

Graph-theoretic strengths of contextuality

Nadish de Silva

Department of Computer Science, University College London, WC1E 6BT London, United Kingdom
(Received 1 February 2016; revised manuscript received 23 September 2016; published 9 March 2017)

Cabello-Severini-Winter and Abramsky-Hardy (building on the framework of Abramsky-Brandenburger) both provide classes of Bell and contextuality inequalities for very general experimental scenarios using vastly different mathematical techniques. We review both approaches, carefully detail the links between them, and give simple, graph-theoretic methods for finding inequality-free proofs of nonlocality and contextuality and for finding states exhibiting strong nonlocality and/or contextuality. Finally, we apply these methods to concrete examples in stabilizer quantum mechanics relevant to understanding contextuality as a resource in quantum computation.

DOI: [10.1103/PhysRevA.95.032108](https://doi.org/10.1103/PhysRevA.95.032108)

I. INTRODUCTION

A fundamental question of quantum computation (QC) is to precisely identify the nonclassical features of quantum mechanics accounting for quantum advantages in computation. This critical problem is often phrased in terms of finding properties characterizing those resource states capable of promoting a computational model to greater power, e.g., universal QC. Understanding of quantum resources will yield both hardware efficiency gains and a theoretical basis for developing novel applications of quantum information. Entanglement [1], superposition [2], and discord [3] have been proposed as candidates but found unsatisfactory as explanations of quantum advantage. Numerous striking results [4–8] have recently established *contextuality* as necessary for magic state distillation (MSD) [9] and measurement-based quantum computation (MBQC) [10].

Our aim in this paper is to carefully delineate the connections between the vastly differing frameworks for contextuality used in these recent results in order to better understand the role of contextuality in QC and to conceptually clarify contextuality. We use this connection to give directly computable graph-theoretic characterizations of the logical strengths of contextuality and apply these tools to examples in stabilizer quantum mechanics relevant to QC.

A. Nonlocality and contextuality

Contextuality is a generalization of *nonlocality*: the notion that the predictions of quantum mechanics do not admit a model that is classical in the sense of being locally causal. That is, one in which, upon conditioning on a causal past, the joint distributions describing experiments performed at different sites factorize into distributions associated with each site [11,12]. Data from a two-site experiment are correlations $p(a,b|A,B)$ where A,B and a,b are measurement settings and outcomes respectively. A locally causal model for such data is a space Λ of hidden variables, a distribution q on Λ , and conditional distributions $r_A(-|\lambda), r_B(-|\lambda)$ on outcomes that account for the data:

$$p(a,b|A,B) = \sum_{\lambda \in \Lambda} q(\lambda) r_A(a|\lambda) r_B(b|\lambda).$$

A *Bell inequality* is a bound on a weighted sum of correlations that is satisfied by all data a locally causal model. Data not admitting any such model exhibit *nonlocality*. Bell gave the first

such inequality and an entangled quantum state that violates it. Decades later, entanglement and nonlocality became the basis of quantum communication protocols, e.g., superdense coding, quantum teleportation, and quantum cryptography [13–15].

There are degrees of strength of nonlocality. The correlations arising from a Bell state violate a Bell inequality. Hardy states [16,17] preclude a locally causal model via logical arguments without the need for inequalities. The Greenberger-Horne-Zeilinger (GHZ) state [18] is, in a sense defined below, maximally nonlocal. Nonlocality has been studied as a resource for communication tasks and it has been recognized that stronger nonlocality yields greater advantages [19–22].

Contextuality is the notion, due to Kochen and Specker [23], that a system’s observable properties cannot all be assigned deterministic outcomes in a manner independent of the method of observation used to ascertain each property. It subsumes nonlocality as a special case since Fine’s theorem [24] tells us that data admitting a locally causal model also admit one in which the distributions $r_A(-|\lambda), r_B(-|\lambda)$ are deterministic. Like nonlocality, contextuality admits a description via inequalities, e.g., Klyachko *et al.*’s *contextuality inequality* [25]. Unlike nonlocality, contextuality can manifest in single-site systems.

B. Contextuality as a resource in QC

Recently, the hypothesis that contextuality plays a critical role in enabling quantum computation—analogue to the role nonlocality plays in superclassical communication—has received considerable attention. For example, Howard *et al.* [9] showed that contextuality is a necessary criterion in finding those quantum states that are suitable for the magic state distillation (MSD) protocol [26] which promotes a scheme of classical computing power to (fault-tolerant) quantum universality.

A natural question is: does stronger contextuality yield greater computational advantages? Positive evidence for this question is given by Raussendorf [10] in the setting of measurement-based quantum computers (MBQCs) [27]. Further progress on the resource theory of contextuality necessitates clarification of the strengths of contextuality.

The aforementioned results on contextuality as a resource rely on mathematical tools developed by Cabello and Severini and Winter (CSW) [28] and Abramsky and Brandenburger (AB) [29] which promote the example-based understanding of nonlocality and contextuality to a higher-level, structural one.

They apply in any experimental scenario and both provide a complete set of Bell and contextuality inequalities. However, they differ vastly in their approach, and how they relate to one another is not immediately clear. The plurality of perspectives and vernaculars for describing contextuality constitutes an obstacle to progress on a computational resource theory of contextuality.

C. Overview

We first review some definitions and results of CSW and AB before we

(1) Precisely detail the links between their two furnished classes of Bell and contextuality inequalities.

(2) Cast the powerful, highly abstract, logical hierarchy of strengths of contextuality due to AB in terms of graph invariants in order to render them more tractable as theoretical tools and enable direct computation using standard software libraries.

(3) Pose the question, does the logical strength of contextuality of a state correlate with its usefulness as a resource?

We provide three computational examples to demonstrate how these tools may be used to gain insight into the contextuality in a QC program.

The first result shows that the sort of contextuality witnessed by logical contradictions (rather than mere violation of contextuality inequalities) may not be useful in identifying the single-qutrit states suitable for magic state distillation. However, we provide evidence that it may be useful in identifying higher-qutrit magic states. The third result addresses the question that Howard *et al.* end their paper with the following:

In the qubit case, it is a pressing open question whether a suitable operationally motivated refinement or quantification of contextuality can align more precisely with the potential to provide a quantum speed-up.

Our result asserts that the quantification sought would necessarily need to discriminate between states that are all maximally contextual. One might hope to achieve this by considering measures specialized to the particular setting of stabilizer QM, e.g., [30].

As we shall see, nonlocality is the special case of contextuality where the observables are compatible (i.e., comensurable) precisely when they correspond to spatially separated local measurements. Thus, in the remainder of this paper, we use the terminology contextuality and contextual inequalities to include the special cases of nonlocality and Bell inequalities.

II. GENERAL APPROACHES TO CONTEXTUALITY

Both the AB and CSW approaches begin by abstractly describing physical experiments without making an assumption of simultaneous comensurability of all properties and by describing how operational data from such an experiment are tabulated. The primary question asked of operational data is whether it can, in principle, be reproduced by a model in which observables simultaneously possess deterministic values.

AB's and CSW's representations of experiments and data differ significantly. Before delineating their connections, we review the key structures and results of each.

A. CSW inequalities

CSW employ graph theory to study nonlocality and contextuality for general experimental scenarios. (The basic graph-theoretic definitions necessary for understanding what follows are found in Appendix A). An experiment is formalized by an *exclusivity graph*. The vertices of such a graph represent events, or, alternatively, answers to propositions about a system answerable by the experiment. Vertices that are adjacent represent *mutually exclusive* events.

Operational data coming from repeatedly performing the experiment on a fixed preparation of the system is tabulated using probabilities p_i for each vertex v_i . CSW inequalities are upper bounds on a linear combination of these probabilities: $\sum_i w_i p_i$ where p_i is the likelihood of observing the i th event and $w_i \geq 0$ is a coefficient. Such a linear combination is compactly represented as a weighted graph (G, w) .

CSW define a *classical model* for operational data represented by a graph G and probabilities p_i as a classical sample space Λ , an event $e_i \subset \Lambda$ for each vertex v_i such that the events corresponding to adjacent vertices are mutually exclusive (i.e., disjoint subsets of Λ), and a probability distribution μ on Λ such that p_i is the probability that μ assigns to e_i .

For a linear combination $\sum_i w_i p_i$, represented by the weighted graph (G, w) , one can ask the question, what is the maximum this sum can achieve if the data p_i has a classical model? CSW show that this maximum is precisely the weighted independence number of (G, w) . Data that violate any one of these bounds, for some choice of coefficients w_i , cannot be explained with a classical model. Therefore, CSW contextuality inequalities are of the form

$$\sum_i w_i p_i \leq \alpha(G, w).$$

A *quantum model* is a Hilbert space \mathcal{H} , projectors $P_i \in \mathcal{B}(\mathcal{H})$ for each vertex v_i such that adjacent vertices are represented by orthogonal projectors, and a pure state $|\psi\rangle \in \mathcal{H}$ such that the probabilities p_i arise as $\langle \psi | P_i | \psi \rangle$. In this case, the maximum $\sum_i w_i p_i$ can achieve over all quantum models (i.e., the Cirel'son-type bound) is bounded above by the Lovász theta number: $\sum_i w_i p_i \leq \vartheta(G, w)$.

Finally, a *generalized model*¹ is simply data p_i such that the sum of probabilities corresponding to all the vertices in a clique is less than or equal to 1. The maximum $\sum_i w_i p_i$ can achieve over all generalized models is given by the fractional packing number: $\sum_i w_i p_i \leq \alpha^*(G, w)$.

B. The AB sheaf-theoretic approach

In the AB approach,² an experiment is formally described by a *measurement scenario*: a pair $(\mathcal{M}, \mathcal{C})$ where \mathcal{M} is a

¹These models are those that satisfy the *consistent exclusivity* or *EI* principle.

²We give an elementary introduction with simplified terminology. Curious readers are encouraged to read [29].

set of abstract labels for measurements and \mathcal{C} is the set of *contexts*. A context is a set $C \subset \mathcal{M}$ representing a maximal set of compatible (i.e., comeasurable) measurements; thus, it is required that if C is a context, no proper subset of C is also a context. It is further required that every measurement $m \in \mathcal{M}$ is contained in at least one context. Measurement of each individual $m \in \mathcal{M}$ yields a value from the outcome set \mathcal{O} (usually $\{0,1\}$).

For example, the standard Bell scenario is described with $\mathcal{M} = \{A_0, A_1, B_0, B_1\}$ and contexts $\mathcal{C} = \{\{A_0, B_0\}, \{A_0, B_1\}, \{A_1, B_0\}, \{A_1, B_1\}\}$.

A *formal event* is a function $e : S \rightarrow \mathcal{O}$ from a set $S \subset \mathcal{M}$ of measurements to outcomes; in other words, a joint outcome for all of the measurements in S . Note that this is a strictly mathematical construction as it is not assumed that S is a subset of a context. In other words, the measurements in S are not assumed to be physically comeasurable. The set of all formal events for a fixed S is denoted by $E(S) = \{e : S \rightarrow \mathcal{O}\}$. Whenever $S' \subset S$, a *coarse-graining* of a formal event $e : S \rightarrow \mathcal{O}$ is defined by forgetting the outcomes for measurements in S but not in S' ; it is denoted by $e|_{S'}$ and explicitly defined by $e|_{S'}(M') = e(M')$ for $M' \in S'$.

The partial distribution sets $D(S)$ are then defined as the set of all probability distributions on the formal event set of S , i.e., $D(S) = \{p : E(S) \rightarrow [0,1] \mid \sum_{e \in E(S)} p(e) = 1\}$. Whenever $S' \subset S$, the *marginal distribution* of $p \in D(S)$ is defined by averaging over the the information about measurements in S but not in S' ; it is denoted by $\mu_{S'}^S(p)$ and explicitly defined, for $e' \in E(S')$, by

$$\mu_{S'}^S(p)(e') = \sum_{e|_{S'}=e'}^{e \in E(S)} p(e).$$

Operational data from repeated experiments on a system in a fixed preparation are tabulated by families of probability distributions $\mathcal{E}_C \in D(C)$ that are indexed by contexts $C \in \mathcal{C}$. They describe the likelihoods of joint outcomes for all the maximal sets of comeasurable measurements. Such data are *nonsignalling* (also known as *nondisturbing*) when they yield common marginal distributions for each intersection of contexts: $\mu_{C \cap C'}^C(\mathcal{E}_C) = \mu_{C \cap C'}^{C'}(\mathcal{E}_{C'})$ for any pair C, C' of contexts. When data obeys this nonsignalling condition, the distributions \mathcal{E}_C constitute an *empirical model* \mathcal{E} .

TABLE I. Bell state correlations (top); PR box correlations (bottom). Each row is a context and a distribution on joint outcomes.

A	B	00	01	10	11
A_0	B_0	1/2	0	0	1/2
A_0	B_1	3/8	1/8	1/8	3/8
A_1	B_0	3/8	1/8	1/8	3/8
A_1	B_1	1/8	3/8	3/8	1/8
A_0	B_0	1/2	0	0	1/2
A_0	B_1	1/2	0	0	1/2
A_1	B_0	1/2	0	0	1/2
A_1	B_1	0	1/2	1/2	0

An empirical model is local and/or noncontextual precisely when its predictions can be accounted for by a locally causal and/or noncontextual hidden variable model. That is, there is a hidden variable space Λ , a distribution q on Λ , and conditional distributions $r_M(-|\lambda)$ on \mathcal{O} for each $M \in \mathcal{M}$ such that, for any context $C = \{M_1, \dots, M_N\}$ and $e : C \rightarrow \mathcal{O}$,

$$\mathcal{E}_C(e) = \sum_{\lambda \in \Lambda} q(\lambda) r_{M_1}(o_1|\lambda) \cdots r_{M_N}(o_N|\lambda),$$

where $o_i = e(M_i)$ is the outcome of the i th measurement. Equivalently, an empirical model \mathcal{E} is noncontextual when there is a joint distribution $J \in D(\mathcal{M})$ yielding the \mathcal{E}_C as marginals: $\mathcal{E}_C = \mu_C^{\mathcal{M}}(J)$. Thus, the hidden variable space can be taken to have the canonical form of $\Lambda = E(\mathcal{M})$ whose *canonical hidden variables* are functions $\lambda : \mathcal{M} \rightarrow \mathcal{O}$ and the $r_M(-|\lambda)$ are the deterministic distributions assigning probability 1 to the outcome $\lambda(M)$ and 0 to all others. That an empirical model that arises as a family of marginals distributions of a single joint probability distribution on all measurements always admits such a canonical model is a vast generalization of Fine’s theorem to all experimental scenarios.

Abramsky-Hardy [31] give a complete set of Bell and contextuality inequalities via logical consistency; AB describe two natural logical strengths of contextuality. We detail these tools, and how they may be understood in terms of graph invariants, below.

III. THE EXCLUSIVITY GRAPH OF A MEASUREMENT SCENARIO

In this section, we build a CSW exclusivity graph from AB’s description of an experiment. Explicitly corresponding the two representations of experiments and data facilitates the synthesis of insights and tools of each approach.

An *observable event* is an $e \in E(C)$ for some context $C \in \mathcal{C}$. They are the elementary events observable in an experiment, or, alternatively, answers to the most refined questions one can ask of a system. Given an empirical model \mathcal{E} , an observable event e is *impossible* when $\mathcal{E}_C(e) = 0$; otherwise, it is *possible*.

Given a probability table for an empirical model, the observable events correspond to the (empty) cells. An empirical model provides the actual probabilities. Two observable events are called *mutually exclusive* or *inconsistent* when they assign different outcomes to a shared measurement.

Definition 1. The *exclusivity graph* $\mathcal{G}(\mathcal{M}, \mathcal{C})$ of the *measurement scenario* $(\mathcal{M}, \mathcal{C})$ has vertices given by the observable events $e : C \rightarrow \mathcal{O}$ where $C \in \mathcal{C}$ is a context. Two vertices $e_1 : C_1 \rightarrow \mathcal{O}$ and $e_2 : C_2 \rightarrow \mathcal{O}$ are adjacent whenever e_1 and e_2 are *inconsistent*; that is, whenever $e_1(M) \neq e_2(M)$ for some measurement M in both C_1 and C_2 .

Acín *et al.* associate a hypergraph with a measurement scenario via *measurement protocols* [32, Definition D.1.4]. The nonsignalling conditions of empirical models are encoded as clique constraints on this hypergraph’s associated nonorthogonality graph [32, Definition 2.3.1]. Definition 1 is isomorphic to the complement of the graph built this way; see Appendix C for a proof. Here, we focus on directly constructing the exclusivity graph of a measurement scenario in order to understand sheaf-theoretic tools in terms of graph invariants.

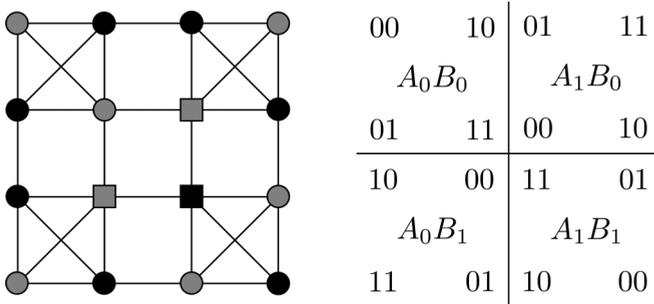


FIG. 1. The exclusivity graph for the standard Bell scenario (left); labels of measurement settings and outcomes (right). Straight lines connect exclusive events. The standard CHSH inequality [33] is derived by giving the grey vertices the weight 1 and the others 0; the independence number is 3. Circular vertices are the possible events of a Hardy model. Grey vertices are the possible events of a PR box [34].

Definition 2. The *support graph* of an empirical model \mathcal{E} is the induced subgraph $\mathcal{G}_{\mathcal{E}}$ of $\mathcal{G}(\mathcal{M}, \mathcal{C})$ given by retaining only the possible events.

These definitions generalize constructions used by Sadiq *et al.* [35] and Fritz *et al.* [36,37] for analyzing nonlocality to the setting of contextuality.

In our proofs, we repeatedly exploit the following key connection between the approaches of AB and CSW:

Lemma 1. For any measurement scenario $(\mathcal{M}, \mathcal{C})$, there is a bijective correspondence between canonical hidden variables $\lambda : \mathcal{M} \rightarrow \mathcal{O}$ and independent sets $I \subset \mathcal{G}(\mathcal{M}, \mathcal{C})$ of size $|\mathcal{C}|$.

As the observable events of a common context form a clique, the size of an independent set in $\mathcal{G}(\mathcal{M}, \mathcal{C})$ can contain at most one observable event from each clique. Thus, the size of an independent set is bounded by the number of contexts. Independent sets whose size achieves this upper bound must contain precisely one event of each context.

Given such an independent set I , one can construct a canonical hidden variable $\lambda_I : \mathcal{M} \rightarrow \mathcal{O}$ by defining $\lambda_I(M) = e(M)$ where $e : \mathcal{C} \rightarrow \mathcal{O}$ is any event in I such that $M \in C$. If $e' : \mathcal{C}' \rightarrow \mathcal{O}$ is another event in I such that $M \in C'$, the fact that e and e' are not adjacent tells us that $e(M) = e'(M)$; thus, λ_I is well defined. Conversely, given a canonical hidden variable $\lambda : \mathcal{M} \rightarrow \mathcal{O}$, the set $I_{\lambda} = \{\lambda|_C : C \rightarrow \mathcal{O} \mid C \in \mathcal{C}\}$ of its coarse-grainings form an independent set.

IV. LOGICAL BELL INEQUALITIES

The *logical Bell inequalities* of $(\mathcal{M}, \mathcal{C})$ are a distinguished class of Bell or contextuality inequalities identified by Abramsky and Hardy [31]. Working in the setting of measurements with binary outcomes³ ($\mathcal{O} = \{0, 1\}$), we can encode observable events e as propositional formulas $f(e)$ by viewing the measurements $M \in \mathcal{M}$ as Boolean variables:⁴

$$f(e) = \bigwedge_{M \in \mathcal{C}} \begin{cases} M & \text{if } e(M) = 1, \\ \neg M & \text{if } e(M) = 0. \end{cases}$$

³The case of measurements with different outcome sets is easily reduced to the binary case.

⁴The symbols \wedge, \vee, \neg represent AND, OR, NOT respectively.

For example, the two grey events of the last row in Table I are individually encoded as $A_1 \wedge \neg B_1$ and $\neg A_1 \wedge B_1$.

Two vertices are adjacent precisely when their formulas are logically inconsistent, i.e., one cannot assign *true* or *false* to the variables in \mathcal{M} in a way making both their formulas true.

Given N Boolean formulas f_i that are true with probability p_i , if the formulas are logically inconsistent, then it follows from elementary probability theory [31, Sec. 1A] that

$$\sum p_i \leq N - 1.$$

A logical Bell inequality is built by choosing a subset of the observable events $E_i \subset E(C_i)$ from each context $C_i \in \mathcal{C}$ and constructing the formulas

$$f_i = \bigvee_{e \in E_i} f(e).$$

When the formulas f_i contain a contradiction, we obtain an inequality: the sum of the likelihoods of all of the events in all of the E_i is less than or equal to $N - 1$.

For example, the grey subset of events of the last row (in either table) of Table I is encoded as $f_4 = (A_1 \wedge \neg B_1) \vee (\neg A_1 \wedge B_1)$ or, equivalently, $A_1 \iff \neg B_1$. The CHSH inequality is derived by noting that the formulas for all four rows ($A_0 \iff B_0, A_0 \iff B_1, A_1 \iff B_0, A_1 \iff \neg B_1$) are contradictory and so the sum of probabilities of all the grey events has a classical bound of 3. The Bell state violates this bound by 1/4. The PR box maximally violates it by 1.

Theorem 1. Logical Bell inequalities can be seen as examples of CSW inequalities on the exclusivity graph $\mathcal{G}(\mathcal{M}, \mathcal{C})$.⁵

The inequalities are constructed by giving the events of E_i the integer weight 1 and all other events weight 0. The CSW bound given by the independence number is tight whereas the AB bound of $N - 1$ need not be.

Abramsky and Hardy also define *extended logical Bell inequalities* by choosing sets of observable events $E_i \subset E(C_i)$ whose formulas f_i are contradictory as above; however, the coefficients in these inequalities are allowed to be any integer $k_i \geq 0$. The upper bound they give, defined in terms of logical consistency, is, in fact, simply the CSW bound given by the weighted independence number. As these extended inequalities have the same integer coefficient for all observable events of a common context, they form a much smaller class of contextual inequalities than all possible CSW inequalities on $\mathcal{G}(\mathcal{M}, \mathcal{C})$ which allow each individual observable event its own coefficient. However, they still form a *complete* class of contextuality inequalities in the sense that knowing that an empirical model satisfies only the extended logical Bell inequalities is sufficient to conclude that it is noncontextual.

The logical interpretation of contextuality inequalities given by Abramsky and Hardy can be extended to all CSW inequalities: the weighted independence number can be seen as maximizing the weighted sum over logically consistent formulas.

V. HARDY'S PARADOX AND LOGICAL CONTEXTUALITY

Considerable work has been done on establishing Bell's theorem without inequalities [17,18,38–41]. The idea is to preclude locally causal models of quantum theory using knowledge of the mere possibility or impossibility of certain

correlated events rather than the precise correlations. While the first proof of Bell’s theorem without inequalities was the GHZ state, Hardy’s paradox [16] better exemplifies the subtle logical contradictions at play. As inequality-free proofs of nonlocality cannot be accomplished with standard Bell states, Hardy’s paradox can be interpreted as a stronger form of nonlocality than the mere statistical one. In the AB framework, this notion is generalized to *logical nonlocality and contextuality*; we exploit this to give a graph-theoretic characterization of operational data that preclude locally causal and noncontextual models without resort to inequalities.

A state is logically nonlocal or contextual if any locally causal or noncontextual model that predicts the occurrence of all empirically observed events must also predict the occurrence of an event that is not empirically observed. More formally: there is a possible event e such that any canonical hidden variable $\lambda : \mathcal{M} \rightarrow \mathcal{O}$ that is compatible with e (i.e., the coarse-graining $\lambda|_C$ is e) must also predict the occurrence of an impossible event (there is a context $C' \in \mathcal{C}$ for which $\lambda|_{C'}$ is impossible). A state that satisfies all logical Bell inequalities cannot be logically contextual.

Finding a graph-theoretic characterisation of logical contextuality requires defining a new graph invariant heretofore unused in the CSW approach:

Definition 3. The *independence degree* of a vertex v in a finite graph G is the size of the largest independent set in G that contains v ; the *minimal independence number* is the minimum independence degree over all vertices.

Note that the standard independence number of a graph is the maximum independence degree over all vertices.

Theorem 2. An empirical model \mathcal{E} is logically nonlocal or contextual (i.e., admits an inequality-free proof of nonlocality or contextuality) if and only if the minimal independence number of its support graph $\mathcal{G}_{\mathcal{E}}$ is less than the number of contexts.⁵

In Fig. 1, that there is no independent set of 4 circular vertices that includes the top-left corner vertex proves that Hardy’s model is logically nonlocal

VI. MAXIMAL NONLOCALITY AND STRONG CONTEXTUALITY

The notion of *maximal nonlocality* was introduced by Elitzur, Popescu, and Rohrlich [42] for states in a standard Bell experiment and extended to more general nonlocality experiments by Barrett *et al.* [43]. AB generalized this notion to maximal contextuality and gave an equivalent logical notion called *strong contextuality* which we exploit to give a simple graph-theoretic criterion.

Well-known examples of maximally nonlocal empirical models include GHZ states and PR boxes. Access to maximally nonlocal resources yields advantages for communication tasks. For example, sharing an unlimited number of PR boxes between two parties renders all communication complexity problems trivial [19].

Strong contextuality is known to play a role in quantum computation: Raussendorf [10] showed that strong contextu-

ality is necessary for allowing certain MBQCs to deterministically compute nonlinear functions.

An empirical model \mathcal{E} can be convexly decomposed into a noncontextual part A and a nonsignalling part Z :

$$\mathcal{E} = \tau A + (1 - \tau)Z.$$

An empirical model is *maximally contextual* if admits no decomposition with nonzero τ .

This is equivalent to strong contextuality: there is no canonical hidden variable $\lambda : \mathcal{M} \rightarrow \mathcal{O}$ such that the observable event $\lambda|_C$ is possible for all contexts $C \in \mathcal{C}$. An empirical model is strongly contextual precisely when it maximally violates (by 1) the logical Bell inequality yielded by choosing E_i to be its possible events in $E(C_i)$. Strongly contextual empirical models are precisely the convex sums of the nonsignalling polytope’s contextual vertices [32,44].

Theorem 3. An empirical model \mathcal{E} is strongly nonlocal or contextual if and only if the independence number of its support graph $\mathcal{G}_{\mathcal{E}}$ is less than the number of contexts.⁵

In Fig. 1, that there is no independent set of four grey vertices proves that the PR box is strongly nonlocal.

VII. APPLICATIONS

We provide computational examples wherein we apply Theorems 2 and 3 to study the contextuality of quantum states relative to all stabilizer operations [45].

Recent work of Howard *et al.* [9] has shown contextuality with respect to two-qudit stabilizer operations to be a necessary condition for MSD in odd-prime-power qudit systems. Specifically, they expressed the inequalities describing the faces of the polytope of quantum states with positive Wigner function (in the Gross representation [46]) as CSW inequalities on the orthogonality graph of two-qudit stabilizer projectors. As a classically efficient simulation algorithm is known for stabilizer quantum theory augmented with nonstabilizer states having positive Wigner representation [6], any quantum state that is noncontextual with respect to stabilizer projectors cannot serve as a resource to promote stabilizer operations to superclassical computational power.

A. Qutrit stabilizer operations

In the odd-prime-power qudit cases, where contextuality is a useful necessary criterion for identifying MSD resource states, we ask whether stronger forms of contextuality are also useful. Following Howard *et al.*, we consider contextuality with respect to two-qutrit stabilizer operations.

Result 1. The states $|M\rangle \otimes |M\rangle$, where $|M\rangle$ is the qutrit magic state $|M_0\rangle$, $|E\rangle$, $|N'\rangle$, or $|\psi_{U_v}\rangle$ [47–49] are neither logically contextual nor strongly contextual with respect to two-qutrit stabilizer operations.

Therefore, logical contextuality of $|M\rangle \otimes |M\rangle$ is not a necessary condition for a qutrit state $|M\rangle$ to serve as a resource promoting stabilizer operations to universal QC.

Result 2. The two-qutrit magic state

$$|CS\rangle = \sum_{j,k=0}^2 e^{\frac{2\pi i}{3}jk^2} |j\rangle \otimes |k\rangle$$

[50] is strongly contextual with respect to two-qutrit stabilizer operations.

⁵The proof is found in Appendix B.

This suggests that stronger forms of contextuality may play a role in identifying higher-qudit magic states; we leave further investigation of this to future work. As $|CS\rangle$ deterministically yields a non-Clifford gate, this result strengthens the connection between strong contextuality and advantage in deterministic computation and communication tasks that is suggested by the essential role strong contextuality plays in MBQC [4,10] and zero-error classical communication [51]. It seems reasonable to conjecture that any magic state exhibiting such a deterministic property must be strongly contextual.

The support graphs of the aforementioned states are induced subgraphs of the orthogonality graph of two-qutrit stabilizer projectors which was computed using Gross' discrete phase space [46]. The phase space for a two-qutrit system is given by $V = (\mathbb{Z}_3 \times \mathbb{Z}_3) \times (\mathbb{Z}_3 \times \mathbb{Z}_3)$: intuitively, a position and momentum variable for each qutrit. The contexts are in correspondence with the Lagrangian subspaces of V whereas the stabilizer states can be computed from a Lagrangian subspace $M \subset V$ and a phase point $v \in V$ using [46, Lemma 8]. The Lagrangian subspaces are easily enumerated by finding all linearly independent pairs of vectors and eliminating those with nonzero symplectic product. Some pairs will generate the same subspace; this redundancy has to be eliminated.

To reach Result 1, the minimal independence numbers of the support graphs were computed to be equal to the number of contexts (40) and thus, by Theorem 2, the states are not logically contextual and thus not strongly contextual. To reach Result 2, the independence number of the support graph of $|CS\rangle$ was computed and found to be 34. Thus, by Theorem 3, $|CS\rangle$ is strongly contextual.

B. Qubit stabilizer operations

It is impossible to use mere contextuality as a sufficient criterion for determining which qubit states are suitable for MSD as the Peres-Mermin argument⁵ [52] identifies all two-qubit states as contextual with respect to two-qubit stabilizer operations. Howard *et al.* posed as an open problem whether a quantification of contextuality could serve as a better criterion. A close analysis of the Peres-Mermin argument shows that it, in fact, identifies all two-qubit states as *strongly* contextual with respect to two-qubit stabilizer operations. Thus, we suggest a negative answer to Howard *et al.*'s question as such a quantification would need to discriminate between states which are all maximally contextual. While it may still be possible to use contextuality as a criterion in identifying qubit MSD resources, it seems that such a measure will have to be specialized to the setting in question. An interesting workaround to state-independent contextuality in qubit MSD is provided by [8].

Howard *et al.* describe noncontextuality with respect to all stabilizer operations in the sense of obeying all CSW inequalities on the orthogonality graph of stabilizer projectors. In order to apply the AB contextuality criteria, it is necessary to describe this with a physical setting with a measurement scenario. We chose as the set of measurements \mathcal{M} the 15 possible tensor pairs of I, X, Y, Z excluding $I \otimes I$. The contexts are the 15 maximal commuting subsets (of size 3) of these measurements.

Constructing the graph $\mathcal{G}(\mathcal{M}, \mathcal{C})$, we see that it is made up of 15 cliques of 2^3 formal events: 120 formal observable

events. However, there are only 60 two-qubit stabilizer states. The resolution of this puzzle is that half of the formal events represent joint outcomes which are actually impossible in quantum theory as they do not respect the algebraic relationships between the quantum measurements.

Result 3. All n -qubit states are strongly contextual with respect to n -qubit stabilizer operations whenever $n > 1$.⁵

Consider $n = 2$. The exclusivity graph \mathcal{G} has as vertices the rank-1 two-qubit stabilizer projectors; edges connect orthogonal projectors. We compute $\alpha(\mathcal{G}) = 12$, agreeing with [53, Table 3], and note that $\alpha(\mathcal{G}_\rho) \leq \alpha(\mathcal{G})$ for the empirical model arising from any quantum state ρ . Since the number of contexts is 15, it follows from Theorem 3 that ρ is strongly contextual. For $n > 2$, an analytic proof that directly extends the Peres-Mermin argument is given in Appendix B.

The above proof applies Theorem 3 for each quantum state individually by noting that the support graph of any quantum state is a subgraph of the support graph of the maximally mixed state. Generalizing, we see that *state-independent strong contextuality* will occur whenever the independence number of the orthogonality graph of a quantum measurement scenario is less than the number of contexts. AB identified the connection between Kochen-Specker-type configurations of vectors and generic strong contextuality; the arguments of [29, Secs. 7.1, 9.2] or the theory of generalized all-versus-nothing arguments [54, Sec. 4.2] can also be used to reach Result 3.

VIII. CONCLUSIONS

The above examples clarify the relationship between the logical strength of contextual resources and computational power. We intend to further investigate the strength of contextuality of higher-qudit magic states. Further development will also require considering other measures of contextuality and investigating the relationship between measures of contextuality and other models of quantum computation. Another important future direction is to clarify the relationship between the above described notion of contextuality with Spekkens' [55,56] notion.

We have shown how to synthesize insights and tools from both the Cabello-Severini-Winter and Abramsky-Brandenburger approaches for understanding contextuality of operational data for very general experimental settings. We repeatedly exploited the correspondence between canonical hidden variables for $(\mathcal{M}, \mathcal{C})$ and independent sets of maximal size of $\mathcal{G}(\mathcal{M}, \mathcal{C})$ which relies on the special structure of exclusivity graphs arising from measurement scenarios. An interesting future direction is understanding arbitrary graphs as *contextual sample spaces*. It will be necessary to understand how to complete an abstract graph to one arising from a measurement scenario. Perhaps additional structure on the graph will be necessary for this. For example, it may be necessary to fix a distinguished vertex cover by disjoint cliques indicating events of a common context; more general hypergraph-theoretic structure [32] might also be required.

ACKNOWLEDGMENTS

The author wishes to thank Eric Cavalcanti, Samson Abramsky, Simone Severini, Rui Soares Barbosa, Kohei

Kishida, Dan Browne, Shane Mansfield, Andrew Simmons, Mark Howard, and Ana Belén Sainz for many helpful discussions. This research was financially supported by the EPSRC via grant EP/N017935/1 (Contextuality as a resource in quantum computation). This work was partially completed at the Simons Institute for the Theory of Computing (supported by the Simons Foundation) at the University of California, Berkeley, during the Logical Structures in Computation programme.

APPENDIX A: BACKGROUND

1. Graph theory

A (simple, undirected) *graph* G is a set of vertices $V = \{v_1, \dots, v_n\}$, pairs of which may be adjacent, i.e., joined by an edge. The edges are represented by a set $E \subset V \times V$ such that $(v, v) \notin E$ for any vertex v and such that $(v, v') \in E$ if and only if $(v', v) \in E$. That is, no edge joins a vertex to itself and v and v' are adjacent whenever v' and v are. Given a subset $V' \subset V$ of vertices, the *induced subgraph* has V' as vertices and retains all edges $(v, v') \in E$ with $v, v' \in V'$.

A weighted graph (G, w) is a graph G together with a function $w : V \rightarrow \mathbb{R}^{\geq 0}$ from vertices to non-negative reals. We denote the weight $w(v_i)$ of a vertex v_i by w_i .

An *independent (vertex) set* is a subset $I \subset V$ such that no two vertices in I are adjacent: $(v, v') \notin E$ whenever $v, v' \in I$. A *clique* is a subset $C \subset V$ such that every two distinct vertices in C are adjacent: $(v, v') \in E$ whenever $v, v' \in C$ and $v \neq v'$.

The *independence number* of a graph G is the size of the largest independent set in G ; in terms of Definition 3, it is the maximal independence degree over all vertices. The independence number of a weighted graph (G, w) is the maximum value of the sum $\sum_{i \in I} w_i$ where I is any independent set.

To define the *Lovász theta number* $\vartheta(G, w)$ of a weighted graph (G, w) we must first define an orthonormal representation of a graph G . This is a choice of a unit vector $|\phi\rangle \in \mathbb{R}^d$ and an assignment to each vertex v_i a unit vector $|\psi_i\rangle \in \mathbb{R}^d$ such that $\langle \psi_i | \psi_j \rangle = 0$ whenever v_i and v_j are adjacent. The choice of dimension d can be arbitrary. The Lovász theta number $\vartheta(G, w)$ is then defined to be the maximum $\sum_{i \in V} w_i |\langle \psi_i | \phi \rangle|^2$ over all possible orthonormal representations of G .

The *fractional packing number* $\alpha^*(G, w)$ is the maximum possible value of the sum $\sum_{i \in V} p_i w_i$ where $\{p_1, \dots, p_n\}$ is a choice of a non-negative real number for each vertex such that $\sum_{i \in C} p_i \leq 1$ for every clique $C \subset V$.

APPENDIX B: PROOFS

Theorem 1. Logical Bell inequalities are derived from CSW inequalities on the exclusivity graph $\mathcal{G}(\mathcal{M}, \mathcal{C})$.

Proof. We are given $E_i \subset C(C_i)$ such that the formulas f_i are logically inconsistent. The corresponding logical Bell inequality is

$$\sum_{i=1, \dots, N} \sum_j p_{ij} \leq N - 1$$

where N is the number of contexts and p_{ij} is the probability of the j th event in E_i . Consider the weighted graph (G, w) with weights 1 for events in E_i and 0 for all others. The

corresponding CSW inequality is

$$\sum_{i=1, \dots, N} \sum_j p_{ij} \leq \alpha(G, w).$$

The quantity $\alpha(G, w)$ is simply the size of the largest independent set of events in $\cup E_i$. This integer must be strictly less than N or else these events together yield an assignment of *true* or *false* to all the variables in \mathcal{M} , which makes all the formulas f_i true, contradicting their assumed logical inconsistency. ■

Theorem 2. An empirical model \mathcal{E} is logically nonlocal or contextual (i.e., admits an inequality-free proof of nonlocality or contextuality) if and only if the minimal independence number of its support graph $\mathcal{G}_{\mathcal{E}}$ is less than the number of contexts.

Suppose an empirical model \mathcal{E} gives rise to a support graph $\mathcal{G}_{\mathcal{E}}$ whose minimal independence number is less than $|\mathcal{C}|$. This means that there is a possible event $e : C \rightarrow \mathcal{O}$ with independence degree less than $|\mathcal{C}|$, the number of contexts. If $\lambda : \mathcal{M} \rightarrow \mathcal{O}$ is a hidden variable compatible with e (that is, $\lambda|_C = e$), then the set of events $I_\lambda = \{\lambda|_C : C \rightarrow \mathcal{O} \mid C \in \mathcal{C}\}$ is an independent subset of $\mathcal{G}(\mathcal{M}, \mathcal{C})$ of size $|\mathcal{C}|$. One of these events cannot be in the support graph $\mathcal{G}_{\mathcal{E}}$ and is therefore impossible. As there is a possible event e such that every compatible hidden variable must predict an impossible event, \mathcal{E} is logically contextual.

Conversely, suppose an empirical model \mathcal{E} gives rise to a support graph $\mathcal{G}_{\mathcal{E}}$ whose minimal independence number is $|\mathcal{C}|$ and e is a possible event. The event e must be contained in an independent subset I of $\mathcal{G}_{\mathcal{E}}$ of size $|\mathcal{C}|$. Construct the hidden variable $\lambda_I : \mathcal{M} \rightarrow \mathcal{O}$ by $\lambda_I(M) = i(M)$ for some $i \in I$ whose context contains M . Every coarse-graining of λ_I is in $I \subset \mathcal{G}_{\mathcal{E}}$ and therefore possible. So, λ_I is a canonical hidden variable that is compatible with e that does not predict an impossible event. We conclude that \mathcal{E} is not logically contextual.

Theorem 3. An empirical model \mathcal{E} is strongly nonlocal or contextual if and only if the independence number of its support graph $\mathcal{G}_{\mathcal{E}}$ is less than the number of contexts.

If the support graph $\mathcal{G}_{\mathcal{E}}$ of an empirical model \mathcal{E} contains an independent set I of size $|\mathcal{C}|$, then, as above, construct the canonical hidden variable λ_I . Each coarse-graining of λ_I is in $I \subset \mathcal{G}_{\mathcal{E}}$ and thus, is possible. Therefore, \mathcal{E} is not strongly nonlocal or contextual.

Conversely, suppose every independent subset of the support graph $\mathcal{G}_{\mathcal{E}}$ has size less than $|\mathcal{C}|$. If $\lambda : \mathcal{M} \rightarrow \mathcal{O}$ is any canonical hidden variable, then the set I_λ of its coarse-grainings $\lambda|_C$ form an independent subset of $\mathcal{G}(\mathcal{M}, \mathcal{C})$ of size $|\mathcal{C}|$ and thus at least one $i \in I$ is impossible. So, \mathcal{E} is strongly nonlocal or contextual.

Result 4. All n -qubit states are strongly contextual with respect to n -qubit stabilizer operations whenever $n > 1$.

Consider the measurement scenario consisting of those n -qubit Pauli operators whose square is the identity as measurements and, as contexts, those subsets which generate maximal Abelian subgroups of the n -qubit Pauli group P_n . The support graph of the maximally mixed state, i.e., those formal events which are possible within quantum theory, is the orthogonality graph of n -qubit stabilizer quantum mechanics. It has rank-1 stabilizer projections as vertices with

edges joining orthogonal projections. We must show that the independence number of this graph is less than the number of contexts $|C_n|$.

Suppose for contradiction that I is an independent vertex subset of size $|C_n|$. I must contain precisely one projector from each context for otherwise it will contain projectors onto two distinct eigenstates for the Paulis of the context which must be orthogonal and thus adjacent. We denote the projector chosen from a context C by I_C .

Thus, I yields a noncontextual value assignment $v : P_n \rightarrow \{\pm 1, \pm i\}$. For an n -Pauli $P \in P_n$ in a context $C \in C_n$, $v(P)$ is eigenvalue associated by P to the state I_C . This eigenvalue is independent of the choice of context for if P is a member of both C and C' , then it associates the same eigenvalue with both I_C and $I_{C'}$; otherwise, I_C and $I_{C'}$ belong to different eigenspaces of P which contradicts the assumption that no pair from I is orthogonal. We can extend v to all Paulis by multiplicativity. However, there exists no function $v : P_n \rightarrow \{\pm 1, \pm i\}$ mapping Paulis to one of their eigenvalues such that $v(-P) = -v(P)$ and $v(PP') = v(P)v(P')$ whenever P and P' commute.

We first reproduce the standard Mermin argument for two-qubit systems. Consider the table of Paulis

$Z \otimes I$	$I \otimes X$	$Z \otimes X$
$I \otimes Z$	$X \otimes I$	$X \otimes Z$
$Z \otimes Z$	$X \otimes X$	$Y \otimes Y$

Each entry has eigenvalues in the set $\{\pm 1\}$. The rows and columns give commuting triples of P_2 . Suppose a valuation v satisfying our hypotheses exists and consider the product of $v(P)$ over all P in each row or column. Since each entry appears in one row and one column, this is the product of $(\pm 1)^2$ nine times: 1. However, if we collect the terms into rows and columns, we find that we are taking the product of $v(I \otimes I)$ five times with $v(-I \otimes I)$ once. Therefore, the product is -1 and we reach a contradiction.

For the general $n > 2$ case, we repeat the same argument only we tensor all of the entries of the table with $I^{\otimes(n-2)}$ to get entries in P_n . This does not change the eigenvalues of the entries as I has only eigenvalue 1 nor does it affect the commutativity of the rows and columns. The product of the rows and columns yields $I^{\otimes n}$ five times and $-I^{\otimes n}$ once.

APPENDIX C: RELATION TO HYPERGRAPH APPROACH

Here, we show that the graph of Definition 1 coincides with the complement of the nonorthogonality graph [32, Definition 2.3.1] of the hypergraph [32, Definition D.1.4] associated by Acín *et al.* with a measurement scenario.

We first recall their relevant definitions before sketching a proof of the result.

Definition 4 (D.1.2). For a measurement $A \in \mathcal{M}$ in a measurement scenario $(\mathcal{M}, \mathcal{C})$, the *induced measurement scenario* $\mathcal{M}\{A\}$ has as measurements all $B \in \mathcal{M}$ such that $A \neq B$ and $\{A, B\}$ is contained within some context $C \in \mathcal{C}$. The contexts are the sets $C \setminus \{A\}$ for all $C \in \mathcal{C}$ such that $A \in C$.

Definition 5 (D.1.3). A *measurement protocol* $T \in MP(\mathcal{M})$ on a measurement scenario $(\mathcal{M}, \mathcal{C})$ is defined recursively as

(1) Base case: $\mathcal{M} = \emptyset$

$T = \emptyset$

(2) Recursive case: $\mathcal{M} \neq \emptyset$

$T = (A, f)$, where $A \in \mathcal{M}$ is a measurement and $f : \mathcal{O} \rightarrow MP(\mathcal{M}\{A\})$ is a function from the outcome set to the set of all measurement protocols on $\mathcal{M}\{A\}$.

Definition 6. A *protocol outcome* $\alpha \in POut(T)$ of a measurement protocol T is defined recursively as

(1) Base case: $T = \emptyset$

$\alpha = \{\emptyset\}$

(2) Recursive case: $T = (A, f)$

$\alpha = (A, a, \alpha')$, where $a \in \mathcal{O}$ and $\alpha' \in POut(f(a))$ is a protocol outcome of the measurement protocol $f(a)$.

A protocol outcome α yields an observable event $s(\alpha) : C(\alpha) \rightarrow \mathcal{O}$ where $C(\alpha) \in \mathcal{C}$, also defined recursively, that maps A to a at each stage.

Definition 7 (D.1.4). The *contextuality scenario* $H[\mathcal{M}]$ associated with a measurement scenario $(\mathcal{M}, \mathcal{C})$ is a hypergraph with all observable events (functions $e : C \rightarrow \mathcal{O}$ from some context $C \in \mathcal{C}$ to outcomes) as vertices. Every measurement protocol T on $(\mathcal{M}, \mathcal{C})$ defines a hyperedge in the hypergraph given by $\{s(\alpha) : \alpha \in POut(T)\}$.

Definition 8 (2.3.1). The *nonorthogonality graph* of a contextuality scenario $H[\mathcal{M}]$ has the same vertices. Two vertices share an edge if and only if they are not contained within a common hyperedge.

Lemma 2. Let $(\mathcal{M}, \mathcal{C})$ be a measurement scenario. For any $A \in \mathcal{M}$ and observable event $e : C \rightarrow \mathcal{O}$ where $A \in C \in \mathcal{C}$, there exists a measurement protocol $T \in MP(\mathcal{M})$ such that $T = (A, f)$ and $e = s(\alpha)$ for some $\alpha \in POut(T)$.

Proof. This follows by applying the recursive hypothesis to $\mathcal{M}\{A\}$, $A' \in (C \setminus A)$, and $e|_{(C \setminus A)} \rightarrow \mathcal{O}$ and choosing f to map $e(A)$ to the resulting induced measurement protocol. ■

Theorem 4. The exclusivity graph $\mathcal{G}(\mathcal{M}, \mathcal{C})$ of a measurement scenario $(\mathcal{M}, \mathcal{C})$ is isomorphic to the graph complement of the nonorthogonality graph of $H[\mathcal{M}]$.

Proof. Suppose $e : C \rightarrow \mathcal{O}$ and $e' : C' \rightarrow \mathcal{O}$ are mutually exclusive events, i.e., there is a measurement $A \in C \cap C'$ such that $e(A) \neq e'(A)$. We must show that they are contained within a common hyperedge by constructing a measurement protocol $T = (A, f)$ for which e and e' arise as $s(\alpha)$ and $s(\alpha')$ for $\alpha, \alpha' \in POut(T)$. This can be done by choosing f to map $e(A)$ to the measurement protocols yielded by applying Lemma 1 to $\mathcal{M}\{A\}$, $B \in (C \setminus A)$ and $e|_{(C \setminus A)} : (C \setminus A) \rightarrow \mathcal{O}$ (and similarly for $e'(A)$).

Conversely, suppose two distinct observable events e and e' are contained within a common hyperedge. There is thus a measurement protocol T for which e and e' arise as $s(\alpha)$ and $s(\alpha')$ for $\alpha, \alpha' \in POut(T)$. The minimal depth at which these protocol outcomes diverge gives the common measurement for which e and e' assign different outcomes. ■

[1] G. Vidal, *Phys. Rev. Lett.* **91**, 147902 (2003).

[2] D. Deutsch, *Proc. R. Soc. A London* **400**, 97 (1985).

[3] A. Datta, A. Shaji, and C. M. Caves, *Phys. Rev. Lett.* **100**, 050502 (2008).

- [4] J. Anders and D. E. Browne, *Phys. Rev. Lett.* **102**, 050502 (2009).
- [5] E. F. Galvão, *Phys. Rev. A* **71**, 042302 (2005).
- [6] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, *New J. Phys.* **14**, 113011 (2012).
- [7] N. Delfosse, P. Allard Guerin, J. Bian, and R. Raussendorf, *Phys. Rev. X* **5**, 021003 (2015).
- [8] R. Raussendorf, D. E. Browne, N. Delfosse, C. Okay, and J. Bermejo-Vega, [arXiv:1511.08506](https://arxiv.org/abs/1511.08506).
- [9] M. Howard, J. Wallman, V. Veitch, and J. Emerson, *Nature (London)* **510**, 351 (2014).
- [10] R. Raussendorf, *Phys. Rev. A* **88**, 022322 (2013).
- [11] J. S. Bell, *Physics* **1**, 195 (1964).
- [12] H. M. Wiseman, *J. Phys. A* **47**, 424001 (2014).
- [13] C. H. Bennett and S. J. Wiesner, *Phys. Rev. Lett.* **69**, 2881 (1992).
- [14] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Phys. Rev. Lett.* **70**, 1895 (1993).
- [15] C. H. Bennett and G. Brassard, in *International Conference on Computer System and Signal Processing, 1984* (IEEE, Piscataway, NJ, 1984), pp. 175–179.
- [16] L. Hardy, *Phys. Rev. Lett.* **68**, 2981 (1992).
- [17] L. Hardy, *Phys. Rev. Lett.* **71**, 1665 (1993).
- [18] D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger, *Am. J. Phys.* **58**, 1131 (1990).
- [19] W. van Dam, Ph.D. thesis, Faculty of Physical Sciences, University of Oxford, 1999.
- [20] J. Barrett, *Phys. Rev. A* **75**, 032304 (2007).
- [21] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, *Phys. Rev. A* **71**, 022101 (2005).
- [22] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, *Rev. Mod. Phys.* **82**, 665 (2010).
- [23] S. Kochen and E. P. Specker, *J. Math. Mech.* **17**, 59 (1967).
- [24] A. Fine, *Phys. Rev. Lett.* **48**, 291 (1982).
- [25] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, *Phys. Rev. Lett.* **101**, 020403 (2008).
- [26] S. Bravyi and A. Kitaev, *Phys. Rev. A* **71**, 022316 (2005).
- [27] R. Raussendorf and H. J. Briegel, *Phys. Rev. Lett.* **86**, 5188 (2001).
- [28] A. Cabello, S. Severini, and A. Winter, *Phys. Rev. Lett.* **112**, 040401 (2014).
- [29] S. Abramsky and A. Brandenburger, *New J. Phys.* **13**, 113036 (2011).
- [30] M. Howard and E. T. Campbell, [arXiv:1609.07488](https://arxiv.org/abs/1609.07488).
- [31] S. Abramsky and L. Hardy, *Phys. Rev. A* **85**, 062114 (2012).
- [32] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, *Commun. Math. Phys.* **334**, 533 (2015).
- [33] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969).
- [34] S. Popescu and D. Rohrlich, *Found. Phys.* **24**, 379 (1994).
- [35] M. Sadiq, P. Badziag, M. Bourennane, and A. Cabello, *Phys. Rev. A* **87**, 012128 (2013).
- [36] T. Fritz, A. B. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín, *Nat. Commun.* **4**, 2263 (2013).
- [37] A. B. Sainz, T. Fritz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín, *Phys. Rev. A* **89**, 032117 (2014).
- [38] A. Cabello, *Phys. Rev. Lett.* **86**, 1911 (2001).
- [39] A. Cabello, *Phys. Rev. A* **65**, 032108 (2002).
- [40] K. Yokota, T. Yamamoto, M. Koashi, and N. Imoto, *New J. Phys.* **11**, 033011 (2009).
- [41] Y. Aharonov, A. Botero, S. Popescu, B. Reznik, and J. Tollaksen, *Phys. Lett. A* **301**, 130 (2002).
- [42] A. C. Elitzur, S. Popescu, and D. Rohrlich, *Phys. Lett. A* **162**, 25 (1992).
- [43] J. Barrett, A. Kent, and S. Pironio, *Phys. Rev. Lett.* **97**, 170409 (2006).
- [44] S. Abramsky, R. S. Barbosa, K. Kishida, L. R., and S. Mansfield, *J. Math. Psychol.* **74**, 58 (2016).
- [45] D. Gottesman, Ph.D. thesis, California Institute of Technology, 1997.
- [46] D. Gross, *J. Math. Phys.* **47**, 122107 (2006).
- [47] H. Anwar, Ph.D. thesis, University College London, 2014.
- [48] H. Dawkins and M. Howard, *Phys. Rev. Lett.* **115**, 030501 (2015).
- [49] M. Howard and J. Vala, *Phys. Rev. A* **86**, 022316 (2012).
- [50] M. Howard (personal communication).
- [51] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, *Phys. Rev. Lett.* **104**, 230503 (2010).
- [52] N. D. Mermin, *Rev. Mod. Phys.* **65**, 803 (1993).
- [53] M. Howard, E. Brennan, and J. Vala, *Entropy* **15**, 2340 (2013).
- [54] S. Abramsky, R. Soares Barbosa, K. Kishida, R. Lal, and S. Mansfield, [arXiv:1502.03097](https://arxiv.org/abs/1502.03097).
- [55] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman, *Phys. Rep.* **506**, 1 (2011).
- [56] R. W. Spekkens, *Phys. Rev. A* **71**, 052108 (2005).