** $\mathcal{H}_{\mathcal{W}}$ is a regular hypergraph satisfying definition 51 equipped with a set of weights $\mathcal{W}$ ascribed to each node such that a weighted hypergraph is written as a triplet $(\mathcal{W}, \mathcal{N}, \mathcal{E})$.

$$\mathcal{W} = \{w_i \mid i \in \mathcal{I}_{\mathcal{N}}, w_i \in \mathbb{R}\}$$

One would say that a particular node n_i carries weight w_i for each $i \in \mathcal{I}_{\mathcal{N}}$.

Definition 55. A **bounded transversal** of a weighted hypergraph $\mathcal{H}_{\mathcal{W}}$ is a transversal $\mathcal{T}$ of the unweighted hypergraph $\mathcal{H}$ and a real number t (denoted $\mathcal{T}_{\leq t}$) such that the sum of the node weights of the transversal is bounded by t.

$$\mathcal{T}_{\leq t} = \{n_i \mid i \in \mathcal{I}_{\mathcal{T}}\} \quad \text{s.t.} \sum_{j \in \mathcal{I}_{\mathcal{T}}} w_j \leq t$$

One can definte analogous **(strictly) upper/lower bounded transversals** by considering modifications of the notation: $\mathcal{T}_{<t}, \mathcal{T}_{\geq t}, \mathcal{T}_{>t}$.

Todo (TC Fraser): Trivial nodes Todo (TC Fraser): Duplication under Deflation Todo (TC Fraser): Memory Todo (TC Fraser): Parallelization Todo (TC Fraser): Simplicial Complex

2. *Advantages*

3. *Avoiding Trivial Inequalities*

A **trivial compatibility inequality** $I_{\mathcal{M}}$ is an inequality that is satisfied by all marginal models $P^{\mathcal{M}}$, not just compatible ones. Trivial compatibility inequalities are of no importance Since the set of all extension cover inequalities Todo (TC Fraser): Targeted Seeds

Definition 56. A **hypergraph transversal generation** is any algorithm that correctly generates the complete set of all minimal transversals of $\mathcal{H}$.

Definition 57. Any hypergraph with strictly non-empty edges $\forall e \neq \emptyset$ will always admit the **trivial transversal** $\mathcal{T}^*$ where all nodes are considered as members $\mathcal{T}^* = \mathcal{N}$. Any hypergraph with empty edges will be called a **degenerate hypergraph** as it admits no transversals[11].

Remark **??** guarantees that all marginal hypergraphs are non-degenerate. There are two distinct approaches to hypergraph transversal generation: *top-down* and *bottom-up*.

Definition 58. A **top-down** hypergraph transversal generation refers to any algorithm that begins with the trivial transversal $\mathcal{T}^*$ and iteratively removes unnecessary nodes from $\mathcal{T}^*$.

Definition 59. A **bottom-up** hypergraph transversal generation refers to any algorithm that begins with the empty set $\emptyset$ and iteratively adds nodes.

One should select a top-down method if the typical size of minimal transversals $|\mathcal{T}|$ is comparable to the number of nodes $|\mathcal{N}|$, otherwise a bottom-up method will perform better. For our purposes we implemented a deep-first transversal algorithm similar to Recalling **??**

---

[11] Most authors require that all hypergraphs be non-degenerate [20].

E. Weighted Hypergraph Transversals

In section VII is was demonstrated that the Fritz distribution is witness-able via a certificate inequality. It is also possible to witness Todo (TC Fraser): Discuss the Inequalities Derived/ Trivial and non-trivial

Todo (TC Fraser): Weighted transversals and Optimizations Todo (TC Fraser): Seeding inequalities (huge advantage here)

IX. DERIVING SYMMETRIC INEQUALITIES

Symmetric compatibility inequalities are useful for a number of reasons. First, Bancal et. al. [21] discuss computational advantages in considering symmetric versions of marginal polytopes mentioned in section III; 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 [11]) can be written in a way that is symmetric under the exchange of parties, demonstrating that quantum-non-trivial inequalities can be recovered from facets of a symmetric polytope. Second, numerical optimizations against inequalities invariant under exchange of parties will lead to one of two interesting cases: either the extremal distribution is invariant under exchange of parties or it is not. The latter case generates a family of incompatible distributions obtained by exchanging parties of the found extremal distribution[12]. In this section we discuss how to achieve computational advantage and its application to the inequality techniques mentioned in section VII and **??**.

A. Causal Symmetry

The aim of this section is to formalize what types of symmetries are present in a particular causal structure $\mathcal{G}$. The group of these causal symmetries will define, for each compatibility inequality $I_{\mathcal{M}}$, a family of inequalities that are also valid incompatibility inequalities.

First consider the permutation group acting on a set of nodes $N \subseteq \mathcal{N}$ denoted $\mathsf{Perm}(N)$ to be the set of all bijective maps φ from N to N.

$$\mathsf{Perm}(N) = \{\varphi \mid \varphi : N \to N\}$$

The action of $\varphi \in \mathsf{Perm}(N)$ on a causal structure $\mathcal{G}$ is defined via an the extension of φ to the corresponding element in $\mathsf{Perm}(\mathcal{N})$ that leaves nodes $n \in \mathcal{N} \setminus N$ invariant,

$$\begin{aligned} \varphi[\mathcal{G}] &\equiv (\varphi[\mathcal{N}], \varphi[\mathcal{E}]) \\ \varphi[\mathcal{N}] &\equiv \{\varphi[n] \mid n \in \mathcal{N}\} = \mathcal{N} \\ \varphi[\mathcal{E}] &\equiv \{\varphi[e] \mid e \in \mathcal{E}\} \end{aligned}$$

To motivate a more general treatment of such symmetries, consider the Triangle Scenario of fig. 1. Due to the rotational and reflective symmetries of the Triangle Scenario, any permutation $\varphi \in \mathsf{Perm}(\mathcal{N}_O)$ of the observable nodes $\mathcal{N}_O = \{A, B, C\}$ creates a new causal structure $\varphi[\mathcal{G}]$ that is *equivalent* to $\mathcal{G}$ up to a relabeling of *latent* nodes. By construction, any valid compatibility inequality $I_{\mathcal{M}}$ for the marginal scenario $\mathcal{M} = \{V_1, \ldots, V_k\}$ is *independent* of the labeling of latent nodes. Therefore applying φ to all distributions in $I_{\mathcal{M}}$ yields $\varphi[I_{\mathcal{M}}] \equiv \varphi[I]_{\varphi[\mathcal{M}]}$, another valid compatibility inequality for the permuted marginal scenario $\varphi[\mathcal{M}]$ defined as,

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

More generically, any permutation φ that preserves the graphical structure of $\mathcal{G}$ can be applied to $I_{\mathcal{M}}$ to generate new and valid compatibility inequalities. Specifically this alludes to the **graph automorphism group** $\mathsf{Aut}(\mathcal{G})$ of $\mathcal{G}$.

$$\mathsf{Aut}(\mathcal{G}) \equiv \{\varphi \in \mathsf{Perm}(\mathcal{N}) \mid \varphi[\mathcal{G}] = \mathcal{G}\}$$

In general, $\mathsf{Aut}(\mathcal{G})$ could include elements φ that map latent nodes to observable nodes[13]. This behaviour is *undesired* for causal inference where the latent nodes will never appear in a marginal scenario. Instead, the subgroup of $\mathsf{Aut}(\mathcal{G})$ that never maps observable nodes to latent nodes is the **set-wise stabilizer** of $\mathsf{Aut}(\mathcal{G})$ for $\mathcal{N}_O$ and will be expressed with a subscript.

---

[12] If the extremal distribution is not invariant under exchange of parties, there is indication that the space of accessible distributions is non-convex.

[13] In practice however, it is rare to be considering a causal structure $\mathcal{G}$ where some latent node $n_L \in \mathcal{N}_L$ takes on an role indistinguishable from some observable node $n_O \in \mathcal{N}_O$, so this won't be an issue.

Definition 60. The **causal symmetry group** for a causal structure $\mathcal{G}$ is the graph automorphism subgroup that stabilizes the observable nodes.

$$\mathsf{Aut}_{\mathcal{N}_O}(\mathcal{G}) = \{\varphi \in \mathsf{Perm}(\mathcal{N}) \mid \varphi[\mathcal{G}] = \mathcal{G}, \forall n \in \mathcal{N}_O : \varphi[n] \in \mathcal{N}_O\}$$

Given any compatibility inequality $I_{\mathcal{M}}$, each element φ of the causal symmetry group defines a new valid compatibility inequality $\varphi[I_{\mathcal{M}}] = \varphi[I]_{\varphi[\mathcal{M}]}$.

The inflation technique discussed in section VI allows one to derive compatibility inequalities for a causal structure $\mathcal{G}$ by considering the marginal problem over the pre-injectable sets of $\mathcal{G}'$ denoted $\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}') = \{V_1, \ldots, V_k\}$. It is important to recall that due to lemma 25, inequalities I' for $\mathcal{G}'$ are *only* transferable to inequalities I for $\mathcal{G}$ if I' is in terms of distributions over the pre-injectable sets $I'_{\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')}$. As a consequence of this observation, we need to consider a subgroup of the causal symmetry group that preserves the pre-injectable sets,

$$\varphi\big[\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')\big] = \mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}') \tag{24}$$

Definition 61. The **restricted causal symmetry group** Φ for a causal structure $\mathcal{G}$ and a marginal scenario $\mathcal{M}$ is the graph automorphism subgroup that stabilizes the marginal scenario $\mathcal{M}$.

$$\Phi \equiv \mathsf{Aut}_{\mathcal{M}}(\mathcal{G}) = \{\varphi \in \mathsf{Perm}(\mathcal{N}) \mid \varphi[\mathcal{G}] = \mathcal{G}, \forall V \in \mathcal{M} : \varphi[V] \in \mathcal{M}\}$$

Given any compatibility inequality $I_{\mathcal{M}}$ each element φ of the causal symmetry group defines a new valid compatibility inequality $\varphi[I_{\mathcal{M}}] = \varphi[I]_{\mathcal{M}}$ over the *same* marginal context $\mathcal{M}$.

The restricted causal symmetry group for the pre-injectable sets of an inflation $\mathcal{G}'$ is the ideal symmetry group for use with the inflation technique. If one obtains a particular inequality $I'_{\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')}$ as well as the restricted causal symmetry group Φ, an entire family of valid compatibility inequalities are also obtained.

$$\left\{\varphi[I']_{\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')} \mid \varphi \in \Phi\right\} \text{ where } \Phi = \mathsf{Aut}_{\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')}(\mathcal{G}')$$

We have obtained Φ for the inflated triangle scenario fig. 2. Φ is an order 48 group with the following 4 generators:

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

It is easy to verify that $\varphi_1, \varphi_2, \varphi_3, \varphi_4$ are all automorphisms of the inflated Triangle Scenario $\mathcal{G}'$ of fig. 2. Moreover they stabilize the observable nodes, and the map pre-injectable sets of eq. (9) to the pre-injectable sets. Finally, in the Triangle Scenario we have that,

$$\Phi = \mathsf{Aut}_{\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')}(\mathcal{G}') \sim \mathsf{Aut}_{\mathcal{N}_O}(\mathcal{G})$$

To see this, φ_1 and φ_4 become the identity element $\mathbb{I}$ in $\mathsf{Aut}_{\mathcal{N}_O}(\mathcal{G})$ upon removal of the copy-index, leaving φ_2 to generate reflections and φ_3 to generate rotations.

B. Symmetric Marginal Polytope

Up until this point, we have discussed how new inequalities can be created from known ones by exploiting causal symmetry. Although very useful, we divert our attention to finding *symmetric* inequalities themselves.

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

In order to do this, we must first define how elements $\varphi \in \Phi$ act on marginal outcomes $O_{\mathcal{M}}$.

Definition 62. Let $\mathcal{J} = \{v_1, \ldots, v_k\}$ be a collection of random variables and let $\varphi \in \mathsf{Perm}(\mathcal{J})$. $\mathcal{J}$ is said to be **label invariant with respect to φ** if the outcome labels for each variable $v \in \mathcal{J}$ are unchanged under the action of φ.

$$\forall v \in \mathcal{J} : O_v = O_{\varphi[v]}$$

Label invariance is a prerequisite for any kind of symmetry amoung random variables.

Definition 63. Let $\mathcal{J} = (v_1, \ldots v_k)$ be a collection of random variables along with a specific outcome $m \in \mathsf{Out}(V)$ where $V \subseteq \mathcal{J}$. The action of $\varphi \in \mathsf{Perm}(\mathcal{J})$ on m is a new outcome $\varphi[m] \in \mathsf{Out}(\varphi[V])$ over $\varphi[V] \subseteq \mathcal{J}$ defined as follows,

$$\forall v \in V : (\varphi[m])(\varphi[v]) \equiv m(v)$$

The idea being that φ modifies m in such a way that permutes its domain while leaving its range invariant.

Example 64. Consider the outcome $m = \{A \mapsto 0, B \mapsto 1, C \mapsto 2\} \in \mathsf{Out}(\{A, B, C\})$ where all variables have 4 possible outcomes $O_A = O_B = O_C = \{0, 1, 2, 3\}$ and the permutation $\varphi = \{A \to D, B \to C, C \to B, D \to A\}$. The *action* of φ on m is as follows,

$$\begin{aligned}
\varphi[m] &= \varphi[\{A \mapsto m(A), B \mapsto m(B), C \mapsto m(C)\}] \\
&= \{\varphi[A] \mapsto m(A), \varphi[B] \mapsto m(B), \varphi[C] \mapsto m(C)\} \\
&= \{D \mapsto m(A), C \mapsto m(B), B \mapsto m(C)\} \\
&= \{D \mapsto 0, C \mapsto 1, B \mapsto 2\} \\
&= \{B \mapsto 2, C \mapsto 1, D \mapsto 0\}
\end{aligned}$$

Evidently, $\varphi[m] \in \mathsf{Out}(\{B, C, D\})$.

Through repeated action of $\varphi \in \Phi$ on marginal outcomes $m \in \mathsf{Out}(\mathcal{M})$ and joint outcomes $j \in \mathsf{Out}(\mathcal{J})$, one can define group orbits of Φ in $\mathsf{Out}(\mathcal{M})$ and $\mathsf{Out}(\mathcal{J})$.

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

Definition 65. The **symmetric incidence matrix** M_{Φ} for a marginal scenario $\mathcal{M}$ and a particular restricted causal symmetry Φ is a contracted version of the incidence matrix M for $\mathcal{M}$. Each row of M_{Φ} corresponds to a marginal orbit $\mathsf{Orb}_{\Phi}(m)$. Analogously each column of M_{Φ} corresponds to a joint orbit $\mathsf{Orb}_{\Phi}(j)$. The entries of M_{Φ} are integers and correspond to summing over the rows and columns of M that belong to each orbit.

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

C. Results

Todo (TC Fraser): How we obtained the desired symmetry group Todo (TC Fraser): Group orbits to symmetric marginal description matrix Todo (TC Fraser): representative antecedents completely solve the marginal problem

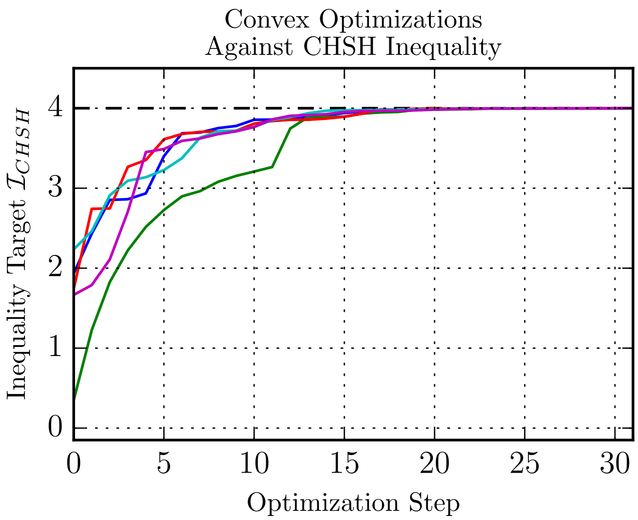

FIG. 3. Convex optimizations against I_{CHSH} recover algebraic violation of 4.

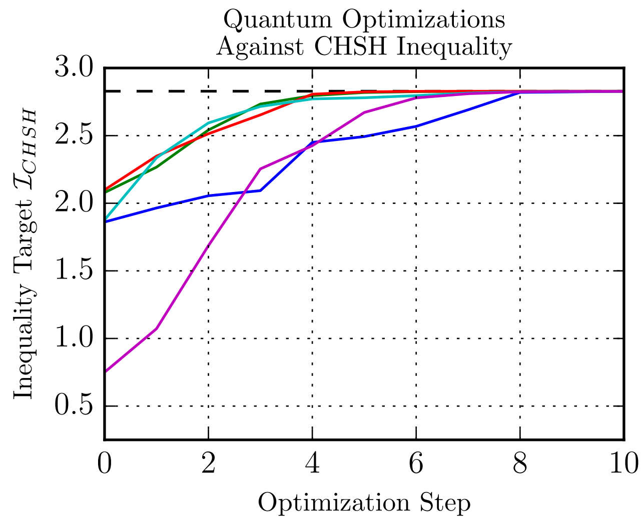

FIG. 4. Quantum optimizations against I_{CHSH} recover maximum violation of $2\sqrt{2}$.

X. NON-LINEAR OPTIMIZATIONS

Todo (TC Fraser): Mention outlook on te importance of classification and determining interesting inequalities [21] Compatibility inequalities for a given causal structure are fantastic for finding incompatible distributions. In the inflation technique, this is no exception. Parameterizing a space distributions using a set of real-valued parameters λ, enables us to perform numerical optimizations against these inequalities in hopes that a particular set of parameters λ is able to generate an incompatible distribution P. To illustrate this generic procedure and it's reliability, we will first examine the popular CHSH inequality.

A. Numerical Violations of The CHSH Inequality

The CHSH inequality [11] can we viewed as a causal compatibility inequality for the iconic Bell Scenario (Fig. 19 of [22], Fig. 11 of [8], Fig. 1 a) of [23], etc.) corresponding to Bell's notion of local causality [22]. It constrains the set of 2-outcome bipartite distributions over local binary measurement settings for each party $P_{AB|S_AS_B} \equiv \{P_{AB|00}, P_{AB|01}, P_{AB|10}, P_{AB|11}\}$. Numerical optimization *should* obtain the algebraic violation associated with the PR-Box correlations [24]. Maintaining full generality, we simply need to parameterize these 4 distributions using eq. (E2), each requiring 4 real-valued parameters. We define the optimization target for the CHSH inequality to be the left-hand-side of eq. (5),

$$I_{\mathrm{CHSH}} = \langle AB|11\rangle + \langle AB|12\rangle + \langle AB|21\rangle - \langle AB|22\rangle$$

Figure 3 demonstrates this optimization for 5 random seed parameters λ^0, each converging to the expected value of 4. Analogously, [12]

Todo (TC Fraser): Demonstrate Quantum, Convexity Todo (TC Fraser): Why Inequalities are great for optimizations Todo (TC Fraser): Non-linearity Todo (TC Fraser): Techniques Used Todo (TC Fraser): Finding maximum violation of CHSH easily Todo (TC Fraser): Unreliance when number of parameters increases Todo (TC Fraser): Issues with local minimum Todo (TC Fraser): Using initial conditions close to fritz, obtain greater violation Todo (TC Fraser): Greater violation shares possibilistic structure of fritz and violates CHSH under definition Todo (TC Fraser): Not realizable with maximally entangled qubit states Todo (TC Fraser): Not realizable with separable measurements Todo (TC Fraser): Many non-trivial inequalities to be tested Todo (TC Fraser): inequality -> dist -> inequality evolution

XI. CONCLUSIONS

Todo (TC Fraser): Inflation technique allows one to witness fritz incompatibility Todo (TC Fraser): Linear optimization induces certificates which are incompatibility witnesses Todo (TC Fraser): There are quantum distributions in the Triangle Scenario that are incompatible and different from fritz in terms of entanglement but not possibilistic structure

XII. OPEN QUESTIONS & FUTURE WORK

The marginal problem has be well studied in various contexts [1, 8, 25, . . .]. Reference [1] contains an excellent summary of its connection to other fields of research including knowledge integration, database theory and coalition games in game theory. Todo (TC Fraser): Lots of stuff

Appendix A: Exemplary Inequalities

Appendix B: Connections to Sheaf-Theoretic Treatment

Appendix C: Computationally Efficient Parametrization of the Unitary Group

Spengler, Huber and Hiesmayr [26] suggest the parameterization of the unitary group $\mathcal{U}(d)$ using a $d \times d$-matrix of real-valued parameters $\lambda_{n,m}$,

$$U = \left[\prod_{m=1}^{d-1}\left(\prod_{n=m+1}^{d}\exp(iP_n\lambda_{n,m})\exp(i\sigma_{m,n}\lambda_{m,n})\right)\right]\cdot\left[\prod_{l=1}^{d}\exp(iP_l\lambda_{l,l})\right] \tag{C1}$$

Where P_l are one-dimensional projective operators,

$$P_l = |l\rangle\langle l| \tag{C2}$$

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

$$\sigma_{m,n} = -i|m\rangle\langle n| + i|n\rangle\langle m|$$

Where $1 \leq m < n \leq d$. Spengler et. al. proved the validity of eq. (C1) in Ref. [26].

For the sake of reference, let us label the matrix exponential terms in eq. (C1) in a manner that corresponds to their affect on an orthonormal basis $\{|1\rangle, \dots, |d\rangle\}$.

$$\begin{aligned} GP_l &= \exp(iP_l\lambda_{l,l}) \\ RP_{n,m} &= \exp(iP_n\lambda_{n,m}) \\ R_{m,n} &= \exp(i\sigma_{m,n}\lambda_{m,n}) \end{aligned} \tag{C3}$$

It is possible to remove the reliance on matrix exponential operations in eq. (C1) by utilizing the explicit form of the exponential terms in eq. (C3). As a first step, recognize the defining property of the projective operators eq. (C2),

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

This greatly simplifies the global phase terms GP_l,

$$GP_l = \exp(iP_l\lambda_{l,l}) = \sum_{k=0}^{\infty}\frac{(iP_l\lambda_{l,l})^k}{k!} = \mathbb{I} + \sum_{k=1}^{\infty}\frac{(i\lambda_{l,l})^k}{k!}P_l^k = \mathbb{I} + P_l\left[\sum_{k=1}^{\infty}\frac{(i\lambda_{l,l})^k}{k!}\right] = \mathbb{I} + P_l\left(e^{i\lambda_{l,l}} - 1\right) \tag{C4}$$

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

$$RP_{n,m} = \cdots = \mathbb{I} + P_n\left(e^{i\lambda_{n,m}} - 1\right) \tag{C5}$$

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

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

One can verify that the following properties hold,

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

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

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

$$R_{m,n} = \mathbb{I} + (|m\rangle\langle m| + |n\rangle\langle n|)(\cos\lambda_{n,m} - 1) + (|m\rangle\langle n| - |n\rangle\langle m|)\sin\lambda_{n,m} \tag{C6}$$

By combining the optimizations of eqs. (C5) to (C4) together we arrive at an equivalent form for eq. (C1) that is computational more efficient.

$$U = \left[\prod_{m=1}^{d-1}\left(\prod_{n=m+1}^{d} RP_{n,m}R_{m,n}\right)\right]\cdot\left[\prod_{l=1}^{d} GP_l\right] \tag{C7}$$

In quantum mechanics, the global phase of a state $|\psi\rangle \in \mathcal{H}^n$ is a *redundant* parameter. Parameterizing unitaries using eq. (C7) is especially attractive since the global phase terms GP_l can be dropped, allowing one to parameterize all unitaries in $\mathcal{U}(d)$ up to this degeneracy [26][14].

$$U_{/GP_l} = \left[\prod_{m=1}^{d-1}\left(\prod_{n=m+1}^{d} RP_{n,m}R_{m,n}\right)\right] \tag{C8}$$

Todo (TC Fraser): Explanation of Computational Complexity $\mathcal{O}(d^3)$ vs. $\mathcal{O}(1)$ using [27] Todo (TC Fraser): Pre-Caching for Fixed dimension d Todo (TC Fraser): Talk about inverse via haar measure

Appendix D: Parametrization of Quantum States & Measurements

Throughout section X, we utilize a variety of parameterizations of quantum states and measurements in order to generate quantum-accessible probability distributions. There are numerous techniques that can used when parameterizing quantum states and measurements [26, 28–31] with applications Todo (TC Fraser): Finish this sentence. For our purposes, we need to parameterize the space of quantum-accessible distributions P_Q that are *realized* on the Triangle Scenario. We have implemented P_Q under the following description.

$$P_{ABC}(abc) = \mathrm{Tr}[\Omega^{\intercal}\rho_{AB}\otimes\rho_{BC}\otimes\rho_{CA}\Omega M_{A,a}\otimes M_{B,b}\otimes M_{C,c}] \tag{D1}$$

1. Quantum States

The bipartite states $(\rho_{AB}, \rho_{BC}, \rho_{CA})$ of eq. (D1) were taken to be two-qubit density matrices acting on $\mathcal{H}^2\otimes\mathcal{H}^2$.[15] The space of all such states corresponds to the space of all 4×4 positive semi-definite hermitian matrices with unitary

[14] In our implementation, we accomplish this by explicitly setting $\lambda_{l,l} = 0$ in eq. (C4)

[15] We also considered qutrit $\mathcal{H}^3$ qutit $\mathcal{H}^4$ states. However for 6 d-dimensional $\mathcal{H}^d$ states, the joint density matrix ρ acts on $\left(\mathcal{H}^d\right)^{\otimes 6}$ making it a $\left(d^6, d^6\right)$ matrix with d^{12} entries. Computationally only $d = 2$ was feasible for our optimization tasks.

trace. Throughout this section, we refer to these bipartite states simply as ρ unless otherwise indicated. There are three distinct techniques that we have considered.

Taking inspiration from [31], we can parameterize all such density matrices ρ using **Cholesky Parametrization** [32]. The Cholesky decomposition allows one to write any hermitian positive semi-definite matrix ρ in terms of a lower (or upper) triangular matrix T using $\rho = T^\dagger T$. Our Cholesky parameterization consists of assigning 16 real-valued parameters λ to the entires of T and generating a unitary trace ρ similar to eq. (4.4) of [31].

$$\rho = \frac{T^\dagger T}{\mathrm{Tr}(T^\dagger T)} \quad T = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ \lambda_2 + i\lambda_3 & \lambda_4 & 0 & 0 \\ \lambda_5 + i\lambda_6 & \lambda_7 + i\lambda_8 & \lambda_9 & 0 \\ \lambda_{10} + i\lambda_{11} & \lambda_{12} + i\lambda_{13} & \lambda_{14} + i\lambda_{15} & \lambda_{16} \end{bmatrix} \tag{D2}$$

Our deviation from exclusiving using eq. (D2) is two-fold. First, eq. (D2) is degenerate in that the normalization indicates only $16 - 1 = 15$ parameters are required for fully generic parameterization of all such states ρ. Removing this degeneracy is possible although difficult. Second, the parameters λ_i carry no physical meaning associated with the state ρ, unlike our next parameterization.

In Spengler, Huber and Hiesmayr's work [26], they discuss how to parameterization density matrices ρ acting on $\mathcal{H}^d$ of rank k through it's spectral decomposition,

$$\rho = \sum_{i=1}^{k} p_i |\psi_i\rangle\langle\psi_i| \quad p_i \geq 0, \sum_i p_i = 1, k \leq d \tag{D3}$$

Where any orthonormal basis $\{|\psi_i\rangle\}$ of $\mathcal{H}^d$ can be transformed into a computational basis $\{|i\rangle\}$ by a unitary $U \in \mathcal{U}(d)$ such that $|\psi_i\rangle = U|i\rangle$. We refer to eq. (D3) as the **Spengler Parametrization**. Without loss of generality we parameterize all full-rank ($k = d$) matrices by simultaneously parameterizing the $d = 4$ eigenvalues p_i of eq. (D3) using eq. (E1) and the unitary group $\mathcal{U}(4)$ up to global phase equivalence using eq. (C8). Parameterizing ρ using the Spengler parameterization requires $3 + 12 = 15$ parameters; admitting no degeneracies.

Finally in cases where we wish to restrict ourselves to *pure* bipartite states $\rho = |\psi\rangle\langle\psi|$, we have the luxury to use a **Schmidt Parametrization**. This is accomplished via a Schmidt decomposition $|\psi_{AB}\rangle = \sum_i \sigma_i |i_A\rangle \otimes |i_B\rangle$ where normalization demands that $\sum_i \sigma_i^2 = 1$, $\{|i_A\rangle\}$ and $\{|i_B\rangle\}$ are orthonormal bases for $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively [33]. Additionally for qubit sources we can *choose* our orthonormal bases to be the computational basis $\{|0\rangle, |1\rangle\}$ and write,

$$|\psi\rangle = \cos^2(\lambda_1)|0\rangle \otimes |0\rangle + \sin^2(\lambda_1)|1\rangle \otimes |1\rangle$$

Where only 1 real-valued parameter $\lambda = \{\lambda_1\}$ is required to parameterize all pure states up to local unitaries. Pure states are also attractive due to their computational advantage in computing eq. (D1). If each state ρ is decomposable into $|\psi\rangle\langle\psi|$, then eq. (D1) can be written as,

$$P_{ABC}(abc) = \langle \psi_{AB}\psi_{BC}\psi_{CA}| \Omega M_{A,a} \otimes M_{B,b} \otimes M_{C,c} \Omega^\intercal |\psi_{AB}\psi_{BC}\psi_{CA}\rangle \tag{D4}$$

Avoiding the expensive matrix multiplications of eq. (D1).

2. Measurements

With full generality, we consider a measurement M to be a **projective-operator valued measure (POVM)** represented by a set of hermitian, positive semi-definite operators $\{M_i\}_{i=1,\ldots,k}$[16] acting on $\mathcal{H}^d$ summing to the identity,

$$\forall |\phi\rangle \in \mathcal{H}^d : \langle\phi|M_i|\phi\rangle \geq 0 \quad \sum_{i=1}^{k} M_i = \mathbb{I}_{\mathcal{H}^d} \tag{D5}$$

When considering $k = 2$ outcome measurements acting on $\mathcal{H}^4$ we parameterize the first POVM element M_1 by using a Cholesky parameterization similar to eq. (D2) without normalizing for trace. Afterwards, M_2 is fully determined by eq. (D5).

$$M_1 = T^\dagger T \quad M_2 = \mathbb{I} - T^\dagger T \tag{D6}$$

[16] The i-th element of M is referenced using a subscript M_i. The set of measurement elements for a particular party X will be written M_X. When both party and element are to be referenced, we write $M_{X,i}$.

However, in order for M_2 to be positive semi-definite, the largest eigenvalue of M_1 has to be less than 1. To see this is a necessary and sufficient constraint, first expand out eq. (D6),

$$\langle\phi|M_2|\phi\rangle = \langle\phi|\mathbb{I} - M_1|\phi\rangle = |\phi|^2 - \langle\phi|\left(\sum_{i=1}^{d} m_1^{(i)}\left|m_1^{(i)}\right\rangle\left\langle m_1^{(i)}\right|\right)|\phi\rangle$$

Next write a generic $|\phi\rangle \in \mathcal{H}^d$ in terms of a linear combination of the eigenvectors of M_1[17].

$$\langle\phi|M_2|\phi\rangle = \sum_j \left|\left\langle\phi\middle|m_1^{(j)}\right\rangle\right|^2 - \sum_i m_1^{(i)}\left|\left\langle\phi\middle|m_1^{(i)}\right\rangle\right|^2 = \sum_i\left(1 - m_1^{(i)}\right)\left|\left\langle\phi\middle|m_1^{(i)}\right\rangle\right|^2 \tag{D7}$$

Since $|\phi\rangle$ is arbitrary, for each i set $|\phi\rangle = \left|m_1^{(i)}\right\rangle$ to see that each eigenvalue of M_1 needs to be less than 1.

$$\left\langle m_1^{(i)}\middle|M_2\middle|m_1^{(i)}\right\rangle = \left(1 - m_1^{(i)}\right) \geq 0 \implies m_1^{(i)} \leq 1 \tag{D8}$$

By eq. (D7) is not difficult to see that eq. (D8) is a sufficient condition. During optimization, eq. (D8) can either be enforced passively as a constraint or directly by normalizing M_1 by its largest eigenvalue $\max\left(m_1^{(i)}\right)$ whenever necessary.

When generalizing the above parameterization to more than 2 outcomes, only necessary conditions were found. Generating k-outcome POVM measurements is doable using rejection sampling techniques such as those used in [34] however a valid parameterization with little to no degeneracy was not found. Upon making this observation, a necessary departure to **projective-valued measures (PVMs)** is warranted[18]. With loss of generality, consider the set of PVMs M satisfying eq. (D5) in addition to orthogonal and projective properties,

$$M_i M_j = \delta_{ij} M_i \quad M_i = |m_i\rangle\langle m_i| \tag{D9}$$

Parameterizing M for k-outcome measurements corresponds to parameterizing the set of all k-th order orthonormal sub-bases of $\mathcal{H}^d$. First note that any such basis $\{|\psi_1\rangle, \ldots, |\psi_k\rangle\}$ can be transformed into the computational basis $\{|1\rangle, \ldots, |k\rangle\}$ by a unitary denoted $U \in \mathcal{U}(d)$,

$$U|\psi_i\rangle = |i\rangle$$

With this observation we just need to parameterize the set of all unitaries $\mathcal{U}(d)$,

$$M = \left\{U|i\rangle\langle i|U^\dagger\right\}_{i\in 1,\ldots,k}$$

Specifically, the projective property each M_i means that the global phase of U is completely arbitrary; one only needs to consider parameterizing unitaries up to global phase eq. (C8). This method was inspired by the *measurement seeding* method of Pál and Vértesi's [36] iterative optimization technique.

Analogously to eq. (D4), projective measurements offer considerable computational advantage as eq. (D1) can be rewritten as,

$$P_{ABC}(abc) = \langle m_{A,a} m_{B,b} m_{C,c}|\Omega^\intercal \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA}\Omega|m_{A,a} m_{B,b} m_{C,c}\rangle \tag{D10}$$

3. Network Permutation Matrix

Finally, we introduce the **network permutation matrix** Ω for the Triangle Scenario of fig. 1. For bipartite qubit states, Ω becomes a 64×64 bit-wise matrix that acts on the measurements M and is depicted in fig. 5. To illuminate its necessity, consider eq. (D1) without Ω.

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

[17] The eigenvectors of M_1 form an orthonormal basis for $\mathcal{H}^d$ because M_1 is Hermitian

[18] Strictly speaking, when the number of outcomes (k) *matches* the Hilbert space dimension (d), eq. (D5) implies eq. (D9) by completeness. When considering 4 outcome measurements on bipartite qubit states in $\mathcal{H}^4$, PVMs are completely general. Moreover, Naimark's dilation theorem guarantees that PVMs acting on $\mathcal{H}^q$ can *emulate* the behaviour of any POVM acting on $\mathcal{H}^d$ provided that q is sufficiently larger than d [35].

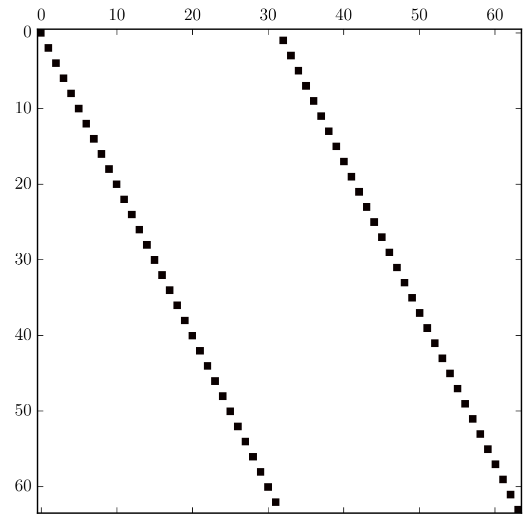

FIG. 5. The network permutation matrix Ω for $\left(\mathcal{H}^2\right)^{\otimes 6}$ realized on the triangle scenario. Black represents a value of 1 and 0 otherwise.

On an operational level, this corresponds to A making a measurement on *both* subsystems of ρ_{AB} and *not* on any component of ρ_{CA}. This is analogously troubling for B and C as well. The network permutation matrix Ω corresponds to *aligning* the underlying 6-qubit joint state ρ with the joint measurement M. To understand its effect, consider its effect on 6-qubit pure state $|q_1\rangle \otimes \cdots \otimes |q_6\rangle = |q_1q_2q_3q_4q_5q_6\rangle$ where $\forall i : |q_i\rangle \in \mathcal{H}^2$.

$$\Omega|q_1q_2q_3q_4q_5q_6\rangle = |q_2q_3q_4q_5q_6q_1\rangle$$

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

$$\Omega \equiv \sum_{|q_i\rangle \in \{|0\rangle,|1\rangle\}} |q_2q_3q_4q_5q_6q_1\rangle\langle q_1q_2q_3q_4q_5q_6|$$

4. Degeneracy

Todo (TC Fraser): Discuss local unitary degeneracy

Appendix E: Convex Parametrization of Finite Probability Distributions

As discussed in section X, there is a need to parameterize the family of all probability distributions P_V over a given set of variables $V = \left(v_1, \ldots, v_{|V|}\right)$. If the cardinality of O_V is finite, then this computationally feasible. The space of probability distributions over $n = |O_V|$ distinct outcomes forms a $n-1$ dimensional convex polytope naturally embedded in $\mathbb{R}^n_{\geq 0}$ [37] that is parameterizable by $n-1$ real value parameters; normalization $\sum_{o[V]\in O_V} P_V(o[V]) = 1$ accounts for the '-1'. An example of a non-degenerate parameterization of P_V consists of $n-1$ parameters $\lambda = (\lambda_1, \ldots, \lambda_{n-1}), \lambda_i \in [0, \pi/2]$ which generate the n probability values p_j using hyperspherical coordinates [26, 29],

$$\begin{aligned} p_j &= \cos^2 \lambda_j \prod_{i=1}^{j-1} \sin^2 \lambda_i \quad \forall j \in 1, \ldots, n-1 \\ p_n &= \prod_{i=1}^{n-1} \sin^2 \lambda_i \end{aligned} \tag{E1}$$

Furthermore due to the periodicity of the parameter space λ, eq. (E1) can be used for either constrained or unconstrained optimization problems. For continuity reasons, unconstrained optimizations are performed whenever possible.

Although non-degenerate, this parameterization suffers from uniformity; a randomly sampled vector of parameters λ *does not* translate to a randomly sampled probability P_V. An easy-to-implement, degenerate parameterization of P_V can be constructed by simply beginning with n real parameters $\lambda = (\lambda_1, \ldots, \lambda_n)$, then making them positive and normalized by their sum[19].

$$p_j = \frac{|\lambda_j|}{\sum_{i=1}^{n} |\lambda_i|} \quad \forall j \in 1, \ldots, n \tag{E2}$$

For various convex optimization tasks sensitive to initial conditions outlined section X, the latter parameterization of eq. (E2) generally performed better than the former eq. (E1).

---

[1] Tobias Fritz and Rafael Chaves, "Entropic inequalities and marginal problems," (2011), 10.1109/TIT.2012.2222863, arXiv:1112.4788.
[2] Samson Abramsky and Adam Brandenburger, "The sheaf-theoretic structure of non-locality and contextuality," (2011), 10.1088/1367-2630/13/11/113036, arXiv:1102.0264.
[3] Bastian Steudel and Nihat Ay, "Information-theoretic inference of common ancestors," (2010), arXiv:1010.5720.
[4] Rafael Chaves, Lukas Luft, and David Gross, "Causal structures from entropic information: geometry and novel scenarios," New Journal of Physics **16**, 043001 (2014).
[5] 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.
[6] 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.
[7] Tobias Fritz, "Beyond bell's theorem: Correlation scenarios," (2012), 10.1088/1367-2630/14/10/103001, arXiv:1206.5115.
[8] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz, "The inflation technique for causal inference with latent variables," (2016).
[9] Yeong-Cherng Liang, Robert W. Spekkens, and Howard M. Wiseman, "Specker's parable of the overprotective seer: A road to contextuality, nonlocality and complementarity," Physics Reports **506**, 1–39 (2011).
[10] Shane Mansfield and Tobias Fritz, "Hardy's non-locality paradox and possibilistic conditions for non-locality," (2011), 10.1007/s10701-012-9640-1, arXiv:1105.1819.
[11] 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).
[12] B. S. Cirel'son, "Quantum generalizations of bell's inequality," Lett Math Phys **4**, 93–100 (1980).
[13] Judea Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, 2009).
[14] Sébastien Lahaie, "How to take the dual of a linear program," (2008).
[15] Bradley, Hax, and Magnanti, *Applied Mathematical Programming* (Addison-Wesley, 1977) Chap. 4, pp. 143–144.
[16] Erling D. Andersen, Computational Optimization and Applications **20**, 171–183 (2001).
[17] L Mančinska and S Wehner, "A unified view on hardy's paradox and the clauser–horne–shimony–holt inequality," Journal of Physics A: Mathematical and Theoretical **47**, 424027 (2014).
[18] Daniel M. Greenberger, "Bell's theorem without inequalities," Am. J. Phys. **58**, 1131 (1990).
[19] Samson Abramsky and Lucien Hardy, "Logical bell inequalities," (2012), 10.1103/PhysRevA.85.062114, arXiv:1203.1352.
[20] Dimitris J. Kavvadias and Elias C. Stavropoulos, "An efficient algorithm for the transversal hypergraph generation," JGAA **9**, 239–264 (2005).
[21] Jean-Daniel Bancal, Nicolas Gisin, and Stefano Pironio, "Looking for symmetric bell inequalities," Journal of Physics A: Mathematical and Theoretical **43**, 385303 (2010).
[22] Christopher J. Wood and Robert W. Spekkens, "The lesson of causal discovery algorithms for quantum correlations: Causal explanations of bell-inequality violations require fine-tuning," (2012), 10.1088/1367-2630/17/3/033002, arXiv:1208.4119.
[23] Armin Tavakoli, "Quantum correlations in connected multipartite bell experiments," (2015), 10.1088/1751-8113/49/14/145304, arXiv:1509.08491.
[24] D. Rohrlich and S. Popescu, "Nonlocality as an axiom for quantum theory," (1995), arXiv:quant-ph/9508009.
[25] N. N. Vorob'ev, "Consistent families of measures and their extensions," Theory of Probability & Its Applications **7**, 147–163 (1962).
[26] Christoph Spengler, Marcus Huber, and Beatrix C. Hiesmayr, "A composite parameterization of unitary groups, density matrices and subspaces," (2010), 10.1088/1751-8113/43/38/385306, arXiv:1004.5252.
[27] 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).

[19] Strictly speaking, eq. (E2) *also* suffers from non-uniformity; being biased toward uniform probability distributions P_V. Todo (TC Fraser): Discuss rejection sampling simplex algorithms

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

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

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

[31] 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.

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

[33] 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.

[34] Dénes Petz and László Ruppert, "Optimal quantum-state tomography with known parameters," Journal of Physics A: Mathematical and Theoretical **45**, 085306 (2012).

[35] Asher Peres, *Quantum Theory: Concepts and Methods (Fundamental Theories of Physics)* (Springer, 1995) Chap. 9.6, p. 285.

[36] 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.

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