

## Causal structures from entropic information: geometry and novel scenarios

**Rafael Chaves, Lukas Luft and David Gross**

Institute for Physics, University of Freiburg, Rheinstrasse 10, D-79104 Freiburg, Germany  
E-mail: [rafael.chaves@physik.uni-freiburg.de](mailto:rafael.chaves@physik.uni-freiburg.de)

Received 4 October 2013, revised 17 January 2014

Accepted for publication 20 February 2014

Published 3 April 2014

*New Journal of Physics* **16** (2014) 043001

doi:[10.1088/1367-2630/16/4/043001](https://doi.org/10.1088/1367-2630/16/4/043001)

### Abstract

Bell's theorem in physics, as well as causal discovery in machine learning, both face the problem of deciding whether observed data is compatible with a presumed causal relationship between the variables (for example, a local hidden variable model). Traditionally, Bell inequalities have been used to describe the restrictions imposed by causal structures on marginal distributions. However, some structures give rise to non-convex constraints on the accessible data, and it has recently been noted that linear inequalities on the observable entropies capture these situations more naturally. In this paper, we show the versatility of the entropic approach by greatly expanding the set of scenarios for which entropic constraints are known. For the first time, we treat Bell scenarios involving multiple parties and multiple observables per party. Going beyond the usual Bell setup, we exhibit inequalities for scenarios with extra conditional independence assumptions, as well as a limited amount of shared randomness between the parties. Many of our results are based on a geometric observation: Bell polytopes for two-outcome measurements can be naturally imbedded into the convex cone of attainable marginal entropies. Thus, any entropic inequality can be translated into one valid for probabilities. In some situations the converse also holds, which provides us with a rich source of candidate entropic inequalities.

**Keywords:** non-locality, marginal scenarios, causal structures, entropic inequalities



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

## 1. Introduction

The starting point of this paper is the question: what can be inferred about the causal relationship of a collection of random variables from a restricted set of observations? To phrase this problem more precisely, we need to introduce the notions of a *marginal scenario* and a *causal structure*—the two pieces of data that specify the instances we will be considering.

A marginal scenario describes which sets of random variables are jointly observable. Joint observations might be constrained for a variety of reasons. In quantum non-locality<sup>1</sup>, these reasons are physical: random variables corresponding to non-commuting observables cannot always be jointly measured. In general, there might also be practical reasons, for instance: we have no access to the variable describing the genetic disposition of a patient to become both a smoker and to develop lung cancer (not least because we do not know whether such a genetic influence exists).

For the purpose of this paper, a causal structure is a list of linear constraints on the (conditional) mutual information between sets of random variables. For example, in the familiar Bell scenario, one commonly demands that measurement choices of, respectively, Alice  $X$  and Bob  $Y$  are independent of the hidden variable  $\lambda$ :  $I(X, Y; \lambda) = 0$ . Relaxing this constraint to demand the correlations be small,  $I(X, Y; \lambda) \leq \epsilon$  would still be linear in the mutual information and thus an element of a causal structure according to our definition. By allowing for arbitrary linear constraints, we go slightly beyond the way the notion of causal structure is commonly formalized in the field of causal inference [2, 3]. There, the combinatorial structure of a direct acyclic graph is used to encode certain sets of conditional mutual information that are assumed to vanish. Our approach subsumes and extends this.

With every given causal structure, we can associate the set of marginal distributions that are compatible with it. If we observe a data point that lies outside that region, we can exclude the presumed causal structure as a valid model for the observed data. This logical structure (characterize the global properties compatible with local observations) is an instance of a marginal problem, which occurs frequently both in classical [4] and in quantum probability [5–8].

In quantum non-locality [1], the focus has traditionally been on settings for which the marginal distributions happen to be convex polytopes. In that case, checking whether an observed marginal distribution is compatible with the causal model reduces to the task of verifying that none of the inequalities associated with the facets of the polytopes are violated—these are the Bell inequalities [9, 10]. However, the non-convex nature of mutual information means that the marginals that appear for more general causal structures are, at best, non-trivial algebraic varieties. A few such examples have been treated in the quantum literature, including bilocality scenarios [11] or scenarios that allow for correlations between Alice’s and Bob’s measurement choices with the hidden variable [12, 13] (cf also section 5).

<sup>1</sup> The notion of quantum non-locality is arguably somewhat unfortunate, as it seems to imply that an unspecified super-luminal mechanism is responsible for the strong correlations observed in quantum experiments. Bell’s theorem only asserts that no local hidden variable (LHV) theory can reproduce these correlations. It does not put forward any alternative mechanism—certainly none involving actions at a distance. It has, however, become customary [1] to refer to statistics that cannot arise in LHV models as ‘non-local’. We prefer sticking to commonly accepted language over inventing a possibly better-suited term that will not be widely understood.

*A priori*, it is unclear whether these more complicated marginal regions allow for an explicit description that is tractable from an analytic and computational point of view. It is this problem that entropic methods greatly simplify. Indeed, as indicated above, (conditional) independence constraints are linear in terms of Shannon entropies. As a result, the image of the marginal regions of general causal structures turns out to possess natural descriptions in terms of linear inequalities.

The set of all joint entropies (without any causal constraints and prior to marginalization) has been analyzed extensively in information theory [14, 15]. While it is known to be a convex cone, its precise form is still not explicitly understood. For practical purposes, it is often replaced by an outer approximation: the convex Shannon cone, which is defined by a finite number of explicit Shannon-type inequalities. All marginals of the Shannon cone can, in principle, be found computationally using linear programming [15]. What is more, causal structures merely amount to further linear constraints and can therefore be included in a natural way. An additional nice feature of entropic inequalities is that they are valid for variables consisting of any number of outcomes. This stands in stark contrast to the usual approach for which increasing the number of outcomes of the marginal scenario increases the dimension and complexity of the correlation polytope, in practice meaning that new inequalities need to be derived and tailored to the specific number of outcomes under consideration.

On the negative side, entropic inequalities provide, in principle, only a necessary condition for the solution of the marginal problem [16]. In spite of this, entropic inequalities are known to be fine enough to distinguish, for example, different causal structures [17, 18] or witness non-locality and contextuality [19–26].

In spite of its potential applications, the entropic approach to the marginal problem has been explored little. In particular, no entropic inequalities are known for Bell scenarios involving more than two parties or many measurement settings. Another problem well suited to be tackled using entropies is the one where the amount of shared randomness between the parties involved in a Bell test is bounded to be below a certain value. Commonly, shared randomness is assumed to be a free and boundless resource, but quantitative considerations about how much of it is actually necessary to reproduce some quantum correlations can give useful insights that would be extremely hard to tackle with the usual approaches.

These are the kind of problems we look at in this paper. In section 2 we start defining the entropic cone described by all Shannon-type inequalities. In section 3 we state known results about convex cones that will be used in section 4 to prove a theorem showing that, for marginal scenarios without statistical independence, any Shannon-type inequality is also valid for the probabilities if a proper translation is made. We also show that the converse is in general not true by providing a counter example showing that not every inequality for probabilities is also valid for Shannon entropies. Inspired by these results, in section 5 we derive the entropic version of the Collins–Gisin (CG) inequalities [27], also considering the effects of bounded shared randomness between the parties. In section 6 we derive a multipartite generalization of the entropic inequality originally derived by Braunstein and Caves for the bipartite case [19], and in section 7 we computationally apply the Fourier–Motzkin (FM) algorithm to derive entropic inequalities for a couple of different scenarios, including marginal models that also include statistical independencies and the effects of bounded shared randomness. We discuss our findings in section 8, while the technical results and proofs can be found in the appendices.

## 2. Characterizing marginal scenarios with Shannon-type inequalities

### 2.1. Marginal scenarios

Given a set of variables  $X_1, \dots, X_n$ , a marginal scenario is a collection of certain subsets of them; those subsets of variables that can be jointly measured. In the case of Bell scenarios, the marginal scenario is achieved by imposing a space-like separation between some of the observables. Clearly, a subset of jointly measurable variables is still a jointly measurable set. Formally (see [20, 22] for further details), see definition 1.

**Definition 1.** A marginal scenario  $\mathcal{M}$  is a collection  $\mathcal{M} = \{S_1, \dots, S_{|\mathcal{M}|}\}$  of subsets  $S_i \subseteq \{X_1, \dots, X_n\}$  such that if  $S \in \mathcal{M}$  and  $S' \subseteq S$ , then also  $S' \in \mathcal{M}$ .

In practice, some joint statistics are measured for every  $S \in \mathcal{M}$ . For example, if the variables  $X_i$  and  $X_j$  are jointly measurable ( $\{X_i, X_j\} =: S \in \mathcal{M}$ ), one can access  $P(X_i = x_i, X_j = x_j)$ , and the probability of obtaining the outcomes  $X_i = x_i$  and  $X_j = x_j$ . These marginal probabilities determine in particular the marginal Shannon entropy:

$$\begin{aligned} H(X_i, X_j) \\ = - \sum_{x_i, x_j} P(X_i = x_i, X_j = x_j) \log_2 P(X_i = x_i, X_j = x_j). \end{aligned}$$

As first noticed in [19], the existence of a joint distribution for all variables  $X_1, \dots, X_n$  implies that marginal Shannon entropies satisfy certain inequalities, which may be violated by measurement statistics originating from quantum experiments. Below, we will recall how to compute these inequalities in general, potentially in the presence of extra causal constraints.

### 2.2. Entropy cones

For the purpose of this section, assume some joint distribution for  $n$  random variables  $X_1, \dots, X_n$ . We denote the set of indices of the random variables by  $[n] = \{1, \dots, n\}$  and its powerset (i.e. set of subsets) by  $2^{[n]}$ . For every subset  $S \in 2^{[n]}$  of indices, let  $X_S$  be the tuple of observables  $(X_i)_{i \in S}$  and  $H(S) := H(X_S)$  the associated marginal entropy. With this convention, the entropy becomes a function

$$H: 2^{[n]} \rightarrow \mathbb{R}, \quad S \mapsto H(S)$$

on the power set. The linear space of all set functions is of course isomorphic to  $\mathbb{R}^{2^n}$  together with a basis  $\{e_S \mid S \subset 2^{[n]}\}$  labeled by subsets. We denote that vector space by  $R_n$  and will henceforth not distinguish between real-valued set functions and the space  $R_n$ . For every vector  $h \in R_n$  and  $S \in 2^{[n]}$ , we denote using  $h_S$  the component of  $S$  with respect to the basis vector  $e_S$ .

The region

$$\{h \in R_n \mid h_S = H(S) \text{ for some entropy function } H\}$$

of vectors in  $R_n$  that correspond to entropies has been researched extensively in information theory [15]. It is known to be a convex cone (cf section 3), but an explicit description has not yet been found. However, several properties of entropy functions are well understood. These are, respectively, monotonicity, sub modularity, and a normalization condition:

$$\begin{aligned} H([n] \setminus \{i\}) &\leq H([n]) \quad \forall i \in [n], \\ H(S) + H(S \cup \{i, j\}) &\leq H(S \cup \{i\}) + H(S \cup \{j\}) \quad \forall S \subseteq [n], i, j \in [n] \setminus S \text{ with } i \neq j, \\ H(\emptyset) &= 0. \end{aligned} \tag{1}$$

The set of inequalities (1) are known as the elementary inequalities in information theory or the polymatroidal axioms. An inequality that follows from the elementary ones is called a Shannon-type inequality. The region defined by the Shannon-type inequalities is the Shannon cone  $\Gamma_n$ , a polyhedral closed convex cone. Clearly, this is an outer approximation to the true entropy cone. Since the latter is not yet fully characterized, we will work for the remainder of this paper solely in terms of the Shannon cone<sup>2</sup>. For future reference, we re-state this definition more formally.

**Definition 2.** The Shannon cone  $\Gamma_n$  is the set of vectors  $h \in R_n$  that are

1. non-negative

$$h_A \geq 0 \text{ (with equality if } A = \emptyset\text{),}$$

2. increasing

$$h_B \leq h_A \text{ for } B \subseteq A,$$

3. sub-modular

$$h_{A \cup B} + h_{A \cap B} \leq h_A + h_B$$

for any  $A, B \in 2^{[n]}$ .

We now return to descriptions involving a marginal scenario  $\mathcal{M}$ . Given a point  $h: 2^{[n]} \rightarrow \mathbb{R}$ , one computes the restriction  $h|_{\mathcal{M}}: \mathcal{M} \rightarrow \mathbb{R}$  dismissing the values  $h(S)$  for all  $S \in [n] \setminus \mathcal{M}$ . The entropic cone bounding the correlations in  $\mathcal{M}$  is a projection of  $\Gamma_n$  along a map  $\mathbb{R}^{2^{[n]}} \rightarrow \mathbb{R}^{|\mathcal{M}|}$  throwing away some of the coordinates, the ones not corresponding to observable quantities. This set is also a convex cone, that we denote by  $\Gamma^{\mathcal{M}}$ . Given an inequality description of  $\Gamma^{\mathcal{M}}$ , deciding if the marginal model can be extended is very simple, since one only needs to check whether it satisfies all the inequalities defining it. In other terms, if a marginal model violates an inequality derived only by the combination of polymatroidal axioms, this implies that this marginal model cannot arise from a joint probability distribution.

To determine the projection  $\Gamma^{\mathcal{M}}$ , a natural possibility would be to calculate the extremal rays of  $\Gamma_n$  and dismiss the irrelevant coordinates of it. However, determining all the extremal rays of the cone  $\Gamma_n$  is a very hard problem, with explicit solutions known only for a few cases [28–30]. To determine  $\Gamma^{\mathcal{M}}$  in practice, we start with the inequality description (1) of  $\Gamma_n$  and then

<sup>2</sup> This relaxation implies that while all the inequalities we will derive below are valid for any true entropy vector, they may fail to be tight.

apply an FM elimination [31], a standard method for calculating the inequality description for the projection of a polyhedral cone.

Note that the framework remains unchanged in the presence of extra homogeneous linear inequalities, e.g. those expressing the constraints implied by a given causal structure. In this case, the new global cone (prior to marginalization) is simply  $\Gamma_n \cap \Gamma_{cc}$ , with  $\Gamma_{cc}$  standing for the cone cut out by the extra linear constraints. In section 5, we will also consider non-homogeneous global constraints. These arise, for example, in scenarios where the degree of shared randomness is allowed to be positive but bounded. The global convex object is then no longer a cone, but a convex polyhedron. This affects some computational issues in a minor way, as discussed in detail in appendix B.

### 2.3. Inequalities for marginal entropies

To illustrate the general method, we begin by considering the simplest non-trivial Bell scenario, corresponding to the CHSH scenario [32] and consisting of two parties, say A and B, who can measure one out of two observables each,  $\{A_0, A_1\}$  and  $\{B_0, B_1\}$ , respectively. This corresponds to a marginal scenario consisting of the following observable variables:  $\mathcal{M} = \{ \{A_0, B_0\}, \{A_0, B_1\}, \{A_1, B_0\}, \{A_1, B_1\} \} = \{A_0, A_1, B_0, B_1, A_0B_0, A_0B_1, A_1B_0, A_1B_1\}$  (in this case  $|\mathcal{M}| = 8$ ). As shown in [20, 22], the only non-trivial Shannon-type entropic inequality (up to symmetries) corresponds to the inequality derived by Braunstein and Caves [19], the entropic CHSH, given by

$$\begin{aligned} CHSH^E = & H_{A_0B_0} + H_{A_0B_1} + H_{A_1B_0} \\ & - H_{A_1B_1} - H_{A_0} - H_{B_0} \geq 0 \end{aligned} \quad (2)$$

where here and in the following we employ the notation  $H(A_i, B_j) = H_{A_iB_j}$  (similarly to any number of variables) to avoid lengthy expressions.

In [19] this inequality was derived using the chain rule of entropies. However, as just discussed, any Shannon-type inequality can be derived from the elemental set of inequalities (1). To illustrate the general procedure, we consider how to obtain the entropic inequality (2), performing an FM elimination of the non-observable variables appearing in the set of elementary inequalities. To derive the CHSH inequality (2), it is sufficient to combine the two sub-modularity inequalities

$$H_{A_0B_0} + H_{A_0B_1} \geq H_{A_0B_0B_1} + H_{A_0} \quad (3)$$

$$H_{A_1B_0} + H_{B_0B_1} \geq H_{A_1B_0B_1} + H_{B_0}. \quad (4)$$

Using  $H_{A_0B_0B_1} \geq H_{B_0B_1}$  and  $H_{A_1B_0B_1} \geq H_{A_1B_1}$  we get exactly (2). Note, however, that these two last monotonicity inequalities are not in the elemental set (1). To obtain, for instance,  $H_{A_0B_0B_1} \geq H_{B_0B_1}$  from the basic ones, we combine

$$H_{A_0A_1B_0B_1} \geq H_{A_1B_0B_1} \quad (5)$$

$$H_{A_1B_0B_1} + H_{A_0B_0B_1} \geq H_{A_0A_1B_0B_1} + H_{B_0B_1}. \quad (6)$$

It is clear that, in general, any monotonicity inequality follows immediately from the elemental ones.



One should note the similarity of  $CHSH^E$  with the usual CHSH inequality in terms of probabilities [27, 32, 33], which can be expressed as (written here in a form closer to that of the CH inequality [34])

$$\begin{aligned} CHSH &= q_{A_0 B_0} + q_{A_0 B_1} + q_{A_1 B_0} \\ &\quad - q_{A_1 B_1} - q_{A_0} - q_{B_0} \leq 0 \end{aligned} \quad (7)$$

with  $q_{A_i B_j}$  being the probability of getting the outcome 0 if the measurement settings  $i, j$  are used, and similarly for the marginals  $q_{A_i}$  and  $q_{B_j}$ . We see that both inequalities are equivalent, if one just makes the simple replacement  $H_{A_i B_j} \rightarrow -q_{A_i B_j}$ . Based on this simple observation, in section 4 we prove a formal explanation to the similarities between the probability and entropic inequalities.

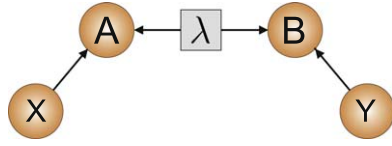
#### 2.4. The role of causal structures

Bell's theorem is usually associated with the incompatibility of quantum correlations with a natural causal structure for space-like separated events. However, in the derivation of the entropic inequality (2), no explicit mention of a causal structure has been made. Inequality (2) is valid for any set of four variables. The only assumption made up to this point is the validity of classical probability theory, or in other terms, the existence of a well-defined joint probability distribution  $p(A_0 = a_0, A_1 = a_1, B_0 = b_0, B_1 = b_1)$ . Bell's theorem can be seen as a recipe for interpreting the variables appearing in (2) or (7) as physically observable quantities.

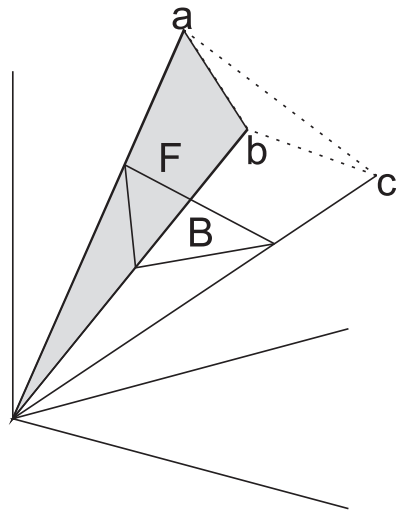
We recall the usual argument: Bell's theorem assumes a description of marginal models where a hidden variable  $\lambda$  exists which subsumes all the information the variables  $A_0, A_1, B_0$  and  $B_1$  may depend on. This is the realism assumption in Bell's construction, ensuring that all the variables have well-defined values prior to any measurement. At each run of the experiment, Alice and Bob independently choose which variable they will locally access, tossing, respectively, uncorrelated coins  $X$  and  $Y$ : if  $X = 0$  Alice measures the observable associated with  $A_0$ , if  $X = 1$  she measures  $A_1$  (similarly to Bob). Because in general  $A_0$  and  $A_1$  (similarly  $B_0$  and  $B_1$ ) are associated with non-commuting observables, quantum mechanics prohibits both to be jointly measurable. The compatibility between  $A_i$  and  $B_j$  is guaranteed by invoking the assumption of locality, stating that space-like events are not causally connected. Note, however, that, for example,  $A_0$  is in principle not an observable quantity, rather what Alice observes is  $A_0$  conditioned by the fact that  $X = 0$ . If  $X$  is correlated with  $\lambda$ , the value of  $A_0$  would potentially be different had Alice chosen to measure  $A_1$ . Here enters the final assumption in Bell's theorem, the statistical independence between local settings  $X$  and  $Y$  and the hidden variable  $\lambda$  (the so-called measurement independence or free will assumption). Together, the three assumptions in Bell's theorem imply the causal structure shown in figure 1.

### 3. Convex cones

In this section, we state several basic facts about closed convex cones and their duals. Detailed background and proofs can be found in [35]. General text books on convexity that also treat



**Figure 1.** A direct acyclic graph representing the causal structure associated with a bipartite Bell experiment [9]. The associated marginal scenario consists of all sets of variables that do not contain two different observables of the same party. For instance, if  $X$  and  $Y$  are both dichotomic with  $x = 0,1$  and  $y = 0,1$ , we have the CHSH scenario [33] that is characterized by the marginal scenario  $\mathcal{M} = \{ \{A_0, B_0\}, \{A_0, B_1\}, \{A_1, B_0\}, \{A_1, B_1\} \}$ .



**Figure 2.** A closed convex cone in  $\mathbb{R}^3$ . The extremal rays are labeled  $a, b, c$ , while  $B$  designates a base. One of three facets is shaded and labeled  $F$ .

cones are [36, 37]. All the cones that appear in this paper are closed and convex, so we will at times drop the attributes.

A closed convex cone  $C$  is a subset of  $\mathbb{R}^n$

1. closed  $\bar{C} = C$ ,
2. convex, and
3. scale-invariant:  $\lambda C = C$  for every  $\lambda > 0$ .

A simple example is given in figure 2.

The simplest types of cones are rays, i.e. sets of the form  $\{\mu v \mid \mu \geq 0\}$  for some vector  $v \in \mathbb{R}^n$ . Let  $L \subset C$  be a ray contained in a closed convex cone  $C$ . It is an extremal ray if it cannot be written as a non-trivial convex combination of elements in  $C$ , i.e. if for all  $x, y \in C$ , whenever  $\frac{1}{2}(x + y) \in L$ , we already have  $x, y \in L$ .

Under a technical assumption, closed convex cones are the convex hull of their extremal rays. To state the assumption, we need to introduce the notion of a base. A base is a convex subset  $B \subset C$  of a convex cone  $C$  such that  $0 \notin B$  and every non-zero element  $v \in C$  is



uniquely of the form  $v = \lambda b$  with  $\lambda > 0$  and  $b \in B$ . Not every cone admits a base ( $C = \mathbb{R}^n \subset \mathbb{R}^n$ , e.g. does not). However, cones which have a compact base are the convex hull of their extremal rays [36]. This will be true for all cones that we will deal with in this paper.

In this sense, it is sufficient to specify the extremal rays in order to specify  $C$ . Thus, cones that have many extremal rays only finitely are of particular interest. A cone has this property if and only if it is the region in  $\mathbb{R}^n$  specified by many finite linear and homogeneous inequalities [35, Chapter 3.4]. Such cones are called polyhedral. The (closure) of all achievable entropy vectors is now known *not* to be polyhedral [38]. However, the cone  $\Gamma_n$  is manifestly defined by many finite inequalities and is hence polyhedral. The same is true for all other cones that we will be working with.

There is a powerful notion of duality for closed convex cones. Let  $C$  be such a cone. The dual cone (also polar cone)  $C^*$  is the set of all homogeneous linear inequalities valid on  $C$ :

$$C^* = \{f \mid \langle f, v \rangle \geq 0 \quad \forall v \in C\}.$$

In this language, the set of Shannon-type inequalities is just the dual cone  $\Gamma_n^*$  to  $\Gamma_n$ . The generating sets in (1) are the extremal rays of  $\Gamma_n^*$ . We will need the following properties of the duality operation.

1. From the Bipolar Theorem,  $(C^*)^* = C$  for every closed convex cone  $C$  [36]. In particular, a cone is completely specified by its duality.
2. Duality reverses inclusions [36]: if  $C, C'$  are closed convex cones and  $C' \subset C$ , then  $C^* \subset (C')^*$ .
3. Dual cones transform ‘contragradiently’: let  $C$  be a closed convex cone and  $D$  a linear map. Then  $C' := D(C)$  is again a convex cone and

$$D^T((C')^*) = \text{range } D^T \cap C^* \subset C^*, \quad (8)$$

where  $D^T$  is the adjoint of  $D$ .

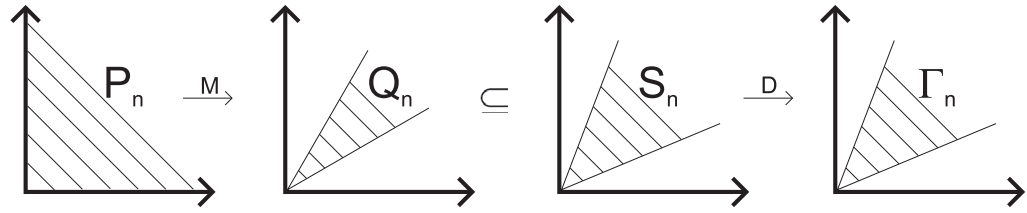
**Proof of property 3.** Let  $C$  be a closed convex cone,  $D$  a linear map, and  $C' = D(C)$ . Then

$$\begin{aligned} D^T((C')^*) &= \{D^T(f) \mid \langle f, v' \rangle \geq 0 \quad \forall v' \in C'\} \\ &= \{D^T(f) \mid \langle f, D(v) \rangle \geq 0 \quad \forall v \in C\} \\ &= \{D^T(f) \mid \langle D^T(f), v \rangle \geq 0 \quad \forall v \in C\} \\ &= \text{range } D^T \cap C^*. \end{aligned}$$

□

#### 4. The correspondence between probabilistic and entropic inequalities

In this section, we will present a simple geometric construction that explains and generalizes the connection, observed above, between the entropic  $CHSH^E$  inequality and the usual  $CHSH$  inequality. We will find that the set of probability distributions for  $n$  binary experiments can be imbedded into the cone  $\Gamma_n$  of set functions fulfilling the polymatroidal axioms. Dually, it follows that every linear inequality valid for  $\Gamma_n$  can be turned into an inequality valid for probability distributions. The linear map that connects the two types of inequalities will turn out to send



**Figure 3.** The various cones appearing in the argument that entropic inequalities can be mapped to Bell inequalities. We start with the positive orthant  $P_n$  of  $R_n$ , which is the cone over the set of probability distributions. The Möbius transform  $M$  sends it linearly and bijectively to a cone  $Q_n$  ( $Q_n^*$ , and all further cones that appear later, happen to be sub-cones of the positive orthant. Thus, in this way, our two-dimensional sketch is faithful). The set of Möbius-transformed distributions is a sub-cone of  $S_n$ , a cone which fulfills a set of ‘inverted’ polymatroidal axioms. The latter cone can be imbedded linearly into the Shannon cone  $\Gamma_n$ . Using cone duality (8), we can invert the chain above and imbed the dual cone  $\Gamma_n^*$  into  $Q_n^*$ . That yields the main claim of this section. Note that the initial Möbius transform is not strictly necessary to arrive at  $D^T(\Gamma_n^*) \subset Q_n^*$ . We have stated this primarily to clarify the geometric nature of  $Q_n$  (i.e. as an orthant, up to a linear isomorphism).

$CHSH^E$  to  $CHSH$ , thus providing a geometric explanation for the observed coincidence. Figure 3 provides a high-level roadmap through the succession of convex cones that appear in the argument.

We start by considering various ways of representing the probability distribution of  $n$  binary random variables  $X_1, \dots, X_n$ . Most naturally, the distribution is given by a function

$$p: \{0, 1\}^n \rightarrow [0, 1]$$

on binary strings of length  $n$  with the interpretation that

$$p(x) = \Pr[X_1 = x_1, \dots, X_n = x_n].$$

Let  $x$  be an  $n$ -bit string. The string is obviously characterized by the set  $A \in 2^{[n]}$  of the positions where it equals 0. Hence, we can equivalently consider  $p$  as a function on the set of subsets of  $[n]$ :

$$\begin{aligned} p: 2^{[n]} &\rightarrow [0, 1], \\ p(A) &= \Pr[X_A = 0 \wedge X_{A^c} = 1], \end{aligned}$$

where, again,  $X_A$  are those components of the random vector  $X$  whose indices appear in the set  $A$ . With this convention,  $p$  can be seen as an element of the vector space

$$R_n = \{f \mid f: 2^{[n]} \rightarrow \mathbb{R}\}$$

of functions on the subsets of  $[n]$ . This vector space comes with a natural basis  $e_A$ ,  $A \in 2^{[n]}$  of ‘delta functions’

$$e_A(B) = \begin{cases} 1 & A = B \\ 0 & A \neq B \end{cases}$$

on a specific subset of  $\{1, \dots, n\}$ . With respect to this basis, every probability distribution  $p$  is an element of the non-negative orthant of  $R_n$ , and every element of that orthant corresponds to a (not necessarily normalized) distribution. We denote the non-negative orthant of  $R_n$  by

$$P_n := \{p \in R_n \mid p(A) \geq 0 \quad \forall A \in 2^{[n]}\}.$$

The reason we found it necessary to elaborate on this rather straightforward correspondence is that the *CHSH* inequality (7) is given in terms of a different parametrization of probability distributions, which we can now explicitly connect to the standard one. Indeed, the quantities appearing in (7) are these:

$$\begin{aligned} q: 2^{[n]} &\rightarrow [0, 1], \\ q(A) &= \Pr[X_A = 0] = \sum_{B \mid A \subseteq B} p(B). \end{aligned} \quad (9)$$

Equation (9) defines a linear map  $M: R_n \rightarrow R_n$  such that  $q = Mp$ . *A priori*, it is not clear that  $M$  is invertible, i.e. that one can specify a distribution in terms of the ‘ $q$ -vector’ above. However, that turns out to be true. In essence, the relation is given by the Möbius inversion formula [39].

**Proposition 3.** [39, equations (6.10, 6.11)] *Let  $F$  be a function on the set of subsets of  $[n]$ . Define  $G$  by*

$$G(A) = \sum_{B \mid B \subseteq A} F(B), \quad (10)$$

*for  $A \subseteq [n]$ . Then the Möbius inversion formula*

$$F(A) = \sum_{B \mid B \subseteq A} (-1)^{|A|-|B|} G(B) \quad (11)$$

*holds.*

Applied to our scenario, we obtain the following.

**Lemma 4.** *The linear map  $M$  defined by*

$$q(A) := (Mp)(A) = \sum_{B \mid A \subseteq B} p(B)$$

*is invertible. Its inverse is given by*

$$p(A) := (M^{-1}q)(A) = \sum_{B \mid A \subseteq B} (-1)^{|A|-|B|} q(B).$$

**Proof.** A few manipulations bring the problem into the standard form of the Möbius transformation. Starting from (9):

$$\begin{aligned} q(A^C) &= \sum_{B \mid A^C \subseteq B} p(B) \\ &= \sum_{B \mid B^C \subseteq A} p(B) \end{aligned} \quad (12)$$

$$= \sum_{B \mid B \subseteq A} p(B^C), \quad (13)$$

where in (12) we only rephrased the condition and in (13) we relabeled ( $B^C \mapsto B$ ). The superscript  $C$  stands, of course, for the set complement within  $[n]$ . With the definitions

$G(A) = q(A^C)$  and  $F(B) = p(B^C)$  this is exactly (10) and thus we can apply the Möbius inversion formula (11).

That way, we can calculate  $p(A)$  and find

$$p(A) = F(A^C) \quad (14)$$

$$= \sum_{B \mid B \subseteq A^C} (-1)^{|A^C| - |B|} G(B) \quad (15)$$

$$= \sum_{B \mid B \subseteq A^C} (-1)^{|A^C| - |B|} q(B^C) \quad (16)$$

$$= \sum_{B \mid B^C \subseteq A^C} (-1)^{|A^C| - |B^C|} q(B) \quad (17)$$

$$= \sum_{B \mid A \subseteq B} (-1)^{|A| - |B|} q(B), \quad (18)$$

which is the stated relation. Here, (14) and (16) are due to the definitions, (15) makes use of the Möbius inversion, in (17) we relabeled ( $B^C \mapsto B$ ) and (18) rephrases the condition and makes use of the fact  $(-1)^{|A| - |B|} = (-1)^{|A^C| - |B^C|}$ .  $\square$

The set of non-negative distributions in  $q$ -representation is thus the Möbius transform of the non-negative orthant. We denote it by

$$Q_n := M(P_n) = \{Mp \mid p \in P_n\}.$$

The significance of  $Q_n$  is that its elements fulfill a set of ‘inverted’ polymatroid axioms. In order to state this precisely, we have to introduce yet another (and final!) cone. In what follows, we will usually switch from a ‘function notation’ to a ‘vector notation’ for elements of  $R_n$ , i.e. we will write  $f_A$  for  $f(A)$ . This will improve readability later on.

**Definition and lemma 5.** *The cone  $S_n$  is the set of vectors  $s \in R_n$  that are*

1. *non-negative*

$$s_A \geq 0 \quad \forall A \in 2^{[n]},$$

2. *decreasing*

$$s_B \geq s_A \text{ for } B \subseteq A,$$

3. *super-modular*

$$h_{A \cup B} + h_{A \cap B} \geq h_A + h_B.$$

It holds that  $Q_n \subset S_n$ .

**Proof.** Let  $q = Mp$  be the Möbius transform of probability distributions. We will verify the properties 1–3 in turn. Since they are obviously invariant under re-scaling by a positive number, this suffices to conclude  $Q_n \subset S_n$ .

Positivity follows directly from the definition of a probability. Property (2) is likewise a straightforward consequence of (9): If  $B \subset A$ , then the probability that all  $X_B$  are

simultaneously zero is certainly larger than or equal to the probability that all  $X_A$  are equal to zero.

As for super-modularity: for any event  $E$ , let  $\delta(E)$  be the ‘indicator function’ that takes the value 1 if  $E$  occurs and else 0. The inequality

$$\delta(X_{A \cup B} = 0) + \delta(X_{A \cap B} = 0) \geq \delta(X_A = 0) + \delta(X_B = 0)$$

holds with probability one. Indeed, as soon as one of the terms on the right-hand side (rhs) is one,  $\delta(X_{A \cap B})$  will also be one; if both terms on the rhs are one, then so are both summands on the lhs. Super-modularity now follows from taking expectations on both sides.  $\square$

The remainder of the argument will proceed as follows: we observe that there is a linear map  $D$  that sends  $S_n$  onto  $\Gamma_n$ . It then follows from elementary convex geometry (section 3) that the dual map  $D^T$  sends linear inequalities valid on  $\Gamma_n$  (i.e. Shannon-type inequalities) to linear inequalities valid on  $S_n$ . Since  $Q_n \subset S_n$ , the inequalities also hold for Möbius-transformed probability distributions. The following statements make this precise.

**Lemma 6.** *Let  $D: R_n \rightarrow R_n$  be defined by*

$$(Ds)_A = s_\emptyset - s_A.$$

*Then  $D(S_n) \subset \Gamma_n$ .*

**Proof.** Let  $A \subseteq B$  and  $s \in S_n$ , then the inequality

$$D(s)_A = s_\emptyset - s_A \leq s_\emptyset - s_B = D(s)_B \quad (19)$$

follows from the fact that vectors  $s$  in  $S_n$  have decreasing components. The inequality

$$D(s)_A + D(s)_B \quad (20)$$

$$= 2s_\emptyset - [s_A + s_B]$$

$$\geq 2s_\emptyset - [s_{A \cup B} + s_{A \cap B}]$$

$$= D(s)_{A \cup B} + D(s)_{A \cap B} \quad (21)$$

follows from super-modularity of  $s \in S_n$ . Next,

$$D(s)_A = s_\emptyset - s_A \geq 0 \quad (22)$$

follows from the fact that  $s$  is decreasing. Finally

$$D(s)_\emptyset = s_\emptyset - s_\emptyset = 0. \quad (23)$$

Inequalities (19), (20), (22) and (23) show, respectively, that  $D(s)$  is monotonously increasing, sub-modular, non-negative and that its  $\emptyset$ -component is zero. These are the defining properties of the Shannon-cone. Hence,  $D(s) \in \Gamma_n$ .  $\square$

We thus find that any Shannon-type inequality can be mapped to an inequality valid for any Möbius-transformed probability distribution.

**Corollary 7.** *Let  $\mathcal{M}$  be a marginal scenario and let  $f \in (\Gamma^{\mathcal{M}})^*$  be a Shannon-type inequality. Then*

$$D^T(f) \in Q_n^*,$$

i.e.  $D^T(f)$  holds for Möbius-transformed probability distributions.

**Proof.** We combine properties 2 and 3 of cone duality as stated in section 3 with lemmas 5, 6 to obtain

$$D^T(\Gamma_n^*) \subseteq S_n^* \subset Q_n^*.$$

Since  $(\Gamma^M)^* \subset \Gamma_n^*$  for any marginal scenario  $M$ , we are done.  $\square$

#### 4.1. Discussion

We return to the relation of the entropic and the usual CHSH inequalities. To this end, it is fruitful to express the map  $D$  with respect to the basis  $\{e_A\}_{A \subset [n]}$  introduced above. If the elements of the basis are ordered in any way starting with  $e_\emptyset$ , it takes the form

$$D = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \dots \\ 1 & -1 & 0 & 0 & 0 & \dots \\ 1 & 0 & -1 & 0 & 0 & \dots \\ 1 & 0 & 0 & -1 & 0 & \dots \\ \vdots & & & & \ddots & \end{pmatrix} \quad (24)$$

from which we can read off the transpose

$$D^T = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & \dots \\ 0 & -1 & 0 & 0 & 0 & \dots \\ 0 & 0 & -1 & 0 & 0 & \dots \\ 0 & 0 & 0 & -1 & 0 & \dots \\ \vdots & & & & \ddots & \end{pmatrix}. \quad (25)$$

The entropic  $CHSH^E$  inequality (2) can be expressed as  $\langle f, v \rangle \geq 0$  for all  $v \in \Gamma_4$  with

$$f := e_{\{A_0, B_0\}} + e_{\{A_0, B_1\}} + e_{\{A_1, B_0\}} - e_{\{A_1, B_1\}} - e_{\{A_0\}} - e_{\{B_0\}}.$$

These explicit representations of  $CHSH^E$  and  $D^T$  make it simple to compute  $D^T f$ , which, by corollary 7 is a vector defining an inequality valid for Möbius-transformed probabilities. Using  $(D^T f)_\emptyset = 0$ , we find

$$D^T f = -f$$

which is the usual  $CHSH$ -inequality as defined in (7). We have thus indeed geometrically explained the coincidence observed initially.

We remark that the inclusion  $D^T(\Gamma_n^*) \subset Q_n^*$  is not tight in general, i.e. it is not the case that all inequalities for  $Q_n$  can be obtained from those of  $\Gamma_n$ . Geometrically, this would be surprising, as  $Q_n$  is just an orthant, while  $\Gamma_n$  seems to be a more complicated geometrical object. It is indeed simple to find explicit counter-examples.

Consider a specific inequality, for instance the Mermin inequality for tripartite correlations [40], and it is possible to gain a better intuition. This is an example of an inequality that is valid on  $q$  functions but can not be translated into an entropic inequality. The reason is that for its derivation, 32 independent inequalities (arising from the positivity of some probability distribution) are needed; one being the positivity of  $q_\Omega \equiv q_{\{A_0, A_1, B_0, B_1, C_0, C_1\}}$ , three corresponding to the decreasing

property, six super-modularities, and another 22 inequalities that cannot be translated into Shannon-type inequalities.

## 5. CG entropic inequalities with and without bounded shared randomness

In this section we derive an entropic version of the CG inequalities [27], concerning a bipartite scenario where each party, say Alice and Bob, can choose between  $m$  measurement settings. We further derive a different version of these inequalities that take into account a bounded amount of shared randomness between the parties.

The CG inequalities are typically written in the form  $I_{mm22} \leq 0$ , where for  $m = 2$  this corresponds to the CHSH inequality [33]. The notation of the inequality  $I_{mm22}$  stresses that each party has access to  $m$  possible measurement settings with two outcomes each. For  $m = 3$  it has been shown that these inequalities are useful since they can detect the non-locality of states that can not be detected by the CHSH inequality [27]. Moreover, as shown in [41], the  $I_{mm22}$  are tight Bell inequalities, that is, they correspond to facets of the local polytope.

The  $I_{mm22}$  inequality can be written compactly using the following matrix notation

$$\begin{pmatrix} q_{A_0} & q_{A_1} & \cdots & q_{A_{m-1}} \\ q_{B_0} & q_{A_0 B_0} & q_{A_1 B_0} & \cdots & q_{A_{m-1} B_0} \\ q_{B_1} & q_{A_0 B_1} & q_{A_1 B_1} & \cdots & q_{A_{m-1} B_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{B_{m-1}} & q_{A_0 B_{m-1}} & q_{A_1 B_{m-1}} & \cdots & q_{A_{m-1} B_{m-1}} \end{pmatrix}. \quad (26)$$

Using this matrix notation, the  $I_{mm22}$  inequality can be written as

$$I_{mm22} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -(m-1) & 1 & 1 & \cdots & 1 & 1 & 1 \\ -(m-2) & 1 & 1 & \cdots & 1 & 1 & -1 \\ -(m-3) & 1 & 1 & \cdots & 1 & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \end{pmatrix} \leq 0. \quad (27)$$

It is important to stress the difference between the way one proves the validity of an entropic inequality and the validity of a probability inequality. In general, to prove that a probability inequality is valid, one uses the information about the extreme points of the local polytope, that is, all the deterministic functions assigning values to the outcomes. In turn, as stressed before, little is known about the extremal rays of the Shannon-type entropic cone (apart from simple cases [28–30]). In the absence of information about the extremal rays of the entropic cone, the only way we can prove that the entropic inequality is valid is to use the linear program approach of Yeung [15]. If the extremal rays are known, a very similar approach to the one used in correlation polytopes [10] can also be used in the entropic case (see appendix B for further details).

We start by considering the case  $m = 3$ . From theorem 6 and the corresponding translation rule  $H_{A_i B_j} \rightarrow -q_{A_i B_j}$ , one could expect that the entropic analogous of  $I_{3322}$ , that we label as  $I_{33}^E$ , could be given by



$$\begin{aligned}
I_{33}^E = & H_{A_0B_0} + H_{A_1B_0} + H_{A_2B_0} + H_{A_0B_1} \\
& + H_{A_1B_1} - H_{A_2B_1} + H_{A_0B_2} - H_{A_1B_2} \\
& - 2H_{B_0} - H_{B_1} - H_{A_0} \geq 0
\end{aligned} \tag{28}$$

This is indeed the case, as this inequality can be obtained by the combination of the following basic inequalities

$$H_{A_0A_1B_0} + H_{A_0A_1B_1} \geq H_{A_0A_1B_0B_1} + H_{A_0A_1} \tag{29}$$

$$H_{A_0B_0} + H_{A_1B_0} \geq H_{A_0A_1B_0} + H_{B_0} \tag{30}$$

$$H_{A_0B_1} + H_{A_1B_1} \geq H_{A_0A_1B_1} + H_{B_1} \tag{31}$$

$$H_{A_2B_0} + H_{B_0B_1} \geq H_{A_2B_0B_1} + H_{B_0} \tag{32}$$

$$H_{A_0B_2} + H_{A_0A_1} \geq H_{A_0A_1B_2} + H_{A_0} \tag{33}$$

together with the following monotonicity inequalities  $H_{A_2B_0B_1} \geq H_{A_2B_1}$ ,  $H_{A_0A_1B_2} \geq H_{A_1B_2}$  and  $H_{A_0A_1B_0B_1} \geq H_{B_0B_1}$  (remember that all monotonicity inequalities can be obtained by the basic inequalities). The notation of the inequality  $I_{mm}^E$  stresses that each party  $A$  and  $B$  has access to  $m$  possible measurement settings with any number of possible outcomes, in contrast to the  $I_{mm22}$  inequalities that are only valid for dichotomic observables. As discussed in the introduction, this outcome size independence is an advantage of the entropic inequalities over the probabilistic ones.

In appendix C it is proven, proceeding with a similar FM elimination as the one sketched above, that the CG inequalities are valid for entropies if one simply applies the transformation rule  $H_{A_iB_j} \rightarrow -q_{A_iB_j}$ , that is,

$$I_{mm}^E = \begin{pmatrix} & -1 & 0 & 0 & 0 & 0 & 0 \\ - (m-1) & 1 & 1 & \dots & 1 & 1 & 1 \\ - (m-2) & 1 & 1 & \dots & 1 & 1 & -1 \\ - (m-3) & 1 & 1 & \dots & 1 & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & -1 & 0 & \dots & 0 & 0 \end{pmatrix} \geq 0, \tag{34}$$

where we have used a similar notation to the one in (26) (replacing  $q_{A_0B_0}$  in the notation of (26) with  $H_{A_0B_0}$ , and similarly for the marginals). From theorem 6, this also implies that the  $I_{mm22}$  inequalities (27) can be derived by relying exclusively on the inverse polymatroidal axioms.

Given the inequality (34), the first question one needs to answer is if it is able to witness non-local correlations. For the usual CG inequality (27), the maximal violation is achieved by the non-local non-signalling distribution

$$p_m(a, b|x, y) = \begin{cases} 1/2, & a \oplus b = 1, x + y = m \\ 1/2, & a \oplus b = 0, x + y \neq m, \\ 0, & \text{otherwise} \end{cases} \tag{35}$$

that can be understood as a generalization of the paradigmatic PR-box [42] for  $m$  measurement settings. If we directly compute the value of  $I_{mm}^E$  for the distribution  $p_m$ , we find no violations.

This is no surprise since entropies are unable to distinguish between correlations and anti-correlations; for example,  $p_m$  is entropically equivalent to the classically correlated distribution

$$p_c(a, b|x, y) = \begin{cases} 1/2, & a \oplus b = 0 \\ 0, & \text{otherwise} \end{cases}. \quad (36)$$

In order to find violations of the entropic inequalities, one needs to find a way of entropically distinguishing correlations from anti-correlations. As shown in [16], one way to do that is to make use of the shared randomness between the parties. Consider two distributions  $p_c$  and  $p_a$  that have, respectively, correlated outputs ( $a \oplus b = 0$ ) and anti-correlated outputs ( $a \oplus b = 1$ ), whatever the inputs. Entropically, both distributions are indistinguishable, but if we allow the parties to make use of some extra shared randomness, then we can tell both distributions apart. For example, mixing the distributions with an independent copy of  $p_c$  with equal probabilities, we see that  $p_c$  remains unchanged while  $p_a$  is turned into an uncorrelated distribution. Similarly if we mix  $\frac{1}{2}p_m + \frac{1}{2}p_c$ , we see that  $I_{mm}^E = m - 1$ , a violation of the entropic inequality that can be proven to be optimal, that is, in some sense (allowing the use of shared randomness) the maximally non-local probability distribution is also the maximally entropically non-local one.

To prove the maximal violation of  $I_{mm}^E \leq 0$ , we first consider the maximum algebraic value that the operator  $I_{mm}^E$  can achieve, which turns out to be the same as the one obtained with the distribution  $\frac{1}{2}p_m + \frac{1}{2}p_c$ . In order to understand the maximal violations  $I_{mm}^E$ , let us rewrite it in an equivalent form  $\tilde{I}_{mm}^E$  where we replace the entropy  $H_{A_i B_k}$  with the mutual information  $I_{A_i:B_k}$ :

$$\tilde{I}_{mm}^E = \begin{pmatrix} & -(m-1) & -(m-2) & -(m-3) & \cdots & -1 & 0 \\ -1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ 0 & 1 & 1 & \cdots & 1 & 1 & -1 \\ 0 & 1 & 1 & \cdots & 1 & -1 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \end{pmatrix} \leq 0, \quad (37)$$

and where we have used the matrix notation

$$\begin{pmatrix} & H_{A_0} & H_{A_1} & \cdots & H_{A_{m-1}} \\ H_{B_0} & I_{A_0:B_0} & I_{A_1:B_0} & \cdots & I_{A_{m-1}:B_0} \\ H_{B_1} & I_{A_0:B_1} & I_{A_1:B_1} & \cdots & I_{A_{m-1}:B_1} \\ \vdots & \vdots & \vdots & \ddots & \cdots \\ H_{B_{m-1}} & I_{A_0:B_{m-1}} & I_{A_1:B_{m-1}} & \cdots & I_{A_{m-1}:B_{m-1}} \end{pmatrix}. \quad (38)$$

Using  $I_{A_i B_j} - H_{A_i} \leq 0$  and  $I_{A_0 B_0} - H_{B_0} \leq 0$ , we see that the maximum value of  $\tilde{I}_{mm}^E$  is given by  $\tilde{I}_{mm}^E = -\sum_{i=0, \dots, m-2} I_{A_{1+i}:B_{m-1-i}} + \sum_{i=0, \dots, m-2} I_{A_{1+i}:B_{m-2-i}} \leq \sum_{i=0, \dots, m-2} H_{A_{1+i}}$ . For dichotomic observables, it turns out that the maximal violation is given by  $\tilde{I}_{mm}^E \leq m - 1$ .

### 5.1. Entropic CG inequality with bounded shared randomness

With the locality and realism assumption, any correlation displayed between  $A$  and  $B$  can only occur through the hidden variable  $\lambda$ . The variable  $\lambda$  is the common ancestor to all the observable quantities. This means that any correlation shown between  $A$  and  $B$  must be screened off if we know the actual value of  $\lambda$ . In other terms,  $\lambda$  provides all the available shared randomness between the distant parties. Mathematically this corresponds to saying that  $I_{A_0 A_1 : B_0 B_1 | \lambda} = 0$  (similarly for all subsets, for example,  $I_{A_0 : B_0 | \lambda} = 0$ ). Or, in other terms,  $H_{A_0 A_1 B_0 B_1 | \lambda} = H_{A_0 A_1 | \lambda} + H_{B_0 B_1 | \lambda}$  which can be rewritten as  $H_{A_0 A_1 B_0 B_1 | \lambda} + H_\lambda = H_{A_0 A_1 \lambda} + H_{B_0 B_1 \lambda}$ . Remember that the mutual information between two variables can be expressed in terms of Shannon entropies as  $I_{A:B} = H_A + H_B - H_{AB}$  and similarly  $I_{A:B|\lambda} = H_{A\lambda} + H_{B\lambda} - H_{AB\lambda} - H_\lambda$ . Note that we allow the parties to have access to local randomness, that is,  $H_{A_0 A_1 | \lambda}$  is not necessarily equal to 0 (and similarly for  $B$ ). Our aim is to derive the entropic constraints that may result from the knowledge of  $H_\lambda$  (or an upper bound of it).

Following the framework introduced in section 2, we start with all the polymatroidal axioms describing the cone  $\Gamma_{n+1}$ , corresponding to all  $n$  variables of the marginal scenario plus the hidden variable  $\lambda$ . We add to the elemental set all the linear constraints that contain information about the causal structure of the experiment, i.e. those saying that  $\lambda$  is the only common ancestor to all the space-like separated variables. Formally, this means we add to the set of elemental inequalities (1), the following equality constraint

$$I_{A:B|\lambda} = 0 \quad (39)$$

where  $A = (A_0, \dots, A_{m-1})$  and  $B = (B_0, \dots, B_{m-1})$ .

The next stage in our approach corresponds to the marginalization. At this point, we have two options: either we consider  $\mathcal{M}_{\text{unbounded}} = \{ \{A_0, B_0\}, \{A_0, B_1\}, \{A_1, B_0\}, \{A_1, B_1\} \}$  or  $\mathcal{M}_{\text{bounded}} = \{ \{\lambda\}, \{A_0, B_0\}, \{A_0, B_1\}, \{A_1, B_0\}, \{A_1, B_1\} \}$ . The first option corresponds, as we marginalize over  $\lambda$ , to the usual Braunstein–Caves scenario for which the only non-trivial inequality bounding the set of allowed entropies is the inequality (2) (and symmetries of it). In the second case, as we do not marginalize over  $\lambda$ , this means that we may have some information on it, for example an upper bound to the available shared randomness between the parties. Note that all the inequalities describing  $\mathcal{M}_{\text{unbounded}}$ , those that do not exactly depend on  $H_\lambda$ , are also going to be valid for  $\mathcal{M}_{\text{bounded}}$ . However, for  $\mathcal{M}_{\text{bounded}}$ , new inequalities (depending explicitly on  $H_\lambda$ ) are going to appear.

Since (39) is the only inequality that contains the causal structure information of the experiment, any non-trivial inequality depending on the amount of shared randomness should appear as the sum of this inequality and some of the other inequalities. To begin with, we now prove that

$$H_A + H_B \leq H_{AB} + H_\lambda, \quad (40)$$

where again  $A = (A_0, \dots, A_{m-1})$  and  $B = (B_0, \dots, B_{m-1})$ .

To obtain (40), we add the independence condition  $I_{A:B|\lambda} = 0$  to one submodularity inequality, one monotonicity inequality:

$$H_{A\lambda} + H_{B\lambda} = H_{AB\lambda} + H_{\lambda}, \quad (41)$$

$$H_{AB\lambda} + H_A \leq H_{AB} + H_{A\lambda}, \quad (42)$$

$$H_B \leq H_{B\lambda}. \quad (43)$$

Note that in the limit that  $H_{\lambda} = 0$ , since  $H_B + H_A \geq H_{AB}$  this implies that  $H_B + H_A = H_{AB}$ , that is, no correlation between  $A$  and  $B$  is possible, as one should expect. We have checked for the CHSH scenario computationally that (40) is the only extra inequality to the usual elemental set one gets after eliminating all variables that depend on  $\lambda$  (for example  $H_{A_0A_1B_0B_1\lambda}$ ) but  $H_{\lambda}$  itself. We believe this is still the case for scenarios with more measurement settings, but we do not have a formal proof of that.

Note that all the terms appearing in (40) involve non-observable quantities and should then be eliminated. Our approach here is to add basic inequalities in such a way that we eliminate all the non-observable quantities. Combining the following basic inequalities

$$H_{A_1B_1} + H_{A_0A_1B_0B_1} \leq H_{A_1B_0B_1} + H_{A_0A_1B_1} \quad (44)$$

$$H_{A_0} + H_{A_0A_1B_1} \leq H_{A_0A_1} + H_{A_0B_1} \quad (45)$$

$$H_{B_0} + H_{A_1B_0B_1} \leq H_{B_0B_1} + H_{A_1B_0} \quad (46)$$

with inequality (40) we obtain

$$BI_{22}^E = -H_{A_0B_1} - H_{A_1B_0} + H_{A_1B_1} + H_{A_0} + H_{B_0} \leq H_{\lambda}, \quad (47)$$

or in terms of mutual information

$$\tilde{BI}_{22}^E = I_{A_0B_1} + I_{A_1B_0} - I_{A_1B_1} \leq H_{\lambda}, \quad (48)$$

which one can regard as the entropic CHSH accounting for the effects of bounded shared randomness. It can be computationally checked (see section 7) that (47) and (2) are the only non-trivial facets describing the entropic cone of  $\mathcal{M}_{\text{bounded}}$ . A violation of (2) implies non-locality, that is, even with an infinite amount of shared randomness we cannot reproduce the observed statistics with the presumed causal structure. In turn, a violation of (47) (with a certain value for  $H_{\lambda}$ ) means that more shared randomness is required to reproduce the observed correlations. Consider for example the local distribution that assigns random outcomes for  $A$  and  $B$  if  $x \oplus y = 0$  and correlated outcomes otherwise. Clearly this distribution does not violate the BC inequality (2) and  $\tilde{BI}_{22}^E = 2$ , that is, at least two bits of shared randomness ( $H_{\lambda} = 2$ ) are required in this case.

For general  $m$ , as proven in the appendix, the following inequality can be regarded as the entropic CG inequality with bounded shared randomness,

$$BI_{mm}^E = \begin{pmatrix} & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 & -1 \\ m-2 & -1 & -1 & \dots & -1 & -1 & 1 \\ m-3 & -1 & -1 & \dots & -1 & 1 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \end{pmatrix} \leq H_{\lambda}. \quad (49)$$

Where once more we have used a matrix notation similar to the one in (26). In terms of the mutual information, the inequality can be written as (using the matrix notation (38))

$$\tilde{B}I_{mm}^E = \begin{pmatrix} -(m-2) & -(m-3) & -(m-4) & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 1 & 1 & \dots & 1 & 1 & -1 \\ 0 & 1 & 1 & \dots & 1 & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & -1 & 0 & \dots & 0 & 0 \end{pmatrix} \leq H_\lambda. \quad (50)$$

In order to understand the violation of the  $\tilde{B}I_{mm}^E$  and its relation to  $\tilde{I}_{mm}^E$ , we first note that  $\tilde{B}I_{m,m}^E = I_{m-1,m-1}^E + I_{A_{m-1},B_0} - I_{A_{m-1},B_1} + H_{B_0} \leq H_\lambda$ . For any local distribution  $\tilde{I}_{m-1,m-1}^E \leq 0$ , and we have  $I_{A_{m-1},B_0} - I_{A_{m-1},B_1} + H_{B_0} \leq H_{B_0} + \min(H_{A_{m-1}}, H_{B_0})$ , which means that any local distribution is bounded by  $\tilde{B}I_{mm}^E \leq H_{B_0} + \min(H_{A_{m-1}}, H_{B_0})$ . Consider for instance dichotomic observables such that for local distributions  $\tilde{B}I_{mm}^E \leq 2$ . This means that independently of how many measurement settings one employs, the inequality will be saturated resorting to not more than two bits of shared randomness.

## 6. Multipartite scenarios

We start by considering the simplest multipartite scenario, consisting of three parties with two measurement settings each. In terms of the correlation polytope, it is known that there are 46 different classes of inequalities [43]. As we discuss in section 7, the FM elimination method to obtain the entropic inequalities bounding the marginal scenario is too demanding and we were not able to finish the computation.

To circumvent this limitation, we proceed to derive a non-trivial inequality using the chain rule for entropies, a similar approach originally employed to derive the entropic CHSH inequality [19]. Remember that a marginal model in accordance with an LHV description ensures the existence of the joint full probability distribution  $p(a_{x=1}, b_{y=1}, c_{z=1}, a_{x=0}, b_{y=0}, c_{z=0})$ , with  $x, y$ , and  $z$  describing the measurement choices, for example  $x = 0$  corresponds to Alice measuring the observable  $A_0$ . The existence of the joint full distribution in turn implies the existence of the joint full entropy  $H(A_1, B_1, C_1, A_0, B_0, C_0)$ . Using the chain rule for the entropies, we have

$$\begin{aligned} H_{A_1 B_1 C_1 A_0 B_0 C_0} &= H_{A_1 B_1 C_1 A_0 B_0 C_0} + H_{B_1 | C_1 A_0 B_0 C_0} \\ &\quad + H_{C_1 | A_0 B_0 C_0} + H_{A_0 | B_0 C_0} + H_{B_0 | C_0} + H_{C_0}, \end{aligned} \quad (51)$$

that in turns implies

$$\begin{aligned} M_3 &= H_{A_1 B_1 C_1} - H_{A_1 B_0 C_0} - H_{A_0 B_1 C_0} - H_{A_0 B_0 C_1} \\ &\quad - H_{A_0 B_0 C_0} + H_{A_0 B_0} + H_{A_0 C_0} + H_{B_0 C_0} \leq 0, \end{aligned} \quad (52)$$

where we have simply used the monotonicity of the Shannon entropy and the fact that conditioning on a variable cannot increase the entropy, that is,  $H_A \leq H_{AB}$  and  $H_{A|BC} \leq H_{A|C}$ . Note that the chain rule for entropies and the two other aforementioned properties are Shannon-type relations, and as so the inequality  $M_3 \leq 0$  can also be derived from the basic set of inequalities (1). Given the inequality (52) the first question one needs to answer is if the inequality is able to detect tripartite non-local correlations. To show this, we consider the two

kinds of tripartite non-local correlations introduced in [44]:

$$p_1(a, b, c|x, y, z) = \begin{cases} 1/4, & a \oplus b \oplus c = xyz \\ 0, & \text{otherwise} \end{cases}, \quad (53)$$

and

$$p_2(a, b, c|x, y, z) = \begin{cases} 1/4, & a \oplus b \oplus c = xy \oplus xz \oplus yz \\ 0, & \text{otherwise} \end{cases}. \quad (54)$$

If we compute the value  $M_3$  for these distributions we find no violations. As mentioned before, this comes as no surprise since entropies are unable to distinguish between correlations and anti-correlations; for example, both distributions  $p_1$  and  $p_2$  are entropically equivalent to the classically correlated distribution

$$p_c(a, b, c|x, y, z) = \begin{cases} 1/4, & a \oplus b \oplus c = 0 \\ 0, & \text{otherwise} \end{cases}. \quad (55)$$

As discussed before, one way to make the distinction between correlation and anti-correlation from the entropic perspective is to use classical shared randomness. If we just mix the distributions  $p_1$  and  $p_2$  with  $p_c$ , for example equally mixing them with the same probability of  $1/2$ , one can straightforwardly compute the value of the operator to be  $M_3 = 1$  for both distributions, a violation of the inequality (52) that therefore witnesses the non-local behaviour of the distributions.

A nice feature of the inequality (52) is that it can be easily generalized for any number of parties  $n$ . Once more, just making use of the chain rule, the monotonicity of the Shannon entropy and the fact that conditioning on a variable cannot increase the entropy, we arrive at

$$\begin{aligned} M_n &= H_{X_1^1 \dots X_1^n} - H_{X_0^1 \dots X_0^n} - P(H_{X_1^1 X_0^2 \dots X_0^n}) \\ &\quad + P(H_{X_0^1 \dots X_0^{n-1}}) \leq 0, \end{aligned} \quad (56)$$

where now we have used the notation  $X_i^j$  to label the  $i$ th observable of the  $j$ th party with  $i = \{0, 1\}$  and  $j = 1, \dots, n$ . The operator  $P$  stands for all the different permutations of the parties, for example for  $n=3$   $P(H_{X_1^1 X_0^2 X_0^3}) = H_{X_1^1 X_0^2 X_0^3} + H_{X_0^1 X_1^2 X_0^3} + H_{X_0^1 X_0^2 X_1^3}$  and  $P(H_{X_0^1 X_0^2}) = H_{X_0^1 X_0^2} + H_{X_0^1 X_0^3} + H_{X_0^2 X_0^3}$ . It is easy to see that (56) is violated by a generalization of the distribution (53) for more parties, given by

$$p_1^n(x^1, \dots, x^n | X^1, \dots, X^n) = \begin{cases} 1/2^{n-1}, & x^1 \oplus \dots \oplus x^n = X^1 \dots X^n \\ 0, & \text{otherwise} \end{cases}, \quad (57)$$

if we just mix it with the classical correlated distribution ( $x^1 \oplus \dots \oplus x^n = 0$ ).

A nice feature of the entropic inequalities is that they can be readily applied to marginal scenarios with an arbitrary number of outcomes. This is in sharp contrast to the usual Bell inequalities approach, where increasing the number of outcomes also increases the complexity and dimension of the correlation polytopes. To our knowledge, very few inequalities have been derived for marginal multipartite Bell scenarios with many outcomes, in particular in [45] tripartite inequalities have been derived for any number of outcomes, but as the authors stress, there is no straightforward generalization of their methods to more parties. (Also note [46], but

there the inequalities involve products of observables from the same party and therefore have no direct application to Bell scenarios.) Entropic inequalities may be proven as a useful tool in such cases. We have briefly explored this possibility by looking for quantum violations of the inequality (56) using multidimensional GHZ states given by

$$|GHZ_d^n\rangle = \sqrt{\frac{1}{d}} \sum_{i=0, \dots, d-1} |i_1 \dots i_n\rangle \quad (58)$$

and employing the Fourier-transformed measurements used in [47]. We have considered  $n = 3, \dots, 10$  and  $d = 2, \dots, 10$  and found that the violation of (56) increases with both  $n$  and  $d$ . As numerically noted in [45], the maximal quantum violation for the inequalities considered there can be reached only using systems with local Hilbert space dimension exceeding the number of measurement outcomes; which suggests that these kinds of inequalities can be used as multipartite dimension witnesses [48]. We believe this is an interesting line of research one may pursue in the entropic approach.

## 7. Computational results

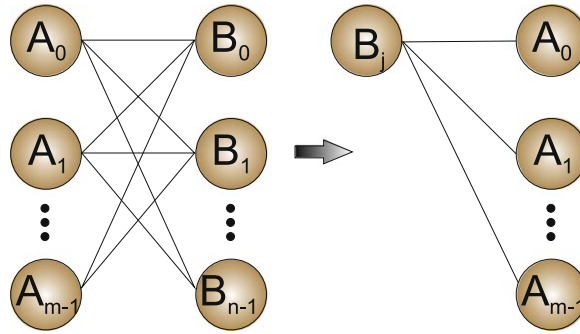
In section 5 we have used a specific combination of the basic inequalities in order to derive the entropic inequalities (37) and (49). However, in principle different combinations could give rise to different classes of entropic inequalities. To understand what other classes of inequalities one gets, we rely in this section on computational results. Using standard software to perform the FM elimination, we computed all classes of entropic inequalities for the simplest marginal models where the computation is expected to finish in a reasonable time.

It turns out that even for very simple scenarios involving more than five variables, the computations are already too large to finish. To go beyond that limitation one needs to further simplify the set of basic inequalities. In order to do that we follow the approach proposed in [49] for usual Bell inequalities. Let us begin by considering a bipartite scenario, with  $m$  measurement settings for Alice and  $n$  for Bob. The existence of a classical description for all pairwise observables is equivalent to the existence of classical descriptions for the  $n$  subsystems,  $\{A_0, \dots, A_{m-1}, B_j\}$  with  $j = 0, \dots, n-1$ , coinciding on  $\{A_0, \dots, A_{m-1}\}$ . To find all the entropic inequalities for the marginal scenario it is then sufficient to start out with the union of the basic set of inequalities defining each one of the  $n$  subsystems (see figure 4). A further simplification is possible since for each of the subsystems (indexed by  $j$ ) it is sufficient to consider only the inequalities involving the subsets of the variables  $\{\{A_i, B_j\}, \{A_0, \dots, A_{m-1}\}\}$  with  $i = 0, \dots, m-1$ . That is, in practice we start with the set of Shannon-type inequalities describing the cone  $\{A_0, \dots, A_{m-1}, B_j\}$  and project it down to the cone describing  $\{\{A_i, B_j\}, \{A_0, \dots, A_{m-1}\}\}$ . With this simplification we were able to fully characterize bipartite marginal models involving up to seven variables, also accounting for the effects of bounded shared randomness.

Using similar simplifications, we also obtain the inequalities for marginal scenarios involving statistical independencies. Details are given in section 7.2.

To characterize the entropic cone of a multipartite marginal model, we can in principle proceed as before, first simplifying the set of basic inequalities. For example, for three parties,





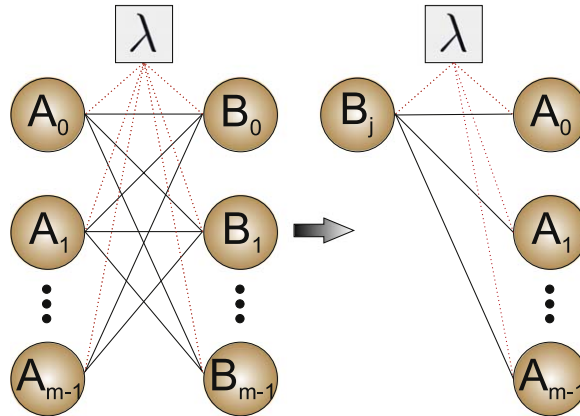
**Figure 4.** Graphical representation (left) of bipartite Bell scenarios: each vertex represents an observable and edge-connected observables that are jointly measurable. The existence of a classical description for all pairwise observables is equivalent to the existence of classical descriptions for the  $n$  subsystems (right),  $\{A_0, \dots, A_{m-1}, B_j\}$  with  $j = 0, \dots, n - 1$ , coinciding on  $\{A_0, \dots, A_{m-1}\}$ .

similar to the bipartite case, we can restrict the initial set of inequalities to the ones describing the following two subsystems,  $\{A_0, A_1, B_0, B_1, C_0\}$  and  $\{A_0, A_1, B_0, B_1, C_1\}$ , where  $A_i, B_j$  and  $C_k$  with  $i, j, k = \{0, 1\}$  describe the measurement choices available to the parties. However, even with that the FM elimination still demanded too many computational resources and we were not able to finish the computation. This highlights the value of analytical derivations like the one in section 6.

### 7.1. Bipartite scenario

In the simplest case, given by the CHSH scenario ( $m_a = m_b = 2$ , that is, two measurement settings for Alice and Bob), it is known [20, 22] that the only class of non-trivial inequalities is given by the entropic CHSH. Using the computational approach described above, it follows that in the case with  $(m_a = 2, m_b = 3)$  there are still only entropic CHSH inequalities, in full analogy with the probabilistic case. However, differences to the probabilistic case already start to appear in the case  $(m_a = m_b = 3)$ . For probabilities, there are only two different classes of non-trivial inequalities, the CHSH and the  $I_{3322}$  inequality [27]. However, for entropies there are four classes of non-trivial tight inequalities, as shown in table A1. Inequalities (3) and (4) correspond, respectively, to the  $I_{22}^E$  and to the  $I_{33}^E$ . By theorem 6 all the inequalities are also valid in the probability space after the proper translation is made, however, inequalities (5) and (6) do not correspond to tight inequalities in the correlation polytope. For the scenario with  $(m_a = 3, m_b = 4)$ , five new classes of inequalities have been found, the inequalities (7) to (11) in table A1.

We have also performed the same computation, but now bounding the amount of shared randomness, following the idea described in section 5.1. First of all we note that as one should expect, the inequalities derived in the absence of any restriction on  $H_\lambda$  still define facets of the entropic cone. For the inequalities depending on  $H_\lambda$  it follows that in the  $(m_a = m_b = 2)$  and  $(m_a = 2, m_b = 3)$  cases, the only inequalities bounding the shared randomness are the ones given in table A2, inequalities (1) to (3). For the case  $(m_a = m_b = 3)$ , it follows that the problem is already too demanding and we were not able to finish the computation. However, a further



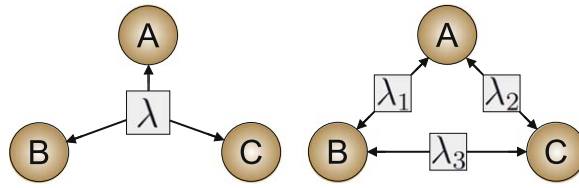
**Figure 5.** Graphical representation (left) of bipartite Bell scenarios: each vertex represents an observable and edge-connect observables that are jointly measurable. The dotted red edges represent the fact that all the correlations between the parties must be mediated by the hidden variable  $\lambda$ , that is,  $I_{A:B|\lambda} = 0$ . The existence of a classical description for all pairwise observables is equivalent to the existence of classical descriptions for the  $n$  subsystems (right),  $\{A_0, \dots, A_{m-1}, B_j, \lambda\}$  with  $j = 0, \dots, n - 1$ , coinciding on  $\{A_0, \dots, A_{m-1}, \lambda\}$ .

simplification is possible as shown in figure 5. We consider the union of the basic inequalities for the sets  $\{A_0, A_1, A_2, B_j, \lambda\}$  and  $\{A_j, B_0, B_1, B_2, \lambda\}$  with  $j = 0, 1, 2$ , coinciding on  $\{A_0, A_1, A_2\}$  and  $\{B_0, B_1, B_2\}$ . For each of these sets of inequalities parameterized by  $j$ , we first project it down to the cones describing  $\{\{A_i, B_j\}, \{A_0, A_1, A_2, \lambda\}\}$  and  $\{\{A_i, B_j\}, \{B_0, B_1, B_2, \lambda\}\}$  with  $i, j = 0, 1, 2$ . Then we proceed to the final FM elimination, eliminating the non-observable variables. With this simplification we were able to finish the computation and 26 new classes of inequalities were found, inequalities (4) to (29) in table A2.

All the inequalities in table A2 have a rather remarkable feature. For all of them it is not difficult to prove that the maximal value achievable by local correlations is given by  $H_{B_0} + \min\{H_{A_0}, H_{B_0}\}$ . That is, up to three measurement settings and considering dichotomic observables, any local distribution needs not more than two bits of shared randomness to be simulated. As we discuss in section 8, it seems improbable that increasing the number of measurement settings, only two bits of shared randomness would suffice to simulate any local distribution, especially the ones arising from entangled states. It would be very interesting to find classes of entropic inequalities that would require in principle more bits of shared randomness to simulate local distributions.

## 7.2. Scenarios with statistical independencies between the hidden variables

Consider three random variables  $A$ ,  $B$  and  $C$ , characterizing for instance some traits of three different languages [17]. From the observed data, one concludes that all three variables are maximally correlated pairwise, for example,  $I_{A:B} = I_{A:C} = I_{B:C} = H_A = H_B = H_C = H_{ABC}$ . Furthermore, no conditional independencies can be inferred from the data. The question is then: are the



**Figure 6.** If the correlations between three random variables are not sufficiently strong, they can be achieved by a causal structure involving no common ancestor to all the variables (right). Sufficiently strong correlations can only be explained by the existence of a common ancestor (left).

observed correlations compatible with a causal structure involving no common ancestor to all three variables (figure 6, right)? Or is a common ancestor needed to explain the data (figure 6, left)? A direct application of causal discovery algorithms [2, 3, 50] would try to distinguish between the two causal structures, but since no conditional independencies are imposed and by the principle of minimality (Occam's razor), the algorithm would return the causal structure on the right as the answer. But clearly this is wrong, because this causal structure implies that (for the observed data) if  $A$  is maximally correlated with  $B$ , then it should be completely uncorrelated with  $C$ .

The entropic approach offers a surprisingly simple solution in this case. It is not difficult to show that the causal structure on the right of figure 6 implies a bound on the correlations given by  $I_{A:B} + I_{A:C} \leq H_A$  (and permutations thereof) [18]. The observed data clearly violates this inequality, meaning that it cannot be explained by the corresponding causal structure. That is, an ancestor common to all three variables is required to explain the observed probability distribution. It is interesting to note that this same scenario has been considered under two very different perspectives, from the purely causal inference one [17] but also from the quantum non-locality point of view [18, 51].

The causal structure depicted on the right of figure 6 implies many statistical independencies. As an example, we have  $I_{\lambda_1:\lambda_2} = I_{\lambda_1:\lambda_3} = I_{\lambda_2:\lambda_3} = 0$  and  $I_{A:Bl\lambda_1} = I_{A:C|l\lambda_2} = I_{B:C|l\lambda_3} = 0$ . Using all available constraints, many of the variables can be eliminated. A final FM elimination gives as a result three different classes of non-trivial entropic inequalities completely characterizing the marginal scenario:

$$H_A + H_B + H_C - H_{AB} - H_{AC} \leq 0 \quad (59)$$

$$3H_A + 3H_B + 3H_C - 3H_{AB} - 2H_{AC} - 2H_{BC} + H_{ABC} \leq 0 \quad (60)$$

$$5H_A + 5H_B + 5H_C - 4H_{AB} - 4H_{AC} - 4H_{BC} + 2H_{ABC} \leq 0 \quad (61)$$

Note that the inequality (59) is exactly the same as the one obtained in [18]. Our derivation shows that indeed this is a tight Shannon-type inequality. However, there are two other inequivalent classes, inequalities (60) and (61) that were not known before. We highlight the fact that, to our knowledge, this is the first complete characterization of the entropic cone in a common ancestor scenario.

Another interesting case is the one of a common ancestor to all the variables, as depicted on the left of figure 6, however, now taking into account the entropy of the common ancestor. In this case we find that the only non-trivial inequality is given by

$$H_{AB} + H_{AC} + H_{BC} - 2H_{ABC} \leq H_\lambda \quad (62)$$

For a distribution fulfilling  $I_{A:B} = I_{A:C} = I_{B:C} = H_A = H_B = H_C = H_{ABC}$ , the maximal value of the expression (62) is  $H_A$ , that is, as expected the distribution needs not more than  $H_A$  bits of shared randomness to be achieved.

## 8. Discussion

In this work we have explored the entropic approach to marginal problems, gathering several results that we believe may pave the way to a better understanding and more systematic application of entropic inequalities in a wide range of applications. In the next paragraphs we summarize and briefly discuss our findings, with special attention given to the open problems and the possibilities that we believe deserve future investigation.

We have shown a correspondence between Shannon-type inequalities and inequalities in the probability space, stating that any Shannon-type inequality is also a valid probabilistic inequality if a very simple translation is made. This correspondence formally explains the similarities observed for the  $n$ -cycle marginal scenario (that has as a particular case the CHSH scenario) [20, 22, 52] between the entropic inequalities and the probabilistic version. For the  $n$ -cycle scenario, all the non-trivial Shannon entropic inequalities have an exact correspondence in the probability space [16, 22], however, this is not true in general, since not all probabilistic inequalities define valid Shannon entropic inequalities, that is, they involve probability inequalities that cannot be translated to a Shannon-type entropic inequality. Also, as mentioned before, not all valid entropic inequalities are of Shannon-type. Could it be that by taking into account non-Shannon-type inequalities, a deeper correspondence between entropic and probabilistic inequalities can be made? The use of non-Shannon-type inequalities is also interesting from a practical perspective, since in principle by taking them into account one may get more restrictive inequalities, bounding the set of allowed correlations more tightly.

Based on the correspondence between entropic and probabilistic inequalities, we analytically proved the entropic version of the CG inequalities [27], valid for a bipartite scenario where each party has access to  $m$  measurements ( $m_a = m_b = m$ ). The entropic inequalities have the advantage of being valid for observables with any number of outcomes, while the CG inequalities are specially tailored for dichotomic ones. Moreover, for the scenarios with  $(m_a = m_b = 2)$ ,  $(m_a = m_b = 3)$  and  $(m_a = 3, m_b = 4)$ , we have computationally (through FM elimination) derived all the entropic inequalities and shown that there are, respectively, (3), (6), (11) inequivalent classes of inequalities. For the computational results, since the FM elimination generally produces a huge list of redundant inequalities as an output, we have also proven a result that allows one to check, given the list of extremal points and half-lines of the convex set, if the inequality corresponds to a facet or not (see appendix B).

We have also considered, for the bipartite case with  $(m_a = m_b = 2)$  and  $(m_a = m_b = 3)$ , the effects of bounded shared randomness and shown that in these cases any local distribution with dichotomic outcomes can be entropically simulated with at most two bits of shared randomness. This is a very interesting point that deserves further investigation. It seems implausible that by increasing the number of measurement settings, any local distribution would still require at most two bits of shared randomness. To further understand that, we have analytically proven a different entropic version of the CG inequalities (for arbitrary  $m$ ), where the effects of bounded shared randomness are taken into account. However, these inequalities still have the surprising property that no more than two bits of shared randomness are necessary to entropically simulate

any local distribution. However, this is only one class of inequalities and there are possibly many more with increasing  $m$ . If one can find other classes of inequalities, one could investigate, for example, the shared randomness requirements needed to simulate the correlations of a Werner state  $\rho_w = \nu|\Phi^+\rangle\langle\Phi^+| + (1 - \nu)\mathbb{I}/4$  parameterized by the visibility  $\nu$  [53]. In the region where the state is known to violate some Bell inequality [54], since the state is non-local it follows that even an infinite amount of shared randomness is not sufficient to reproduce the correlations. There are, however, two interesting regions where not much is known. First, for  $1/3 < \nu < 1/K_G \approx 0.66$ , the state is entangled but local [55]. Since the state is entangled one may expect that more shared randomness would be required, but how much of it? The most interesting region is of course the one for  $1/K_G < \nu < \nu_{\text{ver}} \approx 0.7056$ , where it is not known if the state is non-local or not. What are the shared randomness requirements for correlations obtained in this region?

Working in a generalization of the approach followed in [19], we derived entropic inequalities for multipartite marginal scenarios consisting of any number of parties, each having access to two observables with any possible number of measurement outcomes. Using specific projective measurements, we have numerically shown that the violation of these inequalities for multidimensional multipartite GHZ states increases with both the size and the local dimension of the state. An interesting perspective would be the possible use of these inequalities as multipartite dimension witnesses, similar to what has been suggested in [45].

Finally we have considered a scenario involving conditional independencies, for which the question is to decide if a given correlation for the observable quantities is compatible with a causal structure involving only pairwise common ancestors. A natural question is how to generalize the obtained results to the case of many observable quantities and different configurations of common ancestors [17]. An interesting related problem would be to understand relaxations over the bilocality assumption of entanglement swapping experiments [11], for example, allowing correlations between the hidden variables while keeping the bilocality on the level of the observed quantities. Similarly, one could use entropic inequalities to relax the measurement independence assumption [12, 13], stating that the measurement choice made by the parties is independent of the hidden variable.

## Acknowledgments

It is a pleasure to thank Dominik Janzing for insightful discussions about causal structures. We would also like to thank A Acín and J B Brask for pointing out the potential application of bounded shared randomness inequalities for Werner states, and Elie Wolfe for comments on the manuscript. Our work is supported by the Excellence Initiative of the German Federal and State Governments (Grant ZUK 43) and the Research Innovation fund from the University of Freiburg. The article processing charge was funded by the open access publication fund of the Albert Ludwig University, Freiburg.

## Appendix A. Tables with entropic inequalities

See tables [A1](#) and [A2](#).

**Table A1.** All classes of entropic bipartite inequalities for  $m_a = 2, 3$  and  $m_b = 2, 3, 4$ . We have listed the coefficients of one inequality in each row, and all inequalities are of the form  $\leq 0$ .

Entropic bipartite inequalities																				
# $x / y / xy$	$H(A_x)$				$H(B_y)$				$H(A_x B_y)$											
	0	1	2	3	0	1	2	3	00	01	02	03	10	11	12	13	20	21	22	23
1	1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0
2	-1	0	0	0	-1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	1	0	0	0	1	1	0	1	-1	0	0	0	0	0	0	0
4	1	0	0	0	2	1	0	0	-1	-1	-1	-1	-1	1	-1	1	0	0	0	0
5	1	1	0	0	1	1	0	0	-1	1	-1	0	-1	-1	-1	-1	+1	0	0	0
6	1	1	0	0	2	0	0	0	-1	-1	-1	-1	-1	1	-2	1	0	0	0	0
7	2	1	0	0	1	1	0	0	-1	0	-1	-1	0	-1	-1	1	-1	-1	1	0
8	1	1	0	0	1	1	1	0	-1	-1	1	0	-1	0	-1	-1	0	-1	-1	1
9	1	1	0	0	2	1	1	0	-1	-1	1	-1	-1	1	-1	-1	-1	-1	-1	1
10	2	1	0	0	2	1	0	0	-1	-1	-1	-1	-1	-1	-1	1	-2	1	1	0
11	2	1	0	0	1	1	1	0	-1	-2	0	-1	-1	1	-1	-1	1	-1	-1	1

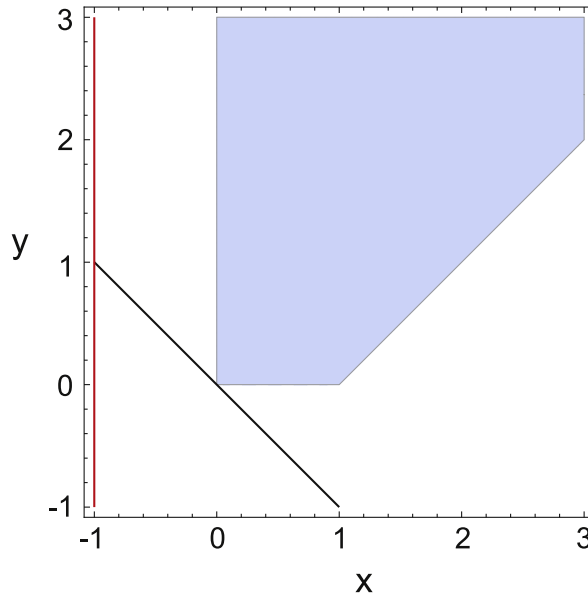
**Table A2.** All classes of entropic bipartite inequalities with bounded shared randomness for  $m_a = 2, 3$  and  $m_b = 2, 3$ . We have listed the coefficients of one inequality in each row, and all inequalities are of the form  $\leq H_\lambda$ .

Entropic bipartite inequalities with bounded shared randomness																
# $x/y/xy$	$H(A_x)$			$H(B_y)$			$H(A_x B_y)$									Bound $H_\lambda$
	0	1	2	0	1	2	00	01	02	10	11	12	20	21	22	
1	1	0	0	1	0	0	-1	0	0	0	0	0	0	0	0	1
2	1	0	0	1	0	0	0	-1	-1	1	0	0	0	0	0	1
3	1	1	0	1	0	0	-1	-1	1	-1	1	-1	0	0	0	1
4	1	1	0	1	0	0	-1	-1	1	0	0	-1	-1	1	0	1
5	1	1	0	1	1	0	-1	1	-1	0	-1	-1	0	-1	1	1
6	2	0	0	1	1	0	-1	0	-2	-1	0	1	1	-1	0	1
7	2	1	0	1	1	0	-1	1	-2	0	-1	-1	-1	-1	2	1
8	2	1	0	1	1	0	-1	0	-2	1	-1	-1	-1	-1	2	1
9	2	1	0	1	1	0	-2	1	-1	1	-1	-1	-1	-1	1	1
10	1	1	1	1	1	0	-1	-2	1	-1	1	-1	1	-1	-1	1
11	2	1	0	1	1	0	-1	1	-2	-1	-1	1	1	-1	0	1
12	2	1	0	1	1	0	-2	1	-1	-2	-1	1	1	-1	0	1
13	2	2	0	2	1	0	-2	1	-2	1	-1	-2	-2	-1	2	1
14	2	1	0	1	1	0	-3	-1	1	-1	1	-1	1	-1	0	1
15	3	1	0	2	1	0	-2	1	-3	1	-1	-1	-2	-1	2	1
16	2	1	0	1	1	0	-1	1	-2	-1	-1	1	1	-1	0	2
17	2	1	0	1	1	0	-2	1	-1	1	-1	-1	0	-1	1	2
18	2	1	0	1	1	0	1	-1	-2	-2	1	0	0	-1	1	2
19	2	2	0	1	1	0	-2	1	-1	1	-1	-2	-1	-1	2	2
20	3	1	0	1	1	0	-2	1	-2	1	-1	-1	-1	-1	2	2
21	2	1	1	1	1	0	-2	1	-1	-2	-1	2	1	-1	-1	2
22	3	1	0	1	1	0	-2	1	-2	-2	1	2	1	-1	0	2
23	3	2	0	2	1	0	1	-2	-2	-2	2	-2	-2	-1	2	2
24	3	1	0	1	1	0	-1	1	-3	1	-1	-1	-1	-1	2	2
25	2	1	1	1	1	0	-1	-3	2	-1	1	-1	1	-1	-1	2
26	2	2	0	1	1	0	-3	-1	2	-1	1	-2	1	-1	0	2
27	3	1	0	1	1	0	-1	1	-3	-2	-1	2	1	-1	0	2
28	3	1	0	2	1	0	1	-1	-3	-1	-1	0	-2	1	1	2
29	3	2	0	2	1	0	-2	2	-3	1	-2	-1	-2	-1	2	2

## Appendix B. Deciding whether an entropic inequality corresponds to a facet of the entropic cone

Every convex set can be expressed in terms of dual representations, either in terms of extremal points and half-lines or in terms of inequalities (half-spaces) defining the facets of the convex set. As discussed in section 2, not much is known about the extremal half-lines/points of the Shannon-type entropic cone, and in order to derive entropic inequalities for a given marginal scenario, one needs to rely on the FM elimination. One problem that arises is that after performing the FM elimination, the set of inequalities will usually contain many (for the





**Figure 7.** The convex set (blue color) generated by the half-lines  $x \geq 0$ ,  $y \geq 0$  and  $-x + y + 1 \geq 0$ , where each of these inequalities generates a facet of the set. The inequalities  $x + 1 \geq 0$  and  $x + y \geq 0$  are indicated, respectively, in red and black, and it is clear that while valid (since they can be obtained as a linear combination of the generating set), they are not facets of the proposed convex set.

scenarios we consider in section 7, typically several thousands) redundant inequalities not corresponding to facets of the cone. That is, among the huge list of inequalities obtained via the FM elimination, we need to find the minimal set of inequalities describing the marginal scenario, the ones corresponding to facets of the entropic cone. One way to find this minimal set of inequalities is to solve a linear problem, i.e. whenever a given inequality can be expressed as a linear combination (with positive coefficients) of other inequalities, it can be safely eliminated. However, given the typical case we face, of sets containing a huge number of inequalities, this approach soon becomes unfeasible. Notwithstanding the difficulty in characterizing the extremal rays/points of the Shannon-type entropic cone, for most of the marginal scenarios we consider computationally, we were also able to get a list of them. In order to derive the minimal set of inequalities and further understand the structure of the entropic cones, we rely instead on the information provided by the extremal points and half-lines.

Given the extreme points, extreme directions and a list of inequalities satisfied by all points of some polyhedral set, it is easy to decide which of these inequalities belongs to a minimal list characterizing the polyhedral set.

Before we prove this fact in general let us first look at the example of the two-dimensional unbounded closed polyhedral set in two-dimensional Euclidean space that is defined by the inequalities

$$x \geq 0, y \geq 0 \text{ and } -x + y + 1 \geq 0 \quad (\text{B.1})$$

displayed in figure 7. Its extreme points are  $p_1 = (0, 0)$  and  $p_2 = (1, 0)$  and the extreme directions are  $v_1 = (0, 1)$  and  $v_2 = (1, 1)$ . Given a list of the three defining inequalities plus, lets

say, the valid inequalities  $x + 1 \geq 0$  and  $x + y \geq 0$ , we want to reduce this to a minimal list containing only the three defining inequalities.

In this two-dimensional example, one can guess that a necessary and sufficient condition for a valid inequality belonging to the minimal list is that it is saturated by a one-dimensional subset of the polyhedral set.

We start with  $x \geq 0$  and see that it is saturated by  $p_1$  and the halfline  $p_1 + \lambda v_1$ ,  $\lambda \geq 0$ . The intersection of our polyhedral set with all points in  $\mathbb{R}^2$  saturating the inequality is therefore the halfline  $p_1 + \lambda v_1$ . Its dimension is one and thus the inequality belongs to the minimal list.

Consider now the inequality  $y \geq 0$ . We find  $p_1$  and  $p_2$  saturating it, and thus the one-dimensional object  $\text{conv}\{p_1, p_2\}$  is the set of points saturating the inequality. The third defining inequality is also saturated by a one-dimensional subset of our set, namely the halfline  $p_2 + \lambda v_2$ ,  $\lambda \geq 0$ .

In turn, the inequality  $x + y \geq 0$  is only saturated by  $p_1$ . So we found a zero-dimensional subset and disregard this inequality. The last inequality  $x + 1 \geq 0$  is not saturated by any point, so we found the  $\emptyset$  and also disregard this inequality. Clearly, the inequalities not belonging to the minimal list are already implied by those belonging to the minimal list.

To begin with, the general case we note that every closed convex set  $P$  in  $\mathbb{R}^n$  that contains no lines is the convex hull of its extreme points and extreme half-lines (corollary 2.6.15 of [56]). In our case, however, after the FM elimination performed with PORTA [57], it is not the extreme points and extreme half-lines we are given, but the extreme points and extreme directions. It is obvious that the set containing all half-lines arising from the combination of all extreme points with all extreme directions contains the set of all extreme half-lines, as a half-line can not be extreme if it arises from the combination of a non-extreme point with an extreme direction. Formally speaking, if  $x = \alpha x_1 + \beta x_2$  with  $\alpha, \beta \geq 0$  and  $\alpha + \beta = 1$ , then  $hl = x + \lambda y = \alpha(x_1 + \lambda y) + \beta(x_2 + \lambda y) = \alpha hl_1 + \beta hl_2$ , with extreme points  $x_1, x_2$ , extreme direction  $y$  and extreme half-lines  $hl_1, hl_2$ .

Every inequality  $\langle x | h_i \rangle \leq C_i$  corresponds to a hyper-plane  $H_i = \{x | \langle x | h_i \rangle = C_i\}$  that divides the space into two closed half-spaces

$$H_i^\pm = \{x | \langle x | h_i \rangle \leq \mp C_i\} \text{ with } P \subseteq H_i^- \text{ for every } i \quad (\text{B.2})$$

**Definition 8.** Let  $P$  be a closed convex set with dimension  $r$  and  $H$  some hyperplane with  $P \subseteq H^-$ , then  $F := P \cap H$  is called an exposed face. An exposed face with dimension  $r - 1$  is called a facet.

Note that the name *exposed face* is justified by the fact that  $F$  indeed is a face of  $P$ .

With the following lemma it is easy to check whether an inequality corresponds to a facet.

**Lemma 9.** Let  $P = \text{conv}\{A\}$  with  $A$  being the set of extreme points and extreme half-lines of  $P$  and  $H$  some hyperplane with  $P \subseteq H^-$  then  $F := P \cap H = \text{conv}\{A \cap H\}$ .

**Proof.**

$$\begin{aligned}
 F &= P \cap H = \text{conv} \{A\} \cap H = \\
 &\left\{ x \left| x = \sum_k \alpha_k y_k, \left\langle \sum_k \alpha_k y_k \right| h \right\rangle = C, \right. \\
 &\left. \alpha_k \geq 0, \sum_k \alpha_k = 1, y_k \in A \right\}
 \end{aligned} \tag{B.3}$$

We know that  $P \subseteq H^-$ , and so  $A \subseteq H^-$ , which means that  $\langle y_k | h \rangle \leq C$  for every  $k$ . Thus the condition

$$\left\langle \sum_k \alpha_k y_k \right| h \rangle = C \tag{B.4}$$

can only be fulfilled for every convex combination if  $\langle y_k | h \rangle = C$  for all  $k$ . And so

$$\begin{aligned}
 F &= P \cap H = \text{conv} \{A\} \cap H = \{x | x = \\
 &\sum_k \alpha_k y_k, \langle y_k | h \rangle = C, \alpha_k \geq 0, \\
 &\sum_k \alpha_k = 1, y_k \in A\} = \text{conv} \{A \cap H\}
 \end{aligned} \tag{B.5}$$

□

For every inequality it is easy to find the set  $A \cap H = \{x | \langle x | h \rangle = C, x \in A\}$ , i.e. the set of all extreme points and half-lines of  $P$  that saturate the inequality.

Still we have to check the dimension of  $F$ . It is known (theorem 4.1.3 of [56]) that  $F$  as the convex combination of its extreme points and extreme half-lines can be written as the Minkowsky sum of the convex combination of its extreme points and the cone of its extreme directions:

$$\begin{aligned}
 F &= \left\{ x \left| x = \sum_{i=1}^n \lambda_i p_i + \sum_{i=1}^m \mu_i v_i \right. \right\} \\
 &= \left\{ x \left| x = p_1 + \sum_{i=2}^n \lambda_i (p_i - p_1) + \sum_{i=1}^m \mu_i v_i \right. \right\}
 \end{aligned} \tag{B.6}$$

with  $\sum_{i=1}^n \lambda_i = 1$ ,  $\lambda_i \geq 0$ ,  $\mu_i \geq 0$ ,  $p_i$  being the  $n$  extreme points and  $v_i$  the  $m$  extreme directions.

We see that  $F$  is some linear combination of the vectors  $(p_i - p_1)$  and  $v_i$ . The dimension of  $F$  is thus equal to the rank of the matrix with columns  $(p_i - p_1)$  and  $v_i$ .

In summary, what one has to do is to find all extreme points and directions that saturate a given inequality and calculate the rank of this matrix. If and only if it is equal to  $(r - 1)$ , the inequality induces a facet.

### Appendix C. Proving the $I_{mm}^E \leq 0$

The proof will consist of two main steps. In the first step we will show how to obtain all the terms of the first two rows and the terms of the last row plus the marginal  $H(A_0)$  in (34). We begin with the SM (sub-modularity) inequality

$$H_{A_0 A_1 \dots A_{m-2} B_0} + H_{A_0 A_1 \dots A_{m-2} B_1} \geq H_{A_0 A_1 \dots A_{m-2} B_0 B_1} + H_{A_0 A_1 \dots A_{m-2}}. \quad (C.1)$$

We then cancel the two NO (non-observable) terms on the lhs by adding the following SM inequalities

$$H_{A_0 B_0} + H_{A_1 A_2 \dots A_{m-2} B_0} \geq H_{A_0 A_1 \dots A_{m-2} B_0} + H_{B_0} \quad (C.2)$$

$$H_{A_0 B_1} + H_{A_1 A_2 \dots A_{m-2} B_1} \geq H_{A_0 A_1 \dots A_{m-2} B_1} + H_{B_1}. \quad (C.3)$$

We continue to cancel the two NO terms on the lhs introduced by the last inequalities by adding the following SM inequalities

$$H_{A_1 B_0} + H_{A_2 A_3 \dots A_{m-2} B_0} \geq H_{A_1 A_2 \dots A_{m-2} B_0} + H_{B_0} \quad (C.4)$$

$$H_{A_1 B_1} + H_{A_2 A_3 \dots A_{m-2} B_1} \geq H_{A_1 A_2 \dots A_{m-2} B_1} + H_{B_1}. \quad (C.5)$$

Note that in each of these steps we get the O (observable) terms  $\{H_{A_i B_0}, H_{A_i B_1}, H_{B_0}, H_{B_1}\}$ . We continue doing that until we add the inequalities

$$H_{A_{m-4} B_0} + H_{A_{m-3} A_{m-2} B_0} \geq H_{A_{m-4} A_{m-3} A_{m-2} B_0} + H_{B_0} \quad (C.6)$$

$$H_{A_{m-4} B_1} + H_{A_{m-3} A_{m-2} B_1} \geq H_{A_{m-4} A_{m-3} A_{m-2} B_1} + H_{B_1}. \quad (C.7)$$

Now we add the inequalities

$$H_{A_{m-3} B_0} + H_{A_{m-2} B_0} \geq H_{A_{m-3} A_{m-2} B_0} + H_{B_0} \quad (C.8)$$

$$H_{A_{m-3} B_1} + H_{A_{m-2} B_1} \geq H_{A_{m-3} A_{m-2} B_1} + H_{B_1}. \quad (C.9)$$

In these  $m - 2$  steps we have added all the terms  $\{H_{A_i B_0}, H_{A_i B_1}\}$  for  $i = 0, \dots, m - 2$  and  $m - 2$   $\{H_{B_0}, H_{B_1}\}$ . To complete the two first rows we add the SM inequality

$$H_{A_{m-1} B_0} + H_{B_0 B_1} \geq H_{A_{m-1} B_1} + H_{B_0}. \quad (C.10)$$

To obtain the last row plus the marginal  $H(A_0)$  we add the SM inequality

$$H_{A_0 B_{m-1}} + H_{A_0 A_1} \geq H_{A_1 B_{m-1}} + H_{A_0}. \quad (C.11)$$

The remaining NO terms are

$$H_{A_0 A_1} + H_{B_0 B_1} \geq H_{A_0 A_1 \dots A_{m-2} B_0 B_1} + H_{A_0 A_1 \dots A_{m-2}}. \quad (C.12)$$

We can cancel two of them by simply adding the M (monotonicity) inequality

$$H_{A_0 A_1 \dots A_{m-2} B_0 B_1} \geq H_{B_0 B_1}. \quad (C.13)$$

The remaining NO terms are then

$$H_{A_0 A_1} \geq H_{A_0 A_1 \dots A_{m-2}}. \quad (C.14)$$

In order to conclude the proof we need to show that the remaining  $m - 3$  rows minus the NO terms above define a valid inequality. To show that, first note that every missing row

( $j = 2, \dots, m-2$ ) is of the form

$$\left( \sum_{i=0, \dots, m-j} (-1)^{\delta_{i,m-j}} H_{A_i B_j} \right) - (m-j-1) H_{B_j}. \quad (\text{C.15})$$

First we are going to prove that

$$\left( \sum_{i=0, \dots, m-j} (-1)^{\delta_{i,m-j}} H_{A_i B_j} \right) - (m-j-1) H_{B_j} + H_{A_0 \dots A_{m-j}} - H_{A_0 \dots A_{m-j-1}} \geq 0. \quad (\text{C.16})$$

To do that we add the following inequalities

$$H_{A_0 \dots A_{m-j}} + H_{A_0 \dots, A_{m-j-1} B_j} \geq H_{A_{m-j} B_j} + H_{A_0 \dots A_{m-j-1}} \quad (\text{C.17})$$

$$H_{A_{m-j-1} B_j} + H_{A_0 \dots A_{m-j-2} B_j} \geq H_{A_0 \dots A_{m-j-1}, B_j} + H_{B_j} \quad (\text{C.18})$$

$$H_{A_{m-j-2} B_j} + H_{A_0 \dots, A_{m-j-3} B_j} \geq H_{A_0 \dots A_{m-j-2} B_j} + H_{B_j} \quad (\text{C.19})$$

$$\dots \quad (\text{C.19})$$

$$H_{A_2 B_j} + H_{A_0 A_1 B_j} \geq H_{A_0 A_1 A_2 B_j} + H_{B_j} \quad (\text{C.20})$$

$$H_{A_0 B_j} + H_{A_1 B_j} \geq H_{A_0 A_1 B_j} + H_{B_j}. \quad (\text{C.21})$$

To finish the proof we only need to add the inequalities (C.16) for  $j = 2, \dots, m-2$ , that is

$$\sum_{j=2, \dots, m-2} \left( \sum_{i=0, \dots, m-j} (-1)^{\delta_{i,m-j}} H_{A_i B_j} \right) - (m-j-1) H_{B_j} + H_{A_0 \dots A_{m-j}} - H_{A_0 \dots A_{m-j-1}} \geq 0 \quad (\text{C.22})$$

$$= \left( \sum_{j=2, \dots, m-2} \left( \sum_{i=0, \dots, m-j} (-1)^{\delta_{i,m-j}} H_{A_i B_j} \right) - (m-j-1) H_{B_j} \right) - H_{A_0 A_1} + H_{A_0 A_1 \dots A_{m-2}} \geq 0. \quad (\text{C.23})$$

This concludes the proof.

#### Appendix D. Proving the $BI_{mm}^E \leq H_\lambda$ inequality

First remember that bounding the shared randomness between the parties implies the following constraint

$$H_{A_0 A_1 \dots A_{m-1}} + H_{B_0 B_1 \dots B_{m-1}} - H_{A_0 A_1 \dots A_{m-1} B_0 B_1 \dots B_{m-1}} \leq H_\lambda \quad (\text{D.1})$$

$$I_{A_0 A_1 \dots A_{m-1} : B_0 B_1 \dots B_{m-1}} \leq H_\lambda. \quad (\text{D.2})$$

In order to prove the inequality  $BI_{mm}^E \leq H_\lambda$  (49), we first note that the rows  $j = 2, \dots, m-2$  are exactly the same as for the  $I_{mm}^E \leq 0$  inequality (34), which fulfill the inequality

$$- \sum_{j=2, \dots, m-2} \sum_{i=0, \dots, m-j} (-1)^{\delta_{i,m-j}} H_{A_i B_j} + (m-j-1) H_{B_j} + H_{A_0 A_1} - H_{A_0 A_1 \dots A_{m-2}} \leq 0. \quad (\text{D.3})$$

If we can prove that the first two rows plus the last row plus the marginal  $H_{A_0}$  in (49) define a valid inequality of the form

$$\begin{aligned}
& H_{B_0} - H_{A_{m-1}B_0} - \sum_{i=0, \dots, m-1} (-1)^{\delta_{i,m-1}} H_{A_i B_j} + (m-2)H_{B_1} \\
& + H_{A_0} - H_{A_0 B_{m-1}} + H_{A_1 B_{m-1}} - H_{A_0 A_1} + H_{A_0 A_1 \dots A_{m-2}} \leq H_\lambda
\end{aligned} \tag{D.4}$$

then we have proven  $BI_{mm}^E \leq H_\lambda$ .

Next we are going to show the proof for  $m$  even, but  $m$  odd follows along similar lines. First add

$$-H_{A_0 B_{m-1}} - H_{A_0 A_1} + H_{A_1 B_{m-1}} + H_{A_0} \leq 0. \tag{D.5}$$

We continue adding the following SM inequalities

$$-H_{A_0 B_1} - H_{A_1 B_1} + H_{A_0 A_1 B_1} + H_{B_1} \leq 0 \tag{D.6}$$

$$\begin{aligned}
& -H_{A_2 B_1} - H_{A_3 B_1} + H_{A_2 A_3 B_1} + H_{B_1} \leq 0 \\
& \vdots
\end{aligned} \tag{D.7}$$

$$-H_{A_{m-4} B_1} - H_{A_{m-3} B_1} + H_{A_{m-4} A_{m-3} B_1} + H_{B_1} \leq 0. \tag{D.8}$$

Add

$$H_{A_0 A_1 \dots A_{m-1}} + H_{B_0 B_1} - H_{A_0 \dots A_{m-1} B_0 B_1} - H_\lambda \leq 0 \tag{D.9}$$

$$H_{A_{m-1} B_0 B_1} + H_{B_0} - H_{B_0 B_1} - H_{A_{m-1} B_0} \leq 0 \tag{D.10}$$

$$H_{A_0 A_1 \dots A_{m-1} B_0 B_1} + H_{A_{m-1} B_1} - H_{A_0 A_1 \dots A_{m-1} B_1} - H_{A_{m-1} B_0 B_1} \leq 0 \tag{D.11}$$

$$H_{A_0 A_1 \dots A_{m-2}} + H_{A_0 A_1 \dots A_{m-1} B_1} - H_{A_0 A_1 \dots A_{m-1}} - H_{A_0 A_1 \dots A_{m-2} B_1} \leq 0. \tag{D.12}$$

At this point the missing terms are (plus the NO terms)

$$\frac{(m-2)}{2} H_{B_1} - H_{A_{m-2} B_1} - H_{A_0 A_1 \dots A_{m-2} B_1} + \sum_{i=0, \dots, m-4} H_{A_i A_{i+1} B_1} \leq 0. \tag{D.13}$$

Add

$$H_{A_0 A_1 \dots A_{m-2} B_1} + H_{B_1} - H_{A_2 A_3 \dots A_{m-2} B_1} - H_{A_0 A_1 B_1} \leq 0 \tag{D.14}$$

$$\begin{aligned}
& H_{A_2 A_3 \dots A_{m-2} B_1} + H_{B_1} - H_{A_4 A_5 \dots A_{m-2} B_1} - H_{A_2 A_3 B_1} \leq 0 \\
& \vdots
\end{aligned} \tag{D.15}$$

$$H_{A_{m-4} \dots A_{m-2} B_1} + H_{B_1} - H_{A_{m-2} B_1} - H_{A_{m-4} A_{m-5} B_1} \leq 0. \tag{D.16}$$

This concludes the proof.

## References

- [1] Brunner N, Cavalcanti D, Pironio S, Scarani V and Wehner S 2013 arXiv:[1303.2849](#)
- [2] Pearl J 2009 *Causality* (Cambridge: Cambridge University Press)
- [3] Spirtes P, Glymour N and Scheines R 2001 *Causation, Prediction, and Search* 2nd edn (Cambridge, MA: MIT)
- [4] Cuadras C M, Fortiana J and Rodra-Guez-Lallena J A 2002 *Distributions With Given Marginals and Statistical Modeling* (Berlin: Springer)

- [5] Klyachko A A 2006 *J. Phys.: Conf. Ser.* **36** 72
- [6] Schilling C, Gross D and Christandl M 2013 *Phys. Rev. Lett.* **110** 040404
- [7] Walter M, Doran B, Gross D and Christandl M 2013 *Science* **340** 1205–8
- [8] Christandl M, Sahinoglu M B and Walter M 2012 arXiv:1210.0463
- [9] Bell J S 1964 *Physics* **1** 195
- [10] Pitowsky I 1989 Quantum probability–quantum logic *Lecture Notes in Physics* (Berlin: Springer)
- [11] Branciard C, Gisin N and Pironio S 2010 *Phys. Rev. Lett.* **104** 170401
- [12] Hall M J W 2010 *Phys. Rev. Lett.* **105** 250404
- [13] Barrett J and Gisin N 2011 *Phys. Rev. Lett.* **106** 100406
- [14] Yeung R W 1997 *IEEE Trans. Inf. Theory* **43** 1924
- [15] Yeung R W 2008 Information theory and network coding *Information Technology–Transmission, Processing, and Storage* (Berlin: Springer)
- [16] Chaves R 2013 *Phys. Rev. A* **87** 022102
- [17] Steudel B and Ay N 2010 arXiv:1010.5720
- [18] Fritz T 2012 *New J. Phys.* **14** 103001
- [19] Braunstein S L and Caves C M 1988 *Phys. Rev. Lett.* **61** 662
- [20] Chaves R and Fritz T 2012 *Phys. Rev. A* **85** 032113
- [21] Kurzyński P, Ramanathan R and Kaszlikowski D 2012 *Phys. Rev. Lett.* **109** 020404
- [22] Fritz T and Chaves R 2013 *IEEE Trans. Inf. Theory* **59** 803
- [23] Pan A K, Sumanth M and Panigrahi P K 2013 *Phys. Rev. A* **87** 014104
- [24] Devi A R U, Karthik H S, Sudha and Rajagopal A K 2013 *Phys. Rev. A* **87** 052103
- [25] Katiyar H, Shukla A, Rao K R K and Mahesh T S 2013 *Phys. Rev. A* **87** 052102
- [26] Kurzynski P and Kaszlikowski D 2013 arXiv:1309.6777
- [27] Collins D and Gisin N 2004 *J. Phys. A: Math. Theor.* **37** 1775
- [28] Shapley L S 1971/72 *Int. J. Game Theory* **1** 11
- [29] Kashiwabara K 2000 *Theor. Comput. Sci.* **235** 239
- [30] Studený M, Bouckaert R R and Kocka T 2000 Extreme supermodular set functions over five variables *Research Report No. 1977* Institute of Information Theory and Automation, Prague
- [31] Williams H P 1986 *Amer. Math. Monthly* **93** 681–95
- [32] Clauser J F, Horne M A, Shimony A and Holt R A 1969 *Phys. Rev. Lett.* **23** 880–4
- [33] Clauser J F, Horne M A, Shimony A and Holt R A 1969 *Phys. Rev. Lett.* **23** 880–4
- [34] Clauser J F and Horne M A 1974 *Phys. Rev. D* **10** 526
- [35] Aliprantis C D and Tourky R 2007 *Cones and Duality* (Providence, RI: American Mathematical Society)
- [36] Barvinok A 2002 *A Course in Convexity* (Providence, RI: American Mathematical Society)
- [37] Rockafellar R 1970 *Convex Analysis, Princeton Mathematical Series* (Princeton, NJ: Princeton University Press)
- [38] Matús F 2007 Infinitely many information inequalities *Proc. 2007 IEEE Int. Symp. on Information Theory* **41–44**
- [39] Brualdi R A 2012 *Introductory Combinatorics* 5th edn (Upper Saddle River, NJ: Pearson Education)
- [40] Mermin N D 1990 *Phys. Rev. Lett.* **65** 1838–40
- [41] Avis D, Imai H, Ito T and Sasaki Y 2005 *J. Phys. A: Math. Theor.* **38** 10971
- [42] Popescu S and Rohrlich D 1994 *Found. Phys.* **24** 379
- [43] Sliwa C 2003 *Phys. Lett. A* **317** 165–8
- [44] Barrett J, Linden N, Massar S, Pironio S, Popescu S and Roberts D 2005 *Phys. Rev. A* **71** 022101
- [45] Grandjean B, Liang Y-C, Bancal J-D, Brunner N and Gisin N 2012 *Phys. Rev. A* **85** 052113
- [46] Arnault F 2011 arXiv:1107.2255v2
- [47] Collins D, Gisin N, Linden N, Massar S and Popescu S 2002 *Phys. Rev. Lett.* **88** 040404
- [48] Brunner N, Pironio S, Acin A, Gisin N, Méthot A A and Scarani V 2008 *Phys. Rev. Lett.* **100** 210503
- [49] Budroni C and Cabello A 2012 *J. Phys. A: Math. Theor.* **45** 385304



- [50] Wood C J and Spekkens R W 2012 arXiv:[1208.4119](https://arxiv.org/abs/1208.4119)
- [51] Branciard C, Rosset D, Gisin N and Pironio S 2012 *Phys. Rev. A* **85** 032119
- [52] Araújo M, Quintino M T, Budroni C, Cunha M T and Cabello A 2013 *Phys. Rev. A* **88** 022118
- [53] Werner R F 1989 *Phys. Rev. A* **40** 4277–81
- [54] Vértesi T 2008 *Phys. Rev. A* **78** 032112
- [55] Acin A, Gisin N and Toner B 2006 *Phys. Rev. A* **73** 062105
- [56] Webster R 1995 *Convexity* (New York: Oxford University Press)
- [57] Christof T and Löbel A 2009 PORTA: Polyhedron Representation Transformation Algorithm Software download available at [http://typo.zib.de/opt-long\\_projects/Software/Porta/](http://typo.zib.de/opt-long_projects/Software/Porta/)