Article:

Cope, Thomas and Colbeck, Roger Andrew orcid.org/0000-0003-3591-0576 (2019) Bell inequalities from no-signaling distributions. Physical Review A. 022114. ISSN 1094-1622

https://doi.org/10.1103/PhysRevA.100.022114

Bell Inequalities From No-Signalling Distributions

Thomas Cope* and Roger Colbeck†
Department of Mathematics, University of York, Heslington, York, YO10 5DD, UK
(Dated: 27th June 2019)

A Bell inequality is a constraint on a set of correlations whose violation can be used to certify non-locality. They are instrumental for device-independent tasks such as key distribution or randomness expansion. In this work we consider bipartite Bell inequalities where two parties have m_A and m_B possible inputs and give n_A and n_B possible outputs, referring to this as the (m_A, m_B, n_A, n_B) scenario. By exploiting knowledge of the set of extremal no-signalling distributions, we find all 175 Bell inequality classes in the $(4, 4, 2, 2)$ scenario as well as providing a partial list of 18277 classes in the $(4, 5, 2, 2)$ scenario. We also use a probabilistic algorithm to obtain 5 classes of inequality in the $(2, 3, 3, 2)$ scenario, which we confirmed to be complete, 25 classes in the $(3, 3, 2, 3)$ scenario, and a partial list of 21170 classes in the $(3, 3, 3, 3)$ scenario. Our inequalities are given in supplementary files. Finally, we discuss the application of these inequalities to the detection loophole problem, and provide new lower bounds on the detection efficiency threshold for small numbers of inputs and outputs.

I. INTRODUCTION

Bell inequalities [1] can be thought of as constraints on the set of correlations realisable at spacelike separation by using shared classical randomness and freely chosen measurements. One of the most counterintuitive features of quantum theory is that by sharing quantum systems instead of classical randomness, these inequalities can be violated, a fact that has been subject to extensive experimental investigation [2–6]. This curious feature has since been used for cryptography [7] and shown to enable device-independent information processing tasks, such as quantum key distribution [8–11] and randomness expansion [12–15]. In essence, that shared classical randomness cannot explain the observed correlations implies that any eavesdropper can have only limited information about them, and so key or random numbers can be distilled.

In spite of their usefulness, relatively little is known about the set of Bell inequalities in all but the simplest cases. In part, this is due to the complexity of finding them and the fact that the number of such inequalities grows rapidly as the number of inputs or outputs is increased. Bell inequalities can be thought of in a geometric way as the facets of the polytope of local (classically realisable) correlations. This insight means that Bell inequalities can in principle be found by facet enumeration, a well-known problem in polytope theory. However, because it quickly becomes intractable, this method has only been applied to enumerate all Bell inequalities in simple cases.

In this paper we propose some alternative algorithms. These algorithms are relatively easy to run, although they have the disadvantage that they don't give a certificate when all the Bell inequalities have been found. In several cases we have been able to enumerate the complete list of Bell inequality classes[1], whilst in other cases we find lower bounds on the total number of classes of Bell inequality, which informs the complexity of doing facet enumeration in these cases. Our approach utilises knowledge of the set of extremal no-signalling distributions, as opposed to techniques based on enumerating the facets of simpler, related polytopes [16, 17], or shelling techniques which travel along the polytope's edges [17].

In particular, we provide a complete list of the 175 classes of Bell inequalities for the $(4, 4, 2, 2)$ scenario, normalized in such a way as to provide easy comparison. The number of inequalities for this scenario was already known [18] but a useable list was not provided, unlike for the $(2, 2, 2, 2)$ [19, 20], $(2, m, 2, n)$ [21], $(3, 3, 2, 2)$ [20, 22, 23], $(3, 4, 2, 2)$ [23], $(3, 5, 2, 2)$ [23, 26] and $(2, 2, 3, 3)$ [24, 25] scenarios, which we summarize in Table I.

In addition we investigate a number of other scenarios. In the $(2, 3, 3, 2)$ scenario we find 5 classes of Bell inequality, which we confirmed to be complete using facet enumeration. In the $(3, 3, 2, 3)$ scenario we find 25 classes of Bell inequality, which may be complete. In the $(4, 5, 2, 2)$ scenario we find 18277 classes of inequality. Finally, in the $(3, 3, 3, 3)$ scenario we find 21170 classes, adding substantially to the set of 19 known classes for this scenario (one was found in [23] and 18 in [27]). In the last two cases, the sets of classes found are incomplete. Explicit representatives

---

*thomas.cope@york.ac.uk
†roger.colbeck@york.ac.uk
[1] Two Bell inequalities are in the same class if they are related by a relabelling.

Scenario	Number of Inequality Classes	Number of Facets	Reference
$(2,2,2,2)$	2	24	[19, 20]
$(2,m_B,2,n_B)$	2	$2(2^{n_B}-2)(m_B^2-2)+4n_Bm_B$	[21]
$(2,2,3,3)$	3	1116	[23–25]
$(3,3,2,2)$	3	684	[20, 22, 23]
$(2,3,3,2)$	5	1260	Sec. IV C
$(3,4,2,2)$	6	12480	[18, 23]
$(3,5,2,2)$	7	71340	[18, 26]
$(4,4,2,2)$	175	36391264	[18], Sec. IV B
$(3,3,2,3)$	≥ 25	≥ 252558	Sec. IV D
$(4,5,2,2)$	≥ 18277		Sec. IV E
$(3,3,3,3)$	≥ 21170		Sec. IV F

TABLE I: Known Bell inequality classes for bipartite scenarios. For all scenarios, one of the classes always corresponds to the trivial positivity inequalities. Those in the upper section are known to be complete, whilst for the final three we provide lower bounds on the number of classes. We suspect the $(3,3,2,3)$ may be complete (see Sec. IV D for details).

of each class are given as supplementary files.

The structure of the paper is as follows. In Section II we give an overview of relevant polytope theory and the related topic of linear programming, before discussing Bell inequalities and detailing our representation of them. In Section III, we discuss how to exploit knowledge of the set of extremal no-signalling distributions to obtain new Bell inequalities, as well as a technique for doing so without such knowledge. Section IV then gives the results. Finally, in Section V we apply these new inequalities to the problem of the detection loophole, presenting some numerical results and list the new inequalities which are the most promising candidates for lowering the detection threshold for small numbers of inputs and outputs.

II. PRELIMINARIES

A. Polytope Theory

A polytope is a convex set that can be described by the intersection of a finite number of half-spaces[2]. Given $A \in \mathbb{R}^{r\times t}$ and $\mathbf{c} \in \mathbb{R}^r$ we can write

$$\mathcal{P} = \left\{ \mathbf{x} \in \mathbb{R}^t \,\middle|\, A\mathbf{x} \geq \mathbf{c} \right\} \tag{1}$$

(each $A_k\mathbf{x} \geq c_k$ describes a half-space). This is called an *H-representation* of the polytope. Polytopes may also be described using a *V-representation*. If a polytope is bounded, then for some set $\{\mathbf{x}_k\}$ with $\mathbf{x}_k \in \mathbb{R}^t$ it can be written

$$\mathcal{P} = \left\{ \mathbf{x} = \sum_k \lambda_k \mathbf{x}_k \,\middle|\, \sum_k \lambda_k = 1,\ \lambda_k \geq 0 \right\}. \tag{2}$$

According to the Minkowski-Weyl theorem, every polytope admits both a V-representation and H-representation. We will always deal with *minimal representations* (in which unnecessary half-spaces or points $\mathbf{x}_i$ are removed). In a minimal representation $\{\mathbf{x}_k\}$ are *vertices*[3]. Given a polytope $\mathcal{P} \subset \mathbb{R}^t$ with dimension $d \leq t$, the intersection of $\mathcal{P}$ with a bounding hyperplane $A_k\mathbf{x} = c_k$ is a *facet* of the polytope if it has dimension $d-1$. The minimal H-representation is exactly the set of facet-defining hyperplanes.

Converting from V-representation to H-representation is known as *facet enumeration* whilst going from H-representation to V-representation is known as *vertex enumeration*. For a polytope of dimension d, given a V-representation with n vertices (or an H-representation with n half-spaces) there are algorithms that can perform this conversion in time $O(ndr)$, with r the number of facets (vertices) enumerated [29][4]. When performing facet enumeration, r is generally not known in advance, and hence the worst case scenario is often used to provide an upper bound.

[2] For a detailed summary of polytope theory, see [28].
[3] For an unbounded polytope the V-representation will also have *rays*.
[4] Some improvements on this have been achieved for specific classes of polytope.

For a given dimension and number of vertices, the so called *cyclic polytope* [30] has the largest possible number of facets, $\binom{n-\lfloor\frac{d-1}{2}\rfloor}{n-d}+\binom{n-\lfloor\frac{d-2}{2}\rfloor}{n-d}$. Using this we obtain complexity $O(n^{\lfloor\frac{d}{2}\rfloor})$ [30]. By contrast the simplest polytope is the simplex, with only $d+1$ facets.

In this work it is convenient to represent points using matrices rather than vectors. In this case, the H-representation of a polytope is based on a set $\{B_i\}$ with $B_i \in \mathbb{R}^{s\times t}$ and a set of real numbers c_i and can be expressed as

$$\left\{\Pi \in \mathbb{R}^{s\times t} \,\middle|\, \mathrm{tr}(B_i^T\Pi) \geq c_i \,\forall\, i\right\}. \tag{3}$$

Likewise the V-representation is formed via a set $\{\Pi_k\}$, $\Pi_k \in \mathbb{R}^{s\times t}$ as

$$\left\{\Pi = \sum_{k=1}^{n} \lambda_k \Pi_k \,\middle|\, \sum_{k=1}^{n} \lambda_k = 1,\ \lambda_k \geq 0\right\}. \tag{4}$$

B. Linear Programming

A linear programming problem involves the optimization of a linear objective function over a set of variables constrained by a finite number of linear equalities and/or inequalities [31]. The canonical form of a linear programming problem is as follows: given a fixed $\mathbf{c} \in \mathbb{R}^n, \mathbf{q} \in \mathbb{R}^d$ and $G \in \mathbb{R}^{d\times n}$,

$$\max_{\mathbf{x}}\ \mathbf{c}^T\mathbf{x} \quad \text{subject to} \quad G\mathbf{x} \leq \mathbf{q},\ \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x} \geq \mathbf{0}\,. \tag{5}$$

For our later considerations it is convenient to rewrite this using $A_i, Q \in \mathbb{R}^{s\times t}$ as

$$\max_{\{x_i\}}\ \sum_i c_i x_i \quad \text{subject to} \quad \sum_i x_i A_i \leq Q,\ x_i \geq 0\,\forall\, i\,. \tag{6}$$

We refer to this as the *primal* form. Any linear programming problem can be written in this way [31].

A linear programming problem is said to be *infeasible*, if there is no $\{x_i\}$ satisfying the constraints. Otherwise we call the domain of $\{x_i\}$ satisfying the constraints the *feasible region*. If the problem admits a finite solution, the problem is *bounded*. If the domain is bounded, then it forms a convex polytope, and the *maximum principle* [32] states that the optimum value is achieved at an extremal point of the polytope and is finite. Every linear program has an associated dual. For a problem written in the form (6), the dual problem is

$$\min_{M}\ \mathrm{tr}(M^TQ) \quad \text{subject to} \quad \mathrm{tr}(M^TA_i) \geq c_i \text{ for all } i,\ M \in \mathbb{R}^{s\times t},\ M \geq 0\,, \tag{7}$$

where the condition $M \geq 0$ should be understood elementwise.

Linear programming problems are *strongly dual*: an optimum solution $\{x_i^*\}$ for the problem (6), and M^* for the problem (7) satisfy $\sum_i c_i x_i^* = \mathrm{tr}((M^*)^TQ)$. If the primal is unbounded, then the dual is infeasible and vice versa. Solutions $\{x_i^*\}, M^*$ also satisfy the *complementary slackness* conditions [31], which in our notation are

$$\mathrm{tr}\left((M^*)^T\left(Q - \sum_i x_i^* A_i\right)\right) = 0, \tag{8}$$

$$\sum_i (\mathrm{tr}((M^*)^TA_i) - c_i)x_i^* = 0\,. \tag{9}$$

Two common approaches for solving linear programming problems are simplex algorithms [33] and interior point methods [34]. Simplex algorithms exploit the fact that the optimum of a linear program is always achieved at a vertex. Such algorithms move between vertices by following edges that improve the value of the objective function until no further improvement is possible. Interior point methods make successive steps towards the optimal solution while remaining in the interior of the feasible region. Since we are interested in finding extremal Bell inequalities, simplex algorithms will be most useful for us.

C. Representing Probability Distributions and Bell Inequalities.

In this work we focus on the bipartite case, although most of the techniques we consider generalise straightforwardly to multipartite scenarios. We consider two spacelike separate measurements, modelling the inputs using random variables X and Y, and the respective outputs A and B. We label the possible values of X by $\{1,\ldots,m_A\}$. Likewise, Y takes values from $\{1,\ldots,m_B\}$, A takes values from $\{1,\ldots,n_A\}$ and B takes values from $\{1,\ldots,n_B\}$. Sometimes we will consider cases in which different inputs have different numbers of outputs. Taking the notation from [35], if the measurements $X=1,2,\ldots$ have numbers of outcomes $n_A^1, n_A^2, \ldots$, and likewise measurements $Y=1,2,\ldots$ have numbers of outcomes $n_B^1, n_B^2, \ldots$ we will label the scenario $[(n_A^1\ n_A^2\ \ldots)\ (n_B^1\ n_B^2\ \ldots)]$. If the measurements $X=x$ and $Y=y$ are performed[5], the joint distribution over the outputs is $P_{AB|xy}$, i.e., for all x,y,a,b we have $0 \le P_{AB|xy}(a,b) \le 1$, and for all x,y we have $\sum_{a,b} P_{AB|xy}(a,b)=1$. A distribution is said to be no-signalling if it satisfies

$$\sum_b P_{AB|xy}(a,b) = \sum_b P_{AB|xy'}(a,b) \quad \forall\, a,x,y,y', \quad \text{and} \quad \sum_a P_{AB|xy}(a,b) = \sum_a P_{AB|x'y}(a,b) \quad \forall\, b,y,x,x'.$$

Since we consider measurements made at spacelike separation, all distributions will be no-signalling.

Using notation from Tsirelson [36], we express the conditional distribution $P_{AB|XY}$ using an $m_A n_A \times m_B n_B$ matrix:

$$\Pi = \left(\begin{array}{ccc|c|ccc}
P_{AB|11}(1,1) & \ldots & P_{AB|11}(1,n_B) & \ldots\ \ldots\ \ldots\ \ldots & P_{AB|1m_B}(1,1) & \ldots & P_{AB|1m_B}(1,n_B) \\
\vdots & \ddots & \vdots & \ldots\ \ldots\ \ldots\ \ldots & \vdots & \ddots & \vdots \\
P_{AB|11}(n_A,1) & \ldots & P_{AB|11}(n_A,n_B) & \ldots\ \ldots\ \ldots\ \ldots & P_{AB|1m_B}(n_A,1) & \ldots & P_{AB|1m_B}(n_A,n_B) \\
\hline
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\
\hline
P_{AB|m_A1}(11) & \ldots & P_{AB|m_A1}(1,n_B) & \ldots\ \ldots\ \ldots\ \ldots & P_{AB|m_Am_B}(1,1) & \ldots & P_{AB|m_Am_B}(1,n_B) \\
\vdots & \ddots & \vdots & \ldots\ \ldots\ \ldots\ \ldots & \vdots & \ddots & \vdots \\
P_{AB|m_A1}(n_A,1) & \ldots & P_{AB|m_A1}(n_A,n_B) & \ldots\ \ldots\ \ldots\ \ldots & P_{AB|m_Am_B}(n_A,1) & \ldots & P_{AB|m_Am_B}(n_A,n_B)
\end{array}\right), \tag{10}$$

where, for clarity, we have added dividing lines to the matrices to indicate the different measurements. This notation makes it convenient to check the no-signalling conditions. However, it has some redundancy: the no-signalling conditions and the normalization condition mean that the true dimension of the space is $(m_A(n_A-1)+1)(m_B(n_B-1)+1)-1$ [36].

A *local deterministic* distribution is one for which $P_{AB|XY} = P_{A|X}P_{B|Y}$ and for which $P_{A|x}(a) \in \{0,1\}$ for all a,x and $P_{B|y}(b) \in \{0,1\}$ for all b,y. There are $(n_A)^{m_A}(n_B)^{m_B}$ such distributions, and we use $P^{\mathrm{L},i}_{AB|XY}$ to denote the i^{th} local distribution for $i=1,\ldots,(n_A)^{m_A}(n_B)^{m_B}$. A *local* distribution is then one that can be written as a convex combination of local deterministic distributions, i.e., $P_{AB|XY} = \sum_i \lambda_i P^{\mathrm{L},i}_{AB|XY}$, where $\lambda_i \ge 0$ and $\sum_i \lambda_i = 1$. We use $\mathcal{L}_{(m_A,m_B,n_A,n_B)}$ for the set of local distributions in each scenario. For all (m_A,m_B,n_A,n_B), these form a polytope (the *local polytope*) with the local deterministic distributions as its vertices.

A *Bell inequality* is a linear inequality that is satisfied if and only if a distribution is local. The most important class of these are the *facet* Bell inequalities, which represent the facets of the local polytope. In principle, these can be found by facet enumeration starting from the local deterministic distributions. We can represent every Bell inequality using a $m_A n_A \times m_B n_B$ matrix, B, such that the Bell inequality can be expressed as $\mathrm{tr}(B^T\Pi) \ge c$, where c is some constant. If Π is local, then it necessarily satisfies every Bell inequality. Given a complete set of facet Bell inequalities, a distribution is local if and only if it satisfies them all. Thus, finding all the facet Bell inequalities is an important task.

Note that the facet Bell inequalities include $m_A m_B n_A n_B$ "trivial" inequalities that correspond to the positivity of each component of Π, and are necessary for Π to represent a probability distribution. For each non-trivial Bell inequality there is some no-signalling distribution that violates it.

Note that there are many ways to express the same Bell inequality. To explain these, we introduce a little terminology. We say that a matrix M is *no-signalling type* if for all Π representing a no-signalling distribution we have $\mathrm{tr}(M^T\Pi)=0$. We say that a matrix M is *identity type* if for all Π representing a valid distribution $\mathrm{tr}(M^T\Pi)=1$. (for example, for $(m_A,m_B,n_A,n_B)=(2,2,2,2)$, $\left(\begin{array}{cc|cc} 1 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$ is no-signalling type and $\left(\begin{array}{cc|cc} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$ is identity

[5] We use upper case to denote random variables and lower case for particular values of these.

type). If B represents the Bell inequality $\mathrm{tr}(B^T\Pi) \geq c$, then for reals s, t, r with $r > 0$, M no-signalling type and I identity type we have that $\tilde{B} = rB + sM + tI$ represents the Bell inequality $\mathrm{tr}(\tilde{B}^T\Pi) \geq rc + t$. Although $\tilde{B} \neq B$, both $\tilde{B}$ and B are representations of the same Bell inequality. For a Bell inequality of the form $\mathrm{tr}(B^T\Pi) \geq c$, any local distribution, $\tilde{\Pi}$, for which $\mathrm{tr}(B^T\tilde{\Pi}) = c$ is said to *saturate* the inequality.

Given a Bell inequality, we can construct others in the same scenario by relabelling inputs and outputs. In addition, if $n_A = n_B$ and $m_A = m_B$ then we can also swap parties to construct others. Two Bell inequalities related by such relabellings are said to be in the same *class*. If $n_A = n_B$ and $m_A = m_B$ then there are $2(n_A!)^{m_A}(n_B!)^{m_B}m_A!m_B!$ ways to relabel (and half as many otherwise), although some relabellings may be equivalent to others.[6] Because relabellings do not change the essential features of a Bell inequality, we focus on acquiring a representative of each Bell inequality class, rather than the full list of inequalities.

Another important property of Bell inequalities is that they may be "lifted" to apply to scenarios with more inputs and/or outputs [37]. We can add an input by setting the coefficients corresponding to the new input to 0. (This means that the new Bell inequality ignores the new input.) To increase the number of outputs, the lifting involves copying the coefficients for one of the existing outputs[7]. This corresponds to treating the new output in the same way as one of the existing outputs. This copying is needed to ensure that the new Bell inequality has the same bound. Note that these methods of lifting have the property that the lifting of a facet Bell inequality always gives a facet Bell inequality [38].

To illustrate the concept of lifting, we present three Bell inequalities: The first (B_{CHSH}) is the CHSH inequality for the $(2, 2, 2, 2)$ scenario. The second (B_I) lifts this to become a $(2, 3, 2, 2)$ inequality, whilst the third (B_O) is a lifting of B_C to a $(2, 2, 2, 3)$ inequality. All three are facet Bell inequalities with the form $\mathrm{tr}(B^T\Pi) \geq 1$.

$$B_{\mathrm{CHSH}} = \left(\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right) \qquad B_I = \left(\begin{array}{cc|cc|cc} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{array}\right) \qquad B_O = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ \hline 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{array}\right)$$

D. The sets of no-signalling and quantum distributions

For a given scenario, the set of all no-signalling distributions, $\mathcal{NS}_{(m_A,m_B,n_A,n_B)}$, forms a polytope whose facets correspond to the positivity of probabilities. The extremal no-signalling distributions can in principle be found by vertex enumeration on these facets. In the general case, we do not know how to express all of these, but local deterministic distributions are always vertices of $\mathcal{NS}_{(m_A,m_B,n_A,n_B)}$. Furthermore, if both parties have only two outputs per measurement, i.e., $n_A = n_B = 2$, it is known that (up to relabelling) all non-local extremal no-signalling distributions have the form [39]

$$\begin{array}{c} \hphantom{S \quad S \quad} \overbrace{\hphantom{S/A \; \ldots \; S/A}}^{m_B-2-g} \; \overbrace{\hphantom{L \; \ldots \; L}}^{g} \\ \left(\begin{array}{ccccccc} S & S & S & \ldots & S & L & \ldots & L \\ S & A & S/A & \ldots & S/A & L & \ldots & L \\ S & S/A & S/A & \ldots & S/A & L & \ldots & L \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ S & S/A & S/A & \ldots & S/A & L & \ldots & L \\ K & K & K & \ldots & K & M & \ldots & M \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ K & K & K & \ldots & K & M & \ldots & M \end{array}\right) \begin{array}{l} \\ \\ \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} m_A - 2 - h \\ \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} h \end{array} \end{array} \tag{11}$$

[6] For example, when $(m_A, m_B, n_A, n_B) = (2, 2, 2, 2)$, there are 32 relabellings, but applying each of these to a CHSH inequality generates only 8 unique inequalities.

[7] The choice of output to copy may vary with each input. In addition we can also lift by only adding an output for one of the inputs, but not the others, e.g., to go from [(3 4) (3 2)] to [(3 4) (4 2)].

where $g \in \{0,1,\ldots,m_B-2\}$, $h \in \{0,1,\ldots,m_A-2\}$ and with the following 2×2 blocks:

$$S = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \qquad A = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \qquad K = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{pmatrix} \qquad L = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \end{pmatrix} \qquad M = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

The set of quantum distributions is a subset of $\mathcal{NS}$. It is convex, but not a polytope. A distribution $P_{AB|XY}$ is quantum if there exist POVMs $\{E_a^x\}_a$ and $\{F_b^y\}_b$ and a bipartite quantum state ρ_{AB} such that $P_{AB|xy}(a,b) = \mathrm{tr}((E_a^x \otimes F_b^y)\rho_{AB})$ for all a, b, x and y. We use $\mathcal{Q}_{(m_A,m_B,n_A,n_B)}$ to denote the set of quantum distributions in each scenario. Given a distribution $P_{AB|XY}$ it is difficult to decide whether it is quantum. To cope with this a series of outer approximations to the quantum set were introduced [40]. For positive integer k, we use $\mathcal{Q}^k_{(m_A,m_B,n_A,n_B)}$ to denote the set of correlations at the k^{th} level. These levels form a hierarchy, in that

$$\mathcal{Q}_{(m_A,m_B,n_A,n_B)} \subseteq \mathcal{Q}^k_{(m_A,m_B,n_A,n_B)} \subseteq \mathcal{Q}^l_{(m_A,m_B,n_A,n_B)}$$

for any positive integers $k > l$. The advantage of using these sets is that testing for membership of $\mathcal{Q}^k_{(m_A,m_B,n_A,n_B)}$ is a semidefinite program, which is in practice tractable for small enough k, m_A, m_B, n_A, n_B.

Given a distribution Π, we use the following measure of its non-locality.

Definition 1 ([41])**.** The *local weight* of a distribution $\tilde{\Pi}$ is the solution to the problem:

$$\begin{aligned} \max_{\{x_i\}} \quad & \sum_i x_i \\ \text{subject to} \quad & \sum_i x_i P^{\mathrm{L},i} \leq \tilde{\Pi} \\ & x_i \geq 0 \text{ for all } i\,, \end{aligned} \tag{12}$$

where $\sum_i x_i P^{\mathrm{L},i} \leq \tilde{\Pi}$ is interpreted component-wise.[8]

If $\tilde{\Pi}$ is local then the local weight is 1, while if $\tilde{\Pi}$ is a non-local extremal no-signalling distribution, its local weight is 0. Note that computing the local weight is a linear program. Its dual can be written

$$\begin{aligned} \min_{M} \quad & \mathrm{tr}(M^T \tilde{\Pi}) \\ \text{subject to} \quad & \mathrm{tr}(M^T P^{\mathrm{L},i}) \geq 1 \text{ for all } i \\ & M \geq 0\,, \end{aligned} \tag{13}$$

where again $M \geq 0$ is treated component-wise, and i runs over all local deterministic distributions. Note that $\mathrm{tr}(M^T P^{\mathrm{L},i}) \geq 1$ for all i implies that $\mathrm{tr}(M^T \Pi) \geq 1$ is a (possibly trivial) Bell inequality. If $\tilde{\Pi}$ is non-local, the matrix M^* that achieves the minimum is a non-trivial inequality violated by $\tilde{\Pi}$.

Definition 2. Let $\{x_i^*\}$ be the argument that achieves the optimum in the definition of the local weight of a distribution. The *local part* of a distribution $\tilde{\Pi}$ is $\sum_i x_i^* P^{\mathrm{L},i}$ and the *non-local part* is $\tilde{\Pi} - \sum_i x_i^* P^{\mathrm{L},i}$.

Note that the local and non-local parts of $\tilde{\Pi}$ are not in general normalized, but can be easily renormalized, for instance by dividing by the sum of all elements and multiplying by $m_A m_B$.

III. GENERATING FACET BELL INEQUALITIES

We can use the insight of the previous section to find Bell inequalities by solving the dual problem (13) for non-local distributions. Note that the Bell inequalities that emerge as solutions have the form where all entries are positive and have local bound 1. It turns out that all Bell inequalities can be represented in such a form (see Lemma 1 in Appendix A). Furthermore, for every non-trivial facet Bell inequality of this form there exists a non-local extremal no-signalling distribution which gives value 0 for this Bell expression (see Lemma 3 in Appendix A). In addition, there

[8] This condition ensures that $\tilde{\Pi} - \sum_i x_i P^{\mathrm{L},i}$ is equal to $(1 - \sum_i x_i)\tilde{P}$ for $(1 - \sum_i x_i) \geq 0$ and with $\tilde{P}$ as a valid distribution.

is a non-local extremal no-signalling distribution achieving this and that takes the form (11) with $g = h = 0$ (see Theorem 4 in Appendix A).

This suggests that we could find all the Bell inequalities by running the dual program for all non-local extremal no-signalling distributions (in cases where these are known). However, there is a hidden subtlety: although 0 is the minimum possible value for any Bell expression of this form (hence no other Bell inequality can have a larger violation) there can be several Bell expressions all of which have value 0 at the same extremal no-signalling distribution. This means that the output Bell inequality may not be a facet inequality, and that some facet inequalities may be missed. To mitigate this problem, we can reduce the degeneracy by mixing extremal no-signalling distributions with local distributions prior to using them as the dual problem's objective function. This is the idea behind our algorithm. In principle one can choose enough local distributions to break this degeneracy completely (see Appendix C); in practice though we only mix with two local distributions to reduce the degeneracy whilst keeping a reasonable runtime.

The idea to use the dual of the local weight problem to find Bell inequalities was used before in [27], where a procedure similar to our Algorithm 2 was used to find 18 Bell inequalities in the $(3, 3, 3, 3)$ scenario.

A. A Linear Programming Algorithm for Bell Inequalities

Our algorithm needs a few sub-algorithms.

1. This is a way to decide whether a Bell inequality B is a facet. To do so, we find the set of local deterministic distributions for which $\mathrm{tr}(B^T P^{\mathrm{L},i}) = 1$ (i.e., those that achieve equality in the Bell inequality). Call the values of i giving equality $a(1), a(2), \ldots, a(t)$. These all lie on a face, which is a facet if its dimension is one less than that of the entire space. In other words we have to check how many dimensions are spanned by $\{P^{\mathrm{L},a(i)} - P^{\mathrm{L},a(1)}\}_{i=2}^{t}$. If this is $(m_A(n_A - 1) + 1)(m_B(n_B - 1) + 1) - 2$ then B represents a facet Bell inequality. (This dimensionality can be found by computing the rank of a matrix whose rows comprise the elements of $P^{\mathrm{L},a(i)} - P^{\mathrm{L},a(1)}$ for $i = 2, \ldots, t$.)

2. This checks whether two matrices B and $\tilde{B}$ are representations of the same inequality. To do so we compute two vectors, v and $\tilde{v}$ with components $v_i = \mathrm{tr}(B^T P^{\mathrm{L},i})$ and $\tilde{v}_i = \mathrm{tr}(\tilde{B}^T P^{\mathrm{L},i})$. By construction, the smallest element of each of these vectors is 1. Let the second smallest value of v be $s > 1$. We perform an affine transformation that maps 1 to 1 and s to 2, (i.e., the function $x \mapsto \frac{1}{s-1}x + \frac{s-2}{s-1}$) to each component of v forming v'. A similar procedure is performed on $\tilde{v}$ forming $\tilde{v}'$. The matrices B and $\tilde{B}$ represent the same inequality if and only if $v' = \tilde{v}'$.

3. Because we are interested in classes of Bell inequality, rather than the inequalities themselves, we also need to check whether two matrices are equivalent up to relabellings. This algorithm checks whether $v' = T_m(\tilde{v}')$ where m runs over all the relabellings, and T_m is the permutation on the entries of $\tilde{v}'$ corresponding to the m^{th} relabelling (a list of such permutations can be computed once before commencing the main algorithm to speed up this check, although for larger cases a lot of memory is required to store them all). Note that if v' and $\tilde{v}'$ do not have the same numbers of each type of entry (the same *tally*) then there cannot be such a permutation. We hence first check for this before running over all the permutations corresponding to relabellings.

Algorithm 1

This algorithm generates new facet inequalities for cases where $n_A = n_B = 2$. It takes input $\epsilon \in (0, 2/3)$, and a list W of known facet inequalities (which could be empty).

1. Set $j = 1$.

2. Set $\tilde{\Pi}$ to be the j^{th} extremal no-signalling distribution of the form (11) with $g = h = 0$.

3. Solve the dual problem (13) for $\tilde{\Pi}$ using a simplex algorithm, giving the matrix M that minimizes $\mathrm{tr}(M^T\tilde{\Pi})$.

4. Generate a list of values of i such that $\mathrm{tr}(M^T P^{\mathrm{L},i}) = 1$. Call these $a(1), a(2), \ldots, a(t)$.

5. Check whether M defines a facet (using subalgorithm 1) and whether it or a Bell inequality in the same class is in the list W (using subalgorithms 2 and 3). If not, add M to W.

6. Choose 2 distinct elements k, l from $\{1, 2, \ldots, t\}$ and form $\Pi' = (1 - \frac{3\epsilon}{2})\tilde{\Pi} + \epsilon P^{\mathrm{L},a(k)} + \frac{\epsilon}{2} P^{\mathrm{L},a(l)}$.

7. Solve the dual problem (13) for Π' using a simplex algorithm, giving the matrix M' that minimizes $\mathrm{tr}((M')^T\Pi')$.

8. Check whether M' defines a facet (using subalgorithm 1) and whether it or a Bell inequality in the same class is in the list W (using subalgorithms 2 and 3). If not, add M' to W.

9. Repeat steps 6–8 running over all distinct pairs k, l from $\{1, 2, \ldots, t\}$.

10. If $j < 2^{(m_A-1)(m_B-1)-1}$ set $j = j + 1$ and return to step 2, otherwise end the algorithm, outputting W.

Algorithm 2

In cases where the complete set of extremal no-signalling vertices is not known, we use another algorithm to find facet Bell inequalities. This works by picking random quantum-realisable distributions instead of extremal no-signalling distributions. This algorithm takes as input a number of iterations, $j_{\max}$, and a list W of known facet inequalities (which could be empty).

1. Set $j = 1$.

2. Randomly choose a pure quantum state of dimension $(\max(n_A, n_B))^2$ and m_A random projective measurements of dimension $\max(n_A, n_B)$ with n_A outcomes and m_B random projective measurements of dimension $\max(n_A, n_B)$ with n_B outcomes.[9] Form the (quantum) distribution $\tilde{\Pi}$ by computing the distribution that would be observed for this state and measurements.

3. Solve the dual problem (13) for $\tilde{\Pi}$ using a simplex algorithm, giving the matrix M that minimizes $\mathrm{tr}(M^T\tilde{\Pi})$.

4. Check whether M defines a facet (using subalgorithm 1) and whether it or a Bell inequality in the same class is in the list W (using subalgorithms 2 and 3). If not, add M to W.

5. If $j < j_{\max}$ set $j = j + 1$ and return to step 2, otherwise end the algorithm, outputting W.

We also consider Algorithm 2′, which is the same as Algorithm 2 except that the following additional step is added between Steps 2 and 3.

2b. If $\tilde{\Pi}$ is non-local, replace $\tilde{\Pi}$ with the renormalized non-local part of $\tilde{\Pi}$.

B. Comments on the algorithms

We use the simplex algorithm as implemented by Mathematica 11.1.1. Unfortunately, full details of this specific implementation are not publicly available (to our knowledge). In particular, for our problem, in spite of the steps taken to break some of the degeneracy, for a given Π' there remain many M' that achieve the optimum for the dual. Which one is given out by the simplex algorithm depends on the details of how it decides which edge to travel along when faced with several possibilities. Mathematica's implementation is deterministic, but for fixed k and l, small changes in ϵ can lead to a different solution based on Π'. It is hence useful to rerun the algorithm for several values of ϵ.

As mentioned above, one disadvantage of the above algorithm is that in many cases the output of the dual program does not correspond to a facet inequality. It is possible to alter the problem such that the solution space is the polar dual of the local polytope, where indeed every solution will be facet. However, by doing this, it is no longer the case that every facet inequality will be given as a solution, regardless of the input no-signalling distribution (i.e., the analogue of Lemma 3 does not hold). This is further discussed in Appendix D. An alternative way to find facet Bell inequalities from lower dimensional faces was used in [16].

Running through all permutations to check whether two Bell matrices are in the same class can take time, so Algorithm 1 can also be run with a modified subalgorithm 3 in which M is added to W if the vectors v' and $\tilde{v}'$ have different tallies. Running the algorithm in this way can generate many classes of Bell inequality, but without

[9] For more details on how we do this, see Appendix B.

also running through all permutations, some classes may be missed. For instance, in the case $(m_A, m_B, n_A, n_B) = (4, 4, 2, 2)$, for $i = 1, 2$, the Bell inequalities $\mathrm{tr}(B_i^T \Pi) \geq 1$ where

$$B_1 = \left(\begin{array}{cc|cc|cc|cc} 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ \hline 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \hline 0 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \end{array}\right) \quad \text{and} \quad B_2 = \left(\begin{array}{cc|cc|cc|cc} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \hline 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \end{array}\right)$$

are in different classes, but the corresponding vectors v' and $\tilde{v}'$ have the same tallies.

Another disadvantage of our algorithm is that we do not have a criteria for deciding when the list of classes found is complete. In cases where we are sure we have found all Bell inequalities, we know this only because the total number had already been found by other means.

Note that because Algorithm 2 is based on quantum distributions, it is unable to find Bell inequalities for which there is no quantum violation [42]. Furthermore, the chosen quantum distribution may be local, in which case the dual program will not output a Bell inequality. To circumvent these disadvantages, other ways to pick non-local distributions can be used. For instance, the renormalized non-local part of a distribution may violate Bell inequalities that the original distribution does not. This motivates our Algorithm $2'$, which may be able to find Bell inequalities without quantum violation (see also Remark 2). If there were an extension of the work of [39] to cases with $n_A > 2$ or $n_B > 2$, we could use Algorithm 1 in these cases. We expect that this would be a quicker way to generate new Bell inequalities.

IV. RESULTS USING THE ALGORITHM

In this section we summarise the results we have obtained using the above algorithm. Due to the large number of inequality classes found we present them in separate files available at [43], along with a file explaining how to import and use them in both Mathematica and Matlab. We give both the version the algorithm found (the "raw" version) and a second version after an affine transformation analogous to that mentioned in subalgorithm 2 has been applied (the "affine" version). These are presented after relabellings that make obvious input/output liftings.

A. $(3, 5, 2, 2)$ scenario

The number of facet classes for this scenario (7) was given in [18]. Six of these are also facet classes of the $(3, 4, 2, 2)$ scenario, and are given in [23]. The remaining class was first found in [26]. We used this case to test Algorithm 1, rederiving the result. We found a representative of the new class to be $\mathrm{tr}(I_{3522}^T \Pi) \geq 1$

$$I_{3522} = \left(\begin{array}{cc|cc|cc|cc|cc} 0 & \frac{2}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 \\ \hline 0 & 0 & 0 & 0 & \frac{2}{3} & 0 & \frac{2}{3} & 0 & 0 & 0 \\ \frac{2}{3} & 0 & 0 & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} \end{array}\right). \tag{14}$$

B. $(4, 4, 2, 2)$ scenario

For this scenario, we have enumerated all 175 inequivalent classes of facet Bell inequality (including the trivial positivity inequality). Whilst the number of classes was already known [18], a complete list was not provided. A partial list of 129 non-trivial inequalities was given in [17]. Our generation of these inequalities was performed on a standard desktop computer and took a few days.

In order to give an idea of the symmetries of this polytope, Table II gives the number of members of each class.

Size of Class	Number of Classes	Notable cases
64	1	Positivity
288	1	CHSH [19]
9216	2	I_{3322} [23]
18432	4	I_{4422} [23]
24576	1	
36864	4	
49152	2	
73728	8	
98304	2	
147456	61	
294912	89	
36391264	175	

TABLE II: The size of each facet class for the $(4, 4, 2, 2)$ local polytope. 294912 is the size of the relabelling symmetry group.

Bell Inequality Class	Number of Faces
Positivity	36
CHSH	216
I^1_{2332}	288
I^2_{2332}	288
I^3_{2332}	432
Total	1260

TABLE III: The facets of the $(2, 3, 3, 2)$ local polytope, sorted into their inequality classes.

C. $(2, 3, 3, 2)$ scenario

The extremal no-signalling distributions are not known for this scenario, but using Algorithm $2'$ we found five classes of Bell inequality: the positivity condition, a lifting of CHSH and three new inequality classes $\mathrm{tr}((I^1_{2332})^T\Pi) \geq 1$ and $\mathrm{tr}((I^2_{2332})^T\Pi) \geq 1$ and $\mathrm{tr}((I^3_{2332})^T\Pi) \geq 1$ with representative matrices

$$I^1_{2332} = \left(\begin{array}{cc|cc|cc} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 1 & 0 \\ 0 & \frac{1}{2} & 1 & 0 & 0 & \frac{1}{2} \\ 1 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \hline \frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right) \qquad I^2_{2332} = \left(\begin{array}{cc|cc|cc} 0 & \frac{1}{2} & 0 & 1 & \frac{1}{2} & 0 \\ 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 1 \\ \hline 0 & 1 & 1 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right) \qquad I^3_{2332} = \left(\begin{array}{cc|cc|cc} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right).$$

For this scenario, we were able to perform facet enumeration of the local polytope, and verify that these five form the complete list of classes. Table III gives the number of inequalities in each class.

D. $(3, 3, 2, 3)$ scenario

In this scenario we again do not know the full list of extremal no-signalling distributions, and we employed Algorithm 2 for 250000 runs, finding 19 inequality classes. We then employed Algorithm $2'$ for 3451207 runs to find 6 more classes (a total of 25). As no new class was found for over 1164000 runs, we speculate this is the full list of classes. We also used Algorithm 4 (detailed in Appendix E) to find the smallest scenario for which each of the inequalities first appears. In Table IV we present this analysis. Surprisingly, only one of the classes found is not a lifting from a smaller scenario.

Pre-Lifting Scenario	Number of Classes Found	Number of Distinct Classes in Pre-Lifting Scenario	Notable Inequality
[(2) (2)]	1	1	Positivity Condition
[(2 2) (2 2)]	1	1	CHSH
[(2 2) (2 3)]	–	–	
[(2 2) (3 3)]	–	–	
[(2 2) (2 2 2)]	–	–	
[(2 2) (2 2 3)]	–	–	
[(2 2) (2 3 3)]	–	–	
[(2 2) (3 3 3)]	–	–	
[(2 2 2) (2 3)]	2	1	
[(2 2 2) (3 3)]	1	1	
[(2 2 2) (2 2 2)]	3	1	I_{3322}
[(2 2 2) (2 2 3)]	8	3	
[(2 2 2) (2 3 3)]	8	4	
[(2 2 2) (3 3 3)]	1	1	

TABLE IV: The inequalities classes found for $(3, 3, 2, 3)$, sorted to show their degree of lifting. Relabelling of inputs has been used to combine scenarios of equivalent dimension e.g., [(2 2 2) (2 2 3)] and [(2 2 2) (2 3 2)]. We use "–" to denote cases for which it is known that no new classes exist. Note that classes that are unique before lifting may define several inequivalent classes in a higher dimension. This is why although I_{3322} is the only new non-trivial inequality class in the [(2 2 2) (2 2 2)] scenario, we found three inequivalent classes corresponding to I_{3322} liftings.

E. $(4, 5, 2, 2)$ scenario

There is no known total of inequivalent classes for this scenario. Using Algorithm 1 we have established the existence of at least 18277 classes and have a representative of each. We expect the total number of classes to be significantly larger[10]. As for this case the size of the relabelling symmetry group is quite large, which slows down the check to see if a previous class member has been found, we use the modified version of subalgorithm 3 discussed in Section III B in which only the tally of v' is checked—this reduces the runtime considerably, but means that we may discard new classes that have the same tally to a class that has already been found.

F. $(3, 3, 3, 3)$ scenario

Again for this scenario we do not know all the extremal no-signalling distributions, so we use Algorithm 2 to find facet inequalities and the modified version of subalgorithm 3 discussed in Section III B in which only the tally of v' is checked. In our initial run, we found 10143 classes with $j_{\max} = 2.5$million. We then ran Algorithm $2'$ with $j_{\max} = 0.6$million (again with the modified version of subalgorithm 3), and found a further 11018 classes. These figures suggest $2'$ is significantly better at finding new classes. We did a small further check to confirm that there are inequivalent inequalities with the same tallies generating 9 further inequalities. We expect the total number of classes to be significantly larger than the figure presented here (see Footnote 10).

Our list of inequalities can be augmented with the results of [27], who present 18 inequalities for this scenario, of which 3 are contained in our list. Again we know that, of these classes, some of them will correspond to liftings of lower dimensional inequalities. Table V partitions our 21170 classes into the scenario for which they first appear, taking into account cases in which different inputs have different numbers of outputs. This uses Algorithm 4, detailed in Appendix E.

---

[10] We stopped the algorithm after a few weeks but new inequality classes were being found regularly.

Pre-Lifting Scenario	Number of Classes Found	Number of Distinct Classes in Pre-Lifting Scenario	Notable Inequality
[(2) (2)]	1	1	Positivity Condition
[(2 2) (2 2)]	2	1	CHSH
[(2 2) (2 3)]	–	–	
[(2 2) (3 3)]	–	–	
[(2 3) (2 3)]	–	–	
[(2 3) (3 3)]	–	–	
[(3 3) (3 3)]	1	1	I_{2233}
[(2 2) (2 2 2)]	–	–	
[(2 2) (2 2 3)]	–	–	
[(2 2) (2 3 3)]	–	–	
[(2 2) (3 3 3)]	–	–	
[(2 3) (2 2 2)]	7	1	
[(2 3) (2 2 3)]	0	0	
[(2 3) (2 3 3)]	0	0	
[(2 3) (3 3 3)]	0	0	
[(3 3) (2 2 2)]	2	1	
[(3 3) (2 2 3)]	5	3	
[(3 3) (2 3 3)]	9	7	
[(3 3) (3 3 3)]	4	4	
[(2 2 2) (2 2 2)]	9	1	I_{3322}
[(2 2 2) (2 2 3)]	27	3	
[(2 2 2) (2 3 3)]	15	4	
[(2 2 2) (3 3 3)]	3	1	
[(2 2 3) (2 2 3)]	154	40	
[(2 2 3) (2 3 3)]	762	337	
[(2 2 3) (3 3 3)]	398	276	
[(2 3 3) (2 3 3)]	2532	1764	
[(2 3 3) (3 3 3)]	6637	6060	
[(3 3 3) (3 3 3)]	10602	10602	

TABLE V: The inequalities classes found for $(3,3,3,3)$, sorted to show their degree of lifting. Due to the additional symmetry, Relabelling of parties/inputs has been used to group classes, e.g. [(2 3) (2 2 2)] and [(2 2 2) (2 3)]. Again we find that unique classes in lower dimensions can be lifted to define several facet classes.

V. APPLICATION TO THE DETECTION LOOPHOLE

In the remainder of this article we investigate whether these inequalities allow us to lower the efficiency required for closing the detection loophole.

To perform a bipartite Bell experiment, entangled photons are sent to two detectors where they are measured. After repeating many times using different randomly chosen measurements we can build up an estimate of the distribution Π, from which we can see whether a particular Bell inequality is violated or not. One of the issues with such an experiment is that real detectors sometimes fail to detect. The question is then how to certify that the setup is non-local in the presence of such no-click events. In particular, we would like to know the minimal detection efficiency at which we can still certify the presence of non-locality. Given that certifying the presence of non-locality is necessary for device-independent tasks, it is important to be able to treat cases with imperfect detectors.

To model this, we assume that each detector has an efficiency, representing the probability that it will click when it should. For simplicity, we consider the case where the efficiencies of each detector are the same and call this parameter η. No-click can be treated as an additional outcome for each measurement that occurs with probability $1-\eta$ (independently for each measurement). Given a probability distribution $P_{AB|XY}$, we use $P^{\eta}_{AB|XY}$ to denote the inefficient detector version with efficiency η, which is formed by adding the possible outcome "N" to each measurement and taking

$$\begin{aligned}P^{\eta}_{AB|xy}(a,b) &= \eta^2 P_{AB|xy}(a,b),\\ P^{\eta}_{AB|xy}(N,b) &= \eta(1-\eta)P_{B|y}(b),\\ P^{\eta}_{AB|xy}(a,N) &= \eta(1-\eta)P_{A|x}(a),\\ P^{\eta}_{AB|xy}(N,N) &= (1-\eta)^2,\end{aligned}$$

for all $a \in \{1,2,\ldots,n_A\}$, $b \in \{1,2,\ldots,n_B\}$, $x \in \{1,2,\ldots,m_A\}$ and $y \in \{1,2,\ldots,m_B\}$. We use Π^{η} to denote the matrix representation of $P^{\eta}_{AB|XY}$. (For a discussion of other ways to deal with no-click events, see [44].)

It is worth noting that for a given $\tilde{\Pi} \in \mathcal{NS}_{(m_A,m_B,n_A,n_B)}$ we have that $\tilde{\Pi}^{\eta} \notin \mathcal{L}_{(m_A,m_B,n_A+1,n_B+1)}$ only if $\tilde{\Pi} \notin \mathcal{L}_{(m_A,m_B,n_A,n_B)}$, and that for $\eta_1 \geq \eta_2$, $\tilde{\Pi}^{\eta_2} \notin \mathcal{L}_{(m_A,m_B,n_A+1,n_B+1)}$ implies $\tilde{\Pi}^{\eta_1} \notin \mathcal{L}_{(m_A,m_B,n_A+1,n_B+1)}$.

We now define the detection threshold for a given Bell inequality, and for a given scenario.

Definition 3. Given a $m_A(n_A+1) \times m_B(n_B+1)$ matrix B such that $\mathrm{tr}(B^T\Pi) \geq c$ for all $\Pi \in \mathcal{L}_{(m_A,m_B,n_A+1,n_B+1)}$, the *detection threshold for* B, is defined by

$$\eta_B := \inf\{\eta \in [0,1] : \exists\, \tilde{\Pi} \in \mathcal{Q}_{(m_A,m_B,n_A,n_B)},\ \mathrm{tr}(B^T\tilde{\Pi}^{\eta}) < c\}\,. \tag{15}$$

This is the smallest detection efficiency such that B can certify non-locality using quantum states and measurements for all higher efficiencies. Note that B is a Bell inequality for the (m_A, m_B, n_A+1, n_B+1) scenario. Some Bell inequalities of this type can be formed from those for the (m_A, m_B, n_A, n_B) scenario by lifting, as discussed in Section II C.

Definition 4. The *detection threshold for the* (m_A, m_B, n_A, n_B) *scenario* is defined by

$$\eta_{(m_A,m_B,n_A,n_B)} := \inf\{\eta \in [0,1] : \exists\, \tilde{\Pi} \in \mathcal{Q}_{(m_A,m_B,n_A,n_B)},\ \tilde{\Pi}^{\eta} \notin \mathcal{L}_{(m_A,m_B,n_A+1,n_B+1)}\}\,. \tag{16}$$

We can also define a detection threshold for a set $\mathcal{S}_{(m_A,m_B,n_A,n_B)} \subseteq \mathcal{NS}_{(m_A,m_B,n_A,n_B)}$.

Definition 5. The *detection threshold for* $\mathcal{S}_{(m_A,m_B,n_A,n_B)}$ is defined by

$$\eta^{\mathcal{S}}_{(m_A,m_B,n_A,n_B)} := \inf\{\eta \in [0,1] : \exists\, \tilde{\Pi} \in \mathcal{S}_{(m_A,m_B,n_A,n_B)},\ \tilde{\Pi}^{\eta} \notin \mathcal{L}_{(m_A,m_B,n_A+1,n_B+1)}\}\,. \tag{17}$$

We will use the fact that if $\mathcal{Q}$ is a subset of $\mathcal{S}$ then the detection threshold for $\mathcal{S}$ is lower than the quantum one, i.e., $\mathcal{Q}_{(m_A,m_B,n_A,n_B)} \subseteq \mathcal{S}_{(m_A,m_B,n_A,n_B)}$ implies $\eta^{\mathcal{S}}_{(m_A,m_B,n_A,n_B)} \leq \eta_{(m_A,m_B,n_A,n_B)}$.

It is known that there exist states and measurements for which this threshold tends to 0 as $d \to \infty$ [45]. However, to achieve this, the number of measurements required scales as $O(2^d)$. For practical reasons we are interested in cases with small numbers of inputs and outputs.

In the $(2,2,2,2)$ scenario, taking $\tilde{\Pi}$ to be quantum correlations that achieve Tsirelson's bound [46] for the CHSH inequality, ($\mathrm{tr}(B^T_{\mathrm{CHSH}}\tilde{\Pi}) = 2-\sqrt{2}$) it has been shown that $\tilde{\Pi}^{\eta}$ is non-local if and only if $\eta > 2(\sqrt{2}-1) \approx 82.8\%$ [47]. A lower value was found by Eberhard [48], who showed that one can use a two-qubit state of the form $\cos\theta|00\rangle + \sin\theta|11\rangle$ and appropriate 2-outcome measurements to give rise to a distribution $\hat{\Pi}_{\theta}$ such that for any $\eta > 2/3$ there exists θ such that $\hat{\Pi}^{\eta}_{\theta}$ is non-local, while for $\eta = 2/3$ the distribution $\hat{\Pi}^{\eta}_{\theta}$ is local for all θ. Somewhat counterintuitively, the state demonstrating non-locality has $\theta \to 0$ as $\eta \to 2/3$. In [49], a $(4,4,2,2)$ inequality (which we refer to as I_{4422} [23]) was considered and a state and measurements on a four dimensional Hilbert space were given demonstrating that $\eta_{I_{4422}} \leq (\sqrt{5}-1)/2 \approx 61.8\%$. The state used has the form $\sqrt{(1-\epsilon^2)/3}\,(|00\rangle + |11\rangle + |22\rangle) + \epsilon|33\rangle$, with the value $(\sqrt{5}-1)/2$ achieved in the limit $\epsilon \to 0$.

In [50], the problem was abstracted away from specific Bell inequalities, instead giving an explicit local-hidden variable construction which can replicate any inefficient no-signalling distribution, provided the detection efficiency is below $(m_A + m_B - 2)/(m_A m_B - 1)$. This hence corresponds to a lower bound on $\eta_{(m_A,m_B,n_A,n_B)}$. In the next subsection we improve on these lower bounds in cases where $n_A = n_B = 2$.

A. A Fundamental Lower Bound on the Detection Threshold

In this section we show how to bound $\eta_{(m_A,m_B,n_A,n_B)}$ using knowledge of the set of no-signalling distributions. As discussed above, since $\mathcal{Q}_{(m_A,m_B,n_A,n_B)} \subseteq \mathcal{NS}_{(m_A,m_B,n_A,n_B)}$, a lower bound for the detection threshold for all no-signalling distributions will apply to the quantum case too.

In cases where we have a complete set of extremal no-signalling distributions, we can obtain an arbitrarily good estimate $\eta^{\mathcal{NS}}_{(m_A,m_B,n_A,n_B)}$ using the following algorithm.

m_A / m_B	2	3	4	5	6	$2/(m_B+1)$
2	2/3	2/3	2/3	2/3	2/3	2/3
3		4/7	5/9	5/9	5/9	1/2
4			1/2	1/2	1/2	2/5
5				4/9	*	1/3

TABLE VI: The maximum detection efficiency such that Π^{η} can be generated classically for any $\Pi \in \mathcal{NS}_{(m_A,m_B,2,2)}$. The * in the $(5,6,2,2)$ case indicates that this was not evaluated due to the high number of non-local extremal no-signalling distributions. The final column is the lower bound of [50]. The main part of the table is populated with fractions, even though the algorithm outputs a decimal. Strictly we should say that the value is consistent with the stated fraction to 7 decimal places.

Algorithm 3

The algorithm takes input $\delta \in (0,1)$, the tolerance we look for in the solutions.

1. Set $j = 1$, $\eta_c = 1$ and $j_{\max}$ to be the total number of extremal no-signalling distributions.
2. Set $\tilde{\Pi}$ to be the j^{th} non-local extremal no-signalling distribution.
3. Set $\eta_{\min} = 0$ and $\eta_{\max} = 1$.
4. Set $\eta' = (\eta_{\min} + \eta_{\max})/2$ and generate $\tilde{\Pi}_{\eta'}$.
5. Find the local weight of $\tilde{\Pi}_{\eta'}$ by solving the linear program (12), setting this to w.
6. If $w = 1$, set $\eta_{\min} = \eta'$, otherwise, set $\eta_{\max} = \eta'$.
7. If $\eta_{\max} - \eta_{\min} > \delta$, go to step 4.
8. If $(\eta_{\min} + \eta_{\max})/2 < \eta_c$, set $\eta_c = (\eta_{\min} + \eta_{\max})/2$.
9. If $j < j_{\max}$, set $j = j + 1$ and return to step 2 otherwise the algorithm ends, outputting η_c.

This algorithm runs through all the non-local extremal no-signalling distributions, and computes at what η they become local. Then by taking the minimum over all such distributions we obtain the no-signalling detection threshold for this scenario.

We have applied this algorithm to the $(m_A, m_B, 2, 2)$ scenario for various m_A and m_B. The results are shown in Table VI, where we have also compared with the lower bound from [50] in order to highlight the improvement we obtain.

B. Bounding Detection Thresholds using the Semidefinite Hierarchy

We can think of these no-signalling values as lower bounds on the quantum detection thresholds. When interpreted this way, we do not expect these lower bounds to be tight, although, somewhat surprisingly, they are in the case where $m_B = 2$ [48]. In order to give better bounds, we can use other supersets of the quantum set, for instance, those based on the semidefinite hierarchy [40]. In other words, we can try to find $\eta^{\mathcal{Q}^k}_{(m_A,m_B,n_A,n_B)}$ for some level $k \in \mathbb{N}$ of the hierarchy.

Since $\mathcal{Q}^k_{(m_A,m_B,n_A,n_B)}$ is not a polytope, we cannot directly use the method of Section V A. Instead, for each value η we perform a semidefinite optimisation over $\mathcal{Q}^k_{(m_A,m_B,n_A,n_B)}$, minimising $\mathrm{tr}(B^T\hat{\Pi}^{\eta})$ over $\hat{\Pi} \in \mathcal{Q}^k_{(m_A,m_B,n_A,n_B)}$ for a specific Bell inequality $\mathrm{tr}(B^T\Pi) \geq 1$ for the $(m_A, m_B, n_A + 1, n_B + 1)$ scenario. If the minimum is at least 1, we can conclude that there is no $\tilde{\Pi} \in \mathcal{Q}_{(m_A,m_B,n_A,n_B)}$ for which $\mathrm{tr}(B^T\tilde{\Pi}^{\eta}) < 1$. Performing the computation for higher levels k of the hierarchy gives successively tighter bounds.

C. Results

All of the results in this section were obtained using the convex optimisation interface CVX [51] for Matlab, with MOSEK [52] as the solver. Unless stated otherwise, tests were run at the default CVX precision. Note also that in the code, a value of $\mathrm{tr}(B^T \Pi^\eta) \geq 1 - \epsilon$ was considered local for $\epsilon \approx 1.49 \times 10^{-8}$. This mitigates against the possibility of concluding a distribution is non-locality because of the solver precision, when in fact it is not. However, it can mean that we sometimes incorrectly conclude a distribution is local. Thus, the values we obtain will be upper bounds of the threshold over the considered set.

In the supplementary files these results are given in the following format: first the inequality considered is given, followed by the lifting choices of Alice, then the lifting choices of Bob, followed by the threshold found.

1. $(4, 4, 2, 2)$ *Scenario*

For this scenario an explicit quantum construction is given in [49] achieving $(\sqrt{5}-1)/2 \approx 0.6180$ (61.80%) requiring shared entangled states with local dimension 4. We were able to test every inequality in this scenario at level 2, with no inequality beating I_{4422}, which was the inequality used in [49]. We were only able to obtain the value 61.83% for the I_{4422} construction, implying we cannot rely on the results beyond 3 s.f., and an alternative lifting of the same inequality was able to achieve 61.8% also. Repeating the analysis for these two with the CVX precision variable set to "high", gave a value of 61.82% for both liftings, which also means we cannot improve the number of significant figures this way.

2. $(3, 5, 2, 2)$ *Scenario*

For the single new inequality $(3, 5, 2, 2)$, we tested it with the CVX precision set to high. Note that we are considering $(3, 5, 3, 3)$ probability distributions, so lifting the inequality by adding an extra output for each input can be done in $2^{3+5} = 256$ ways. For the first level, $\mathcal{Q}^1$, we obtained an optimal threshold of 64.0%, and for $\mathcal{Q}^2, \mathcal{Q}^3$ we obtained a threshold of 66.7%. This, together with the fact that restricting to $m_A \in \{1, 3\}$ and $m_B \in \{3, 5\}$ (with reference to the form of I_{3522} in (14)), reduces to a CHSH inequality for which the threshold is known to be 2/3, suggests that $\eta_{I_{3522}} = 2/3$.

The supplementary file for this case omits the inequality as only one was considered.

3. $(3, 3, 3, 3)$ *Scenario*

For each new inequality there are $3^{3+3} = 729$ possible liftings. Since it would be time consuming to test the detection threshold for all liftings of every inequality, we selected an "intuitively promising" subset of liftings to test for each inequality. For each lifting we summed all coefficients corresponding to detection failure outcomes, and then tested the ten with the lowest sum for each inequality. This is to minimise the impact of the failure sub-distributions, which are entirely local. If more than ten had an equivalent sum, all were tested.

To cut down on computational time further, each chosen lifting was tested with $\eta = 0.65$ and discarded if no non-local distributions $\tilde{\Pi}^{0.65}$ could be found for $\tilde{\Pi} \in \mathcal{Q}^2$. 338 inequality/lifting combinations achieved a threshold below this value; 65 of them obtained value 61.8% — due to the precision of 3 s.f. we are unable to compare them definitively with I_{4422}. Testing these 65 at level 3 of the hierarchy we find 21 of them maintain the value 61.8%. These inequalities can be found in the supplementary files provided at [43]. All of these inequalities match the threshold $(\sqrt{5}-1)/2$ up to 3 s.f., but we have not given an explicit quantum construction achieving this value. If it turns out that the thresholds for I_{4422} and any of these inequalities are the same, this cannot be explained by a set of correlations common to both, unlike in the case of I_{3522} and CHSH, whose thresholds match due to I_{3522} having a CHSH submatrix (cf. Section V C 2).

Added note

These results formed part of TC's Ph.D. thesis [53] submitted to the University of York in September 2018. Since then the work [54] appeared which independently found all the Bell inequalities in the $(4, 4, 2, 2)$ scenario. Ref. [55] took these inequalities and performed some analysis of them.

Acknowledgements

We are grateful to Carl Miller, Yaoyun Shi, Kim Winick and Tamás Vértesi for useful discussions and to Yeong-Cherng Liang for pointers to the literature. TC and RC are supported by EPSRC's Quantum Communications Hub (grant number EP/M013472/1). TC's Ph.D. was supported by a White Rose Studentship. RC is also supported by an EPSRC First Grant (grant number EP/P016588/1).

---

[1] Bell, J. On the Einstein Podolsky Rosen paradox. *Physics* **1**, 195–200 (1964).
[2] Aspect, A., Grangier, P. & Roger, G. Experimental tests of realistic local theories via Bell's theorem. *Physical Review Letters* **47**, 460–463 (1981).
[3] Tittel, W., Brendel, J., Gisin, B., Herzog, T., Zbinden, H. & Gisin, N. Experimental demonstration of quantum correlations over more than 10 km. *Physical Review A* **57**, 3229–3232 (1998).
[4] Giustina, M., Versteegh, M. A. M., Wengerowsky, S., Handsteiner, J., Hochrainer, A., Phelan, K., Steinlechner, F., Kofler, J., Larsson, J. A., *et al.* Significant-loophole-free test of Bell's theorem with entangled photons. *Physical Review Letters* **115**, 250401 (2015).
[5] Hensen, B., Bernien, H., Dréau, A. E., Reiserer, A., Kalb, N., Blok, M. S., Ruitenberg, J., Vermeulen, R. F. L., Schouten, R. N. *et al.* Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015).
[6] Shalm, L. K., Meyer-Scott, E., Christensen, B. G., Bierhorst, P., Wayne, M. A., Stevens, M. J., Gerrits, T., Glancy, S., Hamel, D. R. *et al.* Strong loophole-free test of local realism. *Physical Review Letters* **115**, 250402 (2015).
[7] Ekert, A. K. Quantum cryptography based on Bell's theorem. *Physical Review Letters* **67**, 661–663 (1991).
[8] Mayers, D. & Yao, A. Quantum cryptography with imperfect apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS-98)*, 503–509 (IEEE Computer Society, Los Alamitos, CA, USA, 1998).
[9] Barrett, J., Hardy, L. & Kent, A. No signalling and quantum key distribution. *Physical Review Letters* **95**, 010503 (2005).
[10] Acín, A., Gisin, N. & Masanes, L. From Bell's theorem to secure quantum key distribution. *Physical Review Letters* **97**, 120405 (2006).
[11] Vazirani, U. & Vidick, T. Fully device-independent quantum key distribution. *Physical Review Letters* **113**, 140501 (2014).
[12] Colbeck, R. *Quantum and Relativistic Protocols For Secure Multi-Party Computation.* Ph.D. thesis, University of Cambridge (2007). Also available as https://arxiv.org/abs/0911.3814.
[13] Pironio, S. *et al.* Random numbers certified by Bell's theorem. *Nature* **464**, 1021–1024 (2010).
[14] Colbeck, R. & Kent, A. Private randomness expansion with untrusted devices. *Journal of Physics A* **44**, 095305 (2011).
[15] Miller, C. A. & Shi, Y. Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, 417–426 (ACM, New York, NY, USA, 2014).
[16] Bancal, J.-D., Gisin, N. & Pironio, S. Looking for symmetric Bell inequalities. *Journal of Physics A* **43**, 385303 (2010).
[17] Pál, K. & Vértesi, T. Quantum bounds on Bell inequalities. *Physics Review A* **79**, 022120 (2009).
[18] Deza, M. & Sikirić, M. Enumeration of the facets of cut polytopes over some highly symmetric graphs. *International Transactions in Operational Research* **23**, 853–860 (2015).
[19] Clauser, J., Horne, M., Shimony, A. & Holt, R. Proposed experiment to test local hidden-variable theories. *Physical Review Letters* **23**, 880–884 (1969).
[20] Froissart, M. Constructive generalization of Bell's inequalities. *Il Nuevo Cimento B* **64**, 241–251 (1981).
[21] Pironio, S. *Aspects of Quantum Non-locality.* Ph.D. thesis, Université Libre de Bruxelles (2004).
[22] Śliwa, C. Symmetries of the Bell correlation inequalities. *Physical Letters A* **317**, 165-168 (2003).
[23] Collins, D. & Gisin, N. A relevant two qubit Bell inequality inequivalent to the CHSH inequality. *Journal of Physics A* **37**, 1775–1787 (2004).
[24] Collins, D., Gisin, N., Linden, N., Massar, S. & Popescu, S. Bell inequalities for arbitrarily high-dimensional systems. *Physics Review Letters* **88**, 040404 (2002).
[25] Kaszlikowski, D., Kwek, L. C., Chen, J.-L., Żukowski, M. & Oh, C. H. Clauser-Horne inequality for three-state systems. *Physical Review A* **65**, 032118 (2002).
[26] Quintino, M. T., Vértesi, T. & Brunner, N. Joint measurability, Einstein-Podolsky-Rosen steering, and Bell nonlocality. *Physical Review Letters* **113**, 160402 (2014).
[27] Schwarz S. & Bessire, B. & Stefanov A. & Liang, Y.-C. Bipartite Bell inequalities with three ternary-outcome measurementsfrom theory to experiments. *New Journal of Physics* **18**, 035001 (2016).
[28] Grünbaum, B. *Convex Polytopes* (Springer Science and Business Media New York Academic Press, 1967).
[29] Avis, D. & Fukuda, K. A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Discrete and Computational Geometry* **8**, 295–313 (1992).
[30] McMullen, P. & Shephard, G. *Convex Polytopes and the Upper Bound Conjecture* (Cambridge University Press, 1971).
[31] Walsh, G. *An Introduction to Linear Programming* (Holt, Rinehart and Winston Ltd., 1971).
[32] Rockafellar, R. *Convex Analysis* (Princeton University Press, 1970).

[33] Dantzig, G. Maximization of a linear function of variables subject to linear inequalities. *Activity Analysis of Production and Allocation* 339–347 (1947).
[34] Nocedal, J. & Wright, S. *Numerical Optimization* Springer Series in Operations Research and Financial Engineering (2006).
[35] Rosset, D. & Bancal, J.-D. & Gisin, N. Classifying 50 years of Bell inequalities. *Journal of Physics A* **47**, 424022 (2014).
[36] Tsirelson, B. Some results and problems on quantum Bell-type inequalities. *Hadronic Journal Supplement* **8**, 329–345 (1993).
[37] Peres, A. All the Bell Inequalities. *Foundations of Physics* **29**, 589-614 (1999).
[38] Pironio, S. Lifting Bell inequalities. *Journal of Mathematical Physics* **46**, 062112 (2005).
[39] Jones, N. & Masanes, L. Interconversion of nonlocal correlations. *Physical Review A* **72**, 052312 (2005).
[40] Navascues, M., Pironio, S. & Acín, A. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlation. *New Journal of Physics* **10**, 073013 (2008).
[41] Żukowski, M., Kaszlikowski, D., Baturo, A. & Åke Larsson, J. Strengthening the Bell theorem: Conditions to falsify local realism in an experiment. https://arxiv.org/abs/quant-ph/9910058 (1999).
[42] Almeida, M. *et al.* Guess your neighbor's input: A multipartite nonlocal game with no quantum advantage. *Physical Review Letters* **104**, 230404 (2010).
[43] Cope, T. & Colbeck, R. Supplementary files available at http://www-users.york.ac.uk/~rc973/Bell.
[44] Branciard, C. Detection loophole in Bell experiments: How postselection modifies the requirements to observe nonlocality. *Physical Review A* **83**, 032123 (2011).
[45] Massar, S. Nonlocality, closing the detection loophole, and communication complexity. *Physical Review A* **65**, 032121 (2002).
[46] Cirel'son, B. Quantum generalizations of Bell's inequality. *Letters of Mathematical Physics* **4**, 93 (1980).
[47] Mermin, N. *Techniques and Ideas in Quantum Measurement Theory*, 422–428 (New York Academy of Sciences, 1986).
[48] Eberhard, P. Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment. *Physical Review A* **47**, R747–R750 (1993).
[49] Vértesi, T., Pironio, S., & Brunner, N. Closing the detection loophole in Bell experiments using qudits. *Physics Review Letters* **104**, 060401 (2010).
[50] Massar, S. & Pironio, S. Violation of local realism vs detection efficiency. *Physical Review A* **68**, 062109 (2003).
[51] Grant, M. & Boyd, S. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx (2014).
[52] MOSEK ApS. *The MOSEK optimization toolbox for Matlab manual. Version 8.1.* (2017). URL http://docs.mosek.com/8.1/toolbox/index.html.
[53] Cope, T. *The Role of Entanglement in Quantum Communication, and Analysis of the Detection Loophole*. Ph.D. thesis, University of York (2018). Also available as http://etheses.whiterose.ac.uk/22955/.
[54] Oudot, E., Bancal, J.-D., Sekatski, P., & Sangouard, N. Bipartite nonlocality with a many-body system. https://arxiv.org/abs/1810.05636 (2018).
[55] Cruzeiro, E. Z. & Gisin, N. Complete list of Bell inequalities with four binary settings. https://arxiv.org/abs/1811.11820 (2018).

Appendix A: Main mathematical results

Lemma 1. *Let B be a matrix such that $\mathrm{tr}(B^T\Pi) \geq c$ is a Bell inequality. There is a matrix $\tilde{B}$ whose entries are all non-negative such that $\mathrm{tr}(\tilde{B}^T\Pi) \geq 1$ represents the same Bell inequality.*

Proof. Let $\mathbf{1}$ be the matrix with every entry 1. The matrix $B' = B + (1-c)\mathbf{1}/(m_A m_B)$ is such that $\mathrm{tr}((B')^T\Pi) \geq 1$ is a Bell inequality. If B' has no negative entries we are done. Otherwise, suppose the minimum entry of B' is $-\alpha$. If we choose $\tilde{B} = (B' + \alpha\mathbf{1})/(1 + \alpha m_A m_B)$, so that, by construction, it has no negative entries, then for any local Π we have

$$\mathrm{tr}(\tilde{B}^T\Pi) = \frac{1}{1+\alpha m_A m_B}\mathrm{tr}((B')^T\Pi) + \frac{\alpha}{1+\alpha m_A m_B}\mathrm{tr}(\mathbf{1}\Pi) \geq \frac{1}{1+\alpha m_A m_B} + \frac{\alpha m_A m_B}{1+\alpha m_A m_B} = 1\,.$$

□

Corollary 2. *Let B be a matrix such that $\mathrm{tr}(B^T\Pi) \leq c$ is a Bell inequality. There is a matrix $\tilde{B}$ whose entries are all non-negative such that $\mathrm{tr}(\tilde{B}^T\Pi) \geq 1$ represents the same Bell inequality.*

Proof. The original Bell inequality is equivalent to $\mathrm{tr}((-B)^T\Pi) \geq c$ from which we can apply Lemma 1. □

Lemma 3. *Let B be a matrix with no negative entries and such that $\mathrm{tr}(B^T\Pi) \geq 1$ is a facet Bell inequality. There exists an extremal no-signalling distribution Π^{NS} such that $\mathrm{tr}(B^T\Pi^{NS}) = 0$.*

Proof. Since B represents a violatable Bell inequality, there exists a no-signalling point $\tilde{\Pi}$ such that $\mathrm{tr}(B^T\tilde{\Pi}) < 1$ but $\mathrm{tr}(\hat{B}^T\tilde{\Pi}) \geq 1$ for all other matrices $\hat{B}$ that represent facet Bell inequalities with local bound 1. Thus, if we run the dual program (13) for $\tilde{\Pi}$, the optimum is obtained when M^* corresponds to the same Bell inequality as B (it may be that $M^* \neq B$, but it must represent the same Bell inequality). Using the complementary slackness condition (8), we have

$$\mathrm{tr}\left(B^T\left(\tilde{\Pi} - \sum_i x_i^* P^{\mathrm{L},i}\right)\right) = 0\,, \tag{A1}$$

where $\{x_i^*\}$ achieve the optimum in the primal problem (12). This implies that the non-local part, $\tilde{\Pi} - \sum_i x_i^* P^{\mathrm{L},i}$, of $\tilde{\Pi}$ gives value 0 for the Bell inequality B. This non-local part is a convex combination of extremal non-local no-signalling distributions, $\{\Pi_j^{NS}\}$, each satisfying $\mathrm{tr}(B^T\Pi_j^{NS}) = 0$, as required. □

Remark 1. For the problem in the previous lemma, the second complementary slackness condition (9) gives that

$$\sum_i \left(\mathrm{tr}\left(B^T P^{\mathrm{L},i}\right) - 1\right) x_i^* = 0. \tag{A2}$$

Hence, for all i either $x_i^* = 0$ or $\left(\mathrm{tr}\left(B^T P^{\mathrm{L},i}\right) - 1\right) = 0$ (both values are non-negative). As x_i^* is non-zero if and only if the local distribution $P^{\mathrm{L},i}$ is in the local part of $\tilde{\Pi}$, we can conclude that $\mathrm{tr}\left(B^T P^{\mathrm{L},i}\right) = 1$ for these local distributions, and so the local part of $\tilde{\Pi}$ satisfies the Bell inequality with equality.

Remark 2. Remark 1 provides an insight into why it may be useful to remove the local part of a quantum distribution before using as the dual objective function. If one uses the original distribution, any deterministic distributions in the local part must be saturating points of the Bell inequality obtained by the dual. Removing the local part removes this requirement, which may allow the output Bell inequality to be such that these local points are not saturating. We are not aware of any bipartite Bell inequality for which there is no quantum violation (although these exist for three parties [42]). If such non-violatable inequalities exists, it is possible that they may be found by the dual program whose input is a quantum distribution with its local part removed.

Theorem 4. *Consider the* $(m_A, m_B, 2, 2)$ *scenario and let* B *be a matrix with no negative entries and such that* $\mathrm{tr}(B^T\Pi) \geq 1$ *is a facet Bell inequality. There exists an extremal no-signalling distribution* $\hat{\Pi}$ *that, up to relabellings, takes the form of* (11) *with* $g = h = 0$ *such that* $\mathrm{tr}(B^T\hat{\Pi}) = 0$.

Proof. By Lemma 3, there exists an extremal no-signalling distribution $\tilde{\Pi}$ such that $\mathrm{tr}(B^T\tilde{\Pi}) = 0$. Suppose that $\tilde{\Pi}$ has the form (11) with either $g \neq 0$, $h \neq 0$, or both. Our aim is to show that in these cases there is always another extremal no-signalling distribution $\hat{\Pi}$ of the form (11) with $g = h = 0$ such that $\mathrm{tr}(B^T\hat{\Pi}) = 0$.

Case 1: Suppose $g = 0$, $h \neq 0$. Since the coefficients of B are non-negative, B must have zero entries whenever $\tilde{\Pi}$ has non-zero entries. Hence, $B(0b|xy) = 0$ for $b \in \{0,1\}$, $x \in \{m_A - h + 1, \ldots, m_A\}$ and $y \in \{1, \ldots, m_B\}$.

Suppose there exists a local deterministic distribution, P^1, such that $\mathrm{tr}(B^T P^1) = 1$ and for which $P^1_{A|x'}(1) = 1$ for some $x' \in \{m_A - h + 1, \ldots, m_A\}$. Consider now another local deterministic distribution P^2 that is identical to P^1 except that $P^2_{A|x'}(0) = 1$ (i.e., P^2 is formed by exchanging the row of P^1 corresponding to $P^1_{A|x'}(0)$ with the row corresponding to $P^1_{A|x'}(1)$). It follows that $1 \leq \mathrm{tr}(B^T P_2) \leq \mathrm{tr}(B^T P_1) = 1$, i.e., P^2 saturates the Bell inequality if P^1 does. It also follows that the 2×2 blocks of B corresponding to a measurement of $X = x'$ and any $Y \in \{1, \ldots, m_B\}$ have the form $\begin{pmatrix} 0 & 0 \\ 0 & \gamma \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 \\ \gamma & 0 \end{pmatrix}$ (depending on whether $P^1_{B|y}(0) = 1$ or $P^1_{B|y}(1) = 1$), where γ is an arbitrary non-negative value.

We can therefore replace the 2×2 blocks of $\tilde{\Pi}$ corresponding to $X = x'$, $Y \in \{1, \ldots, m_B\}$ by A or S without affecting the value of $\mathrm{tr}(B^T\tilde{\Pi})$. In other words, if $\mathrm{tr}(B^T P^1) = 1$ and $P^1_{A|x'}(1) = 1$ for some $x \in \{m_A - h + 1, \ldots, m_A\}$, there exists another extremal no-signalling distribution, Π', with a smaller value of h that also has $\mathrm{tr}(B^T\Pi') = 0$.

If we can reduce in this way until $h = 0$, we are done. Alternatively, we reduce to the case where $\mathrm{tr}(B^T P) = 1$ implies $P_{A|x}(0) = 1$ for all $x \in \{m_A - h + 1, \ldots, m_A\}$. The affine span of the local deterministic distributions satisfying $\mathrm{tr}(B^T P) = 1$ is hence the same as the affine span of those satisfying $\mathrm{tr}(\tilde{B}^T\tilde{P}) = 1$, where $\tilde{B}$ and $\tilde{P}$ comprise the first $2(m_A - h)$ rows of B and P respectively. This is at most $((m_A - h)(n_A - 1) + 1)(m_B(n_B - 1) + 1) - 2$, and hence contradicts the assumption that B is a facet inequality.

Case 2: If $g \neq 0$, $h = 0$, we can run an analogous argument.

Case 3: Suppose that both $g \neq 0$ and $h \neq 0$, and consider some $x' \in \{m_A - h + 1, \ldots, m_A\}$ and $y' \in \{m_B - g + 1, \ldots, m_B\}$. We must have $B(0b|x'y) = 0$ for $b \in \{0, 1\}$ and $y \in \{1, \ldots, m_B - g\}$, $B(a0|xy') = 0$ for $a \in \{0, 1\}$ and $x \in \{1, \ldots, m_A - h\}$, and $B(00|x'y') = 0$.

Suppose there exists a local deterministic distribution, P^1, such that $\mathrm{tr}(B^T P^1) = 1$ and for which $P^1_{A|x'}(1) = 1$. Let P^2 be another local deterministic distribution that is identical to P^1 except that $P^2_{A|x'}(0) = 1$ and $P^2_{B|y}(0) = 1$ for all $y \in \{m_B - g + 1, \ldots, m_B\}$. It follows that $1 \leq \mathrm{tr}(B^T P_2) \leq \mathrm{tr}(B^T P_1) = 1$, i.e., P^2 saturates the Bell inequality if P^1 does.

It also follows that

- For $X = x'$ and any $Y \in \{1, \ldots, m_B - g\}$ the corresponding 2×2 blocks of B have the form $\begin{pmatrix} 0 & 0 \\ 0 & \gamma \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 \\ \gamma & 0 \end{pmatrix}$ (depending on whether $P^1_{B|y}(0) = 1$ or $P^1_{B|y}(1) = 1$).
- For $X = x'$ and $Y \in \{m_B - g + 1, \ldots, m_B\}$ the analogous blocks of B have the form $\begin{pmatrix} 0 & \gamma_1 \\ \gamma_2 & 0 \end{pmatrix}$ or $\begin{pmatrix} 0 & \gamma_1 \\ 0 & \gamma_2 \end{pmatrix}$.

We can therefore replace the 2×2 blocks of $\tilde{\Pi}$ corresponding to $X = x'$ by either A, S or L (whichever matches the zeros of B) without changing the value of $\mathrm{tr}(B^T \tilde{\Pi})$. In other words, if there exists a local deterministic distribution, P^1, such that $\mathrm{tr}(B^T P^1) = 1$ and for which $P^1_{A|x'}(1) = 1$, then there exists an extremal no-signalling distribution, Π', with a smaller value of h that also satisfies $\mathrm{tr}(B^T \Pi') = 0$.

If we can reduce in this way until $h = 0$, then either $g = 0$ are we are done, or we can complete the argument using Case 2. Alternatively, we reduce to the case where $\mathrm{tr}(B^T P) = 1$ implies $P_{A|x}(0) = 1$ for all $x \in \{m_A - h + 1, \ldots, m_A\}$, and $P_{B|y}(0) = 1$ for all $y \in \{m_B - g + 1, \ldots, m_B\}$. The affine span of the local deterministic distributions satisfying $\mathrm{tr}(B^T P) = 1$ is hence the same as the affine span of those satisfying $\mathrm{tr}(\tilde{B}^T \tilde{P}) = 1$, where $\tilde{B}$ and $\tilde{P}$ comprise the first $2(m_A - h)$ rows and $2(m_B - g)$ columns of B and P respectively. This is at most $((m_A - h)(n_A - 1) + 1)((m_B - g)(n_B - 1) + 1) - 2$, and hence contradicts the assumption that B is a facet inequality. $\square$

To get the idea of the proof, let us consider an example. Take $\tilde{\Pi} = \begin{pmatrix} S & S & L \\ S & A & L \\ K & K & M \end{pmatrix}$. A Bell inequality with $\mathrm{tr}\left(B^T \tilde{\Pi}\right) = 0$ must have the form

$$B = \left(\begin{array}{cc|cc|cc} 0 & v_1 & 0 & v_2 & 0 & v_3 \\ v_4 & 0 & v_5 & 0 & 0 & v_6 \\ \hline 0 & v_7 & v_8 & 0 & 0 & v_9 \\ v_{10} & 0 & 0 & v_{11} & 0 & v_{12} \\ \hline 0 & 0 & 0 & 0 & 0 & v_{13} \\ v_{14} & v_{15} & v_{16} & v_{17} & v_{18} & v_{19} \end{array}\right),$$

where v_i denotes an arbitrary non-negative entry.

Consider now the local deterministic distributions

$$P^1 = \left(\begin{array}{cc|cc|cc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{array}\right), \quad P^2 = \left(\begin{array}{cc|cc|cc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right), \quad \hat{P}^1 = \left(\begin{array}{cc|cc|cc} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right).$$

If $\mathrm{tr}(B^T P^1) = 1$ then we must have $\mathrm{tr}(B^T P^2) = 1$, and hence $v_{15} = v_{17} = v_{18} = 0$. In this case the distribution $\Pi' = \begin{pmatrix} S & S & L \\ S & A & L \\ S & S & L \end{pmatrix}$ will also satisfy $\mathrm{tr}(B^T \Pi') = 0$. We can then apply the argument of case 2 to this distribution.

Similarly, if $\mathrm{tr}(B^T \hat{P}^1) = 1$ then we must have $\mathrm{tr}(B^T P^2) = 1$, and hence $v_3 = v_9 = v_{15} = v_{17} = v_{19} = 0$. In this case the distribution $\Pi' = \begin{pmatrix} S & S & A \\ S & A & A \\ S & S & S \end{pmatrix}$ will also satisfy $\mathrm{tr}(B^T \Pi') = 0$. Note that by relabelling the outputs for $y = 3$ corresponds to exchanging the A entries for Ss in the final column of Π', bringing us into a form matching (11).

By arguments of this kind it follows that either we can reduce to a case with a lower value of g or h, or all the local deterministic distributions with $\mathrm{tr}(B^T P) = 1$ have zeros in the final row and column ($P_{X|a=3}(0) = 1$ and $P_{Y|b=3}(0) = 1$). In the latter case the dimension of the plane containing the saturating local deterministic distributions is insufficient for B to be a facet Bell inequality.

Appendix B: Generating Quantum Distributions

As stated in the main body, one may use our linear programming algorithm using quantum distributions rather than extremal no-signalling ones. To generate these, one must first fix the dimension of the state d. One then creates a normalised real vector $\boldsymbol{\lambda}$ of d non-zero real elements. We take these as the Schmidt coefficients of the pure entangled state $|\phi\rangle = \sum_{i=1}^{d} \lambda_i |i\rangle \otimes |i\rangle$. We then generate $m_A + m_B$ random unitaries $\{U^{A_1}, U^{A_2}, \ldots, U^{A_{m_A}}, U^{B_1}, U^{B_2}, \ldots, U^{B_{m_B}}\}$, each unitary corresponding to a measurement. Since the columns of each U^i are orthonormal, we can define projection operators $P^i_k := \sum_{k \in \mathcal{S}^i_n} U^i |k\rangle\langle k| (U^i)^\dagger$, where $\{\mathcal{S}^i_n\}$ is a partition of $\{1, 2, \ldots, n_A\}$ or $\{1, 2, \ldots, n_B\}$ as appropriate. These projectors satisfy $\sum_k P^i_k = \left(U^i\right)^\dagger U^i = \mathbb{I}_d$. Thus, we obtain the probability distribution:

$$P_{AB|xy}(a,b) = \langle \phi | (P^{A_x}_a \otimes P^{B_y}_b) | \phi \rangle. \tag{B1}$$

We may then use this as the objective function. Note that this construction does not guarantee a non-local distribution.

Appendix C: Guaranteeing All Inequality Classes

Although for the scenarios in which the number of inequality classes was known, Algorithm 1 was able to generate a representative of every class, due to the degeneracy of the optimal solutions it does not guarantee that this generally will be the case. In this appendix we discuss an alteration to the algorithm that can, in principle, provide this. However, the run time of such an algorithm is prohibitive. We nevertheless state the method here, because it gives the idea behind Algorithm 1.

Suppose we have a specific Bell inequality B we wish to find as a solution to (13). Let $\tilde{\Pi}$ be an extremal no-signalling distribution such that $\mathrm{tr}(B^T \tilde{\Pi}) = 0$ and let $\{P_j\}_{j=1}^d$ be a set of linearly independent local deterministic distributions saturating B (where d is the dimension of the local polytope). If we define $\Pi' = (1-\delta)\,\tilde{\Pi} + \delta \sum_{j=1}^{d} P_j / d$, then $\mathrm{tr}(B^T \Pi') = \delta$.

Furthermore, the matrix B represents the unique Bell inequality that achieves the minimum solution to the dual problem (13) with input Π'. To see this, note that no matrix M for which $\mathrm{tr}(M^T P) \geq 1$ is a Bell inequality can give a lower value because $\mathrm{tr}(M^T \Pi)$ is bounded below by 0 and $\mathrm{tr}(M^T P_i) \geq 1$. In addition, no other matrix $\hat{M}$ with non-negative entries for which $\mathrm{tr}(\hat{M}^T P) \geq 1$ is a different Bell inequality will achieve this value because $\mathrm{tr}(\hat{M}^T \Pi') \geq \delta \sum_{j=1}^{d} \mathrm{tr}(\hat{M}^T P_j)/d$. It cannot be that $\mathrm{tr}(\hat{M}^T P_j) = 1$ for all $j \in \{1, \ldots, d\}$ because this would mean $\{P_j\}_{j=1}^d$ also all lie on the facet formed by $\hat{M}$, which is only possible if the facets are identical.

It follows that a Bell inequality of every class can be generated by considering all possible objective functions of the form $(1-\delta)\,\tilde{\Pi} + \delta \sum_{j=1}^{d} P_j/d$ where $\tilde{\Pi}$ is an extremal no-signalling point of the form in Eq. (11) and $\{P_j\}_{j=1}^d$ are linearly independent deterministic local distributions. However, this algorithm is impractical for all but the smallest cases. In the case $n_A = n_B = 2$, we wish to choose $d = (m_A+1)(m_B+1) - 1$ local deterministic distributions from $2^{m_A+m_B}$ possible choices and the number of ways to do this scales roughly as $2^{m_A m_B (m_A+m_B)}$, which is prohibitively large. This is why in Algorithm 1 we mix each no-signalling point with only two local deterministic distributions. In this algorithm the local deterministic distributions that are mixed with are taken from those that saturate the Bell inequality M found in Step 3. Because of Remark 1, it is possible that the Bell inequality M' that is found in Step 7 is equal to M, since not all degeneracies are broken. In practice though, it turns out this small mixing leads the linear programming algorithm to output different inequalities. For small values of m_A and m_B the algorithm can be run in reasonable time.

Appendix D: Using the Polar Dual

One disadvantage of our algorithm is that not all solutions of the linear programming problem are facet inequalities. This increases the calculation time because we have to check the affine dimension of every solution and discard many non-facet cases. In this appendix we consider an alternative approach which guarantees facet outputs, by taking advantage of the *polar dual* of the local polytope.

Definition 6. Given a polytope $\mathcal{P} \subset \mathbb{R}^{s\times t}$, its *polar dual*[11] is the set of points:

$$\mathcal{P}^\star := \left\{ B \in \mathbb{R}^{s\times t} \,\middle|\, \mathrm{tr}(B^T\Pi) \leq 1 \;\forall\, \Pi \in \mathcal{P} \right\}. \tag{D1}$$

If the co-ordinate origin is interior to $\mathcal{P}$ (i.e., $\Pi = 0$ is a non-boundary element of $\mathcal{P}$), then the polar dual satisfies $(\mathcal{P}^\star)^\star = \mathcal{P}$, and the two are linked by the following [28]

H-representation	V-representation
$\mathcal{P} = \left\{ \Pi \in \mathbb{R}^{s\times t} \,\middle|\, \mathrm{tr}(B_j^T\Pi) \leq 1 \;\forall\, j \right\}$	$\mathcal{P} = \left\{ \Pi = \sum_i \lambda_i \Pi_i \,\middle|\, \sum_i \lambda_i = 1,\ \lambda_i \geq 0 \right\}$ (D2)
$\mathcal{P}^\star = \left\{ B \in \mathbb{R}^{s\times t} \,\middle|\, \mathrm{tr}(B^T\Pi_i) \leq 1 \;\forall\, i \right\}$	$\mathcal{P}^\star = \left\{ B = \sum_j \lambda_j B_j \,\middle|\, \sum_j \lambda_j = 1,\ \lambda_j \geq 0 \right\}$, (D3)

where $\Pi_i, B_i \in \mathbb{R}^{s\times t}$. Hence, there is a one-to-one correspondence between the vertices of the primal and the facets of the dual, and vice versa.

The motivation for considering the polar dual is that by optimizing with the polar dual of the local polytope as the solution space, a simplex algorithm will always give a solution corresponding to a facet of the local polytope.

In order to use this form of the polar dual we require the origin to lie in the interior. We hence perform a translation of coordinates. A natural choice of the new origin is the distribution Π^u whose entries are all $1/n_A n_B$. For example, in the $(2, 2, 2, 2)$ scenario this would map the extremal no-signalling distribution

$$\left(\begin{array}{cc|cc} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \hline \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{array}\right) \quad \text{to} \quad \left(\begin{array}{cc|cc} \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ \hline \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \end{array}\right).$$

Note that this representation of a distribution can have negative entries. The linear programming problem we wish to solve is then

$$\begin{aligned} \max_{M} \quad & \mathrm{tr}(M^T \vec{\tilde{\Pi}}) \\ \text{subject to} \quad & \mathrm{tr}(M^T \vec{P}^{\mathrm{L},i}) \leq 1 \text{ for all } i. \end{aligned} \tag{D4}$$

where $\vec{\tilde{\Pi}}$ refers to distribution $\tilde{\Pi}$ after shifting origin. The canonical form of a linear program requires positive entries. To take care of this, we can write $M = M_+ - M_-$ where $M_+, M_- \geq 0$ component-wise, making the problem

$$\begin{aligned} \max_{M_+, M_-} \quad & \mathrm{tr}(M_+^T \vec{\tilde{\Pi}}) - \mathrm{tr}(M_-^T \vec{\tilde{\Pi}}) \\ \text{subject to} \quad & \mathrm{tr}(M_+^T \vec{P}^{\mathrm{L},i}) - \mathrm{tr}(M_-^T \vec{P}^{\mathrm{L},i}) \leq 1 \text{ for all } i \\ & M_+, M_- \geq 0\,. \end{aligned} \tag{D5}$$

Although the solution space is technically unbounded due to rays of the form $(M_+)_{ij} = (M_-)_{ij}$, these cannot contribute to the objective function, and the problem (D5) is bounded. However, our conversion of the problem into the form (D4) means we no longer have a *known* bound on the objective function[12]. To illustrate the problem

[11] In the study of convex bodies, there are other types of dual, which we do not use here.
[12] In the original form, we had that $\mathrm{tr}(M^T\tilde{\Pi}) \geq 0$.

with this, we perform the optimisation (D5), for all extremal non-local no-signalling distributions. For the $(4,4,2,2)$ scenario, the largest value obtained is $12/5$, and the smallest is 2. (Unlike in problem (13), the optimal value varies depending on the extremal no-signalling point used.)

We now consider a particular $(4,4,2,2)$ Bell inequality of the form:

$$\overrightarrow{B}_{\text{ex}} = \left(\begin{array}{cc|cc|cc|cc} 0 & -\frac{4}{5} & 0 & 0 & 0 & 0 & 0 & -\frac{4}{5} \\ 0 & 0 & -\frac{4}{5} & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & -\frac{4}{5} & 0 & 0 \\ -\frac{4}{5} & 0 & 0 & -\frac{4}{5} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -\frac{4}{5} & 0 & 0 & -\frac{4}{5} & 0 \\ 0 & 0 & 0 & 0 & -\frac{4}{5} & 0 & 0 & 0 \end{array}\right), \quad \operatorname{tr}\left(\overrightarrow{B}_{\text{ex}}^{T}\overrightarrow{\Pi}\right) \leq 1, \tag{D6}$$

and the corresponding problem of maximizing $\operatorname{tr}(\overrightarrow{B}_{\text{ex}}^{T}\overrightarrow{\Pi})$ over all $\overrightarrow{\Pi} \in \overrightarrow{\mathcal{NS}}_{(m_A,m_B,n_A,n_B)}$. Performing this optimisation we find the maximal value to be $9/5 < 2$. We can therefore conclude there is *no* extremal no-signalling point that, when used as the objective function for the problem (D4), will give the solution $M = \overrightarrow{B}_{\text{ex}}$. By using the polar dual, if ϵ is too small there are certain facets that Algorithm 1 (with this new linear programming problem) will never output, and it is not clear whether ϵ can be chosen in such a way that all facets could in principle be output. However, this restriction does not apply to Algorithm 2, for which this translation may be useful.

Appendix E: Finding the Unlifted Form of a Bell Inequality

Given a facet inequality B, we would like to know whether it is lifted from a lower dimensional scenario. Whilst it is sometimes easy to tell a lifting by inspection, because there are many ways to represent the same Bell inequality (as discussed in Section II C), this is not always the case. Here we present an algorithm that finds the smallest scenario from which the inequality has been lifted, and gives the inequality in the smallest scenario. This algorithm uses two subalgorithms, the first of which checks for lifted outputs, and the second of which removes lifted inputs.

The idea here is that output liftings correspond to copying the coefficients corresponding to one of the existing outputs. This means that, in the resulting inequality, two local deterministic distributions that are identical up to the choice of this output give the same Bell value.

1. Given a facet Bell inequality and an input x, this checks whether any of the outputs are lifted. To do so the algorithm runs through all pairs of outputs a_1 and a_2 corresponding to the input x. If for all local deterministic distributions that use the output a_1 for input x the Bell value is the same as for the local deterministic distribution that is identical except that it uses the output a_2 for x, then the output a_2 can be removed for input x. Having checked all pairs of outputs, the algorithm returns the set of outputs that can be removed, O_j.

2. Given a facet Bell inequality for which one of the inputs has only one possible output, this returns a new Bell inequality for the scenario with one fewer input without altering the bound on the Bell inequality and with the property that the input Bell inequality is a lifting of the returned inequality. This works by adding no-signalling type matrices to the matrix form of the Bell inequality in such a way as to make all of the entries corresponding to the input with only one output equal to zero. [This can be achieved with a no-signalling type matrix whose only non-zero entries are in a single row in which the entry corresponding to the input to be removed takes value $-v$ and the entries corresponding to another input with d outputs has entries $(v, v, \ldots, v)$.] Once all the entries are zero the new Bell inequality is formed by removing all entries corresponding to the input.

Algorithm 4

This algorithm starts with a matrix representation B of a facet Bell inequality, and the scenario in which it was found, $S = [(n_A^1\ n_A^2\ \ldots\ n_A^{m_A})\ (n_B^1\ n_B^2\ \ldots\ n_B^{m_B})]$.

1. Set $j = m_A$, $S' = S$ and $B' = B$.

2. Apply subalgorithm 1 to the j^{th} input of Alice.

3. If the output set O_j is non-empty, update B' by removing the entries in the matrix corresponding to the outputs in set O_j for the j^{th} input and update S' by replacing n_A^j by $n_A^j - |O_j|$.

4. If the j^{th} input of Alice has only one possible output, update B' using subalgorithm 2 and update S' by removing n_A^j.

5. Set $j = j - 1$.

6. Repeat steps 2-5 until $j = 0$.

7. Set $j = m_B$.

8. Repeat steps 2-5, but with the roles of Alice and Bob interchanged, until $j = 0$.

9. Return B' and S'.

This algorithm will return a facet Bell inequality B' and a new scenario S' reduced such that B' contains no lifted inputs and outputs, and hence S' is the smallest scenario for which B can be formed from a lifting. Note also that the local bounds for B and B' are the same and that if B has all positive entries, so does B'.