(Non-)Contextuality of physical theories as an axiom

Adán Cabello,[1] Simone Severini,[2] and Andreas Winter[3]
[1] *Universidad de Sevilla*
[2] *University College London*
[3] *University of Bristol, and National University of Singapore*
(Dated: 14 October 2010)

We show that there is of family of inequalities associated to each compatibility structure of a set of events (a graph), such that the bound for noncontextual theories is given by the independence number of the graph, and the maximum quantum violation is given by the Lovász ϑ-function of the graph, which was originally proposed as an upper bound on its Shannon capacity. Probabilistic theories beyond quantum mechanics may have an even larger violation, which is given by the fractional packing number. We discuss the sets of probability distributions attainable by noncontextual, quantum, and generalized models; the latter two are shown to have semidefinite and linear characterizations, respectively. The implications for Bell inequalities are discussed. In particular, we show that every Bell inequality can be recast as a noncontextual inequality within this family.

For details and proofs see arXiv:1010.2163 [quant-ph].

Introduction.—Recently, Klyachko *et al.* (KCBS) [1] have introduced a noncontextual inequality (*i.e.*, one satisfied by any noncontextual hidden variable theory), which is violated by quantum mechanics, and therefore can be used to detect quantum effects. The simplest physical system which exhibits quantum features in this sense is a three-level quantum system or qutrit [2–4]. The KCBS inequality is the simplest noncontextual inequality violated by a qutrit. It can adopt two equivalent forms. Consider 5 *yes-no* questions P_i ($i = 0, \ldots, 4$) such that P_j and P_{j+1} (with the sum modulo 5) are *compatible*: both questions can be jointly asked without mutual disturbance, so, when the questions are repeated, the same answers are obtained; and *exclusive*: not both can be true. One can represent each of these questions as a vertex of a pentagon (*i.e.*, a 5-cycle) where the edges denote compatibility and exclusiveness. What is the maximum number of *yes* answers one can get when asking the 5 questions to a physical system? Clearly, two, because of the exclusiveness condition [6]. If we denote *yes* and *no* by 1 and 0, respectively, then, even if we ask only a single question to each one of an identically prepared collection of systems, and then count the average number of *yes* answers corresponding to each question, then $\beta := \sum_{i=0}^{4} \langle P_i \rangle \leq 2$, if we assume that these answers are predetermined by a hidden variable theory. This is the first form of the KCBS inequality. What has $\beta \leq 2$ to do with noncontextuality? Noncontextual hidden variable theories are those in which the answer of P_j is independent of whether one ask P_j together with P_{j-1} (which is compatible with P_j), or together with P_{j+1} (which is also compatible with P_j). A set of mutually compatible questions is called a *context*. Since, P_{j+1} and P_{j-1} are not necessarily compatible, $\{P_j, P_{j-1}\}$ is one context and $\{P_j, P_{j+1}\}$ is a different one, and they are not both contained in a joint context. The assumption is that the answer to P_j will be the same in both.

Now, let us consider contexts instead of questions, *i.e.*, let us ask individual systems not one but two compatible and exclusive questions. In the pentagon, a context is represented by an edge connecting two vertices, so we have 5 different contexts. In order to study the correlations between the answers to these questions, it is useful to transform each question into a dichotomic observable with possible values -1 (no) or $+1$ (yes), so when both questions give the same answer the product of the results of the observables is $+1$, but when the answers are different then the product of the results of the observables is -1. For instance, this can be done by defining the observables $A_i = 2P_i - 1$. Then, inequality $\beta \leq 2$ is equivalent to the noncontextual correlation inequality, the second form of KCBS, $\beta' := \sum_{i=0}^{4} \langle A_i A_{i+1} \rangle \geq -3$, which can be derived independently based solely on the assumption that the observables A_i have noncontextual results -1 or $+1$. *I.e.*, we do not need to assume exclusiveness to derive it, effectively because the occurrence of correlation functions $\langle A_i A_{i+1} \rangle$ implements a penalty for violating exclusiveness.

For a qutrit, the maximum quantum violation is $\beta_{\text{QM}} = \sqrt{5} \approx 2.236$, which is equivalent to $\beta'_{\text{QM}} = 5 - 4\sqrt{5} \approx -3.94$. The maximum violation occurs for vectors connecting the origin with the vertices of a regular pentagon. These vectors form an orthonormal representation of the 5-cycle.

General compatibility structures.—The KCBS inequality suggests itself a generalization to arbitrary graphs instead of the 5-cycle. Most generally, Kochen-Specker (KS) theorems [4] are about the possibility of interpreting a given structure of compatibility of "events," and additional constraints such as exclusiveness, in a classical or nonclassical probabilistic theory. These events are interpreted as *atomic* events, each of which can occur in different contexts. Formally, the events are labelled by a set V. The set of all valid contexts is a *hypergraph* Γ, which is simply a collection of subsets $C \subset V$; note that for hypergraphs of contexts, with each $C \in \Gamma$, all of the subsets of C are also valid contexts, and hence part of C. The interpretation is that there should exist (deterministic) events in a probabilistic model, one P_i for each $i \in V$, and for each context C a measurement among whose outcomes are the P_i ($i \in C$). The events are hence mutually exclusive, as in the measurement postulated to exist for some $C \in \Gamma$, at most one outcome $i \in C$ can occur. For instance, a *classical (noncontextual) model* would be a measurable space Ω, with each P_i being the indicator function of a measurable set (an event, in fact) such that for all $C \in \Gamma$, $\sum_{i\in C} P_i \leq 1$ (*i.e.*, the supporting sets of the P_i should be pairwise disjoint).

In contrast, a *quantum model* requires a Hilbert space $\mathcal{H}$ and associates projection operators P_i to all $i \in V$, such that for all $C \in \Gamma$, $\sum_{i\in C} P_i \leq \mathbb{1}$ (*i.e.*, the P_i can be thought of as outcomes in a von Neumann measurement).

Thanks to KS we know that quantum models are strictly more powerful that classical ones; but they are still not the most general ones. A *generalized model* requires choosing a generalized probabilistic theory in which the P_i can be interpreted as measurement outcomes: following [7–11], formally it consists of a real vector space $\mathcal{A}$ of observables, with a distinguished unit element $u \in \mathcal{A}$ and a vector space order: the latter is given by the closed convex cone $\mathcal{P} \subset \mathcal{A}$ of positive elements containing u in its interior, such that $\mathcal{P}$ spans $\mathcal{A}$ and is pointed, meaning that, with the exception of 0, $\mathcal{P}$ is entirely on one side of a hyperplane. For two elements $X, Y \in \mathcal{A}$ we then say $X \leq Y$ if and only if $Y - X \in \mathcal{P}$. (We shall only discuss finite dimensional $\mathcal{A}$, otherwise there will be additional topological requirements.) The elements with $0 \leq E \leq u$ are called *effects*. This structure is enough to talk about measurements: they are collections of effects $(E_1, \ldots, E_k)$ such that $\sum_{j=1}^k E_j = u$.

Now, a generalized model for the hypergraph Γ is the association of an effect $P_i \in \mathcal{A}$ to each $i \in V$, such that each P_i is a sum of normalized extremal effects, and for all $C \in \Gamma$, $\sum_{i\in C} P_i \leq u$. The latter condition ensures that the family $(P_i : i \in C)$ can be completed to a measurement, possibly in a larger space $\widetilde{\mathcal{A}} \supset \mathcal{A}$. We finally demand that this can be done such that also $u - \sum_{i\in C} P_i$ is a sum of normalized extremal effects.

Notice that in all of the above we never require that any particular context should be associated to a complete measurement: the conditions only make sure that each context is a subset of outcomes of a measurement and that they are mutually exclusive. Thus, unlike the original KS theorem, it is clear that every context hypergraph Γ has always a classical noncontextual model, besides possibly quantum and generalized models. This is where noncontextual inequalities come in: note that all of the above types of models allow for the choice of a state (be it a probability density, a quantum density operator, or generalized state), under which all expectation values $\langle P_i \rangle$ make sense, and hence also the expression $\beta = \sum_{i\in V} \langle P_i \rangle$. Moreover, all probabilities $\langle P_i \rangle$ are independent of the context in which P_i occurs, as they depend only on the effect P_i and the underlying state. Since this is the condition underlying Gleason's theorem, we call it the *Gleason property*.

We can then ask for the set of all attainable vectors $\big(\langle P_i \rangle\big)_{i\in V}$ for given hypergraph Γ, over all models of a given sort (classical noncontextual, quantum mechanical, or generalized probabilistic theory) and states within it. These are evidently convex subsets in $[0,1]^V \subset \mathbb{R}^V$; we denote the sets of noncontextual, quantum and generalized expectations by $\mathcal{E}_{\mathrm{C}}(\Gamma)$, $\mathcal{E}_{\mathrm{QM}}(\Gamma)$ and $\mathcal{E}_{\mathrm{GPT}}(\Gamma)$, respectively. The central task of the present theory is to characterize these convex sets and to compare them for various Γ. This is because a point $\vec{p} \in \mathcal{E}_{\mathrm{X}}(\Gamma)$ in any of these sets describes the outcome probabilities of any compatible set of events (*i.e.*, any context). Note that all of them are *corners* in the language of [12]: if $0 \leq q_i \leq p_i$ for all $i \in V$, then $\vec{p} \in \mathcal{E}_{\mathrm{X}}(\Gamma)$ implies also $\vec{q} \in \mathcal{E}_{\mathrm{X}}(\Gamma)$.

In particular, the extreme values of β over these sets are denoted $\beta_{\mathrm{C}}(\Gamma)$, $\beta_{\mathrm{QM}}(\Gamma)$, and $\beta_{\mathrm{GPT}}(\Gamma)$, respectively. It is clear that

$$\beta_{\mathrm{C}}(\Gamma) \leq \beta_{\mathrm{QM}}(\Gamma) \leq \beta_{\mathrm{GPT}}(\Gamma) \tag{1}$$

by definition.

Maximum values.—For given hypergraph Γ, we can define the adjacency graph G on the vertex set V: two $i, j \in V$ are joined by an edge if and only if there exists a $C \in \Gamma$ such that both $i, j \in C$. Then,

$$\beta_{\mathrm{C}}(\Gamma) = \alpha(G), \quad \beta_{\mathrm{QM}}(\Gamma) = \vartheta(G), \tag{2}$$

where $\alpha(G)$ is the independence number of the graph, *i.e.* the maximum number of pairwise disconnected vertices, and $\vartheta(G)$ is the Lovász ϑ-function of G [12–14], defined as follows: First, an *orthonormal representation* (OR) of a graph is a set of unit vectors associated to the vertices such that two vectors are orthogonal if the corresponding vertices are adjacent. Then,

$$\vartheta(G) := \max \sum_{i=1}^{n} |\langle \psi | v_i \rangle|^2, \tag{3}$$

where the maximum is taken over all unit vectors $|\psi\rangle$ (in Eucledian space) and ORs $\{|v_i\rangle : i = 1, \ldots, n\}$ of G [15]. Furthermore, $\vartheta(G)$ is given by a semidefinite program (SDP) [13], which explains the key importance of this number for combinatorial optimization and zero-error information theory – indeed $\vartheta(G)$ is an upper bound to the Shannon capacity of a graph [13].

Observe that this says in particular that when discussing classical and quantum models, we never need to consider contexts of more than two events. Indeed, it is a (nontrivial) property of these models that if in a set of events any pair is compatible and exclusive, then so is the whole set; more generalized probabilistic theories do not have this property, cf. [16].

To prove Eq. (2), we notice that, for a given probabilistic model, the expectation is always maximized on an extremal, *i.e.* pure, state. In the classical case, this amounts to choosing a point $\omega \in \Omega$, so that $w_i := P_i(\omega)$ is a 0-1-valuation of the set V. By definition, it has the property that, in each hyperedge $C \in \Gamma$, at most one element is marked 1, and β is simply the number of marked elements. It is clear that the marked elements form an independent set in Γ (and equivalently in the graph G). In the quantum case, let the maximizing state be given by a unit vector $|\psi\rangle$, and for each i, $\langle\psi|P_i|\psi\rangle = |\langle\psi|v_i\rangle|^2$, for $|v_i\rangle := P_i|\psi\rangle/\sqrt{\psi|P_i|\psi\rangle}$. This clearly is an orthogonal representation of G, in fact the projectors $|v_i\rangle\langle v_i|$ form another quantum model of Γ, with the same maximum value of β, which by the definition we gave earlier is just Lovász' $\vartheta(G)$.

Each graph G where $\alpha(G) < \vartheta(G)$ thus exhibits a limitation of classical noncontextuality, which can be witnessed in experiments with an appropriate set of projectors, and on an appropriate state. In this sense, each such graph provides a proof of the KS theorem.

Taking $n \geq 5$ odd and applying a result from [13] to $G = C_n$, the n-cycle, one obtains the same noncontextual and quantum bounds recently obtained in [16].

It is known that $\vartheta(G)$ can be much larger than $\alpha(G)$; in particular, it is known that (for appropriate, arbitrarily large n) there are graphs G with $\vartheta(G) \approx \sqrt{n}$ but $\alpha(G) \approx 2\log n$, and others with $\vartheta(G) \approx \sqrt[4]{n}$ but $\alpha(G) = 3$ [17]. Hence, the quantum violation of noncontextual inequalities can be arbitrarily large.

Description of the probability sets.—We now show that arbitrary linear functions can be optimized over $\mathcal{E}_{\rm QM}(\Gamma)$ as semidefinite programs: for an arbitrary vector $\vec{\lambda} \in R^V$, let

$$\vec{\lambda}(\mathcal{E}_{\rm QM}(\Gamma)) = \max \sum_i \lambda_i p_i \text{ s.t. } \vec{p} \in \mathcal{E}_{\rm QM}(\Gamma). \tag{4}$$

First of all, without loss of generality, all λ_i are nonnegative; this follows because $\mathcal{E}_{\rm QM}(\Gamma)$ is a corner and hence $\vec{\lambda}(\mathcal{E}_{\rm QM}(\Gamma))$ is unchanged when we replace all negative λ_i by 0. Using the ideas of [13], this can be recast as the following SDP

$$\begin{aligned}\vec{\lambda}(\mathcal{E}_{\rm QM}(\Gamma)) &= \max \operatorname{tr} \Lambda T \\ &\quad \text{s.t. } T \geq 0,\ \operatorname{tr} T = 1,\ i \sim j \Rightarrow T_{ij} = 0.\end{aligned} \tag{5}$$

The value $\vec{\lambda}(\mathcal{E}_{\rm QM}(\Gamma))$ is known as a *weighted Lovász number* (or *ϑ-function*) [12].

The previous discussion implies that not only function optimization, but also membership in $\mathcal{E}_{\rm QM}(\Gamma)$ is an efficient convex problem: there is a polynomial-time algorithm that, given a vector $\vec{p}$, tests whether it is in $\mathcal{E}_{\rm QM}(\Gamma)$ or not. This follows from general considerations of convex optimisation [19–21].

Does there exist such a nice and efficient description also for the classical set $\mathcal{E}_{\rm C}(\Gamma)$? The fact that the maximum of β over it is the independence number $\alpha(G)$, which is well-known to be NP complete, means that the answer is "no." In fact, $\mathcal{E}_{\rm C}(\Gamma)$ encodes the independence numbers $\alpha(G|_S)$ of all induced subgraphs of G on subsets $S \subset V$, and the best description that we have is as the following 0-1-polytope:

$$\mathcal{E}_{\rm C}(\Gamma) = \operatorname{conv}\big\{\vec{\sigma} : \sigma_i \in \{0,1\},\ i \sim j \Rightarrow \sigma_i\sigma_j = 0\big\}. \tag{6}$$

The difficulty in evaluating $\beta_{\rm GPT}(\Gamma)$ lies in capturing the constraint that the P_i have to be sums of extremal, normalized effects in the generalized probabilistic theory. If we relax this condition simply to P_i having to be an effect, we arrive at what we would like to call a *fuzzy model*, which formalizes the notion that all $\{P_i : i \in C\}$ are compatible, but not necessarily exclusive events: so we are left with Gleason's constraints $0 \leq \langle P_i\rangle \leq 1$ and for all $C \in \Gamma$, $\sum_{i\in C}\langle P_i\rangle \leq 1$. Denote the (convex) set of all expectations $\big(\langle P_i\rangle\big)_{i\in V}$ when varying over models and their states by $\mathcal{E}_{\rm F}(\Gamma)$. Then,

$$\beta_{\rm GPT}(\Gamma) = \beta_{\rm F}(\Gamma) = \alpha^*(\Gamma), \tag{7}$$

where $\alpha^*(\Gamma)$ is the *fractional packing number* of the hypergraph Γ, defined by the following intuitive linear program:

$$\begin{aligned}\alpha^*(\Gamma) &= \max \sum_{i\in V} w_i \\ &\quad \text{s.t. } \forall i\ 0 \leq w_i \leq 1 \text{ and } \forall C \in \Gamma\ \sum_{i\in C} w_i \leq 1.\end{aligned} \tag{8}$$

The vectors $\vec{w}$ are *fractional packings of* Γ.

Conversely, given a fractional covering $\vec{w}$, we can indeed show that there is an appropriate generalized probabilistic model with effects P_i and a state, such that $w_i = \langle P_i\rangle$.

This proves actually that $\mathcal{E}_{\rm GPT}(\Gamma) = \mathcal{E}_{\rm F}(\Gamma)$, the set of fractional packings. This means that *any* linear function of expectation values can be optimized over $\mathcal{E}_{\rm GPT}(\Gamma)$ as a linear program; likewise, checking whether $\vec{p}$ is in $\mathcal{E}_{\rm GPT}(\Gamma)$ is a linear programming feasibility problem. □

For an example, for the n-cycles above, $\alpha^*(C_n) = n/2$, regardless of the parity of n, which is strictly larger than $\vartheta(C_n)$ for all odd $n \geq 5$. Again, we know of arbitrarily large separations: there are hypergraphs Γ such that the adjacency graph G is the complete graph K_n, hence $\alpha(G) = \vartheta(G) = 1$, yet $\alpha^*(\Gamma) \gg 1$ [18].

Bell inequalities.—Where does nonlocality come into this? After all, Bell inequalities exploit locality in the form that one party's measurement is compatible with another party's, and that the former's outcomes are independent of the latter's choices (*i.e.*, insensitive to different contexts). We can model this also in our setting, by going to the atomic events, which are labelled by a list of settings and outcomes for each party. For instance, for bipartite scenarios, let Alice and Bob's settings be $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, respectively, and their respective outcomes be $a \in \mathcal{A}$ and $b \in \mathcal{B}$. Then, we construct a graph with vertex set $V = \mathcal{A} \times \mathcal{B} \times \mathcal{X} \times \mathcal{Y}$ and edges $abxy \sim a'b'x'y'$ if and only if ($x = x'$ and $a \neq a'$) or ($y = y'$ and $b \neq b'$), encoding precisely that two events in V are connected in the graph if and only if they are compatible and mutually exclusive (as events in the Bell experiment as a whole). Let Γ be the hypergraph of all cliques in G.

We can now discuss classical noncontextual, quantum and generalized models for this graph, and hence also noncontextual inequalities, restricting as above to linear functions $\vec{\lambda}\vec{p}$ of the vector of the probabilities $p_{ab|xy} = \langle P_{abxy} \rangle$, with with non-negative coefficient vector $\vec{\lambda}$. Note that any Bell inequality can always be rewritten in such a form, by removing negative coefficients using the identity $-p_{ab|xy} = -1 + \sum_{a'b' \neq ab} p_{a'b'|xy}$ for all x, y, a, and b. These equations are not automatically realized in the sets $\mathcal{E}_{\mathrm{X}}(\Gamma)$, $\mathrm{X} = \mathrm{C}, \mathrm{QM}, \mathrm{GPT}$ – as indeed in the underlying (classical, quantum or generalized) model it needs not hold that $\sum_{ab} P_{abxy}$ is the unit element, for any x, y. Hence, define for any class of models $\mathrm{X} = \mathrm{C}, \mathrm{QM}, \mathrm{GPT}$,

$$\mathcal{E}^1_{\mathrm{X}}(\Gamma) := \mathcal{E}_{\mathrm{X}}(\Gamma) \cap \left\{ \vec{p} : \forall xy \ \sum_{ab} p_{ab|xy} = 1 \right\}, \quad (9)$$

the set of probability assignments consistent with the contextuality structure Γ, and in addition satisfying normalization.

We can prove that $\mathcal{E}^1_{\mathrm{C}}(\Gamma)$ is the set of correlations explained by local hidden variable theories, and that $\mathcal{E}^1_{\mathrm{GPT}}(\Gamma)$ are exactly the no-signalling correlations. Furthermore, to calculate the local hidden variable value Ω_c of a given Bell inequality with non-negative coefficient vector $\vec{\lambda}$, it holds that $\Omega_c = \vec{\lambda}(\mathcal{E}^1_{\mathrm{C}}(\Gamma)) = \vec{\lambda}(\mathcal{E}_{\mathrm{C}}(\Gamma))$. In this sense, any Bell inequality is at the same time a noncontextual inequality for the underlying graph G.

With classical and no-signalling correlations taken care of, we turn our attention to the quantum case. We can also prove that the following subset of $\mathcal{E}^1_{\mathrm{QM}}(\Gamma)$ is precisely the set of correlations obtainable by local quantum measurements on a bipartite state:

$$\mathcal{E}^{\mathbb{1}}_{\mathrm{QM}}(\Gamma) = \left\{ \left(\langle P_{abxy} \rangle \right)_{abxy} : \forall xy \ \sum_{ab} P_{abxy} = \mathbb{1} \right\}. \quad (10)$$

This means that, for a given Bell inequality with coefficients $\vec{\lambda}$, the maximum quantum value is $\Omega_q = \vec{\lambda}(\mathcal{E}^{\mathbb{1}}_{\mathrm{QM}}(\Gamma))$. For the time being, we do not know whether the set of quantum correlations, *i.e.* $\mathcal{E}^{\mathbb{1}}_{\mathrm{QM}}(\Gamma)$, is efficient to characterize. It follows, however, from the above considerations and the general theory of convex optimization [19–21] that the – potentially larger – set $\mathcal{E}^1_{\mathrm{QM}}(\Gamma)$ can be decided efficiently. In fact, we shall see directly that the maximum values $\vec{\lambda}(\mathcal{E}^1_{\mathrm{QM}}(\Gamma))$ are computed to arbitrary precision by semidefinite programming, thus providing efficient upper bounds to Ω_q. Implementing this for example for the CHSH inequality [5], we recover the Tsirelson bound $2\sqrt{2}$ [22]. For the I_{3322} inequality [23] the method yields the upper bound 0.251 47 on the quantum value; the currently best upper bound is slightly smaller [24], from which we conclude that $\mathcal{E}^{\mathbb{1}}_{\mathrm{QM}}(\Gamma)$ is strictly contained in $\mathcal{E}^1_{\mathrm{QM}}(\Gamma)$.

[1] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Phys. Rev. Lett. **101**, 020403 (2008).
[2] A. M. Gleason, J. Math. Mech. **6**(6), 885 (1957).
[3] J. S. Bell, Rev. Mod. Phys. **38**, 447 (1966).
[4] S. Kochen and E. P. Specker, J. Math. Mech. **17**, 59 (1967).
[5] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[6] R. Wright, in *Mathematical Foundations of Quantum Mechanics*, edited by A. R. Marlow (Academic Press, San Diego, 1978), p. 255.
[7] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, New York, 1963).
[8] G. Ludwig, Z. Phys. **181**(3), 233 (1964).
[9] G. Ludwig, Comm. Math. Phys. **4**(5), 331 (1967).
[10] A. S. Holevo, *Statistical Structure of Quantum Theory* (Springer, Berlin, 2001).
[11] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
[12] D. Knuth, Elec. J. Comb. **1**, 1 (1994).
[13] L. Lovász, IEEE Trans. Inf. Theory **25**, 1 (1979).
[14] J. Körner and A. Orlitsky, IEEE Trans. Inf. Theory, **44**, 2207 (1998).
[15] L. Lovász, *Geometric Representations of Graphs*, http://www.cs.elte.hu/~lovasz/geomrep.pdf
[16] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman, arXiv:1010.1273 [quant-ph].
[17] R. Peeters, Combinatorica **16**, 417 (1996).
[18] T. S. Cubitt, D. W. Leung, W. Matthews, and A. Winter, arXiv:1003.3195 [quant-ph].
[19] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization* (Springer, Berlin, 1988).
[20] D. Bertsimas and S. Vempala, J. ACM **51**, 540 (2004).
[21] Y.-K. Liu, arXiv:0712.3041 [quant-ph].
[22] B. S. Cirel'son [Tsirelson], Lett. Math. Phys. **4**, 93 (1980).
[23] N. Brunner and N. Gisin, Phys. Lett. A **327**, 3162 (2008).
[24] K. F. Pál and T. Vértesi, arXiv:1006.3032 [quant-ph].