Correlations constrained by composite measurements

John H. Selby[1], Ana Belén Sainz[1], Victor Magron[2], Łukasz Czekaj[1], and Michał Horodecki[1]

[1]International Centre for Theory of Quantum Technologies, University of Gdańsk, 80-308 Gdańsk, Poland

[2]LAAS-CNRS and Institute of Mathematics, University of Toulouse, LAAS, 7 Avenue du Colonel Roche, 31077 Toulouse Cédex 4, France

How to understand the set of correlations admissible in nature is one outstanding open problem in the core of the foundations of quantum theory. Here we take a complementary viewpoint to the device-independent approach, and explore the correlations that physical theories may feature when restricted by some particular constraints on their measurements. We show that demanding that a theory exhibits a composite measurement imposes a hierarchy of constraints on the structure of its sets of states and effects, which translate to a hierarchy of constraints on the allowed correlations themselves. We moreover focus on the particular case where one demands the existence of a correlated measurement that reads out the parity of local fiducial measurements. By formulating a non-linear Optimisation Problem, and semidefinite relaxations of it, we explore the consequences of the existence of such a parity reading measurement for violations of Bell inequalities. In particular, we show that in certain situations this assumption has surprisingly strong consequences, namely, that Tsirelson's bound can be recovered.

Contents

1 Introduction

Bell nonclassicality is a well-known phenomenon featured by quantum theory, and attests that correlations observed in nature are not always compatible with a classical common cause shared among the distant wings of an experiment [1]. That is, non-classical common causes are necessary to explain our observational data [2]. Bell's theorem not only teaches us a valuable lesson about the foundational aspects of nature, but also underpins a variety of current technological applications. For example, non-classical correlations enable cryptographic applications, such as key distribution [3–8] and randomness generation [9–12], and provide an information-theoretic advantage in other families of so-called non-local games [13–15].

Understanding quantum correlations – in particular their limitations – is therefore an important open problem within quantum information theory. Research on these lines has recently been carried out within the device-independent formalism, that is, where the only information used to reason about nature are the classical variables that denote measurement choices and their outcomes, together with the observed outcome statistics. Within this paradigm, quantum correlations are studied "from the outside", by exploring the constraints that physical or information-theoretical principles impose on the observed correlations [16–22]. In the device-independent framework, hence, such proposed constraints are therefore formulated at the level of the correlations themselves.

In this work we take a complementary viewpoint to the problem of characterising quantum correlations, by examining the possible correlations that may arise when constraints are imposed on the underlying physical theory. From this perspective, hence, one asks how various elements of the physical theory constrain or enable particular correlations. The particular objects we are interested in here are the *measurements* that the theory may feature. Even though in principle one could also impose constraints on the states as well, our curiosity on measurements arises from the results of Ref. [23] – if demanding such parity constraints on measurement outcomes at the level of the statistics yields such substantial constraints on correlations (see below), what can we expect if we open the box and demand conditions on the measurement themselves? On the one hand, it is well known that the theory known colloquially as 'Boxworld' [24], which was

formulated in order to realise arbitrary no-signalling correlations, only features local measurements and wirings thereof. That is, Boxworld does not display entangled measurements. This is in contrast to quantum theory, where entangled measurements are ubiquitous – you may for instance think of the so-called Bell measurements. A natural question then arises: is there any relationship between the types of measurements a theory features and the correlations it may produce. Even seemingly simple at first sight, this question is far from trivial: by enlarging the set of allowed measurements one necessarily needs to shrink down the set of allowed states, since the state space featured by a given physical theory is constrained by the dual space of the effects[1] space. How the set of allowed correlations (measurements on states) changes in consequence is therefore not straightforward.

Progress on this question was made in Ref. [23], where it was shown that demanding the existence of a particular entangled effect would constrain the correlations admissible in a bipartite Bell scenario to those realisable by an entangled pair of qubit quantum systems. That is, by demanding that the theory features a particular entangled measurement, it was shown that the allowed correlations in the so-called Clauser-Horne-Shimony-Holt (CHSH) scenario [25] was indeed the set of quantum ones.

In this paper we explore what types of constraints the existence of bipartite effects impose on the possible correlations that a theory may feature. The framework we use to describe the underlying physical theory is that of General Probabilistic Theories (GPTs) [24, 26–34] . First, we take a compositional perspective and show how the existence of one (arbitrary) bipartite effect imposes not one but an infinite hierarchy of constraints which must be satisfied by the states and effects of the GPT. This hierarchy of constraints on bipartite states immediately translates to a hierarchy of constraints on the correlations realisable within the theory. Inspired by Ref. [23], we then consider a particular setup where we demand that there exists a measurement in the GPT which can measure the parity of fiducial measurements (or a subset thereof), which we call a *(partial) parity reading measurement*. We say that the observables which appear in such partial parity reading measurement are *parity-readable observables*. We then define an Optimisation Problem that computes the maximum violation of a Bell inequality by the corresponding GPT, provided that such a parity reading measurement exists. Such an Optimisation Problem provides a way to characterise the set of correlations allowed by such a GPT when the parties choose among these parity-readable fiducial measurements in the Bell tests. The solution to this optimisation problem is, however, computationally complex, given that the problem itself is polynomial on the optimised variables. We hence present a series of relaxations that upper-bound the solution to the Optimisation problem. We finish by applying our techniques to a variety of Bell inequalities, and discussing the necessity of the Local Tomography assumption.

Inspired by our results, we moreover formulate a conjecture:

Conjecture 1. *Under the assumption of local tomography, the local observables that are* **parity-readable** *satisfy Tsirelson's bound, i.e., they cannot violate Bell inequalities better than quantum mechanics does (with arbitrary measurements).*

The inspiration came from the fact that various of our numerical explorations do indeed satisfy this property. Moreover, we have a counterexample demonstrating the necessity of the assumption of local tomography within the conjecture. That is, if we have a GPT which does not satisfy local tomography, then it is possible to create a PR box using observables which are parity-readable.

---

[1]The so-called effects in a physical theory may be thought of as its dichotomic measurements.

(a) Measuring parity of local observables.

(b) Reading parity of two pairs of observables: XX and ZZ.

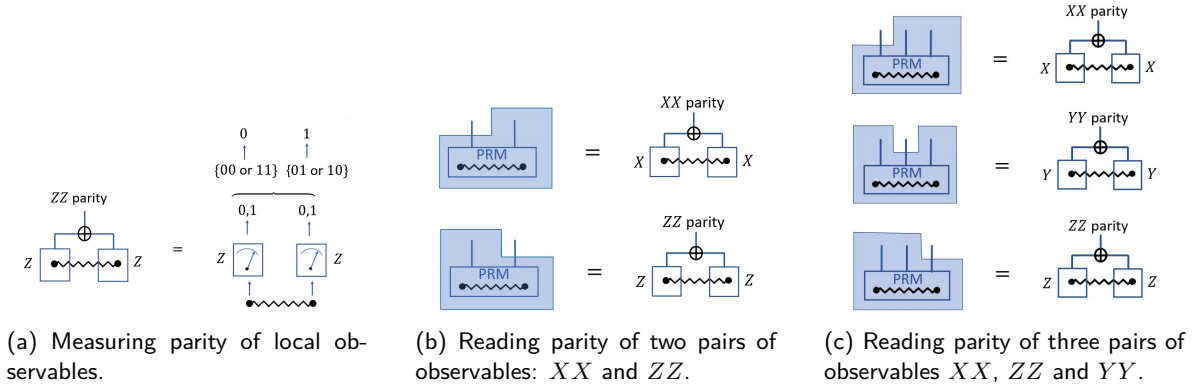(c) Reading parity of three pairs of observables XX, ZZ and YY.

Figure 1: The idea of a measurement that reads the parity of the parity-readable observables (denoted by PRM): a composite measurement that measures parities of several pairs of local observables in one go.

Finally, as we comment in the Discussion section, our work opens the door to further conjecturing that Quantum Theory, within the landscape of possible locally-tomographic physical theories, is the theory that features the best balance between allowed states and effects, in the sense that it yields the largest violation of any Bell inequality by parity-readable fiducial measurements.

2 Descriptive summary of the results

Suppose that two parties, the ubiquitous Alice and Bob, each have three binary observables, X, Y, and Z, that they can measure on some shared system. Moreover, suppose that there exists some joint measurement that they could perform on their joint system if they were to get together, whose outcome would determine the parity of XX and ZZ. That is, the joint measurement does not necessarily reveal the values that they would have obtained had they measured X and Z individually, but just whether or not their XX and ZZ measurements would have been correlated or not.[2] We say that such observables are *parity-readable*, and illustrate this idea in Fig. 1.

In this manuscript, we consider the impact that parity readability has on the correlations that can be generated in a Bell scenario when measuring parity-readable observables. For example, is it possible to create PR-box correlations via parity-readable observables?

It is clear that, even within the standard quantum mechanical formalism, this imposes a restriction on the correlations which can be observed – not all observables are parity-readable, and some correlations can only be achieved by those that are not. However, what about if we go beyond quantum mechanics?

We conjecture that, within this landscape of parity-readable observables, quantum theory is always optimal. That is, any correlation that can be generated by parity-readable observables, independently of which underlying (tomographically-complete) physical theory they belong to, can also be generated by parity reading observables within quantum theory.

If this conjecture is true, then this would be in stark contrast to the landscape of arbitrary observables, in which there are correlations which cannot be realised within our quantum world. It would therefore show, for the first time, a way in which quantum theory is an optimal physical theory for an information theoretic task.

---

[2]Such is the case in quantum theory, where X and Z are incompatible measurements, but their parity can nonetheless be measured by performing a suitable coarse graining of a standard Bell measurement.
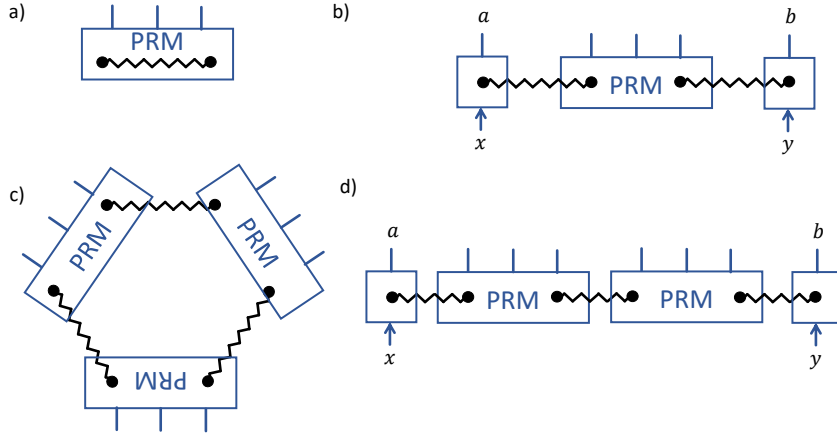
Figure 2: The existence of a measurement that reads the parity of the parity-readable observables, implies an infinite hierarchy of positivity constraints. For instance, the probabilities of outcomes should be positive on (a) a single copy of a state, (b) all pairs of products of steered states that can be obtained from the original state, (c) three copies of the state in a triangle network, and (d) a single copy of the state times any two pairs of steered states.

Utilising techniques and insight coming from the field of generalised probabilistic theories we specify the constraints that the existence of a parity-reading measurement imposes on the possible correlations those observables may generate in a Bell test. These constraints are formulated as a hierarchy of convex optimisation problems which can be tackled using standard numerical methods. We apply this technique to and numerically explore various Bell scenarios and Bell inequalities, whose results lead us to formulating the conjecture discussed above. More precisely, we consider scenarios in which each party has at most three observables X, Y and Z, and in which either two or three of these are parity-readable. The Bell inequalities explored include the CHSH [25], AMP [35], and AQ [36] inequalities. Further development of both the numerical methods, as well as analytical convex optimisation techniques, are necessary to explore this conjecture further.

It is worth highlighting that the main technique that we develop and apply here, relies on demanding the parity-readable observables to yield valid probabilities when applied both to a composite state as well as to products of steered states that can be generated from it (see Figs. 2a, 2b.). These constraints are indeed phrased as positivity conditions on variables we optimise over. However, to capture the full power of the constraints that these parity-readable observables impose, one needs to take into account infinitely many conditions (examples of which are presented in Figs. 2c and 2d), described in the article, and which deserve further exploration.

3 Warm-up: non-existence of Popescu-Rohrlich correlations for parity measurable X and Z observables.

While the general problem of finding a bound for Bell inequalities for parity-readable observables is a complex one (as we will see further in this paper), one can relatively easily show that the CHSH inequality cannot achieve its maximal algebraic bound. Namely, we shall show that parity-readable observables cannot exhibit so called Popescu-Rohrlich correlations [16].

In this section we shall present such reasoning, as a simple warm-up exercise in anticipation of the rest of the paper. In this warm-up we will take a device-independent approach, in the

sense of relying only on the conditional probability distributions for the argument (these black boxes are the only information we leverage for describing the underlying states of the system). In the remainder of the paper we will utilise the language of Generalised Probabilistic Theories (GPTs) and, in particular, we will express the question using a formal diagrammatic language. Full justification of some formulas will be found later in the paper. For self-consistency of the main part of the paper, we will repeat there some definitions used here.

3.1 States

The so called *PR-box* [16] is defined as the set of conditional probabilities:

$$p(ab|xy) = \begin{cases} \frac{1}{2} & \text{if} \quad a \oplus b = xy \\ 0 & \text{else} \end{cases} , \tag{1}$$

where $x, y \in \{0, 1\}$ are inputs, and $a, b \in \{0, 1\}$ are outputs. The observables X, Z from the previous section are now encoded by the classical variables $x = 0, 1$ for Alice, and $y = 0, 1$ for Bob. The PR-box is a maximally nonlocal non-signaling box, since it violates the CHSH inequality up to its maximal algebraic value. Moreover, PR-box correlations have the feature that the pairs of observables XX, XZ, and ZX are perfectly correlated, while the pair ZZ is perfectly anticorrelated.

The principles of Local Tomography (see Sec. 4) and No Signaling allow one to characterise any bipartite state s by means of the following table:

$$\mathbf{p_s} = \begin{bmatrix} p_s(00|00) & p_s(00|01) & p_s^{(1)}(0|0) \\ p_s(00|10) & p_s(00|11) & p_s^{(1)}(0|1) \\ p_s^{(2)}(0|0) & p_s^{(2)}(0|1) & 1 \end{bmatrix} , \tag{2}$$

where (i) specifies the party (i.e., (1) for Alice and (2) for Bob). This is indeed similar to the representation of non-signalling correlations in the CHSH scenario known as Collins-Gisin [37], which shows how nine parameters are enough to fully specify the 16 components of the full conditional probability distribution.

Since a PR box has perfect correlations or anticorrelations for each pair of observables, it is natural that the steered states obtained from it are all the states that have well defined value for both observables. Thus, we have that, for each party, there are four steered state, the state s_1 with $X = 0, Z = 0$, the state s_2 with $X = 0, Z = 1$, the state s_3 with $X = 1, Z = 0$ and the state s_4 with $X = 1, Z = 1$. These four states are depicted in Fig. 3. We shall present

$$X = 0, Z = 1 \qquad\qquad X = 1, Z = 1$$

$$s_2 \qquad\qquad s_4$$

$$s_1 \qquad\qquad s_3$$

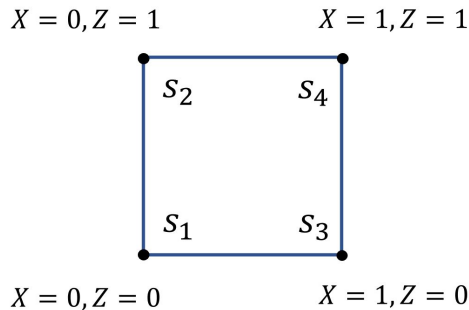$$X = 0, Z = 0 \qquad\qquad X = 1, Z = 0$$

Figure 3: Steered states coming from PR box.

here the four pairs of steered states which we will use in the proof. These are products of all
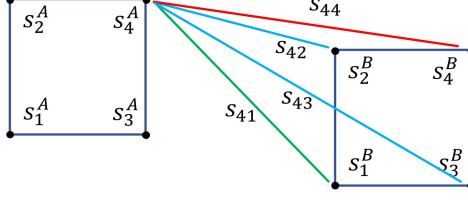
Figure 4: Selected pairs of steered states.

four steered states on Bob's site and a single fixed state on Alice's side, namely the state with $X_A = Z_A = 1$, as is depicted Fig. 4.

We denote these four states as:

$$s_{41} = s_4^A \otimes s_1^B, \quad s_{42} = s_4^A \otimes s_2^B, \quad s_{43} = s_4^A \otimes s_3^B, \quad s_{44} = s_4^A \otimes s_4^B. \tag{3}$$

In the above matrix notation these are characterised by:

$$s_{41} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad s_{42} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

$$s_{43} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad s_{44} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

3.2 Parity Reading Measurement.

Now we consider the measurement that measures the parities of XX and ZZ in a single shot, that is, X and Z are *parity-readable* observables. Specifically, this parity reading measurement (PRM) outputs a pair of bits: the first one reports the parity of XX and the second one reports the parity of ZZ. This is expressed by the following pair of conditions:

$$R_{XX}(\mathbf{p_s}) \equiv p_s(00|\text{PRM}) + p_s(01|\text{PRM}) = p_s(00|00) + p_s(11|00), \tag{5}$$

$$R_{ZZ}(\mathbf{p_s}) \equiv p_s(00|\text{PRM}) + p_s(10|\text{PRM}) = p_s(00|11) + p_s(11|11). \tag{6}$$

where we denote by $p_s(rq|\text{PRM})$ the probability of obtaining the pair of outcomes rq when measuring the PRM on state $\mathbf{p_s}$. Notice that R_{XX} (R_{ZZ}) have the interpretation of the probability of XX (ZZ) being correlated (i.e., the probability of obtaining the same outcomes by Alice and Bob). In addition, note that the probabilities of the outcomes of the PRM satisfy the normalisation condition:

$$p_s(00|\text{PRM}) + p_s(01|\text{PRM}) + p_s(10|\text{PRM}) + p_s(11|\text{PRM}) = 1. \tag{7}$$

We therefore have a system of three linear equations with four unknown quantities $p_s(rq|\text{PRM})$. Hence, a solution may be found with one quantity remaining an independent variable, for example the expression:

$$C(\mathbf{p_s}) = p_s(00|\text{PRM}) - p_s(01|\text{PRM}) - p_s(10|\text{PRM}) + p_s(11|\text{PRM}). \tag{8}$$

Due to local tomography, the probabilities of the outcomes of any measurement for a particular state can be written as a linear combination of state parameters, $\mathbf{p_s}$. Hence, as C is a linear combination of such probabilities, it can be computed via:

$$C(\mathbf{p_s}) = \mathcal{C} \cdot \mathbf{p_s}, \tag{9}$$

where $\cdot$ denotes Frobenius matrix product, and $\mathcal{C}$ is the matrix of, at the moment, unspecified parameters describing the PRM:

$$\mathcal{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}. \tag{10}$$

Combining Eqs. (5), (7), (8) and (9), we obtain the formulas for computing the probabilities of PRM outcomes, expressed in terms of the state parameters, as well as the free parameter $\mathcal{C}$:

$$p_s(00|\text{PRM}) = \frac{1}{4}\left(2R_{XX}(\mathbf{p_s}) + 2R_{ZZ}(\mathbf{p_s}) + \mathcal{C}\cdot\mathbf{p_s} - 1\right),$$

$$p_s(01|\text{PRM}) = \frac{1}{4}\left(2R_{XX}(\mathbf{p_s}) - 2R_{ZZ}(\mathbf{p_s}) - \mathcal{C}\cdot\mathbf{p_s} + 1\right),$$

$$p_s(01|\text{PRM}) = \frac{1}{4}\left(-2R_{XX}(\mathbf{p_s}) + 2R_{ZZ}(\mathbf{p_s}) - \mathcal{C}\cdot\mathbf{p_s} + 1\right),$$

$$p_s(11|\text{PRM}) = \frac{1}{4}\left(-2R_{XX}(\mathbf{p_s}) - 2R_{ZZ}(\mathbf{p_s}) + \mathcal{C}\cdot\mathbf{p_s} + 1\right). \tag{11}$$

3.3 Proof of nonexistence of the PR-box.

We shall now argue that, for any arbitrary choice of $\mathcal{C}$, the probability of at least one PRM output will necessarily be negative on some of the pairs of steered states. This proves that parity-readable observables cannot feature PR-box correlations, and hence cannot violate the CHSH inequality up to its maximal algebraic bound.

We shall first impose positivity of $p_s(qr|\text{PRM})$ for each of the four states in Eq. (3). Note that, for the state s_{44}, by definition, both XX and ZZ are perfectly correlated, since $X_A = X_B = 1$ and $Z_A = Z_B = 1$. Hence, $R_{XX}(s_{44}) = R_{ZZ}(s_{44}) = 1$ for the state s_{44}. An analogous reasoning for the remaining three states yields:

$$\begin{aligned} R_{XX}(s_{41}) &= 0, & R_{ZZ}(s_{41}) &= 0, \\ R_{XX}(s_{42}) &= 0, & R_{ZZ}(s_{42}) &= 1, \\ R_{XX}(s_{43}) &= 1, & R_{ZZ}(s_{43}) &= 0, \\ R_{XX}(s_{44}) &= 1, & R_{ZZ}(s_{44}) &= 1. \end{aligned} \tag{12}$$

Of course, we might alternatively compute these values from the matrix form of the states, and the definition of R_{XX}, R_{ZZ}. Inserting them into Eq.(11) for each of the four states, yields the following conditions for the positivity of PRM probabilities:

$$\begin{aligned} p_{s_{41}}(qr|\text{PRM}) \geq 0 &\quad \Leftrightarrow \quad \mathcal{C}\cdot s_{14} = 1, \\ p_{s_{42}}(qr|\text{PRM}) \geq 0 &\quad \Leftrightarrow \quad \mathcal{C}\cdot s_{24} = -1, \\ p_{s_{43}}(qr|\text{PRM}) \geq 0 &\quad \Leftrightarrow \quad \mathcal{C}\cdot s_{34} = -1, \\ p_{s_{44}}(qr|\text{PRM}) \geq 0 &\quad \Leftrightarrow \quad \mathcal{C}\cdot s_{44} = 1. \end{aligned} \tag{13}$$

We then use the matrix form of our states of Eq. (4) and the form of $\mathcal{C}$ of Eq. (10) and rewrite these conditions as:

$$c_{31} + c_{32} + c_{33} = 1, \quad c_{31} + c_{33} = -1, \quad c_{32} + c_{33} = -1, \quad c_{33} = 1. \tag{14}$$

We see that this set of equalities does not have any real solutions. We therefore conclude that there does not exist a Parity Reading Measurement whose outcomes give legitimate probabilities

8

for the above four pairs of steered states. Since the existence of a PR-box allows for such steered states, we conclude that existence of Parity Reading Measurement excludes PR-box correlations.

The remainder of this work explores this curious property in more detail. In particular: on the one hand, can we go beyond simply ruling out the PR-box and see how PRMs constrain the set of correlations; and, on the other hand, can we go beyond the assumptions involved in this derivation, namely, that of local tomography and that there are simply two observables per party.

4   Generalised probabilistic theories

In order to explore the previously presented questions, we will work within the framework of generalised probabilistic theories (GPTs) [24, 26]. This framework was developed in order to be able to describe essentially arbitrary conceivable theories of nature – taking quantum and classical theory as just two particular points within a broad landscape of potential physical theories. The GPT framework is based on the idea that, ultimately, the way that we characterise physical devices is by the probabilities that they give rise to in experiments. From this simple observation, one can build a rich mathematical structure which any GPT must have.

GPTs have already proved a vital tool in the study of computation [38–46] and cryptography [24, 47–52] beyond quantum theory. Moreover, recently tools from convex optimisation theory have been used to gain new insight into GPTs [47–49, 53–55]. In particular, the generalisation of quantum theory to GPTs is analogous to the generalisation from semi-definite to conic programmes. These optimisation tools will be vital for developing a complete understanding of how the structures of the GPT impact on the realisable correlations of the GPT.

For simplicity of the presentation, in this manuscript we will focus on a particular class of GPTs, namely, those that satisfy the principle of Local Tomography. A GPT is locally-tomographic if any state of a composite system can be uniquely determined by the information obtained from performing local measurements on its constituents (see, e.g., Ref. [26] for a full formal definition). GPTs that satisfy local tomography tend to have very useful properties, and in particular they admit a useful parametrisation of their state and effect vectors, which will come in handy in various stages in this manuscript. The majority of our results, such as the formulation of the hierarchy of constraints, however, do not require this principle to hold, hence we will highlight the instances where the assumption is indeed necessary.

In this paper we take a categorical approach to tomographically local GPTs. This is an intrinsically compositional approach, which allows us to describe arbitrary experimental scenarios. Moreover, the diagrammatic representation in terms of string diagrams, which comes from this approach, provides an intuitive way to reason about these complex situations. We provide a brief technical introduction to the formalism in Appendix B, and refer the reader to Refs. [32–34, 56–58] for more extensive introductions to these tools.

4.1   Constraints on states and effects

We can see how the state and effect spaces constrain one another when demanding that scalars are probabilities. For some system V (see the Appendix for details on notation), the states can be thought of as vectors $s \in \Omega_V$ living inside V, and effects as linear functionals, $e \in \mathcal{E}_V$ living in the dual space V^*. Any pair of an effect and a state must satisfy:

$$\begin{array}{c} \triangle e \\ | V \\ \triangledown s \end{array} \in [0, 1] . \tag{15}$$

The geometric consequences of this for local and composite states are presented in App. C.

It was noted in Ref. [59], however, that even if a pair of state and effect spaces satisfy the standard constraints discussed in App. C (i.e., Eqs. (169) and (170)), it is not straightforward that they actually define a valid GPT, at least when the No-restriction hypothesis is not assumed (i.e., when it is not required that $\Omega = \mathcal{E}^*$). An important condition that must also be checked, as shown in Theorem 9 of Ref. [59], is that the steered states are also valid states within the theory. That is, any bipartite state s must satisfy:

$$
\begin{array}{ccc}
\vcenter{\hbox{}} \in \Omega_V & \text{and} & \vcenter{\hbox{}} \in \Omega_W
\end{array}
\tag{16}
$$

for all $e_v \in \mathcal{E}_V$ and $e_w \in \mathcal{E}_W$. This constraint can be interpreted in many forms:

- as a constraint on the bipartite state space, namely, that a bipartite state must lead to valid steered states,
- as a constraint on the local state spaces, namely, it forces them to include all of these steered states as valid local states,
- as a constraint on the local effect spaces, namely, an effect is only allowed if it leads to valid steered states when composed with any bipartite state.

However, we believe that the constraints of Eq. (16) are probably best viewed not from any of these individual perspectives, but instead just as a compatibility condition between local states and effects, and bipartite states.

Similarly, we can consider bipartite effects e and note that these have a similar compatibility condition together with local states and effects:

$$
\begin{array}{ccc}
\vcenter{\hbox{}} \in \mathcal{E}_V & \text{and} & \vcenter{\hbox{}} \in \mathcal{E}_W
\end{array}
\tag{17}
$$

for all $s_v \in \Omega_V$ and $s_w \in \Omega_W$.

One may be inclined to think that these constraints on bipartite states/effects, steered states, and steered effects, are sufficient to characterise a valid GPT. However, there are only the tip of the iceberg – a whole plethora of further compatibility constraints lie underneath the surface. For example, consider a normalised bipartite state s and a bipartite effect e. By taking two copies of each, one should be capable of wiring them as follows and obtain a valid probability:

$$
\vcenter{\hbox{}} \in [0,1].
\tag{18}
$$

In addition, if one takes two copies of s and one of e, one should be capable of wiring them as follows and obtain a valid bipartite state:

$$
\vcenter{\hbox{}} \in \Omega_{V \otimes W}.
\tag{19}
$$

Other types of compatibility constraints include diagrams like the following, which arise due to symmetry in the special case when all the local systems have the same type:

$$
\text{(diagram)} \in [0,1], \qquad \text{(diagram)} \in [0,1]. \tag{20}
$$

One can readily see how these belong to an infinite family of constraints, each featuring the same (but arbitrary) number of normalised bipartite states s and bipartite effects e being connected in this "braided" fashion.

Even if the bipartite states consist of local systems of the same type, it is not necessary that they are symmetric under a swap operation of the local systems. Hence, a different hierarchy of braided-type constraints will arise by requiring consistent probability assignments to diagrams of the form:

$$
\text{(diagram)} \in [0,1]. \tag{21}
$$

In Section 6.1 we will see how to formalise these types of hierarchies, and how they can be used to constrain the potential correlations in a GPT.

4.2 Correlations in a GPT

To describe correlations in a GPT we must first introduce classical systems. Here we describe a classical system by classical random variable, which can take values from a set, such as X, Y, A, B. We denote these classical systems by thin gray wires (to distinguish them from GPT wires). Correlations, in this formalism, are then viewed as no-signalling stochastic maps, $\mathsf{N} : X \times Y \to A \times B$, between these random variables. Diagrammatically, we denote these no-signalling boxes as:

$$
\text{(diagram } \mathsf{N}\text{)} , \tag{22}
$$

which must satisfy the no-signalling constraints:

$$
\text{(diagram)} = \text{(diagram)} \quad \text{and} \quad \text{(diagram)} = \text{(diagram)} . \tag{23}
$$

This is equivalent to the standard view of correlations [60] as being described by a conditional probability distribution $\Pr(A, B|X, Y) = \{p(ab|xy)\}_{\{a \in A, b \in B, x \in X, y \in Y\}}$, which can be seen by defining:

$$
p(ab|xy) := \text{(diagram)} , \tag{24}
$$

11

and checking that the no-signalling conditions of Eqs. (23) are equivalent to the standard no-signalling conditions for the conditional probability distribution. To do so it is useful to note that, for example:

$$
\frac{\equiv}{\big|X} \quad := \quad \sum_{x \in X} \frac{\triangle x}{\big|X} \quad .
\tag{25}
$$

Then, in order to understand the possible correlations in a GPT, it is useful to describe measurements as transformations from a GPT to a classical system, where the choice of measurement is controlled by another classical system. These controlled measurements must satisfy the constraint:

$$
\frac{\frac{\equiv}{\big|A}}{\boxed{M}\diagdown}_{\big|X \ \diagdown V} \quad = \quad \frac{\equiv}{\big|X} \ \frac{\equiv}{\big|V} \quad .
\tag{26}
$$
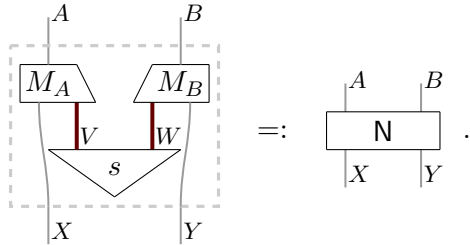
Correlations that can be generated in a Bell experiment are hence of the form:

$$\tag{27}$$
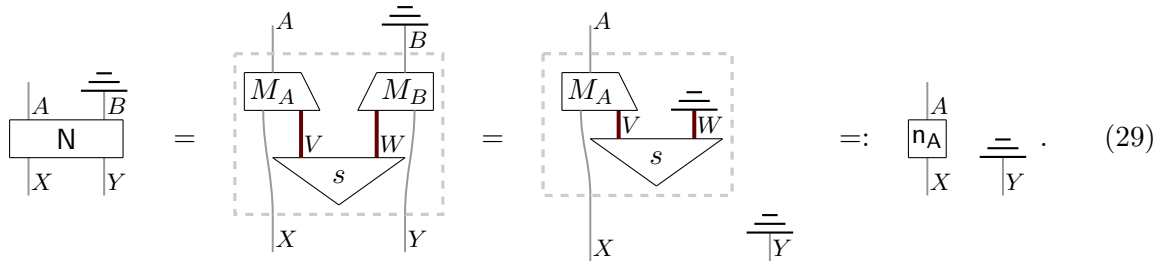
where the local controlled measurements M_A and M_B (for Alice and Bob respectively) are performed on a bipartite system on state s, with local system types V, W in the GPT. These local measurements are controlled on the input classical variable and have an outcome recorded in the output classical variable.

In any GPT, such a diagram corresponds to a no-signalling stochastic map:

$$\tag{28}$$

It can then be shown that the constraint on measurements of Eq. 26 immediately implies the relevant no-signalling conditions, for example:

$$\tag{29}$$

A more general version of this proof first appeared in Ref. [61] and was generalised to arbitrary causal structures in Ref. [62].

Bell inequalities [1, 60] are then a particular class of linear functionals from this space of stochastic maps to the reals. A linear functional corresponding to a Bell inequality, hereon denoted by $\mathcal{I}$, can be diagrammatically denoted as:

$$
\begin{array}{ccccc}
\mathcal{I} & :: & \mathsf{N} & \mapsto & \mathcal{I}[\mathsf{N}] \in \mathbb{R}\,.
\end{array}
\tag{30}
$$

Note that $\mathcal{I}$ should not be interpreted as a process within the GPT – $\mathcal{I}$ is simply some linear functional, and can lead to negative values.

The value of a Bell inequality on a stochastic map N – realised within the GPT as per Eq. (27) – is given by:

$$
\tag{31}
$$

The maximal value of a Bell inequality $\mathcal{I}$ achievable by correlations within a given GPT $\mathcal{G}$, is therefore given by the following optimisation problem:

$$
\mathcal{I}_{\max} := \sup \left\{ \; \middle| \; V, W, M_A, M_B, s \in \mathcal{G} \right\}\,.
\tag{32}
$$

Notice that the optimisation is carried over the types of systems V and W present in $\mathcal{G}$, as well as over the local measurements M_A and M_B, and bipartite states s. The solution to this optimisation problem will of course depend on the properties of the GPT being studied. However, we readily see that when maximising over states and measurements from the theory, the compatibility constraints we discussed in the previous section will play a crucial role. Indeed, if the GPT admits some bipartite effect e, then the above mentioned hierarchy of constraints (see Eqs. (18), (19), (20), and (21)) will restrict the sets of states that the value of $\mathcal{I}$ is optimised over. In other words, the existence of bipartite effects e within the GPT will impose a hierarchy of constraints on the correlations that such GPT may feature. In the next sections we elaborate on this fact with a concrete example.

5   Parity reading measurement

In this section we will explore the constraints on the correlations that a GPT may feature, given that bipartite effects associated to a particular measurement – which we call a *Parity reading measurement* (PRM) – exist within the GPT.

Suppose we have a controlled measurement, M, for a system V, with a setting variable labeled by the set $\boldsymbol{\eta} := \{0, ..., n-1\}$ such that $|\boldsymbol{\eta}| = n$, and a binary outcome variable $\boldsymbol{\beta} := \{0, 1\}$

as an outcome. This is diagrammatically denoted by:

$$\begin{array}{c} |\boldsymbol{\beta} \\ \overline{M} \\ |\boldsymbol{\eta} \; |V \end{array} \quad . \tag{33}$$

Recall that, as this is a measurement, it must satisfy:

$$\begin{array}{c} \overline{\overline{=}} \\ |\boldsymbol{\beta} \\ \overline{M} \\ |\boldsymbol{\eta} \; |V \end{array} \quad = \quad \begin{array}{cc} \overline{=} & \overline{=} \\ |X & |V \end{array} \quad , \tag{34}$$

which ensures that the correlations it can generate are no-signalling.

Then we can define a measurement $\mathcal{P}[M]$ which reads out the parity of such a measurement M as follows:

Definition 5.1. *Parity Reading Measurement.*–
A parity reading measurement *for M, denoted by $\mathcal{P}[M]$, is a bipartite measurement on $V \otimes V$ with n binary variables as outputs:*

$$\begin{array}{c} \overbrace{\qquad\qquad}^{n} \\ |\boldsymbol{\beta}_0 \cdots \; |\boldsymbol{\beta}_{n-1} \\ \boxed{\mathcal{P}[M]} \\ |V \qquad |V \end{array} \quad , \tag{35}$$

such that tracing out all but the i-th outcome gives the parity of the i-th setting for M:

$$\forall i \in \boldsymbol{\eta} \quad , \qquad \begin{array}{c} \overline{=} \quad \overline{=} \; |\overline{\overline{=}}\; \quad \overline{=} \\ \diagdown\boldsymbol{\beta}_0 \cdots \diagdown\boldsymbol{\beta}_{i-1}\; |\boldsymbol{\beta}_i\; \diagdown\boldsymbol{\beta}_{i+1}\cdots \diagdown\boldsymbol{\beta}_{n-1} \\ \boxed{\mathcal{P}[M]} \\ |V \qquad |V \end{array} \quad = \quad \begin{array}{c} |\boldsymbol{\beta} \\ \oplus \\ \boldsymbol{\beta}\diagup \quad \diagdown\boldsymbol{\beta} \\ \boxed{M}\;\;\boxed{M} \\ \boxed{\boldsymbol{\eta}}\;\;\;\boxed{\boldsymbol{\eta}} \\ \underset{i}{\bigtriangledown}|V\;\;\underset{i}{\bigtriangledown}|V \end{array} \quad . \tag{36}$$

We will also be interested in situations in which we have a measurement which can only read the parity of a certain subset of the setting variable $\boldsymbol{\iota} \subseteq \boldsymbol{\eta}$:

Definition 5.2. *Partial Parity Reading Measurement.*–
A partial parity reading measurement *for the settings $\boldsymbol{\iota} \subseteq \boldsymbol{\eta}$ of measurement M, denoted by $\mathcal{P}[M]_{\boldsymbol{\iota}}$ is simply a parity reading measurement for the measurement:*

$$\begin{array}{c} |\boldsymbol{\beta} \\ \overline{M_{\boldsymbol{\iota}}} \\ |\boldsymbol{\iota} \; |V \end{array} \quad := \quad \begin{array}{c} |\boldsymbol{\beta} \\ \overline{M} \\ \boxed{\boldsymbol{\eta}} \\ \bullet \\ |\boldsymbol{\iota} \; |V \end{array} \tag{37}$$

where $\bullet$ is the canonical embedding of $\boldsymbol{\iota}$ into $\boldsymbol{\eta}$[3]. Using this notation we can succinctly define these partial partity reading measurements by:

$$\mathcal{P}[M]_{\boldsymbol{\iota}} := \mathcal{P}[M_{\boldsymbol{\iota}}] \, . \tag{38}$$

---

[3]That is, it maps $\boldsymbol{\iota}$ viewed as a set in its own right into $\boldsymbol{\iota}$ viewed as a subset of $\boldsymbol{\eta}$

Note that when $\iota = \boldsymbol{\eta}$ then we recover the notion of a PRM.

In order to see how the existence of a (partial) PRM constraints the correlations that the GPT may feature, we will fist discuss the concepts of a *Fiducial Measurement* and *Fiducial effects*.

Let n be the affine dimension of the normalised state space, that is, $n = |V| - 1$. A *fiducial measurement*, $\mathcal{F}$, is a controlled measurement with n settings (described by the set $\boldsymbol{\eta}$) and binary outcomes (described by the set $\boldsymbol{\beta}$):

$$
\begin{array}{c}
|\boldsymbol{\beta} \\
\boxed{\mathcal{F}} \\
|\boldsymbol{\eta} \; |V
\end{array}
\tag{39}
$$

$\mathcal{F}$ is called a fiducial measurement if all of the fiducial effects can be obtained from such a measurement. Fiducial effects, in turn, form a (minimal) spanning set for the effect space of the GPT. As an example, consider the case where $n = 2$: here $\mathcal{F}$ will have a binary input system, and the three fiducial effects will be given by

$$
\left\{
\begin{array}{ccc}
\overbrace{0}^{} & \overbrace{0}^{} & \overline{\overline{=}} \\
\boxed{\mathcal{F}} & \boxed{\mathcal{F}} & \boxed{\mathcal{F}} \\
\underbrace{0}_{} & \underbrace{1}_{} & \underbrace{0}_{}
\end{array}
\;,\; = \; \overline{\overline{=}}_{V} \; = \;
\begin{array}{c}
\overline{\overline{=}} \\
\boxed{\mathcal{F}} \\
\underbrace{1}_{V}
\end{array}
\right\}
,
\tag{40}
$$

where the equality for the third effect comes from the fact that $\mathcal{F}$ is a valid controlled measurement and hence satisfies:

$$
\begin{array}{c}
\overline{\overline{\overline{=}}} \\
|\boldsymbol{\beta} \\
\boxed{\mathcal{F}} \\
|\boldsymbol{\eta} \; |V
\end{array}
\quad = \quad
\begin{array}{cc}
\overline{\overline{=}} & \overline{\overline{=}} \\
|\boldsymbol{\eta} & |V
\end{array}
\;.
\tag{41}
$$

Coming back to the case of an arbitrary n, notice that the fact that fiducial effects,

$$
\left\{
\begin{array}{c}
\overbrace{e_j}^{} \\
|V
\end{array}
\right\}_{j=0:n}
\tag{42}
$$

span the corresponding vector space, means that any state s can be uniquely characterised by the vector of probabilities:

$$
\mathbf{p}_s \; := \;
\left(
\begin{array}{c}
\overbrace{e_0} \\
|V \\
\underbrace{s}
\end{array}
\;, \; \ldots \;,\;
\begin{array}{c}
\overbrace{e_n} \\
|V \\
\underbrace{s}
\end{array}
\right)^T
\;.
\tag{43}
$$

For example, going back to the case where $n = 2$, this vector of probabilities will be given by:

$$
\mathbf{p}_s \; := \;
\left(
\begin{array}{ccc}
\overbrace{0} & \overbrace{0} & \overline{\overline{=}} \\
\boxed{\mathcal{F}} & \boxed{\mathcal{F}} & |V \\
\underbrace{0}_{V} & \underbrace{1}_{V} & \underbrace{s} \\
\underbrace{s} & \underbrace{s} &
\end{array}
\right)^T
\;.
\tag{44}
$$

Now we can briefly state the case we will explore in this section: GPTs that have a PRM for a Fiducial measurement, and where a bipartite Bell experiment is carried by Alice and Bob performing this controlled fiducial measurement in each wing.

A PRM $\mathcal{P}[\mathcal{F}]$ for the fiducial measurement $\mathcal{F}$ will satisfy the following constraints:

$$\forall i \in \boldsymbol{\eta}. \qquad (45)$$

It is worth mentioning that a PRM $\mathcal{P}[\mathcal{F}]$ is not necessarily uniquely singled out by these constraints – more than one PRM may qualify as potential candidates for the role. We denote by $\mathsf{ParMeas}[\mathcal{F}]$ the set of all PRM $\mathcal{P}[\mathcal{F}]$ that satisfy Eq. (45) for the given $\mathcal{F}$.

5.1 Examples

Qubits. Let us conclude this discussion with an example from quantum theory. Consider the case of qubit systems, where the affine local dimension is $n = 3$. A fiducial measurement corresponds to measuring the three Pauli observables X, Y, and Z. A PRM is given by (a suitable post-processing of) the Bell measurement

$$\left\{ \left| \phi^+ \right\rangle\!\left\langle \phi^+ \right|, \left| \phi^- \right\rangle\!\left\langle \phi^- \right|, \left| \psi^+ \right\rangle\!\left\langle \psi^+ \right|, \left| \psi^- \right\rangle\!\left\langle \psi^- \right| \right\}, \qquad (46)$$

with $\left| \phi^\pm \right\rangle = \frac{\left| 00 \right\rangle \pm \left| 11 \right\rangle}{\sqrt{2}}$ and $\left| \psi^\pm \right\rangle = \frac{\left| 01 \right\rangle \pm \left| 10 \right\rangle}{\sqrt{2}}$. Denoting the four element set of outcomes of the Bell measurement as $\mathbf{B} := \{0, 1, 2, 3\}$, and the qubit system by Q_2, we can diagrammatically represent this as:

$$(47)$$

To see that a post-processing is necessary for this measurement to fit into our definition of a PRM is easy: the measurement of Eq. (47) is a four outcome measurement, but, a PRM for X, Y, and Z should have three binary outcomes. The required post-processing can be described diagrammatically as:

$$(48)$$

where the white dot first makes three copies of the outcome, and the processes C_X, C_Y and C_Z correspond to the three different equal bipartitions of $\mathbf{B}$. For example:

$$(49)$$

16

To read the parity of observable ZZ, we hence apply post-processing $\{0,1\} \to 0$ and $\{2,3\} \to 1$ depicted in Eq. (49), which quantum mechanically comes from

$$
\begin{aligned}
|\phi_+\rangle\langle\phi_+| + |\phi_-\rangle\langle\phi_-| &= |00\rangle\langle00| + |11\rangle\langle11|, \\
|\psi_+\rangle\langle\psi_+| + |\psi_-\rangle\langle\psi_-| &= |01\rangle\langle01| + |10\rangle\langle10|, \\
\sigma_z \otimes \sigma_z &= |00\rangle\langle00| + |11\rangle\langle11| - (|01\rangle\langle01| + |10\rangle\langle10|).
\end{aligned}
\tag{50}
$$

All three required post-processings, and the observables whose parity they read, are presented below:

$$
\begin{array}{ll|c}
\text{post-processing} & & \text{parity} \\
\hline
C_X: & \{\phi_+, \psi_+\} \to 0, \{\phi_-, \psi_-\} \to 1 & XX \\
C_Y: & \{\phi_-, \psi_+\} \to 0, \{\phi_+, \psi_-\} \to 1 & YY \\
C_Z: & \{\phi_+, \phi_-\} \to 0, \{\psi_+, \psi_-\} \to 1 & ZZ
\end{array}
\tag{51}
$$

With this we conclude the argument for why the measurement given by

$$
\tag{52}
$$

is a parity reading measurement for the X, Y and Z observables.

Mirror quantum correlations. Such correlations have been considered in Ref. [63]. To introduce them, let us first notice a feature of the Bell measurement:

- the effect corresponding to ψ_- appears only in the bipartitions of **B** that give rise to anticorrelations, i.e., a value of 1 for the classical system $\boldsymbol{\beta}$,
- the other three effects appear each only once in a bipartition that measures anticorrelations.

The following table summarises this feature, where we specify, for each observable whose parity we want to read, whether each effect belongs to the correlation (0) or anticorrelation (1) bipartitions:

$$
\begin{array}{c|ccc}
\text{effect} & XX & YY & ZZ \\
\hline
\phi_+ & 0 & 1 & 0 \\
\phi_- & 1 & 0 & 0 \\
\psi_+ & 0 & 0 & 1 \\
\psi_- & 1 & 1 & 1
\end{array}
\tag{53}
$$

We then say that quantum parity reading measurement has *signature*

$$
\{(+, -, +), (-, +, +), (+, +, -), (-, -, -)\}.
\tag{54}
$$

In the case of *mirror quantum mechanics* [63] the table in Eq. (53) does not hold anymore, and instead the following are satisfied:

$$
\begin{array}{c|ccc}
\text{effect} & XX & YY & ZZ \\
\hline
(\phi_+)^{PT} & 1 & 0 & 1 \\
(\phi_-)^{PT} & 0 & 1 & 1 \\
(\psi_+)^{PT} & 1 & 1 & 0 \\
(\psi_-)^{PT} & 0 & 0 & 0
\end{array}
\tag{55}
$$

where PT stands for partial transposition[4].

Thus in mirror quantum case we have signature:

$$\{(-,+,-),(+,-,-),(-,-,+),(+,+,+)\}. \tag{56}$$

This case then corresponds to the following post-processing of the measurement outcomes:

post-processing		parity
$C_X :$	$\{0,2\} \to 0, \{1,3\} \to 1$	XX
$C_Y :$	$\{0,3\} \to 0, \{1,2\} \to 1$	YY
$C_Z :$	$\{2,3\} \to 0, \{0,1\} \to 1$	ZZ

$$\tag{57}$$

Classical theory. In the quantum case, the parity reading measurement was entangled. It had to be so, because it measured parities of observables that are not jointly measurable. In classical theory we can consider three bit system, described by X, Y, Z which are now jointly measurable. Then the PRM just amounts to post-process the joint measurement of all the 6 observables (three per party).

5.2 $\mathcal{P}[\mathcal{F}]$ and maximal violations of Bell inequalities for fiducial measurements $\mathcal{F}$

So, how does the existence of a $\mathcal{P}[\mathcal{F}]$ in the GPT constrain the correlations we may observe in a Bell test? More precisely, what are the constraints on the correlations that a bipartite system of a certain type can produce when there exists a PRM for the fiducial measurements on that system type? In this section we aim at optimising the value of a Bell inequality $\mathcal{I}$ when the measurements that the parties perform are given by $\mathcal{F}$ on a system of type V.

Notice that, in general, the cardinality of the input settings for the Bell test need not coincide with the number of settings for the fiducial measurement $\mathcal{F}$. That is, $|X| = |\boldsymbol{\eta}| = |Y|$ does not necessarily hold. In this manuscript we will work with the case where $|X| \leq |\boldsymbol{\eta}|$ and $|Y| \leq |\boldsymbol{\eta}|$, and hence only some of the settings of the controlled measurement $\mathcal{F}$ might be used for the Bell test. In such a case, then, we will focus on the constraints that a partial PRM imposes when its existence is demanded on the settings of $\mathcal{F}$ used in the Bell test. For simplicity in the discussion, in this section we will present the case where $|X| = |\boldsymbol{\eta}| = |Y|$, but the most general case follows similarly. We will return to the optimisation problems for partial PRM later on in the manuscript.

The optimisation problem that we focus on then reads:

$$\tag{58}$$

where X, Y, A, B are all binary variables, and we are using the shorthand notation:

$$\tag{59}$$

---

[4]In mirror quantum theory, these effects are partially transposed effects of Bell measurement.

Given a particular GPT $\mathcal{G}$ – and, in particular, given a specification of its state and effect spaces – this optimisation problem reduces to a type of cone program which has been explored in recent literature [47, 48, 64] regarding their relationship to GPTs. That is, there is some convex spanning cone of states $K_{V \otimes V} \subset V \otimes V$, which s belongs to, and some normalisation constraint on s, $u_V \otimes u_V(s) = 1$ so the above problem can be rewritten as:

$$\mathcal{I}_{\max} := \sup \left\{ \vcenter{\hbox{}} \; \middle| \; s \in K_{V \otimes V}, u_V \otimes u_V(s) = 1 \right\}. \tag{60}$$

Here, however, we do not consider a particular GPT $\mathcal{G}$ which we optimise over. What we carry out here is an optimisation over the space of GPTs which have the relevant structure – those which admit a PRM for the fiducial measurement. This optimisation problem is much more complex than that of Eq. (60), as we will now explain.

In Section 4.1 we elaborated on the types of compatibility constraints between states and effects that a GPT must feature. Here, we will demand that the GPT admits the fiducial local measurement $\mathcal{F}$ and a PRM $\mathcal{P} \in \mathsf{ParMeas}[\mathcal{F}]$. By imposing the compatibility constraints motivated in Section 4.1, we hence restrict the possible cones of states $K[\mathcal{P}]$ that such GPT could feature. If we have a characterisation of $K[\mathcal{P}] \subset V \otimes V$, then the optimisation problem becomes:

$$\mathcal{I}_{\max} := \sup \left\{ \vcenter{\hbox{}} \; \middle| \; \mathcal{P} \in \mathsf{ParMeas}[\mathcal{F}], u_V \otimes u_V(s) = 1, s \in K[\mathcal{P}] \right\}. \tag{61}$$

This turns out to be a non-linear optimisation problem, as we will show next.

5.3 The cone K of bipartite states

The key question here is: how to characterise the cones of states $K[\mathcal{P}]$? Here we will take the types of diagrammatic constraints motivated in Section 4.1, and define a systematic hierarchy of conditions that the existence of $\mathcal{P}[\mathcal{F}]$ imposes on the cone K. This hierarchy of conditions will be specified in terms of the number of copies of the bipartite state s featured in the diagram. Framing these constraints in the form of a hierarchy is useful because, as we will see, interesting results can be obtained without needing to impose all of the constraints. For example, in our case we will be interested in possible violations of a given Bell inequality, and, we can obtain upper bounds on this by simply working at the second level of the hierarchy.

Hierarchy constraints – Level 1.
Given the fiducial measurement $\mathcal{F}$ and the PRM $\mathcal{P} \in \mathsf{ParMeas}[\mathcal{F}]$, the normalised bipartite states

$s \in K[\mathcal{P}]$ *must satisfy:*

$$\geq 0 \quad , \tag{62}$$

$$\geq 0 \quad , \tag{63}$$

$$\geq 0 \quad , \tag{64}$$

where by ≥ 0 we mean that every matrix element is non-negative.

Notice that the constraints that come from the deterministic effect u – in particular, the normalisation condition $u_V \otimes u_V(s) = 1$ – together with these positivity constraints, ensures that diagrams in Eqs. 62, 63, and 64 are stochastic maps.

Notice moreover that the constraint of Eq. (64) has the same structure as that of Eq. (63) but applied instead to the swapped state:

$$. \tag{65}$$

We see then that there is a certain structure emerging: (i) there are two layers – one corresponding to the state s and one to the measurements $\mathcal{F}$ and $\mathcal{P}$ –, and (ii) we can vary the order in which the output wires of the state are plugged into the measurements of the second layer. This motivates the definition for the remaining levels of the hierarchy, which relies on the concept of a *wiring*, which we explain next.

Definition 5.3 (Wiring).
A process which describes how a collection of input systems are connected to a collection of output systems is here referred to as a wiring, *and denoted usually by W.*
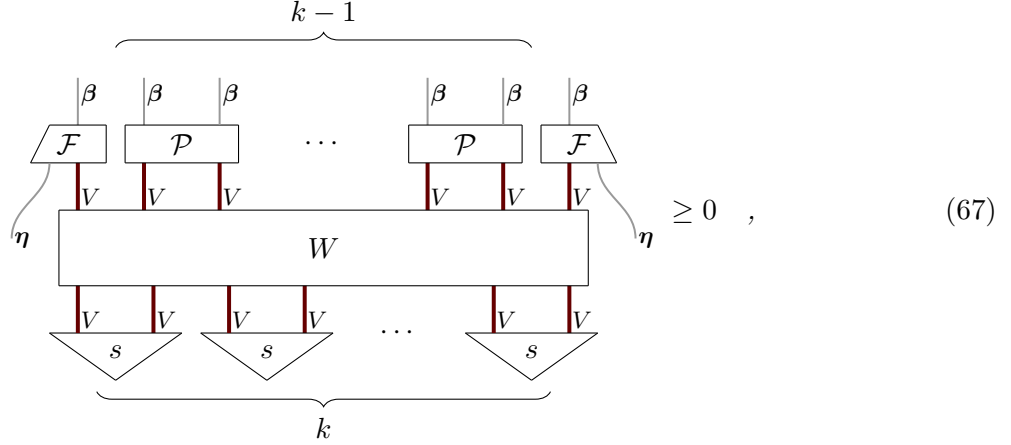
When all the input systems are of the same type – a case we focus on here – wirings reduce to permutations of the systems, e.g.:
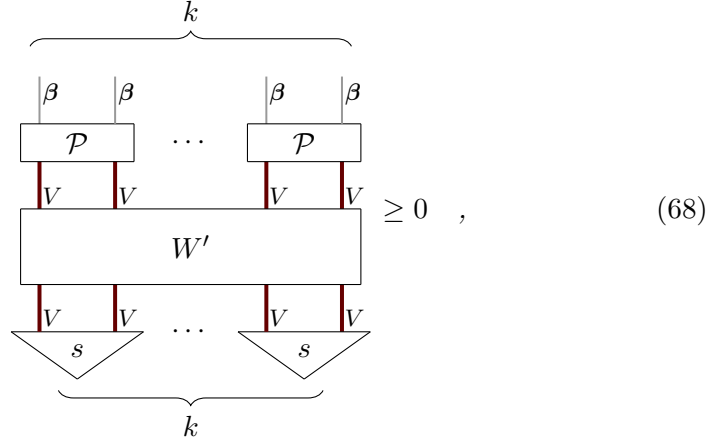
$$. \tag{66}$$

Hierarchy constraints – Level k.

Given the fiducial measurement $\mathcal{F}$ and the PRM $\mathcal{P} \in \mathsf{ParMeas}[\mathcal{F}]$, the normalised bipartite states $s \in K[\mathcal{P}]$ must satisfy the constraints imposed by Hierarchy Level k' for all $k' < k$, as well as the following:

$$\geq 0 \quad , \tag{67}$$

$$\geq 0 \quad , \tag{68}$$

for all distinct totally connected wirings W and W'.

   Here, two wirings are distinct *if they do not give the same diagram under some permutation of the external wires. A wiring is* totally connected *if the diagram it gives rise to does not factorise.*

   First, notice that if the constraints of Level k+1 are satisfied, then the constraints of Level k are also by definition satisfied. Moreover, the constraint that the wirings are distinct ensures that there is no redundancy within a particular level in the hierarchy. In addition, the condition that the wirings are totally connected ensures that the constraints they impose do not reduce to constraints at lower levels in the hierarchy. We discuss the convergence of this hierarchy in App. E. Enumerating and finding a simple description of the distinct totally connected wirings is left as an interesting open problem.

   We see that the constraints that each level k imposes are then of two types: (i) one where the process in the second layer is the product of k copies of $\mathcal{P}$ – Eq. (68) –, and (ii) one where the process in the second layer is the product of $k-1$ copies of $\mathcal{P}$ and two copies of $\mathcal{F}$ – Eq. (67). Note that if we had more copies of $\mathcal{F}$ (and so fewer of $\mathcal{P}$) then it would be impossible to have a totally connected wiring, hence we need consider at most two copies of $\mathcal{F}$.

One particular example of a of constraint imposed by the hierarchy is

$$
\begin{array}{c}
\text{(diagram)}
\end{array}
\geq 0 \, . \tag{69}
$$

This condition, imposed first in Level 2, can be given the following interpretation: the PRM $\mathcal{P}$ must give valid probabilities on products of two steered states, each constructed from s.

6 The optimisation problems

The hierarchy of constraints presented in the previous subsection ultimately defines some convex cone. To see this, suppose σ_1 and σ_2 satisfy the compatibility constraints of Eqs. (67) and (68) for all k. Then, $r_1\sigma_1 + r_2\sigma_2$ will satisfy the constraints for all $r_1, r_2 \in \mathbb{R}^+$.

The optimisation problem of Eq. (61) is therefore carried out over Σ, the union of the cones $K[\mathcal{P}]$:

$$
s \in \bigcup_{\mathcal{P} \in \mathsf{ParMeas}[\mathcal{F}]} K[\mathcal{P}] \quad =: \quad \Sigma. \tag{70}
$$

This allows us to cast the optimization problem in a deceptively simple form:

Optimisation Problem 1.

$$
\mathcal{I}_{\max} := \sup \left\{ \left. \begin{array}{c} \text{(diagram)} \end{array} \right| s \in \Sigma, u_V \otimes u_V(s) = 1 \right\}. \tag{71}
$$

While this form of the optimisation may appear simple, determining membership of the set Σ is computationally extremely difficult. Indeed, Σ is defined as the union of a (potentially infinite) set of cones, each of which is defined by an infinite hierarchy of constraints. In the remaining of the paper we will see how to relax these constraints, to make the optimisation problem computationally tractable, and by so compute upper bounds to $\mathcal{I}_{\max}$. Note that, since the objective function is linear, we can make this a convex optimisation problem by optimising over the convex closure of Σ.

In addition, one may wish to optimise the value of a Bell inequality where the cardinality of the input variables in the Bell test does not coincide with the number of settings in the fiducial measurement, i.e., $|X| \leq |\boldsymbol{\eta}|$ and/or $|Y| \leq |\boldsymbol{\eta}|$. In this case, only a subset $\boldsymbol{\iota} \subset \boldsymbol{\eta}$ of the control settings are of interest, and the relevant constraint is the existence of a partial PRM for $\boldsymbol{\iota}$. In this case, the Optimisation problem becomes:

Optimisation Problem 1′.

$$\mathcal{I}_{\max} := \sup\left\{\ \cdots\ \middle|\ s \in \Sigma_{\boldsymbol{\iota}},\, u_V \otimes u_V(s) = 1 \right\}, \tag{72}$$

where $\Sigma_{\boldsymbol{\iota}}$ is the set of potential states compatible with the existence of a partial PRM $\mathcal{P}[\mathcal{F}]_{\boldsymbol{\iota}}$ for the settings $\boldsymbol{\iota} \subseteq \boldsymbol{\eta}$.

6.1 A relaxation to Optimisation Problem 1

In this section we will specify a particular subset of constraints imposed by the hierarchy that defines $K[\mathcal{P}]$. We will focus on a particular set of minimum requirements to demand to the GPT, which are colloquially stated as:

- Valid states under local fiducial measurements – Eq. (62),
- $\mathcal{P}[\mathcal{F}]$ is a valid measurement on generic composite states – Eq. (63),
- $\mathcal{P}[\mathcal{F}]$ is a valid measurement on products of steered states – Eq. (69),
- $\mathcal{P}[\mathcal{F}]$ is a PRM – Eq. (45).

Implementing these conditions is already computationally demanding, since we are optimising over the possible PRM $\mathcal{P}[\mathcal{F}]$ and the possible bipartite states s of the GPT.

The optimisation problem to solve therefore reads as follows:

Optimisation Problem 2.

$$\mathcal{I}_{\max}^{\mathrm{R}} = \sup_{\{s,\mathcal{P}\}}\ \cdots\ , \tag{73}$$

subject to ($s.t.$):

- $\quad \cdots = 1 \qquad [normalisation]\,,$
- $\quad \cdots \geq 0\,,\qquad \cdots \geq 0\,,\qquad \cdots \geq 0\,,$
- $\quad \cdots = \cdots \qquad \forall i \in \boldsymbol{\eta}\,.$

23

It is readily seen how the Optimisation Problem 2 is a relaxation of the Optimisation Problem 1 – a solution $\mathcal{I}_{\max}^{R}$ to the former will yield an upper bound to the solution $\mathcal{I}_{\max}$ of the latter.

To be able to implement this sort of optimisation problem on a computer, we must switch from the high level diagrammatic description of the processes in a GPT, to a lower level tensorial representation. The method for doing this is presented in App. D, together with a spelled-out example for a particular scenario.

The constraints that appear in Optimisation Problem 2 can then be recast by means of the tensor representation as constraints on real vectors. This lower level form of Eqs. (62), (63), (69), and (45), will be used when coding the scripts to carry out the numerical calculations of the next section.

Example 1.
Consider the case where $n = 2$. Here, as we discussed in Section 5, a normalised state s can be fully parametrised as in Eq. (44) by the vector of probabilities:

$$\mathbf{p}_s = (p_s(0|0), \, p_s(0|1), \, 1)^T \,, \tag{74}$$

where $p_s(a|x)$ is the probability that outcome a is obtained when the fiducial measurement x is performed on a system on state s. In a locally tomographic GPT, a bipartite system can be parametrised as follows:

$$\mathbf{p}_s = \left(p_s(00|00), \, p_s(00|10), \, p_s^{(2)}(0|0), \, p_s(00|01), \, p_s(00|11), \, p_s^{(2)}(0|1), \, p_s^{(1)}(0|0), \, p_s^{(1)}(0|1), \, 1 \right)^T, \tag{75}$$

where $p_s^{(j)}(a|x)$ denotes the marginal conditional probability of subsystem j, and $p_s(ab|xy)$ denotes the joint conditional probabilities. Note that these parameters are also precisely those required to characterise a no-signalling box with binary inputs and outputs. This is not a coincidence – indeed, this is precisely the no-signalling box that we will obtain when we measure this state with the fiducial measurement $\mathcal{F}$ on both systems. Hence, this parameterisation of the bipartite state and the form of $\mathcal{F}$ ensure that the observed correlations are no-signalling.

Using the tensorial notation described in App. D and, in particular, the above parameterisation of the composite state (Eq. (75)), the constraints in Optimisation Problem 2 can be recast as follows.

The first one, i.e., Eq. (62), reads:

$$p_s(ab|xy) \geq 0 \ \forall a,b,x,y \in \{0,1\} \,, \quad \sum_{a,b=0:1} p_s(ab|xy) = 1 \ \forall x,y \in \{0,1\} \,, \tag{76}$$

$$\sum_{a=0:1} p_s(ab|xy) = p_s(b|y) \quad \sum_{b=0:1} p_s(ab|xy) = p_s(a|y) \quad \forall a,b,x,y \in \{0,1\} \,. \tag{77}$$

That is, the outcome statistics of fiducial measurements on a state s are a well-defined no-signalling normalised conditional probability distribution.

The second constraint, i.e., Eq. (63), reads

$$\sum_{vw} \mathcal{P}_{vw}^{qr} \, s^{vw} \geq 0 \quad \forall q,r \in \{0,1\} \,, \tag{78}$$

where v and w are the indices associated to the two GPT vector spaces V, and q and r are the indices associated to the classical outcomes of the parity reading measurement, that is,

each corresponds to one of the parities which is being read. Since s^{vw} is represented by a 9-dimensional probability vector $\mathbf{p}_s$, $\mathcal{P}_{vw}^{qr}$ may be represented by a 4×9 matrix, $[\mathcal{P}]$.

The third constraint, i.e., Eq. (69), reads:

$$\sum_{v_1,w_1,v_2,w_2} \mathcal{F}_{xv_1}^a \, \mathcal{P}_{w_1v_2}^{rq} \, \mathcal{F}_{yw_2}^b \, s^{v_1 w_2} \, s^{v_2 w_2} \geq 0 \quad \forall x,y,q,r,a,b \in \{0,1\}, \tag{79}$$

where the indices x and y are the indices associated to the measurement settings of the two fiducial measurements, and the indices a and b to their outcomes. Notice that the tensors $\mathcal{F}_{xv_1}^a$ and $\mathcal{F}_{yw_2}^b$ correspond to the definition of the fiducial effects for system of type V. Specifically, we can write that:

$$e_{a|x} = (\mathcal{F}_{x0}^a, \mathcal{F}_{x1}^a, \mathcal{F}_{x2}^a) \quad \text{and} \quad e_{b|y} = (\mathcal{F}_{y0}^b, \mathcal{F}_{y1}^b, \mathcal{F}_{y2}^b). \tag{80}$$

Hence, this equation can be further written as:

$$[\mathcal{P}] \circ \left((e_{a|x} \otimes \mathbb{1}_V) \circ s \right) \otimes \left((\mathbb{1}_V \otimes e_{b|y}) \circ s \right) \geq 0 \quad \forall x,y,a,b \in \{0,1\}, \tag{81}$$

where by ≥ 0 we mean that every element of the matrix must be ≥ 0. This equivalent form of Eq. (69) makes it clear to see that it indeed imposes that $\mathcal{P}[\mathcal{F}]$ is a valid measurement on products of steered states, which are steered by fiducial measurements. We can denote the (subnormalised) steered states explicitly by

$$s_{b|y}^{(1)} := (\mathbb{1}_V \otimes e_{b|y}) \circ s \tag{82}$$

$$s_{a|x}^{(2)} := (e_{a|x} \otimes \mathbb{1}_V) \circ s. \tag{83}$$

Then, the condition (69) can be finally written as

$$[\mathcal{P}] \circ s_{b|x}^{(1)} \otimes s_{a|x}^{(2)} \geq 0 \quad \forall x,y,a,b \in \{0,1\}, \tag{84}$$

that is, the parity reading measurement must give valid probabilities on products of steered states.

The last constraint, given by Eq. (45), can be recast in the $n = 2$ case as follows:

$$\sum_{q=0:1} \mathcal{P}_{vw}^{qr} = \mathcal{F}_{1v}^0 \mathcal{F}_{1w}^r + \mathcal{F}_{1v}^1 \mathcal{F}_{1w}^{1 \oplus r} \quad \forall r \in \{0,1\}, \tag{85}$$

$$\sum_{r=0:1} \mathcal{P}_{vw}^{qr} = \mathcal{F}_{0v}^0 \mathcal{F}_{0w}^q + \mathcal{F}_{0v}^1 \mathcal{F}_{0w}^{1 \oplus q} \quad \forall q \in \{0,1\}, \tag{86}$$

where $\oplus$ denotes sum mod 2. These tensorial equations indeed correspond to equality constraints between 9-dimensional covectors:

$$u_{\boldsymbol{\beta}} \otimes \vec{r} \circ [\mathcal{P}] = e_{0|1} \otimes e_{r|1} + (u_V - e_{0|1}) \otimes (u_V - e_{r|1}) \quad \forall r \in \{0,1\}, \tag{87}$$

$$\vec{q} \otimes u_{\boldsymbol{\beta}} \circ [\mathcal{P}] = e_{0|0} \otimes e_{q|0} + (u_V - e_{0|0}) \otimes (u_V - e_{q|0}) \quad \forall q \in \{0,1\}. \tag{88}$$

which can be straightforwardly verified by noting that:

$$u_{\boldsymbol{\beta}} = (1,1), \quad \vec{0} = (1,0), \quad \vec{1} = (0,1), \tag{89}$$

and that

$$u_V = (\mathcal{F}_{x0}^0 + \mathcal{F}_{x0}^1, \ \mathcal{F}_{x1}^0 + \mathcal{F}_{x1}^1, \ \mathcal{F}_{x2}^0 + \mathcal{F}_{x2}^1) \quad \forall \ x \in \{0,1\}. \tag{90}$$

$\blacksquare$

Optimisation Problem 2, despite being a relaxation of Optimisation Problem 1 , still shares a common feature with the latter: they are both nonlinear optimisation problems. Indeed, we can see clearly in the formulation of Optimisation Problem 2 how the constraints feature products of the variables being optimised over. Solutions to such polynomial optimisation problems may be approximated by standard techniques in the literature. Here, we will consider the hierarchy of semidefinite relaxations to polynomial optimisation problems given by Lasserre [65]. Each level of such hierarchy will give an upper bound to the solution $\mathcal{I}_{\max}^{\mathrm{R}}$ of Optimisation Problem 2.

Optimisation Problem 2 is formulated for the situations where the cardinality of the input variables in the Bell test match the number of settings in the fiducial measurement, i.e., $|X| = |Y| = |\boldsymbol{\eta}|$. However, as we mentioned in Section 5.2, this is not always necessarily the case. We will therefore next reformulate Optimisation Problem 2 to encompass the case of partial PRMs. This adjusted version of the optimisation problem will come in handy when exploring quantum correlations, since for example it allows the study of Bell inequalities with two measurement settings per wing (see, e.g., the CHSH scenario in which $|X| = |Y| = 2$) on qubits (whose affine dimension is 3 rather than 2). In addition, importantly, this adjusted version of the optimisation problem might allow us to make device-independent studies of the results[5], since do not require full knowledge of the dimension of the local systems to impose the constraint of existence of partial PRMs. Namely, we hope that the constraints for correlations obeyed by set of local observables imposed by existence of PRM persists, regardless of the dimension of the system.

---

[5]As mentioned in Sec. 3, in such a black-box approach to describing the states of the systems we only want to use the information provided by the correlations read out in the experiment. Hence, we cannot make any assumptions on what the dimension of the underlying physical systems is.

Optimisation Problem 2′.

Let $\iota \subset \eta$.

$$\mathcal{I}_{\max}^{\mathrm{RP}} \quad = \quad \sup_{\{s,\mathcal{P}_\iota\}} \quad \boxed{\begin{array}{c}\mathcal{I}\\ \mathcal{F}_\iota \quad \mathcal{F}_\iota \\ V \quad V \\ s \end{array}} \quad , \tag{91}$$

$$\text{s.t.} \quad \begin{cases} \dfrac{\overline{V}\ \overline{V}}{s} = 1\,, & [\textit{normalisation}] \\[2ex] \begin{array}{c}\beta \quad \beta \\ \mathcal{F} \quad \mathcal{F} \\ V \quad V \\ s \\ \eta \quad \eta \end{array} \geq 0\,, \quad \begin{array}{c}\beta\cdots\beta \\ \mathcal{P}_\iota \\ V \quad V \\ s \end{array} \geq 0\,, \quad \begin{array}{c}\beta \quad \beta\cdots\beta \quad \beta \\ \mathcal{F}\quad \mathcal{P}_\iota \quad \mathcal{F} \\ V\ V \quad V\ V \\ \eta\ \ s \quad s\ \ \eta \end{array} \geq 0\,, \\[4ex] \begin{array}{c}\beta_i \\ \overline{\beta_0}\cdots\overline{\beta_{i-1}}\ \overline{\beta_{i+1}}\cdots\overline{\beta_{|\iota|-1}} \\ \mathcal{P}_\iota \\ V \quad V \end{array} = \begin{array}{c}\beta \\ \oplus \\ \beta \quad \beta \\ \mathcal{F}_\iota \quad \mathcal{F}_\iota \\ \iota_i \quad \iota_i \\ V \quad V \end{array} \quad \forall i \in \iota \subseteq \eta\,. \end{cases}$$

Notice finally that Optimisation Problem 2′ is indeed a relaxation of Optimisation Problem 1′, in the same way that Optimisation Problem 2 is a relaxation of Optimisation Problem 1.

Example 2.

Consider the case where $n = 3$. Here, as we discussed in Section 5, a normalised state s can be fully parameterised as in Eq. (44) by the vector of probabilities:

$$\mathbf{p}_s = (p_s(0|0)\,,\, p_s(0|1)\,,\, p_s(0|2)\,,\, 1)^T\,, \tag{92}$$

where $p_s(a|x)$ is the probability that outcome a is obtained when the fiducial measurement x is performed on a system on state s. In a locally tomographic GPT, a bipartite system can be parameterised as follows:

$$\mathbf{p}_s = \Big(p_s(00|00)\,,\, p_s(00|10)\,,\, p_s(00|20)\,,\, p_s^{(2)}(0|0)\,,\, p_s(00|01)\,,\, p_s(00|11)\,,\, p_s(00|21)\,,\, p_s^{(2)}(0|1)\,,$$

$$p_s(00|02)\,,\, p_s(00|12)\,,\, p_s(00|22)\,,\, p_s^{(2)}(0|2)\,,\, p_s^{(1)}(0|0)\,,\, p_s^{(1)}(0|1)\,,\, p_s^{(1)}(0|2)\,,\, 1 \Big)^T\,, \tag{93}$$

where $p_s^{(j)}(a|x)$ denotes the marginal probability of subsystem j.

We will furthermore consider the case where $\iota = \{0,1\} \subset \{0,1,2\} = \eta$. With this choice, using the tensorial notation of App. D, the constraints in Optimisation Problem 2′ can be recast

as follows. The first one, i.e., Eq. (62), reads:

$$p_s(ab|xy) \geq 0 \ \forall a, b \in \{0,1\}, \ x, y \in \{0,1,2\}, \qquad \sum_{a,b=0:1} p_s(ab|xy) = 1 \ \forall x, y \in \{0,1,2\}, \qquad (94)$$

$$\sum_{a=0:1} p_s(ab|xy) = p_s(b|y) \qquad \sum_{b=0:1} p_s(ab|xy) = p_s(a|y) \qquad \forall a, b \in \{0,1\}, x, y \in \{0,1,2\}. \qquad (95)$$

That is, the outcome statistics of fiducial measurements on a state s are a well-defined no-signalling normalised conditional probability distribution.

The second constraint, i.e., Eq. (63), reads

$$\sum_{vw} \mathcal{P}_{\iota vw}^{qr} \, s^{vw} \geq 0 \quad \forall q, r \in \{0,1\}. \qquad (96)$$

Since s^{ij} is represented by a 16-dimensional probability vector $\mathbf{p}_s$, $\mathcal{P}_{\iota vw}^{qr}$ may be represented by a 4×16 matrix. Notice that the fact that $\mathcal{P}_\iota$ is a partial PRM is captured by the fact that it only has two output systems – hence its matrix representation has four rows.

The third constraint, i.e., Eq. (69), reads:

$$\sum_{v_1,w_1,v_2,w_2} \mathcal{F}_{xv_1}^a \, \mathcal{P}_{\iota w_1 v_2}^{rq} \, \mathcal{F}_{yw_2}^b \, s^{v_1 w_2} \, s^{v_2 w_2} \geq 0 \quad \forall a, b, q, r \in \{0,1\}, \ x, y \in \{0,1,2\}. \qquad (97)$$

Notice that the tensors $\mathcal{F}_{xv_1}^a$ and $\mathcal{F}_{yw_2}^b$ actually correspond to the definition of the fiducial effects for system of type V. If we represent $\mathcal{P}_{\iota vw}^{qr}$ by a 4×16 matrix $[\mathcal{P}_\iota]$, hence, this equation can be further written as:

$$[\mathcal{P}_\iota] \circ \left((e_{a|x} \otimes \mathbb{1}_V) \circ s \right) \otimes \left((\mathbb{1}_V \otimes e_{b|y}) \circ s \right) \geq 0 \quad \forall a, b \in \{0,1\}, \ x, y \in \{0,1,2\}, \qquad (98)$$

where by ≥ 0 we mean that every element of the matrix must be ≥ 0. This equivalent form of Eq. (69) makes it clear to see that it indeed imposes that $\mathcal{P}_\iota$ is a valid measurement on products of steered states, which are steered by fiducial measurements.

The last constraint, given by Eq. (45), can be recast in the $n = 3$ case as follows:

$$\sum_{q=0:1} \mathcal{P}_{\iota vw}^{qr} = \mathcal{F}_{1v}^0 \mathcal{F}_{1w}^r + \mathcal{F}_{1v}^1 \mathcal{F}_{1w}^{1 \oplus r} \quad \forall r \in \{0,1\}, \qquad (99)$$

$$\sum_{r=0:1} \mathcal{P}_{\iota vw}^{qr} = \mathcal{F}_{0v}^0 \mathcal{F}_{0w}^q + \mathcal{F}_{0v}^1 \mathcal{F}_{0w}^{1 \oplus q} \quad \forall q \in \{0,1\}, \qquad (100)$$

where $\oplus$ denotes sum mod 2. These tensorial equations indeed correspond to equality constraints between 16-dimensional covectors:

$$u_\beta \otimes \vec{r} \circ [\mathcal{P}] = e_{0|1} \otimes e_{r|1} + (u_V - e_{0|1}) \otimes (u_V - e_{r|1}) \quad \forall r \in \{0,1\}, \qquad (101)$$

$$\vec{q} \otimes u_\beta \circ [\mathcal{P}] = e_{0|0} \otimes e_{q|0} + (u_V - e_{0|0}) \otimes (u_V - e_{q|0}) \quad \forall q \in \{0,1\}. \qquad (102)$$

$\blacksquare$

7 Approximating $\mathcal{I}_{\max}$ for various Bell inequalities

In Ref. [23] it was shown that existence of the Bell measurement for two systems, each with three observables (i.e., for $|\boldsymbol{\eta}| = 3$) imposes that there are no post-quantum correlations. In our work we want to pose the more general problem of whether parity-readable observables can lead to post-quantum correlations. In order to tackle this question, we defined a hierarchy of

constraints that the existence of parity-readable observables imposes on the states and effects of the underlying GPT, and therefore on the sets of correlations that the GPT allows. We explored the boundary of the allowed correlations by tackling the Optimization Problems defined in the previous section. The results of these numerical explorations led us to further formulate a conjecture, re-stated below:

Conjecture 1. *Suppose we have a bipartite system in a GPT that satisfies local tomography and no-signaling. The, the observables for which there exists a parity reading measurement cannot be used to violate any Bell inequality more than quantum mechanics does.*
[Old below:]
Suppose that a bipartite system satisfies local tomography and no-signaling. Moreover, suppose that product of states is a valid state. Then the observables for which there exists parity reading measurement do not violate any Bell inequality more than quantum mechanics does.

Formally, the conjecture, if true, means that for any Bell inequality the Optimization Problem $1'$ returns at most the quantum bound. Note, however, that we do not always expect OP$1'$ to actually reach the quantum bound, as the maximal quantum value is not necessarily achieved by observables which are parity-readable within quantum theory.

In this section we present numerical computations towards upper-bounding the solution to OP2. Indeed, our numerical explorations focus on the relaxed problems OP2 and OP$2'$, depending on the cardinalities of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$. As mentioned in the previous section, we will approximate the solution to OP$2'$ by means of a hierarchy of semidefinite relaxations formulated by Lasserre [65], using mostly the Lasserre hierarchy levels 1+AB and 2. A brief explanation of what these two levels mean is presented below, and refer the reader to Ref. [65] for a thorough exposition. Throughout the next subsection we also discuss the relation between the numerical results and the conjecture we formulated.

Finally, in this section we provide an analytical proof that GPTs which violate local tomography admit PR-box correlations under the constraints of OP2 – that is, OP2 may yield a value of $\frac{1}{2}$ for the CHSH inequality (in the notation of Eq. (103)) within non-tomographically local GPTs. This highlights the relevance and impact of the assumption of local tomography. We also discuss how the violation of local tomography by a GPT may impact whether its correlations satisfy or not the conjecture. In particular, we discuss how the result we show does not necessarily imply that non-tomographically local GPTs violate Conjecture 1, since the actual optimisation problem to be solved – OP$1'$ – imposes additional constraints to those appearing in OP2.

Before moving on to presenting the numerical results, let us briefly comment on the so-called Lasserre hierarchy. Each Lasserre hierarchy level is related to the semidefiniteness of a matrix (whose definition we will not give here), and the rows and columns of this matrix have particular labels depending on what level we are focusing on. Let Υ be the set that contains the variables we are optimising over plus the element 1. In the first level of the Lasserre hierarchy, the matrix under study has row and columns labelled by the elements of Υ. In the second level of the hierarchy, however, the matrix under study is of much larger size, and its rows and columns are labelled by the elements of $\Upsilon \times \Upsilon$, where $\times$ denotes the Cartesian product of sets. The so-called 1+AB level lies in between the first and the second – the matrix corresponding to 1+AB is a sub-matrix of that of level 2, and the matrix corresponding to level 1 is a sub-matrix of that of 1+AB. In particular, the rows and columns of the matrix corresponding to level 1+AB are labelled by the elements of the set $\Upsilon \times \Upsilon \setminus \{(v, v) | v \in \Upsilon\}$.

The numerical computations from Sec. 7.1 were performed with Python 3.7. The SDP relaxation of polynomial programming was calculated with the package Ncpol2sdpa [66]. The SDP problem was solved using SDPA [67]. All other numerical computations were carried out

by a sparsity-adapted SDP relaxation of the polynomial optimization problem modeled with the TSSOS [68] algorithm. For more details on the modeling syntax, we refer the interested reader to the tutorial from [69, Appendix B.2] and the online website https://github.com/wangjie212/TSSOS. Each SDP problem was solved using Mosek [70].

7.1 CHSH inequality

In a bipartite Bell scenario featuring two dichotomic measurements per party, the most studied inequality is the Clauser, Horne, Shimony, Holt (CHSH) inequality [25], which, using the notation of Eq. (75), reads:

$$I_{\text{CHSH}}(\mathbf{p}_s) = -p_s^{(1)}(0|0) - p_s^{(2)}(0|0) + p_s(00|00) + p_s(00|10) + p_s(00|01) - p_s(00|11)\,. \quad (103)$$

This inequality is bounded from above, and the corresponding classical, quantum, and non-signalling bounds are:

$$\beta_{\text{CHSH}}^{\text{C}} = 0\,, \quad \beta_{\text{CHSH}}^{\text{Q}} = \frac{\sqrt{2}-1}{2} \sim 0.2071\,, \text{ and } \quad \beta_{\text{CHSH}}^{\text{NS}} = \frac{1}{2}\,. \quad (104)$$

Note that in this section we shall make a slight abuse of notation, and denote Alice and Bob's observables by X, Y, Z, which is not to be confused with the use of X and Y to denote the sets of inputs.

The numerical results presented in this subsection are summarized in Fig. 6.

Case $|\boldsymbol{\iota}| = |\boldsymbol{\eta}| = 2$. This is the simplest possible problem, where there are only two observables per party (that is, Alice has two observables X, Z, and the same for Bob) and the PRM measures the two parities XX and ZZ. We approximated the solution of OP2 applied to the CHSH inequality, by applying a Lasserre SDP relaxation with hierarchy level 'a bit lower than 1+AB'. We will specify shortly what this means, but will first elaborate on the specific parameterisation we chose for OP2. The state is described by 8 parameters coming from local tomography (see Eq. (75)) which we recall here in a more compact, matrix notation:

$$\mathbf{p_s} = \begin{bmatrix} p_s(00|00) & p_s(00|01) & p_s^{(1)}(0|0) \\ p_s(00|10) & p_s(00|11) & p_s^{(1)}(0|1) \\ p_s^{(2)}(0|0) & p_s^{(2)}(0|1) & 1 \end{bmatrix}\,. \quad (105)$$

The (unnormalized) Alice states steered by Bob given by Eq. (82) are expressed as:

$$\begin{aligned} s_{0|0}^{(1)} &= (p_s(00|00), p_s(00|10), p^{(2)}(0|0))^T\,, \\ s_{1|0}^{(1)} &= (p^{(1)}(0|0) - p_s(00|00), p^{(1)}(0|1) - p_s(00|10), 1 - p^{(2)}(0|0))^T\,, \\ s_{0|1}^{(1)} &= (p_s(00|01), p_s(00|11), p^{(2)}(0|1))^T\,, \text{ and} \\ s_{1|1}^{(1)} &= (p^{(1)}(0|0) - p_s(00|01), p^{(1)}(0|1) - p_s(00|11), 1 - p^{(2)}(0|1))^T\,. \end{aligned} \quad (106)$$

Bob's states steered by Alice have the same form as in Eq. (106) but exchanging $XZ \leftrightarrow ZX$ and $(1) \leftrightarrow (2)$. Now, the constraint that PRM measures parity (given by Eq. (85)) reads

$$\begin{aligned} \mathcal{P}^{00} + \mathcal{P}^{01} &= \begin{bmatrix} 2 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}\,, \\ \mathcal{P}^{10} + \mathcal{P}^{11} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\,. \end{aligned} \quad (107)$$

This can be obtained from Eq. (85) as follows. First we have, for example,

$$(\mathcal{P}^{00} + \mathcal{P}^{01}) \cdot \mathbf{p}_s = p_s(00|00) + p_s(11|00), \tag{108}$$

where $\cdot$ represents the Frobenius inner product of the two matrices. Then using

$$p_s(00|00) + p_s(01|00) = p^{(1)}(0|0),$$
$$p_s(11|00) + p_s(01|00) = p^{(2)}(0|0),$$
$$p_s(00|00) + p_s(01|00) + p_s(10|00) + p_s(11|00) = 1, \tag{109}$$

we get

$$(\mathcal{P}^{00} + \mathcal{P}^{01}) \cdot \mathbf{p}_s = 1 + 2p_s(00|00) - p^{(1)}(0|0) - p^{(2)}(0|0), \tag{110}$$

which leads to the form above. Preserving probability by PRM reads as

$$\mathcal{P}^{00} + \mathcal{P}^{01} + \mathcal{P}^{10} + \mathcal{P}^{11} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbb{1}. \tag{111}$$

Thus, the condition that $\mathcal{P}$ is a PRM is captured by the following free parameters:

$$\mathcal{P}^{00} - \mathcal{P}^{01} - \mathcal{P}^{10} + \mathcal{P}^{11} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \equiv \mathcal{C}. \tag{112}$$

Finally, the constraints that we still need to impose are the positivity of PRM effects both on the state, as well as on tensor products on all pairs of steered states that can be obtained from it. We see then that OP2 requires us to optimise over the free parameters (state s given by Eq. (105) and $\mathcal{C}$), under the positivity constraints of the previous sentence.

Now, to approximate the solution to OP2, we apply a particular level of the Lasserre hierarchy, which is slightly lower than the previously described 1+AB, and which we will denote by 1+AB*. Let Υ_p denote the set of free parameters given by given by Eq. (105), and $\Upsilon_{\mathcal{C}}$ that given by the free parameters in $\mathcal{C}$. Here, $\Upsilon = \Upsilon_p \bigcup \Upsilon_{\mathcal{C}} \bigcup \{1\}$. However, the matrix under study in the level we consider here has rows and columns labeled by the elements of $(\Upsilon \times \Upsilon) \setminus (\Upsilon_p \times \Upsilon_p) \setminus (\Upsilon_{\mathcal{C}} \times \Upsilon_{\mathcal{C}})$ – that is, it is a submatrix of that considered in level 1+AB.

The upper bound to OP2 given by the 1+AB* level of the Lasserre hierarchy, gives a value of ~ 0.2071, which agrees up to numerical precision with $\mathcal{I}^{\mathrm{R}}_{\max} = \frac{\sqrt{2}-1}{2}$. This equality follows from recalling that the Tsirelson's bound value can be achieved within Quantum theory by a Bell measurement, and hence yields a lower bound to $\mathcal{I}^{\mathrm{R}}_{\max}$. In other words, here we recover Tsirelson's bound for the CHSH inequality.

Case $|\boldsymbol{\iota}| = 2$, $|\boldsymbol{\eta}| = 3$. Here we still assume that the PRM measures just two parities (i.e., those of XX and ZZ), but now Alice and Bob have one more additional observable (i.e., Y). This is an important case, as it allows for the possibility that the constraints imposed by the existence of a PRM are sensitive to the dimension of the local systems. Notice that if the constrains stemming from the existence of a PRM turn out to be independent of the dimensions of the local systems, one can then take a device-independent approach to the problem and only rely on the black-box statistics to make assessments on the possible violations of the Bell inequality.

Our numerical results show that in this case Tsirelson's bound is also not violated. The results are computed exactly as in the previous case with $|\boldsymbol{\eta}| = 2$: we upperbound the value of $\mathcal{I}^{\mathrm{RP}}_{\max}$ via the 1+AB* level of the Lasserre hierarchy, which agrees up to numerical precision with Tsierlson's bound for the CHSH inequality as the solution to OP2$'$.

Case $|\boldsymbol{\iota}| = 3$, $|\boldsymbol{\eta}| = 3$. Here, both parties have three observables, and the PRM measures the parity of all three of them. In this case, there is no need to run numerics to approximate the solution to $\mathcal{I}_{\max}^{\mathrm{R}}$. On the one hand, notice that the optimisation to be carried out is the same as that for the case with $|\boldsymbol{\iota}| = 2$ and $|\boldsymbol{\eta}| = 3$, with some additional constraints given by the requirement that the PRM reads out the parity of the extra pair of fiducial measurements (since now $|\boldsymbol{\iota}| = 3$). Hence, $\mathcal{I}_{\max}^{\mathrm{R}}[|\boldsymbol{\iota}| = 3\,,\,|\boldsymbol{\eta}| = 3] \leq \mathcal{I}_{\max}^{\mathrm{RP}}[|\boldsymbol{\iota}| = 2\,,\,|\boldsymbol{\eta}| = 3]$. Given the results from above it follows that, up to numerical precision, $\mathcal{I}_{\max}^{\mathrm{R}}[|\boldsymbol{\iota}| = 3\,,\,|\boldsymbol{\eta}| = 3] \leq \frac{\sqrt{2}-1}{2}$, the latter being Tsirelson's bound for CHSH.

On the other hand, quantum theory is an example of a generalised probabilistic theory that yields the value $CHSH = \frac{\sqrt{2}-1}{2}$ while complying with the constraints from OP1. Indeed, taking the three fiducial measurements per side to be the Pauli observables, one can construct a PRM that does the trick. Hence, $\mathcal{I}_{\max}^{\mathrm{R}}[|\boldsymbol{\iota}| = 3\,,\,|\boldsymbol{\eta}| = 3] \geq \frac{\sqrt{2}-1}{2}$. From this follows that, up to numerical precision, $\mathcal{I}_{\max}^{\mathrm{R}}[|\boldsymbol{\iota}| = 3\,,\,|\boldsymbol{\eta}| = 3] \sim \frac{\sqrt{2}-1}{2}$.

7.2 AMP inequalities

In a bipartite Bell scenario featuring two dichotomic measurements per party, a relevant family of inequalities was defined by Acín, Massar, and Pironio (AMP) [35]. These correspond to tilted CHSH inequalities, and have been found to be useful for randomness 'generation' [35]. In the traditional language, the value assigned to the linear functional associated to the inequality reads:

$$I_{\alpha,\gamma}^{\mathrm{AMP}} = \langle \gamma\, A_0 + \alpha\, A_0 B_0 + \alpha A_0 B_1 + A_1 B_0 - A_1 B_1 \rangle\,, \tag{113}$$

where the parameters α and γ satisfy: $\alpha \geq 1$, $\gamma \geq 0$, and $\gamma < 2$. These inequalities are bounded from above, and their corresponding classical, quantum, and non-signalling bounds when $\alpha\gamma \leq 2$ are:

$$\beta_{\mathrm{AMP}}^{\mathrm{C}}(\alpha,\gamma) = 2\alpha + \gamma, \tag{114}$$

$$\beta_{\mathrm{AMP}}^{\mathrm{Q}}(\alpha,\gamma) = 2\sqrt{(1+\alpha^2)\left(1 + \frac{\gamma^2}{4}\right)}\,, \text{ and} \tag{115}$$

$$\beta_{\mathrm{AMP}}^{\mathrm{NS}}(\alpha,\gamma) = 2 + 2\alpha\,. \tag{116}$$

In our notation, that is, in terms of the probabilities, the AMP inequialities are equivalently captured by the following linear functional:

$$\begin{aligned} I_{\alpha,\gamma} = -\,&(2\alpha + \gamma)\, p_s^{(1)}(0|0) - (1+\alpha)\, p_s^{(2)}(0|0) + 2\alpha\, p_s(00|00) + 2\alpha\, p_s(00|01) \\ +\,&(1-\alpha)\, p_s^{(2)}(0|1) + 2\, p_s(00|10) - 2\, p_s(00|11)\,, \end{aligned} \tag{117}$$

whose corresponding classical, quantum, and non-signalling bounds when $\alpha\gamma \leq 2$ are:

$$\beta_{\alpha,\gamma}^{\mathrm{C}}(\alpha,\gamma) = 0\,, \quad \beta_{\alpha,\gamma}^{\mathrm{NS}}(\alpha,\gamma) = 1 - \frac{\gamma}{2}\,, \text{ and} \tag{118}$$

$$\beta_{\alpha,\gamma}^{\mathrm{Q}}(\alpha,\gamma) = \sqrt{(1+\alpha^2)\left(1 + \frac{\gamma^2}{4}\right)} - \frac{2\alpha + \gamma}{2}\,, \tag{119}$$

To explore the case of these inequalities in this section, we have only considered the case of $|\boldsymbol{\iota}| = |\boldsymbol{\eta}| = 2$.

The values of α and γ that we considered are quite varied. On the one hand, we took α from the set $\{1, 3, 5, 7, 9, 11\}$ and then, for each such α, considered six equally-spaced values for γ

(see Sec. A in the Appendix). On the other hand, we wanted to explore the transition between $\alpha = 1$ and $\alpha = 3$ more deeply, hence we explored the linear functionals $I_{\alpha,\gamma}$ also for α taken from the set $\{1.01, 1.05, 1.1, 1.2, 1.5, 2, 2.2, 2.4, 2.6, 2.8, 2.9, 2.95\}$ whilst keeping $\gamma = 0$. The motivation for this will hopefully become clear later on.

For each of the linear functionals $I_{\alpha,\gamma}$ defined by the above-mentioned values of α and γ, we asked what the value of $\mathcal{I}_{\max}$ – the solution to the optimisation problem OP1 – is. Here we computed an upper bound to $\mathcal{I}_{\max}$ for each inequality, by applying two relaxations to OP1:

- First, instead of demanding that $s \in \Sigma$, we only request that the state s belongs to the cone $K[\mathcal{P}]$ that satisfies the second level of our hierarchy (see Eqs. (62), (63), (64), (67), and (68)).
- Second, by solving the associated level 3 of the Lasserre hierarchy, we upper bound the solution to the relaxation to OP1 defined in the previous item.

Our numerical calculations show that, in the cases where $\alpha \geq 3$ the upper bound for $\mathcal{I}_{\max}$ is smaller than the inequality's Tsirelson's bound (see Fig. 5). Indeed, up to numerical precision $\mathcal{I}_{\max} \leq 0$, where 0 is the classical bound of the inequality. For the case where $\alpha = 1$, the inequality becomes the CHSH inequality plus a extra term corresponding to the single-party observable A_0. Beyond the case $\gamma = 0$ (which corresponds to the traditional CHSH inequality), other values of γ give an upper bound to $\mathcal{I}_{\max}$ that is larger than yet close to the inequality's Tsirelson's bound (see Fig. 5). Finally, for the values of $\alpha \in \{1, 1.01, 1.05, 1.1, 1.2, 1.5, 2, 2.2, 2.4, 2.6, 2.8, 2.9, 2.95, 3, 5, 7, 9, 11\}$ and $\gamma = 0$ one observes that the upper bound to $\mathcal{I}_{\max}$ drops to 0 when α goes from 1 to 3. Reading into the data of Fig. 5, one can notice some additional interesting behavior. For instance, in a few cases the upper bound to $\mathcal{I}_{\max}$ is equal to Tsirelson's bound, at least, up to the numerical precision. In particular, this happens for the cases where $\alpha\gamma = 2$ explored in this manuscript. The numerical results presented in this subsection are further summarized in Fig. 6.

The cases in which the PRM bounds the value of the AMP inequality to be smaller than the maximal quantum value (in contrast to CHSH in which the exact quantum bound was obtained) are likely to be cases in which the quantum bound is achieved for quantum observables for which there does not exist a PRM. This suggests that the observables which allow for a PRM may feature some particular properties regarding them being maximally complementary. Understanding the scope of parity-readable observables within quantum theory, and the correlations which they can realise, is therefore an important topic for future work.

Now, what does this all mean for the purpose of our conjecture? Well, no conclusive statement can be drawn from the numerics run for $\alpha = 1$, $\gamma \neq 0$. However all other cases are consistent with (and hence support) Conjecture 1.

7.3 AQ inequality

In Ref. [36] an inequality was provided, which is violated by so called "almost quantum" correlations [36], but is not violated by any quantumly realisable correlations. Here we refer to this inequality as AQ inequality. In our notation it is given by

$$I_{\mathrm{AQ}}(\mathbf{p}_s) = \frac{30}{31}p^{(1)}(0|0) - \frac{167}{9}p^{(1)}(0|1) + \frac{30}{31}p^{(2)}(0|0) - \frac{74}{11}p(00|00) + \frac{174}{11}p(00|10)$$
$$- \frac{167}{9}p^{(2)}(0|1) + \frac{174}{11}p(00|01) + \frac{244}{23}p(00|11)\,. \tag{120}$$
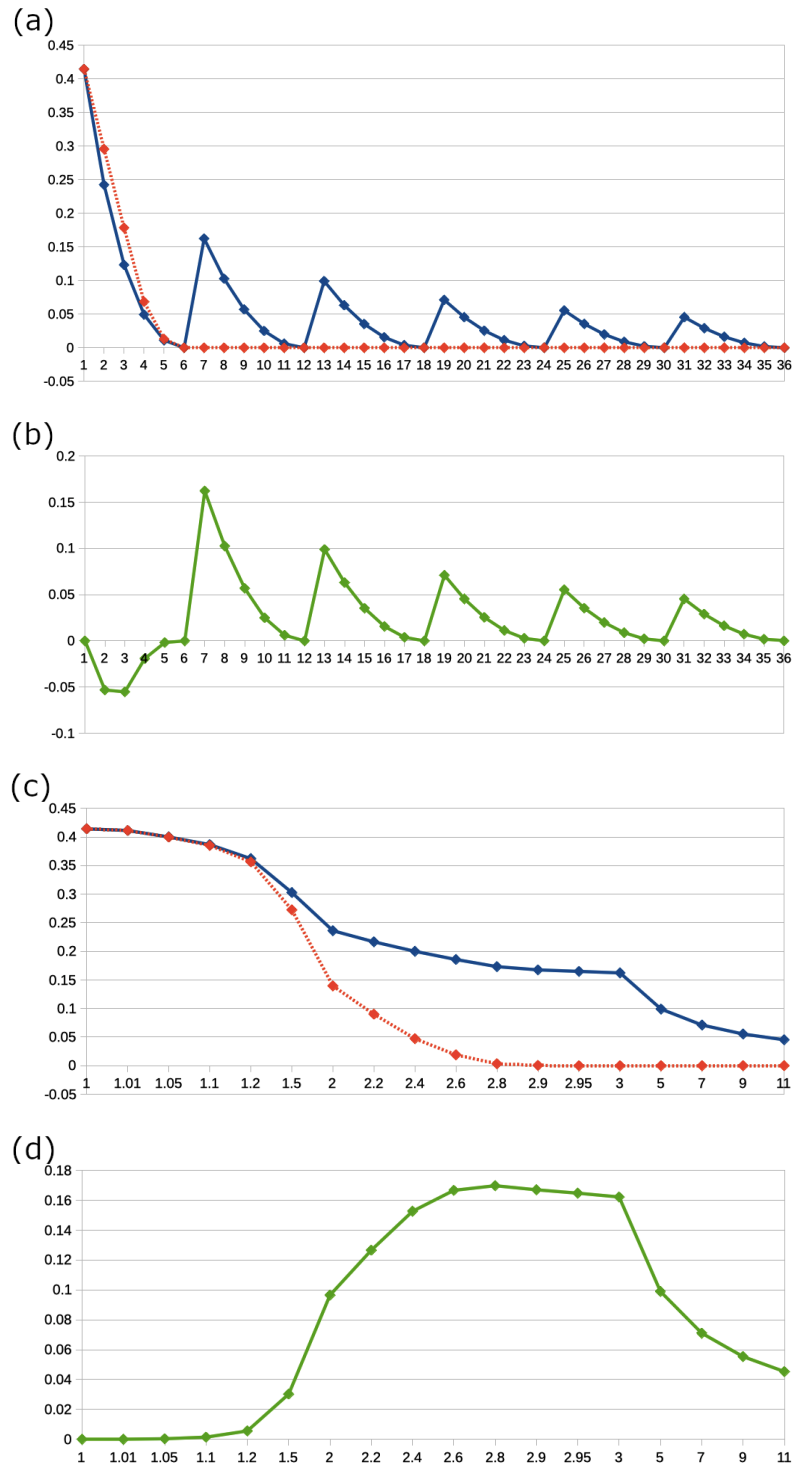
Figure 5: Numerical results for the AMP inequalities. (a) Each integer value, x in the horizontal axis corresponds to a different choice of pair of parameters (α, γ), i.e., a different AMP inequality (see Table in Eq. (150)). The vertical axis plots, for each inequality, both the numerical approximation to OP1 (dashed red line) and the Tsirelson's bound of the inequality (solid blue line). (b) The horizontal axis is the same as for the case (a). The vertical axis plots the difference between the numerical approximation to OP1 and the Tsirelson's bound of the inequality. (c) The value of the horizontal axis corresponds to the value of α. The value of γ is always 0. The vertical axis plots, for each inequality, both the numerical approximation to OP1 (dashed red line) and the Tsirelson's bound of the inequality (solid blue line). (d) The horizontal axis is the same as for the case (c). The vertical axis plots the difference between the numerical approximation to OP1 and the Tsirelson's bound of the inequality. In (b) and (c), witnessing a nonnegative value in the plot means that $\mathcal{I}_{\max}^{R}$ lies below Tsirelson's bound for that particular inequality. In all these figures, we approximate OP1 by first relaxing OP1 and then using the third level of the Lasserre hierarchy – see main text for details.

This inequality is bounded from above, and the corresponding classical, quantum, almost-quantum, and non-signalling bounds are:

$$\beta_{\text{AQ}}^{\text{C}} = \frac{30}{31} \sim 0.9677 \,, \quad \beta_{\text{AQ}}^{\text{Q}} < 1 \,, \tag{121}$$

$$\beta_{\text{AQ}}^{\text{AQ}} = 1.0232 \,, \text{ and } \quad \beta_{\text{AQ}}^{\text{NS}} = 3.5347 \,. \tag{122}$$

Let us first consider the case $|\boldsymbol{\iota}| = 2$, $|\boldsymbol{\eta}| = 2$. Similarly to the case for the AMP inequalities, we upper-bound $\mathcal{I}_{\max}$ by the solution to OP1 provided by the second level of the PRM-hierarchy of constraints presented in this paper. This solution is moreover estimated (i.e., upper bounded) by using the third level of the Lasserre hierarchy for polynomial optimisation problems. In this case, we obtain $\mathcal{I}_{\max} \leq 1.387818418422242$. This upper bound to $\mathcal{I}_{\max}$ is quite larger than the quantum bound, and hence not much can be concluded. Going to higher levels in the PRM-hierarchy might be the most promising step to take, however our current computational capabilities cannot handle the number of constraints and hence we defer this option for future work.

Next, we considered the case of $|\boldsymbol{\iota}| = 3$, $|\boldsymbol{\eta}| = 3$. Approximating the solution of OP2 via the 1+AB level of the Lasserre hierarchy gives $\mathcal{I}_{\max}^{\text{R}} < 1.7$. This number is substantially larger than the quantum bound for the inequality, and hence we are in a similar situation to the case presented before.

In this case, however, one can further explore the specific cases where some extra properties are required of the PRM being optimised over. This is similar to what we discussed in Sec. 5.1. So let us remind ourselves first of what these two specific types of PRM we focus on are. Notice that since $|\boldsymbol{\iota}| = 3$ and $|\boldsymbol{\eta}| = 3$ the outcome of a PRM is of the form $(\pm, \pm, \pm | 1, 2, 3)$, where $\pm_j$ tells whether the pair of fiducial measurements (j, j) is correlated $(+)$ or anti-correlated $(-)$. A PRM is of 'quantum' type if it assigns non-zero probability only to the outcomes

$$\{(-, -, -|1, 2, 3), (-, +, +|1, 2, 3), (+, -, +|1, 2, 3), (+, +, -|1, 2, 3)\} \,, \quad (\text{'quantum'}) \,. \tag{123}$$

The motivation behind the name is that a PRM of Pauli measurements satisfies this constraints. In turn, a PRM is of 'mirror quantum' type if it assigns non-zero probability only to the outcomes

$$\{(-, -, +|1, 2, 3), (-, +, -|1, 2, 3), (+, -, -|1, 2, 3), (+, +, +|1, 2, 3)\} \,, \quad (\text{'mirror quantum'}) \,. \tag{124}$$

The motivation behind the name is that it is exactly the complement of the 'quantum'-type PRM.

With this in mind, we approximated the solution of OP2 via the second level of the Lasserre hierarchy. The results we obtained are:

$$\mathcal{I}_{\max}^{\text{R}} < 0.8782940363666021 \,, \quad (\text{'quantum'}) \,, \tag{125}$$

$$\mathcal{I}_{\max}^{\text{R}} < 1.3953137950470862 \,, \quad (\text{'mirror quantum'}) \,. \tag{126}$$

We see that restricting the optimisation to PRMs that are of 'quantum' type give quite a strong constraint on the possible value of $\mathcal{I}_{\max}$, which here happens to be below the quantum bound (even below the classical bound) of the inequality. We believe that demanding that a PRM of 'quantum' type exists somehow forces the fiducial measurements to display some complementarity properties, and hence are not ideal for maximising the value of the linear functional I_{AQ}.

The numerical results presented in this subsection are summarized in Fig. 6.

Inequality	Size	Type of PRM	PRM Hierarchy level	Lasserre Hierarchy level	Upper bound to I_max	Quantum bound (Tsirelson's)
CHSH	$|\eta|=2,\ |\iota|=2$	Arbitrary	OP2	1+AB*	$\sim \dfrac{\sqrt{2}-1}{2}$	$\dfrac{\sqrt{2}-1}{2}$
	$|\eta|=3,\ |\iota|=2$	Arbitrary	OP2	1+AB*	$\sim \dfrac{\sqrt{2}-1}{2}$	$\dfrac{\sqrt{2}-1}{2}$
	$|\eta|=3,\ |\iota|=2$	$|\eta|=3,\ |\iota|=2$	OP2	1+AB*	$\sim \dfrac{\sqrt{2}-1}{2}$	$\dfrac{\sqrt{2}-1}{2}$
AMP	$|\eta|=2,\ |\iota|=2$	Arbitrary	OP1 — level 2	3	See Fig. 5	$2\sqrt{(1+\alpha^2)\left(1+\dfrac{\gamma^2}{4}\right)}$
AQ	$|\eta|=2,\ |\iota|=2$	Arbitrary	OP1 — level 2	3	~ 1.387818	
	$|\eta|=3,\ |\iota|=3$	Arbitrary	OP2	1+AB	~ 1.7	$0.9677 \leq \beta^{Q}_{AQ} < 1$
		'Quantum'	OP2	2	~ 0.878294	
		'Mirror Quantum'	OP2	2	~ 1.395314	

Figure 6: Bounds for Bell inequalities from parity reading measurement. "Arbitrary PRM" means that we do not restrict it in any way. In particular, for $|\iota| = 3$ it means that the PRM has 8 outcomes.

7.4 Necessity of local tomography

In this section we show that, if we give up on the assumption of local tomography in Optimization Problem 2, then Tsirelson's bound is violated. Moreover, it is violated in an extreme way, namely, that PR-box correlations can be achieved. Recall that the PR-box is a nosignaling box that reaches the nosignaling bound, that is, the algebraic maximum, for the CHSH inequality (i.e., in our notation of Eq. (103) it achieves the value of $1/2$). The PR box can be defined by the fact that it exhibits perfect correlations for XX, XZ, and ZX observables, and perfect anticorrelations for ZZ.

We now assume that local tomography does not hold, and, in particular, that the states are described by one extra non local, "holistic", parameter, which we denote by w_{NL}. Our parameterisation of the state, that is, the equivalent of Eq. (75), now takes the form:

$$
\begin{aligned}
\mathbf{p}_s &= (\mathbf{p}_s^{LT}, w_{NL}) \\
&= \big(p_s(00|00),\, p_s(00|10),\, p_s^{(2)}(0|0),\, p_s(00|01),\, p_s(00|11),\, p_s^{(2)}(0|1), \\
&\qquad\qquad\qquad\qquad\qquad p_s^{(1)}(0|0),\, p_s^{(1)}(0|1),\, 1, w_{NL} \big) .
\end{aligned}
\tag{127}
$$

The parameter w_{NL} is described as a holistic degree of freedom, as products of local observables are independent of its value. In general, however, a PRM will not be simply a product of local observables, and hence, it is possible that it will indeed depend on this holistic parameter.

For the remaining of this section it is more convenient to use a more compact matrix notation for bipartite states and effects, given by:

$$
\mathbf{p_s} =
\left[
\begin{array}{ccc|c}
p_s(00|00) & p_s(00|01) & p_s^{(1)}(0|0) & \\
p_s(00|10) & p_s(00|11) & p_s^{(1)}(0|1) & \\
p_s^{(2)}(0|0) & p_s^{(2)}(0|1) & 1 & \\
\hline
 & & & w_{NL}
\end{array}
\right] .
\tag{128}
$$

In this notation, the state which realises a PR box looks as follows

$$
\mathbf{p}_s^{\text{PR}} = \left[
\begin{array}{ccc|c}
\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \\
\frac{1}{2} & 0 & \frac{1}{2} & \\
\frac{1}{2} & \frac{1}{2} & 1 & \\
\hline
 & & & w_{NL}^{\text{PR}}
\end{array}
\right],
\tag{129}
$$

where w_{NL}^{PR} can be an arbitrary value. By definition of a PRM, the sums $\mathcal{P}^{00}+\mathcal{P}^{01}$ and $\mathcal{P}^{00}+\mathcal{P}^{10}$ depend only on local parameters – hence, their holistic parameter is zero and we get

$$
\mathcal{P}^{00} + \mathcal{P}^{01} = \left[
\begin{array}{ccc|c}
2 & 0 & -1 & \\
0 & 0 & 0 & \\
-1 & 0 & 1 & \\
\hline
 & & & 0
\end{array}
\right] \equiv \mathcal{R}^0, \quad
\mathcal{P}^{00} + \mathcal{P}^{10} = \left[
\begin{array}{ccc|c}
0 & 0 & 0 & \\
0 & 2 & -1 & \\
0 & -1 & 1 & \\
\hline
 & & & 0
\end{array}
\right] \equiv \mathcal{R}^1.
\tag{130}
$$

Preserving probability by PRM reads as

$$
\mathcal{P}^{00} + \mathcal{P}^{01} + \mathcal{P}^{10} + \mathcal{P}^{11} = \left[
\begin{array}{ccc|c}
0 & 0 & 0 & \\
0 & 0 & 0 & \\
0 & 0 & 1 & \\
\hline
 & & & 0
\end{array}
\right] \equiv \mathbb{1}.
\tag{131}
$$

Using this, we can write the free parameters for our optimisation problem as follows:

$$
\mathcal{P}^{00} - \mathcal{P}^{01} - \mathcal{P}^{10} + \mathcal{P}^{11} = (\mathcal{C}_{LT}, c_{NL}) = \left[
\begin{array}{ccc|c}
c_{11} & c_{12} & c_{13} & \\
c_{21} & c_{22} & c_{23} & \\
c_{31} & c_{32} & c_{33} & \\
\hline
 & & & c_{NL}
\end{array}
\right] \equiv \mathcal{C}.
\tag{132}
$$

Here, c_{NL} corresponds to the holistic parameter.

We can then express our parity reading effects in terms of these matrices $\mathcal{R}^0$, $\mathcal{R}^1$ and $\mathbb{1}$ and the free parameters $\mathcal{C}$, as:

$$
\begin{aligned}
\mathcal{P}^{00} &= \frac{1}{4}\left(2\mathcal{R}^0 + 2\mathcal{R}^1 + \mathcal{C} - \mathbb{1}\right), \\
\mathcal{P}^{01} &= \frac{1}{4}\left(2\mathcal{R}^0 - 2\mathcal{R}^1 - \mathcal{C} + \mathbb{1}\right), \\
\mathcal{P}^{10} &= \frac{1}{4}\left(-2\mathcal{R}^0 + 2\mathcal{R}^1 - \mathcal{C} + \mathbb{1}\right), \\
\mathcal{P}^{11} &= \frac{1}{4}\left(-2\mathcal{R}^0 - 2\mathcal{R}^1 + \mathcal{C} + 3\mathbb{1}\right).
\end{aligned}
\tag{133}
$$

In particular, the nonlocal parameter for PRM effects amounts to

$$
\mathcal{P}_{NL}^{00} = \mathcal{P}_{NL}^{11} = \frac{c_{NL}}{4}, \quad \mathcal{P}_{NL}^{01} = \mathcal{P}_{NL}^{10} = -\frac{c_{NL}}{4}.
\tag{134}
$$

Let us now write the requirement of PRM effects to be positive on the PR-box state. First, notice that

$$
\mathcal{R}^0 \cdot \mathbf{p}_s^{\text{PR}} = 1, \quad \mathcal{R}^1 \cdot \mathbf{p}_s^{\text{PR}} = 0,
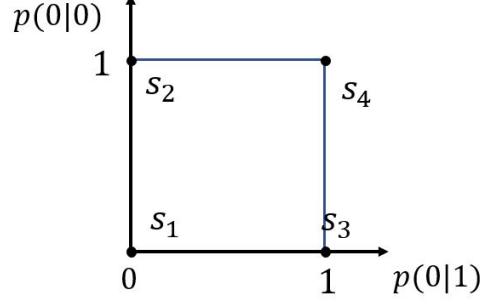\tag{135}
$$

Figure 7: Steered states coming from PR box. Here we are drawing the hyperplane of normalised vectors defined by $(0, 0, 1) \cdot v = 1$.

(as it should be, since PR box has perfect XX correlations and perfect ZZ anticorrelations). We thus get

$$
\begin{aligned}
\mathcal{P}^{00} \cdot \mathbf{p}_s^{PR} &= \frac{1}{4} \left(1 + \mathcal{C} \cdot \mathbf{p}_s^{PR} \right), \\
\mathcal{P}^{01} \cdot \mathbf{p}_s^{PR} &= \frac{1}{4} \left(3 - \mathcal{C} \cdot \mathbf{p}_s^{PR} \right), \\
\mathcal{P}^{10} \cdot \mathbf{p}_s^{PR} &= \frac{1}{4} \left(-1 - \mathcal{C} \cdot \mathbf{p}_s^{PR} \right), \\
\mathcal{P}^{11} \cdot \mathbf{p}_s^{PR} &= \frac{1}{4} \left(1 + \mathcal{C} \cdot \mathbf{p}_s^{PR} \right).
\end{aligned}
\tag{136}
$$

We see that the positivity of a PRM effect on the PR-box is equivalent to the following condition:

$$
\mathcal{C} \cdot \mathbf{p}_s^{PR} = -1.
\tag{137}
$$

Let us now explore the conditions that follow from products of steered states. The form of the unnormalized steered state is given by Eq. (82). Thus, the normalized ones arising from the PR-box state (for each party) are given by

$$
s_1 = (0, 0, 1)^T, \quad s_2 = (0, 1, 1)^T, \quad s_3 = (1, 0, 1)^T, \quad s_4 = (1, 1, 1)^T.
\tag{138}
$$

We see that these define the vertices of the so-called square bit, as can be seen in Fig. 7. The tomographically local degrees of freedom for the products of steered states are $\mathbf{p}_{s,LT}^{ij} = s_i^{(1)} \otimes s_j^{(2)}$, $i, j = 1, \ldots 4$. We then denote:

$$
\mathbf{p}_s^{ij} = (\mathbf{p}_{s,LT}^{ij}, w_{NL}^{ij}) = (s_i^{(1)} \otimes s_j^{(2)}, w_{NL}^{ij}).
\tag{139}
$$

Note that, here, the bracket does not mean scalar product, but rather indicates two groups of parameters: the group of locally tomographic ones, and the group consisting of one nonlocal parameter. For steered states, the nonlocal parameter w_{NL}^{ij} must be a linear combination of the local parameters:

$$
w_{NL}^{ij} = \mathbf{h} \cdot \mathbf{p}_{s,LT}^{ij}.
\tag{140}
$$

This follows from noting that:
  i) the way that local states are combined to give product states must be given by a bilinear function from the local vector spaces into the global vector space;
  ii) the universal property of the tensor product means that this can be written as a linear function from the tensor product space of the local vector spaces into the global vector space;

38

iii) the local parameters are simply the tensor product space;

iv) this means that the value of the non-local parameter is given by a linear functional on the local parameters (i.e., a linear map from the local vector spaces into the reals);

v) finally, the Riesz representation theorem means that we can write this as the dot product with some vector $\mathbf{h}$ in the local parameter space.

We thus have

$$\mathcal{P}^{qr} \cdot \mathbf{p}_s^{ij} = \mathcal{P}_{LT}^{qr} \cdot \mathbf{p}_{s,LT}^{ij} + \mathcal{P}_{NL}^{qr} \mathbf{p}_{s,NL}^{ij} = (\mathcal{P}_{LT}^{qr} + \mathcal{P}_{NL}^{qr} \mathbf{h}) \cdot \mathbf{p}_{s,LT}^{ij}, \tag{141}$$

where we recall, that $\mathcal{P}_{NL}^{qr}$ are numbers (the values of the nonlocal parameter for PRM effects) given by Eq. (134). Using the above equation together with Eq. (133), we obtain

$$\mathcal{P}^{00} \cdot \mathbf{p}_s^{ij} = \frac{1}{2}(\mathcal{R}^0 + \mathcal{R}^1)\mathbf{p}_s^{ij} + \frac{1}{4}(-1 + \mathbf{g} \cdot \mathbf{p}_s^{ij})$$

$$\mathcal{P}^{01} \cdot \mathbf{p}_s^{ij} = \frac{1}{2}(\mathcal{R}^0 - \mathcal{R}^1)\mathbf{p}_s^{ij} + \frac{1}{4}(1 - \mathbf{g} \cdot \mathbf{p}_s^{ij})$$

$$\mathcal{P}^{10} \cdot \mathbf{p}_s^{ij} = \frac{1}{2}(-\mathcal{R}^0 + \mathcal{R}^1)\mathbf{p}_s^{ij} + \frac{1}{4}(1 - \mathbf{g} \cdot \mathbf{p}_s^{ij})$$

$$\mathcal{P}^{11} \cdot \mathbf{p}_s^{ij} = \frac{1}{2}(-\mathcal{R}^0 - \mathcal{R}^1)\mathbf{p}_s^{ij} + \frac{1}{4}(3 + \mathbf{g} \cdot \mathbf{p}_s^{ij}), \tag{142}$$

where we have denoted

$$\mathbf{g} = \mathcal{C}_{LT} + c_{NL}\mathbf{h}, \tag{143}$$

with $\mathcal{C}_{LT}$ being the locally tomographic part of $\mathcal{C}$, and c_{NL} the holistic part of $\mathcal{C}$. Now, the positivity of PRM effects on products of steered states means that we require all four terms to be positive for all $i, j = 1, \ldots, 4$. By using Mathematica [71] we find that positivity is satisfied for only one choice of $\mathbf{g}$:

$$\mathbf{g} = \begin{bmatrix} 4 & 0 & -2 \\ 0 & 4 & -2 \\ -2 & -2 & 4 \end{bmatrix}. \tag{144}$$

To summarise, positivity conditions of the PRM on the PR-box states and products of its steered states reduce to:

$$\text{positivity on PR box state:} \quad \mathcal{C}_{LT} \cdot \mathbf{p}_{s,LT}^{PR} + c_{NL}w_{NL}^{PR} = -1 \tag{145}$$

$$\text{positivity on steered states:} \quad \mathcal{C}_{LT} + c_{NL}\mathbf{h} = \begin{bmatrix} 4 & 0 & -2 \\ 0 & 4 & -2 \\ -2 & -2 & 4 \end{bmatrix}. \tag{146}$$

To prove our original claim, the idea is to choose values for $\mathcal{C}_{LT}, c_{NL}, \mathbf{h}$ and w_{NL}^{PR} such that the above two constraints hold. Our choice is the following:

$$\mathcal{C}_{LT} = \mathbf{0}, \quad c_{NL} = 1, \quad w_{NL}^{PR} = -1, \quad \mathbf{h} = \begin{bmatrix} 4 & 0 & -2 \\ 0 & 4 & -2 \\ -2 & -2 & 4 \end{bmatrix}. \tag{147}$$

These values for $\mathcal{C}_{LT}$ and c_{NL} fix the PRM to take the form:
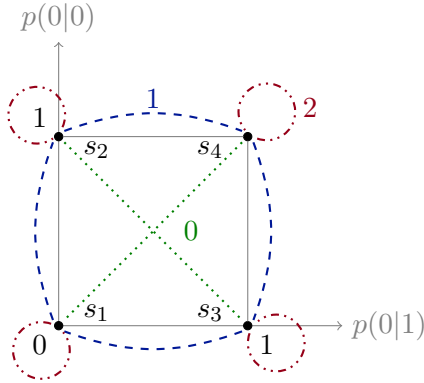
Figure 8: The value of the nonlocal parameter for the pairs of steered states. Each dashdotted-red loop denotes a pair of the same steered states. Each dotted-green or dashed-blue line means two pairs depending on which state goes to Alice and which to Bob. This gives rise to 16 pairs of steered states. Dashdotted-red pairs have the same value for both observables and dashed-blue pairs have the same value for one observable, whilst for dotted-green pairs both observables have opposite values.

$$\mathcal{P}^{00} = \frac{1}{4} \left[\begin{array}{ccc|c} 4 & 0 & -2 & \\ 0 & 4 & -2 & \\ -2 & -2 & -1 & \\ \hline & & & 1 \end{array} \right], \quad \mathcal{P}^{01} = \left[\begin{array}{ccc|c} 4 & 0 & -2 & \\ 0 & -4 & 2 & \\ -2 & 2 & 1 & \\ \hline & & & -1 \end{array} \right]$$

$$\mathcal{P}^{10} = \frac{1}{4} \left[\begin{array}{ccc|c} -4 & 0 & 2 & \\ 0 & 4 & -2 & \\ 2 & -2 & 1 & \\ \hline & & & -1 \end{array} \right], \quad \mathcal{P}^{11} = \left[\begin{array}{ccc|c} -4 & 0 & 2 & \\ 0 & -4 & 2 & \\ 2 & 2 & 3 & \\ \hline & & & 1 \end{array} \right]. \quad (148)$$

In addition, our choice of $\mathbf{h}$ defines the value of the nonlocal parameter for steered states to be

$$w_{NL}^{ij} = \mathbf{h} \cdot \mathbf{p}_{s,LT}^{ij}. \quad (149)$$

We see then that the PR-box state is consistent with the existence of a PRM that satisfies the constraints of OP2. Since performing fiducial measurements on a PR-box state yields PR-box correlations, this shows that $\mathcal{I}_{\max}^{R} = \frac{1}{2}$ for the CHSH inequality, as per Eq. (103). With this we conclude the proof of our claim.

Let us make a final comment on an interesting interpretation for the values of the nonlocal parameter for the pairs of steered states: they count the number of correlations. If both observables have the same value for a given pair of steered states (which happens when Alice and Bob's steered states are the same) then the parameter takes the value 2. When only one of the observables has the same value, then it takes the value 1, and when both observables have the opposite value, then it takes the value 0. This is depicted in Fig. 8.

8   Discussion

In this paper we have shown that postulating within a theory the existence of particular bipartite measurements has a surprisingly rich set of consequences for the structure of the theory itself. Indeed, we showed that this leads to an infinite hierarchy of constraints on the possible bipartite states. These conditions translate analogously into constraints on the statistical correlations

allowed by the theory. In other words, the maximum violation of any Bell inequality by the correlations among fiducial measurements featured by the theory will be subjected to an infinite hierarchy of constraints.

We further explored the consequences of this rich structure for the particular case where there exists a bipartite measurement that can read out the parity of local fiducial measurements. For the case of tomographically-local GPTs, we found that these constraints on the structure of bipartite sates are enough to recover (up to numerical precision) Tsirelson's bound for various inequalities in the CHSH scenario. In addition, we also showed that non-tomographically local GPTs may still reach the maximum algebraic violation of such inequalities (i.e., go beyond Tsirelson's bound) when only the first levels of the hierarchy of constraints are considered. We also noticed that, for inequalities where the maximum quantum violation is not achieved by measuring complementary observables, our technique may also yield values below Tsirelson's bound.

Our initial numerical results led us to formulate a conjecture on the constraints that the existence of a Parity Reading Measurement may yield for tomographically local GPTs:

Conjecture 1. *Under the assumption of local tomography, the local observables that are* **parity-readable** *satisfy Tsirelson's bound, i.e., they cannot violate Bell inequalities better than quantum mechanics does (with arbitrary measurements).*

It is worth mentioning that, after formulating the conjecture, we ran further numerics in other scenarios (all presented in this manuscript) which did not disprove the conjecture.

From looking at Conjecture 1 one can take a step back and further conjecture that quantum theory, among the landscape of GPTs that are locally tomographic, is the theory that displays the necessary balance between its allowed states and effects to feature the following property:

Conjecture 2. *Quantum Theory yields the largest violation of any Bell inequality by parity-readable fiducial measurements, within the landscape of possible locally-tomographic physical theories.*

Notice that in this conjecture we are comparing correlations obtained from *parity-readable* measurements in quantum theory vs. in other more generic (yet locally-tomographic) GPTs, and state that quantum theory will always produce correlations that are more non-classical. This is in contrast to Conjecture 1 which compared correlations obtained from *arbitrary* measurements in quantum theory vs. those that are achieved with parity-readable measurements in an arbitrary tomographically-local GPT. The significance of this is that not all quantum correlations can be achieved with parity-readable measurements alone. However, if measurements beyond parity-readable ones may be used, then it is possible that other GPTs beyond quantum (e.g., Boxworld) can generate correlations that are more non-classical than any that quantum theory may produce. Now, whether Conjecture 2 is true, how to formally express it, and what its consequences are, comprise a topic for future work.

Going beyond the CHSH scenario or GPTs with affine local dimension ≥ 3 is a computationally demanding task. Indeed, the complexity of the optimisation problems to be solved rises considerably with the number of settings and dimension. A complete understanding of the reach of the constraints imposed by parity reading measurements require the further development of analytical and numerical techniques, which are deferred to future work.

Moving forward, one may apply our technique to explore the constraints that entangled measurements beyond parity reading ones may impose. Indeed, Optimisation Problems 2 and 2′ may be straightforwardly adapted to study other bipartite measurements. It would be interesting to see if there is a relation between the properties of bipartite entangled measurements and those of the Bell inequalities whose Tsirelson's bound they recover.

More ambitiously, there is the natural question of multi-partite entangled measurements. Would the structure they impose on multi-partite state spaces have special features that we cannot envision from the phenomenology at the bipartite level? We hope such explorations will bring new insight into the structure of states and effect spaces in GPTs, and their non-classical properties.

Acknowledgements

References

[1] John S Bell. "On the Einstein-Podolsky-Rosen paradox". Physics **1**, 195–200 (1964).

[2] Elie Wolfe, David Schmid, Ana Belén Sainz, Ravi Kunjwal, and Robert W. Spekkens. "Quantifying Bell: the Resource Theory of Nonclassicality of Common-Cause Boxes". Quantum **4**, 280 (2020).

[3] Jonathan Barrett, Lucien Hardy, and Adrian Kent. "No Signaling and Quantum Key Distribution". Phys. Rev. Lett. **95**, 010503 (2005).

[4] Antonio Acín, Nicolas Gisin, and Lluis Masanes. "From Bell's Theorem to Secure Quantum Key Distribution". Phys. Rev. Lett. **97**, 120405 (2006).

[5] Valerio Scarani, Nicolas Gisin, Nicolas Brunner, Lluis Masanes, Sergi Pino, and Antonio Acín. "Secrecy extraction from no-signaling correlations". Phys. Rev. A **74**, 042339 (2006).

[6] A. Acín et al.. "Device-Independent Security of Quantum Cryptography against Collective Attacks". Phys. Rev. Lett. **98**, 230501 (2007).

[7] Umesh Vazirani and Thomas Vidick. "Fully Device-Independent Quantum Key Distribution". Phys. Rev. Lett. **113**, 140501 (2014).

[8] Jędrzej Kaniewski and Stephanie Wehner. "Device-independent two-party cryptography secure against sequential attacks". New J. Phys. **18**, 055004 (2016).

[9] Roger Colbeck and Renato Renner. "Free randomness can be amplified". Nat. Phys. **8**, 450 EP – (2012).

[10] S. Pironio et al.. "Random numbers certified by Bell's theorem". Nature **464**, 1021 EP – (2010).

[11] Matej Pivoluska and Martin Plesch. "Device Independent Random Number Generation". Acta Physica Slovaca **64**, 600–663 (2015).

[12] Chirag Dhara, Giuseppe Prettico, and Antonio Acín. "Maximal quantum randomness in Bell tests". Phys. Rev. A **88**, 052116 (2013).

[13] Anne Broadbent and André Allan Méthot. "On the power of non-local boxes". Theo. Comp. Sci. **358**, 3 – 14 (2006).

[14] Carlos Palazuelos and Thomas Vidick. "Survey on nonlocal games and operator space theory". J. Math. Phys. **57**, 015220 (2016).

[15] Nathaniel Johnston, Rajat Mittal, Vincent Russo, and John Watrous. "Extended non-local games and monogamy-of-entanglement games". Proc. Roy. Soc. A **472**, 20160003 (2016).

[16] Sandu Popescu and Daniel Rohrlich. "Quantum nonlocality as an axiom". Foundations of Physics **24**, 379–385 (1994).

[17] Gilles Brassard, Harry Buhrman, Noah Linden, André Allan Méthot, Alain Tapp, and Falk Unger. "Limit on nonlocality in any world in which communication complexity is not trivial". Physical Review Letters **96**, 250401 (2006).

[18] Marcin Pawłowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Żukowski. "Information causality as a physical principle". Nature **461**, 1101–1104 (2009).

[19] Miguel Navascués and Harald Wunderlich. "A glance beyond the quantum model". Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **466**, 881–890 (2010).

[20] Joe Henson and Ana Belén Sainz. "Macroscopic noncontextuality as a principle for almost-quantum correlations". Physical Review A **91**, 042114 (2015).

[21] Tobias Fritz, Ana Belén Sainz, Remigiusz Augusiak, Jonatan Bohr Brask, Rafael Chaves, Anthony Leverrier, and Antonio Acín. "Local orthogonality as a multipartite principle for quantum correlations". Nature communications **4**, 1–7 (2013).

[22] Noah Linden, Sandu Popescu, Anthony J. Short, and Andreas Winter. "Quantum nonlocality and beyond: Limits from nonlocal computation". Phys. Rev. Lett. **99**, 180502 (2007).

[23] L. Czekaj, M. Horodecki, and T. Tylec. "Bell measurement ruling out supraquantum correlations". Phys. Rev. A **98**, 032117 (2018).

[24] Jonathan Barrett. "Information processing in generalized probabilistic theories". Physical Review A **75**, 032304 (2007).

[25] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. "Proposed Experiment to Test Local Hidden-Variable Theories". Phys. Rev. Lett. **23**, 880–884 (1969).

[26] Lucien Hardy. "Quantum theory from five reasonable axioms" (2001). arXiv:quant-ph/0101012.

[27] G. Ludwig. "An axiomatic basis of quantum mechanics. 1. derivation of hilbert space". Springer-Verlag. (1985).

[28] E Brian Davies and John T Lewis. "An operational approach to quantum probability". Communications in Mathematical Physics **17**, 239–260 (1970).

[29] CH Randall and DJ Foulis. "An approach to empirical logic". The American Mathematical Monthly **77**, 363–374 (1970).

[30] C. Piron. "Axiomatique quantique". Helvetia Physica Acta **37**, 439–468 (1964).

[31] G. W. Mackey. "The mathematical foundations of quantum mechanics". W. A. Benjamin. New York (1963).

[32] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. "Probabilistic theories with purification". Physical Review A **81**, 062348 (2010).

[33] Lucien Hardy. "Reformulating and reconstructing quantum theory" (2011). arXiv:1104.2066.

[34] David Schmid, John H Selby, Matthew F Pusey, and Robert W Spekkens. "A structure theorem for generalized-noncontextual ontological models" (2020). arXiv:2005.07161.

[35] Antonio Acín, Serge Massar, and Stefano Pironio. "Randomness versus Nonlocality and Entanglement". Phys. Rev. Lett. **108**, 100402 (2012). arXiv:1107.2754.

[36] Miguel Navascués, Yelena Guryanova, Matty J. Hoban, and Antonio Acín. "Almost quantum correlations". Nature Communications **6**, 6288 (2015). arXiv:1403.4621.

[37] Daniel Collins and Nicolas Gisin. "A relevant two qubit bell inequality inequivalent to the chsh inequality". Journal of Physics A: Mathematical and General **37**, 1775 (2004).

[38] Marius Krumm and Markus P Müller. "Quantum computation is the unique reversible circuit model for which bits are balls". npj Quantum Information **5**, 7 (2019).

[39] Howard Barnum, Ciarán M Lee, and John H Selby. "Oracles and query lower bounds in generalised probabilistic theories". Foundations of physics **48**, 954–981 (2018).

[40] Andrew JP Garner. "Interferometric computation beyond quantum theory". Foundations of Physics **48**, 886–909 (2018).

[41] Jonathan Barrett, Niel de Beaudrap, Matty J Hoban, and Ciarán M Lee. "The computational landscape of general physical theories". npj Quantum Information **5**, 41 (2019).

[42] Ciarán M Lee and John H Selby. "Deriving Grover's lower bound from simple physical principles". New Journal of Physics **18**, 093047 (2016).

[43] Ciarán M Lee and Matty J Hoban. "Bounds on the power of proofs and advice in general physical theories". Proc. R. Soc. A **472**, 20160076 (2016).

[44] Ciarán M Lee and John H Selby. "Generalised phase kick-back: the structure of computational algorithms from physical principles". New Journal of Physics **18**, 033023 (2016).

[45] Ciarán M Lee and Jonathan Barrett. "Computation in generalised probabilisitic theories". New Journal of Physics **17**, 083001 (2015).

[46] Ciarán M Lee and John H Selby. "Higher-order interference in extensions of quantum theory". Foundations of Physics **47**, 89–112 (2017).

[47] Jamie Sikora and John Selby. "Simple proof of the impossibility of bit commitment in generalized probabilistic theories using cone programming". Physical review A **97**, 042302 (2018).

[48] John H Selby and Jamie Sikora. "How to make unforgeable money in generalised probabilistic theories". Quantum **2**, 103 (2018).

[49] Ludovico Lami, Carlos Palazuelos, and Andreas Winter. "Ultimate data hiding in quantum mechanics and beyond". Communications in Mathematical Physics **361**, 661–708 (2018).

[50] Howard Barnum and Alexander Wilce. "Information processing in convex operational theories". Electronic Notes in Theoretical Computer Science **270**, 3–15 (2011).

[51] Howard Barnum, Oscar CO Dahlsten, Matthew Leifer, and Ben Toner. "Nonclassicality without entanglement enables bit commitment". In Information Theory Workshop, 2008. ITW'08. IEEE. Pages 386–390. IEEE (2008).

[52] Jonathan Barrett, Lucien Hardy, and Adrian Kent. "No signaling and quantum key distribution". Physical Review Letters **95**, 010503 (2005).

[53] Samuel Fiorini, Serge Massar, Manas K Patra, and Hans Raj Tiwary. "Generalized probabilistic theories and conic extensions of polytopes". Journal of Physics A: Mathematical and Theoretical **48**, 025302 (2014).

[54] Anna Jenčová and Martin Plávala. "Conditions on the existence of maximally incompatible two-outcome measurements in general probabilistic theory". Physical Review A **96**, 022113 (2017).

[55] Joonwoo Bae, Dai-Gyoung Kim, and Leong-Chuan Kwek. "Structure of optimal state discrimination in generalized probabilistic theories". Entropy **18**, 39 (2016).

[56] Bob Coecke and Aleks Kissinger. "Picturing quantum processes: A first course in quantum theory and diagrammatic reasoning". Cambridge University Press. (2017).

[57] Stefano Gogioso and Carlo Maria Scandolo. "Categorical probabilistic theories". EPTCS **266**, 367 (2018).

[58] John H. Selby, Carlo Maria Scandolo, and Bob Coecke. "Reconstructing quantum theory from diagrammatic postulates". Quantum **5**, 445 (2021).

[59] Peter Janotta and Raymond Lal. "Generalized probabilistic theories without the no-restriction hypothesis". Physical Review A **87**, 052131 (2013).

[60] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. "Bell nonlocality". Reviews of Modern Physics **86**, 419 (2014).

[61] Bob Coecke. "Terminality implies non-signalling". EPTCS **172**, 27 (2014).

[62] Aleks Kissinger, Matty Hoban, and Bob Coecke. "Equivalence of relativistic causal structure and process terminality" (2017). arXiv:1708.04118.

[63] Borivoje Dakic and Caslav Brukner. "Quantum Theory and Beyond: Is Entanglement Special?" (2009). arXiv:0911.0695.

[64] Jamie Sikora and John H. Selby. "Impossibility of coin flipping in generalized probabilistic theories via discretizations of semi-infinite programs". Phys. Rev. Res. **2**, 043128 (2020).

[65] Jean B Lasserre. "Global optimization with polynomials and the problem of moments". SIAM Journal on optimization **11**, 796–817 (2001).

[66] Peter Wittek. "Algorithm 950: Ncpol2sdpa—sparse semidefinite programming relaxations for polynomial optimization problems of noncommuting variables". ACM Transactions on Mathematical Software (2015). arXiv:1308.6029.

[67] "http://sdpa.sourceforge.net/".

[68] Jie Wang, Victor Magron, and Jean-Bernard Lasserre. "Tssos: A moment-sos hierarchy that exploits term sparsity". SIAM Journal on Optimization **31**, 30–58 (2021).

[69] Victor Magron and Jie Wang. "Sparse polynomial optimization: theory and practice". Series on Optimization and Its Applications, World Scientific Press. (2023).

[70] Erling D Andersen and Knud D Andersen. "The mosek interior point optimizer for linear programming: an implementation of the homogeneous algorithm". High performance optimizationPages 197–232 (2000).

[71] Wolfram Research, Inc. "Mathematica, Version 12.1". Champaign, IL, 2020.

[72] Bob Coecke and Eric Oliver Paquette. "Categories for the practising physicist". In New structures for physics. Pages 173–286. Springer (2010).

[73] Saunders Mac Lane. "Categories for the working mathematician". Volume 5. Springer Science & Business Media. (2013).

A Results for AMP inequalities

In Table (150) we present the values we chose for the pairs of parameters (α, γ), used to define AMP inequalities that we examine in our numerical optimization. Here, the variable x corresponds to the value of the x-axis in Fig. 5.

x	α	γ	x	α	γ	x	α	γ
1	1	0	13	5	0	25	9	0
2	1	0,4	14	5	0,08	26	9	0,04444444
3	1	0,8	15	5	0,16	27	9	0,088888889
4	1	1,2	16	5	0,24	28	9	0,133333333
5	1	1,6	17	5	0,32	29	9	0,177777778
6	1	2	18	5	0,4	30	9	0,222222222
7	3	0	19	7	0	31	11	0
8	3	0,133333333	20	7	0,057142857	32	11	0,036363636
9	3	0,266666667	21	7	0,114285714	33	11	0,072727273
10	3	0,4	22	7	0,171428571	34	11	0,109090909
11	3	0,533333333	23	7	0,228571429	35	11	0,145454545
12	3	0,666666667	24	7	0,285714286	36	11	0,181818182

$$(150)$$

B Introduction our GPT formalism

In this paper we take a categorical approach to GPTs, which we summarise in this Section. The reader unfamiliar with the mathematics of category theory is referred to Ref. [72] for a physicist friendly introduction to the topic, and to Ref. [73] as the classic mathematics textbook on the subject.

The particular formalism that we use here is based on the observation that any tomographically local GPT can be thought of as a particular symmetric monoidal subcategory of the symmetric monoidal category $\mathbf{Vect}_{\mathbb{R}}$ of real vector spaces and linear maps (see, e.g., Ref. [34]). In particular, such a subcategory has the following properties:

1. objects are finite dimensional;
2. the scalars are the unit interval;
3. the hom-sets[6] are closed under convex combinations;
4. the points (and copoints) for an object span the vector space (resp. dual vector space);
5. there is a unique 'deterministic' copoint, u_V, for each object V. This is defined by the constraint that for any other copoint e for the object V there exists a copoint $e^{\perp}$ such that $e + e^{\perp} = u_V$.

One of the benefits of this categorical approach to generalised probabilistic theories is that there is a faithful diagrammatic representation using string diagrams, which we will use throughout the paper. This diagrammatic notation moreover immediately suggests the correct interpretation of the abstract categorical definition given above. For example, consider the following diagram (read bottom to top):

---

[6] A hom-set is the set of transformations from one object to another, in this case, this will be a subset of the set of linear maps from one vector space to another.

$$(151)$$

We view the wires in the above diagram, corresponding to objects (i.e. finite dimensional real vector spaces), as representing physical systems. Then, points, such as, $S : \mathbb{R} \to V \otimes V$, represent physical states, general morphisms, such as $T_1 : V \otimes U \to V \otimes W$ and $T_2 : V \otimes V \to W$ represent physical transformations, and copoints, such as $E : W \to \mathbb{R}$, correspond to physical effects. Closed diagrams, such as:

$$(152)$$

that is, elements of the unit interval, are interpreted as the probability of observing effect E given the system was prepared in state S. We denote the unique deterministic effect as:

$$(153)$$

which defines the normalised states S as those satisfying:

$$(154)$$

Note that for a set of effects $\{E_i\}_{i \in I}$ to describe a measurement it must be the case that:

$$(155)$$

One can then see that (finite dimensional) quantum theory defines such a GPT by noting that the set of Hermitian operators for some Hilbert space $\mathcal{H}$ forms a real vector space $\mathcal{B}(\mathcal{H})$, and that completely positive trace non-increasing (CPTNI) maps between these spaces are a particular class of linear maps between these vector spaces. The other constraints are simple to verify. Similarly, classical stochastic dynamics can be represented as such a GPT. To see this note that stochastic dynamics from some (finite) set X to another (finite) set A can be represented as a particular class of linear maps from the finite dimensional vector space $\mathbb{R}^X$ to the finite dimensional vector space $\mathbb{R}^A$.

We will work with the representation of GPTs in which this classical GPT is included as a subtheory. To distinguish it, we will represent the classical systems by thin gray wires, and, for convenience, we will simply label them by the finite set X, A, ..., rather than the vector spaces $\mathbb{R}^X$, $\mathbb{R}^A$, ... This is convenient because it allows us to explicitly represent measurement outcomes and setting variables within the diagrammatic representation. For example, a controlled

47

measurement of system V with setting variable X and outcome variable A is denoted as:

$$\begin{array}{c} A \\ \boxed{M} \\ X \quad V \end{array} \quad , \tag{156}$$

which must satisfy the constraint:

$$\begin{array}{c} = \\ A \\ \boxed{M} \\ X \quad V \end{array} \quad = \quad \begin{array}{cc} = & = \\ X & V \end{array} \quad . \tag{157}$$

The situation where we perform this measurement M on the system V prepared in some normalisted state S is denoted by:

$$\begin{array}{c} A \\ \boxed{M} \\ V \\ S \\ X \end{array} \quad , \tag{158}$$

and is simply a stochastic map from the setting variable X to the outcome variable A. The probabilities of obtaining a particular outcome $a \in A$ given a setting $x \in X$ can be extracted from this map via:

$$\begin{array}{c} a \\ A \\ \boxed{M} \\ V \\ X \quad S \\ x \end{array} \quad = \quad p_S(a|x) \; . \tag{159}$$

We will also find it useful to use certain processes which live in $\mathbf{Vect}_{\mathbb{R}}$ but which are not part of the subtheory describing the GPT. To visually distinguish these 'non-physical' processes we draw them as shaded objects:

$$\begin{array}{c} V \\ \boxed{L} \\ U \end{array} \quad . \tag{160}$$

Finally, we will define a particular type of linear functionals I. The objects these act on are linear maps from one vector space U to a vector space V. We diagrammatically denote them as:

$$\begin{array}{c} V \\ \boxed{} I \\ U \end{array} \quad . \tag{161}$$

Such a linear functional, I, maps some linear map $L : U \to V$ to a real number by:

$$\begin{array}{c} V \\ I \\ U \end{array} \quad :: \quad \begin{array}{c} V \\ \boxed{L} \\ U \end{array} \quad \mapsto \quad \begin{array}{c} V \\ \boxed{L} \quad I \\ U \end{array} \quad . \tag{162}$$

48

Note that, as $\mathbf{Vect}_{\mathbb{R}}$ is a compact closed category, it can be readily verified that these linear functionals can always be written as:

$$\tag{163}$$

for some vector space ζ_I, vector v_I and covector c_I.

C Geometric constraints on state and effect spaces

We define the dual of a set of vectors $\mathcal{V} \subseteq V$ by:

$$\mathcal{V}^* := \{ w \in V^* | w(v) \in [0,1]\ \forall v \in \mathcal{V} \} . \tag{164}$$

If we then denote the set of states by Ω_V and the set of effects by $\mathcal{E}_V$ then the constraint on state-effect pairs implies[7] the pair of constraints:

$$\mathcal{E}_V \subseteq \Omega_V^* \quad \text{and} \quad \Omega_V \subseteq \mathcal{E}_V^*. \tag{165}$$

That is, the effect space is constrained by the state space and vice versa.

Now, if we consider the special case of bipartite systems $V \otimes W$ then this means that:

$$\mathcal{E}_{V \otimes W} \subseteq \Omega_{V \otimes W}^* \quad \text{and} \quad \Omega_{V \otimes W} \subseteq \mathcal{E}_{V \otimes W}^*. \tag{166}$$

Hence, introducing some bipartite effects for the theory (i.e., enlarging $\mathcal{E}_{V \otimes W}$) will induce a constraint on the bipartite state space (since $\mathcal{E}_{V \otimes W}^*$ will potentially be smaller).

This constraint, however, whilst necessary is not sufficient to ensure that we will end up with a valid GPT. Considerations of compositionality and convexity further constrain our state spaces. For example, it follows from compositionality and convexity, that any state of the form:

$$\sum_i p_i \quad \tag{167}$$

where $s_v^{(i)} \in \Omega_V$ and $s_w^{(i)} \in \Omega_W$, $p_i \in \mathbb{R}^+$, and $\sum_i p_i = 1$, is a valid state for the composite system. This condition – that the bipartite state space contains all separable states – means that $\Omega_V \otimes_{\min} \Omega_W \subseteq \Omega_{V \otimes W}$, where, the so called 'min tensor product' is defined as the set of separable states. The same is also true for effects – compositionality and convexity mean that any effects of the form:

$$\sum_j q_j \quad \tag{168}$$

where $e_v^{(j)} \in \mathcal{E}_V$ and $e_w^{(j)} \in \Omega_W$, $q_j \in \mathbb{R}^+$, and $\sum_j q_j = 1$, is a valid effect for the composite system. This means that $\mathcal{E}_V \otimes_{\min} \mathcal{E}_W \subseteq \mathcal{E}_{V \otimes W}$.

---

[7]Where we identify $V^{**} \cong V$

In conjunction with condition Eq. (166), we can use this to obtain an upper bound on the state space as follows:

$$\Omega_V \otimes_{\min} \Omega_W \subseteq \Omega_{V \otimes W} \subseteq (\mathcal{E}_V \otimes_{\min} \mathcal{E}_W)^* =: \mathcal{E}_V^* \otimes_{\max} \mathcal{E}_W^*. \tag{169}$$

That is, the bipartite state space is bound between the min-tensor product of the local state spaces and the max-tensor of the duals of the local effect spaces. Similarly for the bipartite effect space we obtain:

$$\mathcal{E}_V \otimes_{\min} \mathcal{E}_W \subseteq \mathcal{E}_{V \otimes W} \subseteq (\Omega_V \otimes_{\min} \Omega_W)^* =: \Omega_V^* \otimes_{\max} \Omega_W^*. \tag{170}$$

D Tensor representation

Essentially this representation boils down to picking a suitable basis (and dual basis) for each vector space.

We have already seen, via Eq. (43), how a local state of V can be represented as a $n + 1$-dimensional vector. Next we will see how to extend this to arbitrary processes. The simplest way to do so is to introduce a decomposition of the identity into orthogonal rank-1 projectors for each system. There are actually only three relevant systems (and their composites) in the above problem, the two classical systems, $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$, which decompose as:

$$\tag{171}$$

and the GPT system V which decomposes as:

$$\tag{172}$$

$$\tag{173}$$

The e_i are physically realisable effects, however, the v_i are simply vectors in V which satisfy $e_i(v_j) = \delta_{ij}$. It is important that we do not demand that the v_i are physically realisable states, as, for any non-classical GPT, there are insufficient perfectly distinguishable states to span the vector space.

A remark on notation: in the following sections we will also denote the unit effect for the $\boldsymbol{\beta}$ systems as u_β, and their fiducial effects by $\vec{0}$ and $\vec{1}$.

Now, to obtain a tensorial representation of any diagram we simply decompose all of the internal identities in the diagram and attach e_i and v_j to the free inputs and outputs, for

example:

$$ \qquad\qquad\qquad\qquad \mapsto \quad \left[\; \sum_l \;\right]^{ik}_{\;\;\,j} \qquad\qquad\qquad (174) $$

$$ = \quad \sum_l \left[\;\right]^{i}_{\;jl} \left[\;\right]^{lk} \qquad\qquad\qquad (175) $$

$$ =: \quad \sum_l \mathcal{F}^i_{jl}\, s^{lk} \qquad\qquad\qquad (176) $$

A bipartite state, such as s in the above diagram, is therefore represented by a two-index tensor. If this bipartite state is a product state, then it is easy to see that this two-index tensor is simply the Kronecker product of the one-index tensors associated to the two components:

$$ (s_1 \otimes s_2)^{ij} = \left[\begin{array}{cc} e_i & e_j \\ s_1 & s_2 \end{array}\right]^{ij} = \left[\begin{array}{c} e_i \\ s_1 \end{array}\right]^{i} \left[\begin{array}{c} e_j \\ s_2 \end{array}\right]^{j} = s_1^i s_2^j \;. \qquad (177) $$

D.1 Example

Consider the case where $\boldsymbol{\beta} = \{0,1\}$, $\boldsymbol{\eta} = \{0,1\}$, and $|\boldsymbol{\iota}| = 2$. Moreover, take the dimension of the local GPT system to be 3.

Define $\mathcal{F}^a_{xv}$ real tensors where $a, x \in \{0,1\}$ $v \in \{0,1,2\}$ as follows:

$$ \mathcal{F}^a_{xv} := \begin{array}{c|ccc} & v = 0 & v = 1 & v = 2 \\ \hline a = 0, x = 0 & 1 & 0 & 0 \\ a = 1, x = 0 & -1 & 0 & 1 \\ a = 0, x = 1 & 0 & 1 & 0 \\ a = 1, x = 1 & 0 & -1 & 1 \end{array} \qquad (178) $$

We then define our variables s_{vw}, $\mathcal{P}^{pq}_{vw}$ as real tensors with $v, w \in \{0,1,2\}$ $p, q \in \{0,1\}$. The tensors $\mathcal{F}$, $\mathcal{P}$ and s will determine local fiducial measurements, parity reading measurement, and state, respectively.

Given a particular linear functional I^{xy}_{ab} we optimise:

$$ \beta := \sup_{s,\mathcal{P}} \; \sum_{abxy} I^{xy}_{ab} \sum_{v_1,v_2} \mathcal{F}^a_{xv_1} \mathcal{F}^b_{yv_2} s_{v_1 v_2} \qquad (179) $$

subject to the following constraints. Note that in these constraints any sum is implicitly taken over its whole range and there is an implict $\forall$ for any index which is not contracted:

51

- Parity Reading:

$$\sum_p \mathcal{P}_{vw}^{pq} = \mathcal{F}_{1v}^0 \mathcal{F}_{1w}^q + \mathcal{F}_{1v}^1 \mathcal{F}_{1w}^{1\oplus q} \tag{180}$$

$$\sum_q \mathcal{P}_{vw}^{pq} = \mathcal{F}_{0v}^0 \mathcal{F}_{0w}^p + \mathcal{F}_{0v}^1 \mathcal{F}_{0w}^{1\oplus p} \tag{181}$$

- Probabilities for fiducial measurements:

$$\sum_{v_1,v_2} \mathcal{F}_{xv_1}^a \mathcal{F}_{yv_2}^b s_{v_1 v_2} \geq 0 \tag{182}$$

$$1 = \sum_{a,b,v_1,v_2} \mathcal{F}_{xv_1}^a \mathcal{F}_{yv_2}^b s_{v_1 v_2} \tag{183}$$

- Hierarchy L1:

$$\sum_{v_1,v_2} \mathcal{P}_{v_1,v_2}^{p_1 p_2} s_{v_{\pi(1)} v_{\pi(2)}} \geq 0 \tag{184}$$

for all π where π is a permutation of $\{1,2\}$.
- Hierarchy L2:

$$\sum_{v_1,v_2,v_3,v_4} \mathcal{P}_{v_1,v_2}^{p_1 p_2} \mathcal{P}_{v_3,v_4}^{p_3 p_4} s_{v_{\pi(1)} v_{\pi(2)}} s_{v_{\pi(3)} v_{\pi(4)}} \geq 0 \tag{185}$$

$$\sum_{v_1,v_2,v_3,v_4} \mathcal{P}_{v_1,v_2}^{p_1 p_2} \mathcal{F}_{xv_3}^a \mathcal{F}_{yv_3}^b s_{v_{\pi(1)} v_{\pi(2)}} s_{v_{\pi(3)} v_{\pi(4)}} \geq 0 \tag{186}$$

for all π where now π is a permutation of $\{1,2,3,4\}$
- Hierarchy L3:

$$\sum_{v_1,v_2,v_3,v_4} \mathcal{P}_{v_1,v_2}^{p_1 p_2} \mathcal{P}_{v_3,v_4}^{p_3 p_4} \mathcal{P}_{v_5,v_6}^{p_5 p_6} s_{v_{\pi(1)} v_{\pi(2)}} s_{v_{\pi(3)} v_{\pi(4)}} s_{v_{\pi(5)} v_{\pi(6)}} \geq 0 \tag{187}$$

$$\sum_{v_1,v_2,v_3,v_4,v_5,v_6} \mathcal{P}_{v_1,v_2}^{p_1 p_2} \mathcal{P}_{v_3,v_4}^{p_3 p_4} \mathcal{F}_{xv_5}^a \mathcal{F}_{yv_6}^b s_{v_{\pi(1)} v_{\pi(2)}} s_{v_{\pi(3)} v_{\pi(4)}} s_{v_{\pi(5)} v_{\pi(6)}} \geq 0 \tag{188}$$

$$\tag{189}$$

for all π where now π is a permutation of $\{1,2,3,4,5,6\}$.

Notice that the constraints (182) and (183) just say that the

$$p(ab|xy) = \sum_{v_1,v_2} \mathcal{F}_{xv_1}^a \mathcal{F}_{yv_2}^b s_{v_1 v_2} \tag{190}$$

are probability distributions for each fixed x, y.

Also, the form of tensor $\mathcal{F}$ of Eq. (178) further says that the parametrization state

$$s = \left(s_{00}, s_{10}, s_{2,0}, s_{01}, s_{11}, s_{21}, s_{02}, s_{12}, s_{22}\right) \tag{191}$$

has the following interpretation in terms of the box $p(ab|xy)$:

$$s = \left(p_s(00|00), p_s(00|10), p_s^{(2)}(0|0), p_s(00|01), p_s(00|11), p_s^{(2)}(0|1), p_s^{(1)}(0|0), p_s^{(1)}(0|1), 1\right). \tag{192}$$

Here $p^{(1)}$ and $p^{(2)}$ are marginals obtained from $p(ab|xy)$, e.g., $p^{(1)}(a|x) = \sum_b p(ab|xy)$. Note that the latter does not depend on y due to no-signaling of $p(ab|xy)$, which in turn is enforced by the form of Eq. (190) and the definition of $\mathcal{F}$ given in Eq. (178).

E   Convergence of state cone hierarchy

In this appendix we demonstrate that the hierarchy that we define does indeed converge to the cone $K[\mathcal{P}]$.

The cone $K[\mathcal{P}]$ is characterised by the condition: $s \in K[\mathcal{P}]$ if and only if every diagram with only classical inputs and outputs formed from a finite number of processes must be non-negative.

We now show that this condition is equivalent to our hierarchy.

To begin with, note that any diagram in our theory is constructed by wiring together a finite number of each of: i) the bipartite state, s, ii) the controlled fiducial measurement, $\mathcal{F}$, and iii) the parity reading measurement, $\mathcal{P}$. We call these the generating processes. We can therefore classify diagrams by first representing them in terms of the generating processes, and then counting the number of copies, k, of s that appear.

Next, note that if a diagram has only classical inputs and outputs, then any copy of V that appears in the diagram must have a start point and an end point in the diagram. There is only one generating process which can serve as a start point, namely, the bipartite state s, and either $\mathcal{F}$ or $\mathcal{P}$ can serve as the end point.

If we have k copies of the state s within the diagram, then these must therefore be wired into the measurements $\mathcal{F}$ and $\mathcal{P}$. Every such diagram will factorise into totally connected subdiagrams. Note then, that nonnegativity of the full diagram is guaranteed by nonnegativity of the component subdiagrams. That is, to ensure nonnegativity for every diagram (with only classical inputs and outputs) we must only demand nonnegativity of totally connected diagrams (with only classical inputs and outputs).

It is then simple to see that the totally connected diagrams with k copies of s come in two forms. Firstly, those in which there are k copies of $\mathcal{P}$ which the states s are wired to, and secondly, those in which there are $k - 1$ copies of $\mathcal{P}$ and two copies of $\mathcal{F}$. If there were more than two copies of $\mathcal{F}$ then the diagram would necessarily not be totally connected. Clearly, the first of these is captured by condition (68) and the second by the condition (67) of level k in the hierarchy. Therefore, our hierarchy of constraints fully charaterises the cone $K[\mathcal{P}]$.