Bounding the quantum dimension with contextuality

Otfried Guehne, Costantino Budroni, Adan Cabello, Matthias Kleinmann and Jan-Åke Larsson

Linköping University Post Print

Bounding the quantum dimension with contextuality

Otfried Gühne,[1] Costantino Budroni,[1] Adán Cabello,[2] Matthias Kleinmann,[1] and Jan-Åke Larsson[3]

[1]*Naturwissenschaftlich-Technische Fakultät, Universität Siegen, Walter-Flex-Str. 3, D-57068 Siegen*
[2]*Departamento de Física Aplicada II, Universidad de Sevilla, E-41012 Sevilla, Spain*
[3]*Institutionen för Systemteknik och Matematiska Institutionen, Linköpings Universitet, SE-581 83 Linköping, Sweden*

We show that the phenomenon of quantum contextuality can be used to certify lower bounds on the dimension accessed by the measurement devices. To prove this, we derive bounds for different dimensions and scenarios of the simplest noncontextuality inequalities. Some of the resulting dimension witnesses work independently of the prepared quantum state. Our constructions are robust against noise and imperfections, and we show that a recent experiment can be viewed as an implementation of a state-independent quantum dimension witness.

I. INTRODUCTION

The recent progress in the experimental control and manipulation of physical systems at the quantum level opens new possibilities (e.g., quantum communication, computation, and simulation), but, at the same time, demands the development of novel theoretical tools of analysis. There are already tools which allow us to recognize quantum entanglement and certify the usefulness of quantum states for quantum information processing tasks [1,2]. However, on a more fundamental level, there are still several problems which have to be addressed. For example, how can one efficiently test whether measurements actually access all the desired energy levels of an ion? How to certify that the different paths of photons in an interferometer can be used to simulate a given multidimensional quantum system? Similar questions arise in the analysis of experiments with orbital angular momentum, where high-dimensional entanglement can be produced [3,4], or in experiments with electron spins at nitrogen-vacancy centers in diamond, where the quantumness of the measurements should be certified [5].

The challenge is to provide lower bounds on the *dimension* of a quantum system only from the statistics of measurements performed on it. More precisely, to certify lower bounds on the dimension of the underlying Hilbert space where the measurement operators act on. Such bounds can be viewed as lower bounds on the complexity and the number of levels accessed by the measurement devices: If the measurement operators act nontrivially only on a small subspace, then all measurements results can be modeled by using a low-dimensional quantum system only. Note that this is not directly related to the rank of a density matrix. In fact, a pure quantum state acting on a one-dimensional subspace only can still give rise to measurement results, which can only be explained assuming a higher-dimensional Hilbert space.

The problem of estimating the Hilbert space dimension has been considered in different scenarios, and slightly different notions of dimension were involved. Brunner and co-workers introduced the concept of quantum "dimension witnesses" by providing lower bounds on the dimension of composite systems from the violation of Bell inequalities [6,7]. The nonlocal properties of the correlations produced are clearly the resource used for this task. As a consequence, even if the experimenter is able to access and manipulate many levels of her systems locally, but she is not able to entangle those levels,

the above test fails to certify such a dimension. Such a task can therefore be interpreted as a test of the type of entanglement and correlations produced, namely, how many levels or degrees of freedom the experimenter is able to entangle.

In a complementary scenario, several different states of a single particle are prepared and different measurements are carried out [8–10]. This approach has also recently been implemented using photons [11,12]. In this situation, the dimension of the system can be interpreted as the dimension of the set of states the experimenter is able to prepare.

As a third possibility, also the continuous time evolution can be used to bound the dimension of a quantum system [13]. In this case, the relevant notion of dimension is that of the set of states generated by the dynamical evolution of the system.

In this paper, we focus on sequential measurements on a single system, a type of measurements used in tests of quantum contextuality, and we show how they can be used for bounding the dimension of quantum systems. Quantum contextuality is a genuine quantum effect leading to the Kochen-Specker theorem, which states that quantum mechanics is in contradiction to noncontextual hidden variable (NCHV) models [14–18]. In fact, already in the first formulation of the theorem, the dimension plays a central role [14].

We derive bounds for the several important noncontextuality (NC) inequalities for different dimensions and scenarios. The experimental violation of these bounds automatically provides a lower bound on the dimension of the system, showing that NC inequalities can indeed be used as dimension witnesses. Remarkably, contextuality can be used as a resource for bounding the dimension of quantum systems in a state-independent way.

This illustrates clearly the difference with the existing schemes: Dimension witnesses derived according to Refs. [9,10] certify the minimum classical or quantum dimension spanned by a set of preparations. They distinguish between classical and quantum dimension d, but, in general, not between quantum dimension d and classical dimension $d + 1$. They require at least $d + 1$ preparations to certify a dimension d. On the other side, dimension witnesses based on Bell's theorem or contextuality certify the minimum quantum dimension accessed by the measurement devices acting on a system prepared in a single state. Contrary to the Bell scenario [6,7], in our approach, in some cases, the initial state

and its nonlocal properties play no role, and the result of our test can directly be interpreted as the minimal number of levels accessed and manipulated by the measurement apparatus.

The paper is organized as follows. In Sec. I, we discuss the case of state-dependent noncontextuality inequalities, specifically, the Klyachko-Can-Binicioğlu-Shumovsky (KCBS) inequality [19]. In Sec. II, we discuss what happens when the sequences of measurements contain noncompatible measurements. In Secs. IV and V, we apply the same analysis to the case of state-independent noncontextuality inequalities, specifically, the Peres-Mermin (PM) inequality [20–22]. In Sec. VI, we discuss the case of imperfect measurements, then in Sec. VII we show how a recent experimental test of contextuality can be viewed as an implementation of our dimension witness.

II. KCBS INEQUALITY

We first turn to the state-dependent case. The simplest system showing quantum contextuality is a quantum system of dimension three [14]. The simplest NC inequality in three dimensions is the one introduced by Klyachko, Can, Binicioğlu, and Shumovsky (KCBS) [19]. For that, one considers

$$\langle \chi_{\text{KCBS}} \rangle = \langle AB \rangle + \langle BC \rangle + \langle CD \rangle + \langle DE \rangle + \langle EA \rangle, \quad (1)$$

where $A, B, C, D,$ and E are measurements with outcomes -1 and 1, and the measurements in the same mean value $\langle \ldots \rangle$ are compatible [23], i.e., are represented in quantum mechanics by commuting operators. The mean value itself is defined via a sequential measurement: For determining $\langle AB \rangle$, one first measures A and then B on the same system, multiplies the two results, and finally averages over many repetitions of the experiment.

The KCBS inequality states that

$$\langle \chi_{\text{KCBS}} \rangle \overset{\text{NCHV}}{\geqslant} -3, \quad (2)$$

where the notation "$\overset{\text{NCHV}}{\geqslant} -3$" indicates that -3 is the minimum value for any NCHV theory. Here, noncontextuality means that the theory assigns to any observable (say, B) a value independent of which other compatible observable (here, A or C) is measured jointly with it.

In quantum mechanics, a value of $\langle \chi_{\text{KCBS}} \rangle = 5 - 4\sqrt{5} \approx -3.94$ can be reached on a three-dimensional system, if the observables and the initial state are appropriately chosen. This quantum violation of the NCHV bound does not increase in higher-dimensional systems [18,24], and the violation of the KCBS inequality has been observed in recent experiments with photons [25,26].

Given the fact that quantum contextuality requires a three-dimensional Hilbert space, it is natural to ask whether a violation of Eq. (2) implies already that the system is not two dimensional. The following observation shows that this is the case:

Observation 1. Consider the KCBS inequality where the measurements act on a two-dimensional quantum system and are commuting, i.e., $[A,B] = [B,C] = [C,D] = [D,E] =$

$[E,A] = 0$. Then, the classical bound holds:

$$\langle \chi_{\text{KCBS}} \rangle \overset{\text{2D,com.}}{\geqslant} -3. \quad (3)$$

Proof of Observation 1. First, if two observables A and B are compatible, then $|\langle A \rangle \pm \langle AB \rangle| \leqslant 1 \pm \langle B \rangle$. This follows from the fact that A and B have common eigenspaces and the relation holds separately on each eigenspace. Second, in two dimensions, if $[A,B] = 0 = [B,C]$, then either $B = \pm \mathbb{1}$ or $[A,C] = 0$. The reason is that, if B is not the identity, then it has two one-dimensional eigenspaces. These are shared with A and C, so A and C must be simultaneously diagonalizable.

Considering the KCBS operator χ_{KCBS}, the claim is trivial if $A, \ldots, E$ are all compatible because then the relation holds separately on each eigenspace. It is only possible that not all of them commute if there are two groups in the sequence $\{A, B, C, D, E\}$ of operators separated by identity operators. Without loss of generality, we assume that the groups of commuting operators are $\{E, A\}$ and $\{C\}$ so that $B = b\mathbb{1} = \pm \mathbb{1}$ and $D = d\mathbb{1} = \pm \mathbb{1}$. This gives

$$\begin{aligned} \langle \chi_{\text{KCBS}} \rangle &= b\langle A \rangle + b\langle C \rangle + d\langle C \rangle + d(\langle E \rangle + d\langle EA \rangle) \\ &\geqslant b\langle A \rangle + b\langle C \rangle + d\langle C \rangle - 1 - d\langle A \rangle \\ &= (b-d)\langle A \rangle + (b+d)\langle C \rangle - 1 \geqslant -3 \quad (4) \end{aligned}$$

and proves the claim. In this argumentation, setting observables proportional to the identity does not change the threshold, but in general it is important to consider this case, as this often results in higher values. ∎

It should be added that Observation 1 can also be proved using a different strategy: Given two observables on a two-dimensional system, one can directly see that if they commute, then either one of them is proportional to the identity, or their product is proportional to the identity. In both cases, one has a classical assignment for some terms in the KCBS inequality and then one can check by exhaustive search that the classical bound holds. Details are given in Appendix A 1.

Furthermore, Observation 1 can be extended to generalizations of the KCBS inequality with more than five observables [24]: For that, one considers

$$\langle \chi_N \rangle = \sum_{i=1}^{N-1} \langle A_i A_{i+1} \rangle + s \langle A_N A_1 \rangle, \quad (5)$$

where $s = +1$ if N is odd and $s = -1$ if N is even. For this expression, the classical bound for NCHV theories is given by $\langle \chi_N \rangle \geqslant -(N-2)$. In fact, the experiment in Ref. [25] can also be viewed as measurement of $\langle \chi_6 \rangle$.

The discussion of the possible mean values $\langle \chi_N \rangle$ in quantum mechanics differs for even and odd N. If N is odd, the maximal possible quantum mechanical value is $\langle \chi_N \rangle = \Omega_N \equiv -[3N \cos(\pi/N) - N]/[1 + \cos(\pi/N)]$ and this value can already be attained in a three-dimensional system [18,24]. The proof of Observation 1 can be generalized in this case, implying that for two-dimensional systems the classical bound $\langle \chi_N \rangle \geqslant -(N-2)$ holds. So, for odd N, the generalized KCBS inequalities can be used for testing the quantum dimension.

If N is even, the scenario becomes richer: First, quantum mechanics allows us to obtain values of $\langle \chi_N \rangle = \Omega_N \equiv -N \cos(\pi/N)$, but this time this value requires a four-dimensional system [24]. For two-dimensional quantum

systems, the classical bound $\langle\chi_N\rangle \geqslant -(N-2)$ holds. For three-dimensional systems, one can show that if the observables A_i in a joint context are different ($A_i \neq \pm A_{i+1}$) and not proportional to the identity, then still the classical bound holds (for details, see Appendix A 2). However, if two observables are the same, e.g., $A_1 = -A_2$, then $\langle A_1 A_2\rangle = -1$ and $\langle\chi_N\rangle = -1 + \langle\chi_{N-1}\rangle$. In summary, for even N, we have the following hierarchy of bounds:

$$\langle\chi_N\rangle \overset{\text{2D,com.}}{\geqslant} -(N-2) \overset{\text{3D,com.}}{\geqslant} -1 + \Omega_{N-1} \overset{\text{4D,com.}}{\geqslant} \Omega_N. \quad (6)$$

Here, the notation $\overset{\text{2D,com.}}{\geqslant}$, etc., means that this bound holds for commuting observables in two dimensions. All these bounds are tight. This shows that extended KCBS inequalities are even more sensitive to the dimension than the original inequality.

III. KCBS INEQUALITY WITH INCOMPATIBLE OBSERVABLES

In order to apply Observation 1, the observables must be compatible. Since this condition is not easy to guarantee in experiments [27], we should ask whether it is possible to obtain a two-dimensional bound for the KCBS inequality when the observables are not necessarily compatible. We can state the following:

Observation 2. If the observables $A, \ldots, E$ are dichotomic observables but not necessarily commuting, then, for any two-dimensional quantum system,

$$\langle\chi_{\text{KCBS}}\rangle \overset{\text{2D}}{\geqslant} -\frac{5}{4}(1 + \sqrt{5}) \approx -4.04. \quad (7)$$

This bound is tight and can be attained for suitably chosen measurements.

The strategy of proving this bound is the following: If the observables are not proportional to the identity, one can write $A = |A^+\rangle\langle A^+| - |A^-\rangle\langle A^-|$ and $B = |B^+\rangle\langle B^+| - |B^-\rangle\langle B^-|$, and express $|A^+\rangle\langle A^+|$ and $|B^+\rangle\langle B^+|$ in terms of their Bloch vectors $|\mathfrak{a}\rangle$ and $|\mathfrak{b}\rangle$. Then, one finds that

$$\langle AB\rangle = 2|\langle A^+|B^+\rangle|^2 - 1 = \langle\mathfrak{a}|\mathfrak{b}\rangle. \quad (8)$$

This property holds for all projective measurements on two-dimensional systems and is, together with a generalization below [see Eq. (15)], a key idea for deriving dimension witnesses. Note that it implies that the sequential mean value $\langle AB\rangle$ is independent of the initial quantum state and also of the temporal order of the measurements [28]. Equation (8) allows us to transform the KCBS inequality into a geometric inequality for three-dimensional Bloch vectors. Additional details of the proof are given in Appendix A 3.

Observation 2 shows that the bound for NCHV theories can be violated already by two-dimensional systems, if the observables are incompatible. This demonstrates that experiments, which aim at a violation of Eq. (2), also have to test the compatibility of the measured observables, otherwise the violation can be explained without contextuality.

It must be added that Observation 2 can not be used to witness the quantum dimension since one can show that Eq. (7) holds for all dimensions [29]. As we see in the following, this

difficulty can be surmounted by considering NC inequalities in which quantum mechanics reaches the algebraic maximum.

IV. PERES-MERMIN INEQUALITY

In order to derive the state-independent quantum dimension witnesses, let us consider the sequential mean value [20]

$$\langle\chi_{\text{PM}}\rangle = \langle ABC\rangle + \langle bca\rangle + \langle\gamma\alpha\beta\rangle + \langle A\alpha a\rangle + \langle bB\beta\rangle - \langle\gamma cC\rangle, \quad (9)$$

where the measurements in each of the six sequences are compatible. Then, for NCHV theories the bound

$$\langle\chi_{\text{PM}}\rangle \overset{\text{NCHV}}{\leqslant} 4 \quad (10)$$

holds. In a four-dimensional quantum system, however, one can take the following square of observables, known as the Peres-Mermin square [21,22]:

$$\begin{aligned}
A &= \sigma_z \otimes \mathbb{1}, \quad B = \mathbb{1} \otimes \sigma_z, \quad C = \sigma_z \otimes \sigma_z, \\
a &= \mathbb{1} \otimes \sigma_x, \quad b = \sigma_x \otimes \mathbb{1}, \quad c = \sigma_x \otimes \sigma_x, \quad (11) \\
\alpha &= \sigma_z \otimes \sigma_x, \quad \beta = \sigma_x \otimes \sigma_z, \quad \gamma = \sigma_y \otimes \sigma_y.
\end{aligned}$$

These observables lead for any quantum state to a value of $\langle\chi_{\text{PM}}\rangle = 6$, demonstrating state-independent contextuality. The quantum violation has been observed in several recent experiments [30–32]. Note that the sequences in Eq. (9) are defined such that each observable occurs either always in the first or always in the second or always in the third place of a measurement a sequence. This difference to the standard version does not matter at this point (since the observables in any row or column commute), but it will become important below.

The PM inequality is of special interest for our program since it is violated up to the algebraic maximum with four-dimensional quantum systems and the violation is state independent. Therefore, this inequality is a good candidate for dimension witnesses without assumptions on the measurements. First, we can state the following:

Observation 3. If the measurements in the PM inequality are dichotomic observables on a two-dimensional quantum system and if the measurements in each mean value are commuting, then one can not violate the classical bound

$$\langle\chi_{\text{PM}}\rangle \overset{\text{2D, com.}}{\leqslant} 4. \quad (12)$$

If one considers the same situation on a three-dimensional system, then the violation is bounded by

$$\langle\chi_{\text{PM}}\rangle \overset{\text{3D, com.}}{\leqslant} 4(\sqrt{5} - 1) \approx 4.94. \quad (13)$$

These bounds are tight.

The idea for proving this statement is the following: If one considers the three commuting observables in each mean value and assumes that they act on a three-dimensional system, then three cases are possible: (a) one of the three observables is proportional to the identity or (b) the product of two observables is proportional to the identity or (c) the product of all three observables is proportional to the identity. One can directly show that if case (c) occurs in some mean value, then the classical bound $\langle\chi_{\text{PM}}\rangle \leqslant 4$ holds. For the cases (a) and (b), one can simplify the inequality and finds that it always reduces

to a KCBS-type inequality, for which we discussed already the maximal quantum values in different dimensions [see Eq. (6)]. Details are given in Appendix A 4.

V. PM INEQUALITY WITH INCOMPATIBLE OBSERVABLES

Let us now discuss the PM inequality, where the observables are not necessarily compatible. Our results allow us to obtain directly a bound:

Observation 4. Consider the PM operator in Eq. (9), where the measurements are not necessarily commuting projective measurements on a two-dimensional system. Then, we have

$$\langle \chi_{\mathrm{PM}} \rangle \overset{\mathrm{2D}}{\leqslant} 3\sqrt{3} \approx 5.20. \tag{14}$$

Proof of Observation 4. One can directly calculate as in the proof of Observation 2 that for sequences of three measurements on a two-dimensional system

$$\langle ABC \rangle = \langle A \rangle \langle BC \rangle \tag{15}$$

holds. Here, $\langle A \rangle = \mathrm{tr}(\varrho A)$ is the usual expectation value, and $\langle BC \rangle$ is the state-independent sequential expectation value given in Eq. (8). With this, we can write

$$\langle \chi_{\mathrm{PM}} \rangle = \langle A \rangle (\langle BC \rangle + \langle \alpha a \rangle) + \langle b \rangle (\langle ca \rangle + \langle B\beta \rangle)$$
$$+ \langle \gamma \rangle (\langle \alpha\beta \rangle - \langle cC \rangle). \tag{16}$$

Clearly, this is maximal for some combination of $\langle A \rangle = \pm 1$, $\langle b \rangle = \pm 1$, and $\langle \gamma \rangle = \pm 1$. But, for any of these choices, we arrive at an inequality that is discussed in Lemma 7 in Appendix A 3. Note that due to Eq. (15), the order of the measurements matters in the definition of $\langle \chi_{\mathrm{PM}} \rangle$ in Eq. (9). This motivates our choice; in fact, for some other orders (e.g., $\langle \tilde{\chi}_{\mathrm{PM}} \rangle = \langle ABC \rangle + \langle bca \rangle + \langle \beta\gamma\alpha \rangle + \langle A\alpha a \rangle + \langle \beta bB \rangle - \langle \gamma cC \rangle$), Eq. (14) does not hold, and one can reach $\langle \tilde{\chi}_{\mathrm{PM}} \rangle = 1 + \sqrt{9 + 6\sqrt{3}} \approx 5.404$. ∎

The question arises as to whether a high violation of the PM inequality also implies that the system can not be three dimensional and whether a similar bound as Eq. (14) can be derived. While the computation of a bound is not straightforward, a simple argument shows already that measurements on a three-dimensional system can not reach the algebraic maximum $\langle \chi_{\mathrm{PM}} \rangle = 6$ for any quantum state: Reaching the algebraic maximum implies that $\langle ABC \rangle = 1$. This implies that the value of C is predetermined by the values of A and B and the value A of determines the product BC. As this holds for any quantum state, it directly follows that A,B,C (and all the other observables in the PM square) are diagonal in the same basis and commute, so the bound in Observation 3 holds. From continuity arguments it follows that there must be a finite gap between the maximal value of $\langle \chi_{\mathrm{PM}} \rangle$ in three dimensions and the algebraic maximum.

VI. IMPERFECT MEASUREMENTS

In actual experimental implementations, the measurements may not be perfectly projective. It is therefore important to discuss the robustness of our method against imperfections.

Notice that since we are considering sequential measurements, another possibility for maximal violation of the above

inequalities is the use of a classical device with memory, able to keep track of the measurement performed and adjust the outcomes of the subsequent measurements accordingly in order to obtain perfect correlations or anticorrelations. However, as proved in Ref. [28] and also discussed in Ref. [29], such a classical device can not be simulated in quantum mechanics via projective measurements; more general positive-operator-valued measures (POVMs) are necessary.

We therefore limit our analysis to some physically motivated noise models. A noisy projective measurement A may be modeled by a POVM with two effects of the type $E^+ = (1 - p)\mathbb{1}/2 + p|A^+\rangle\langle A^+|$ and $E^- = (1 - p)\mathbb{1}/2 + p|A^-\rangle\langle A^-|$. Then, the probabilities of the POVM can be interpreted as coming from the following procedure: With a probability of p one performs the projective measurement and with a probability of $(1 - p)$ one assigns a random outcome. For this measurement model, one can show that Observation 4 is still valid. Details and a more general POVM are discussed in Appendix A 5. We add that the proof strongly depends on the chosen measurement order in $\langle \chi_{\mathrm{PM}} \rangle$ and that in any case assumptions about the measurement are made, so the dimension witnesses are not completely independent of the measurement device.

The above discussion shows that it is extremely important to test the extent to which the measurements are projective and whether they are compatible. This can be achieved by performing additional tests. For instance, one can measure observable A several times in a sequence $\langle AAA \rangle$ to test whether the measurement is indeed projective. In addition, one may measure the sequence $\langle ABA \rangle$ and compare the results of the two measurements of A, to test whether A and B are compatible. For NC inequalities, it is known how this information can be used to derive correction terms for the thresholds [27], and similar methods can also be applied here.

VII. EXPERIMENTAL RESULTS

To stress the experimental relevance of our findings, let us discuss a recent ion-trap experiment [30]. There, the PM inequality has been measured with the aim to demonstrate state-independent contextuality. For our purpose, it is important that in this experiment also all permutations of the terms in the PM inequality have been measured. This allows also to evaluate our $\langle \chi_{\mathrm{PM}} \rangle$ with the order given in Eq. (9). Experimentally, a value $\langle \chi_{\mathrm{PM}} \rangle = 5.36 \pm 0.05$ has been found. In view of Observation 3, this shows that the data can not be explained by commuting projective measurements on a three-dimensional system. Furthermore, Observation 4 and the discussion above prove that, even if the measurements are noisy and noncommuting, the data can not come from a two-dimensional quantum system.

VIII. GENERALIZATIONS

Generalizations of our results to other inequalities are straightforward: Consider a general noncontextuality inequality invoking measurement sequences of length two and three. For estimating the maximal value for two-dimensional systems (as in Observations 2 and 4), one transforms all sequential measurements via Eqs. (8) and (15) into expressions with

three-dimensional Bloch vectors, which can be estimated. Also, noise robustness for the discussed noise model can be proven, as this follows also from the properties of the Bloch vectors (cf. Proposition 12 in Appendix A 5). In addition, if a statement as in Observation 3 is desired, one can use the same ideas as those presented here since they rely on general properties of commuting observables in three-dimensional space. Consequently, our methods allow us to transform most of the known state-independent NC inequalities (for instance, the ones presented in Refs. [20,33–35]) into witnesses for the quantum dimension.

IX. DISCUSSION AND CONCLUSION

We have shown that the two main noncontextuality inequalities, the KCBS inequality (Observation 1) and the Peres-Mermin inequality (Observations 3 and 4), can be used as dimension witnesses. In particular, Observation 4 shows that the the Peres-Mermin inequality can be used to certify the dimension of a Hilbert space independently of the state preparation and in a noise robust way. Our methods allow the application of other inequalities, showing that contextuality can be used as a resource for dimension tests of quantum systems. Our tests are state independent, in contrast to the existing tests. This can be advantageous in experimental implementations, moreover, it shows that one can bound the dimension of quantum systems without using the properties of the quantum state. We hope that our results stimulate further research to answer a central open question: For which tasks in quantum information processing is quantum contextuality a useful resource?

APPENDIX

1. Alternative proof of Observation 1

For an alternative proof of Observation 1, we need the following lemma:

Lemma 5. If two dichotomic measurements on a two-dimensional quantum system commute $[A_i, A_{i+1}] = 0$, then either

(a) one of the observables is proportional to the identity $A_i = \pm \mathbb{1}$ or $A_{i+1} = \pm \mathbb{1}$ or

(b) the product of the two observables is proportional to the identity $A_i A_{i+1} = \pm \mathbb{1}$.

Proof of Lemma 5. This fact can easily be checked: the observables A_i and A_{i+1} are diagonal in the same basis and the entries on the diagonal can only be ± 1. Then, only the two cases outlined above are possible. ∎

Alternative proof of Observation 1. With the help of Lemma 5, one can consider each term of the KCBS inequality and make there six possible replacements. For instance, the term $\langle AB \rangle$ may be replaced by $\langle AB \rangle \mapsto \pm \langle B \rangle$ (if one sets

$A \mapsto \pm \mathbb{1}$) or $\langle AB \rangle \mapsto \pm \langle A \rangle$ (if one sets $B \mapsto \pm \mathbb{1}$) or $\langle AB \rangle \mapsto \pm 1$. This results in a finite set of $6^5 = 7776$ possible replacements. Some of them are contradictory and can be disregarded, e.g., if one sets $B \mapsto \mathbb{1}$ from the term $\langle AB \rangle$ and $C \mapsto \mathbb{1}$ from the term $\langle CD \rangle$, then one can not set $\langle BC \rangle \mapsto -1$ anymore. For the remaining replacements, one can directly check with a computer that the $\langle \chi_{\text{KCBS}} \rangle$ reduces to the classical bound. ∎

2. Detailed discussion of the generalized KCBS inequalities

First, we prove the following statement:

Lemma 6. Consider the generalized KCBS operator

$$\langle \chi_N \rangle = \sum_{i=1}^{N-1} \langle A_i A_{i+1} \rangle - \langle A_N A_1 \rangle \tag{A1}$$

for N even, where the A_i are dichotomic observables on a three-dimensional system, which are not proportional to the identity. Furthermore, the commuting pairs should not be equal, that is, $A_i \neq A_{i+1}$. Then, the bound

$$\langle \chi_N \rangle \geqslant -(N-2) \tag{A2}$$

holds.

Proof of Lemma 6. From the conditions, it follows that the observables have to be of the form $A_i = \pm(\mathbb{1} - 2|a_i\rangle\langle a_i|)$ with $\langle a_i | a_{i+1} \rangle = 0$. This implies that the sequential measurements can be rephrased via $A_i A_{i+1} = \pm(\mathbb{1} - 2|a_i\rangle\langle a_i| - 2|a_{i+1}\rangle\langle a_{i+1}|)$. Let us first assume that the signs in front of the A_i are alternating, that is, $A_i = +(\mathbb{1} - 2|a_i\rangle\langle a_i|)$ for odd i and $A_i = -(\mathbb{1} - 2|a_i\rangle\langle a_i|)$ for even i. Then, a direct calculation leads to

$$\langle \chi_N \rangle = -(N-2) + 4\left\langle \sum_{k=2}^{N-1} |a_k\rangle\langle a_k| \right\rangle. \tag{A3}$$

From this, $\langle \chi_N \rangle \geqslant -(N-2)$ follows since the operator in the sum is positive semidefinite.

A general distribution of signs for the A_i results in a certain distribution of signs for the $A_i A_{i+1}$. If I denotes the set of index pairs $(k, k+1)$, where $A_k A_{k+1} = +(\mathbb{1} - 2|a_k\rangle\langle a_k| - 2|a_{k+1}\rangle\langle a_{k+1}|)$, then I has always an odd number of elements. We can then write

$$\langle \chi_N \rangle = -(N-2) + 2(|I| - 1) + 4\left\langle \sum_{k=1}^{N} \alpha_k |a_k\rangle\langle a_k| \right\rangle, \tag{A4}$$

where $\alpha_k = 1$ if both $(k, k+1) \notin I$ and $(k-1, k) \notin I$, $\alpha_k = 0$ if either $(k, k+1) \in I, (k-1, k) \notin I$ or $(k, k+1) \notin I, (k-1, k) \in I$, and $\alpha_k = -1$ if both $(k, k+1) \in I$ and $(k-1, k) \in I$.

It remains to show that the last two terms are non-negative. The main idea to prove this is to use the fact that an operator such as $X = \mathbb{1} - |a_i\rangle\langle a_i| - |a_{i+1}\rangle\langle a_{i+1}|$ is positive semidefinite since $|a_i\rangle$ and $|a_{i+1}\rangle$ are orthogonal.

More explicitly, let us first consider the case where the index pairs in I are connected and distinguish different cases for the number of elements in I. If $|I| = 1$, there are no k with $\alpha_k = -1$, so $2(|I| - 1) + 4\langle \sum_{k=1}^{N} \alpha_k |a_k\rangle\langle a_k| \rangle \geqslant 0$. If $|I| = 2$, then $I = \{(i-1, i), (i, i+1)\}$ and there is a single $\alpha_i = -1$. In this case, one has $2|I| + 4\langle \sum_{k=1}^{N} \alpha_k |a_k\rangle\langle a_k| \rangle \geqslant 0$. This is not yet the desired bound, but it will be useful later.

If $|I| = 3$, then $I = \{(i-1,i),(i,i+1),(i+1,i+2)\}$ and we have $\alpha_i = \alpha_{i+1} = -1$. But now, the fact that $X = \mathbb{1} - |a_i\rangle\langle a_i| - |a_{i+1}\rangle\langle a_{i+1}| \geqslant 0$ directly implies that $2(|I| - 1) + 4\langle\sum_{k=1}^{N}\alpha_k|a_k\rangle\langle a_k|\rangle \geqslant 0$. If $|I| = 4$, there are three $\alpha_k = -1$ and we can use $X \geqslant 0$ two times, showing that again $2|I| + 4\langle\sum_{k=1}^{N}\alpha_k|a_k\rangle\langle a_k|\rangle \geqslant 0$. All this can be iterated, resulting in two different bounds, for $|I|$ odd and $|I|$ even.

To complete the proof, we have to consider a general I which does not necessarily form a single block. One can then consider the different blocks and, since $|I|$ is odd, at least one of the blocks contains an odd number of index pairs. Then, summing up the bound for the single blocks leads to $2(|I| - 1) + 4\langle\sum_{k=1}^{N}\alpha_k|a_k\rangle\langle a_k|\rangle \geqslant 0$. ∎

Finally, in order to justify Eq. (6) in the main text for the three-dimensional case, we have to discuss what happens if one of the observables is proportional to the identity. However, then the mean value $\langle\chi_N\rangle$ reduces to inequalities which will be discussed later (see Lemma 9 in Appendix A 4).

3. Detailed proof of Observation 2

For computing the minimal value in two-dimensional systems, we need the following lemma. Note that the resulting value has been reported before [36], so the main task is to prove rigorously that this is indeed optimal.

Lemma 7. Let $|\mathfrak{a}_i\rangle \in \mathbb{R}^3$ be normalized real three-dimensional vectors and define

$$\chi_N = \sum_{i=1}^{N}\langle\mathfrak{a}_i|\mathfrak{a}_{i+1}\rangle \text{ for } N \text{ odd,} \tag{A5a}$$

$$\chi_N = -\langle\mathfrak{a}_1|\mathfrak{a}_2\rangle + \sum_{i=2}^{N}\langle\mathfrak{a}_i|\mathfrak{a}_{i+1}\rangle \text{ for } N \text{ even.} \tag{A5b}$$

Then, we have

$$\chi_N \geqslant -N\cos\left(\frac{\pi}{N}\right). \tag{A6}$$

Proof of Lemma 7. We write $|\mathfrak{a}_i\rangle = \{\cos(\alpha_i), \sin(\alpha_i)\cos(\beta_i), \sin(\alpha_i)\sin(\beta_i)\}$ and then we have

$$\chi_N = \sum_{i=1}^{N}[\pm][\cos(\alpha_i)\cos(\alpha_{i+1})$$
$$+ \cos(\beta_i - \beta_{i+1})\sin(\alpha_i)\sin(\alpha_{i+1})], \tag{A7}$$

where the symbol $[\pm]$ denotes the possibly changing sign of the term with $i = 1$. Let us first explain why the minimum of this expression can be obtained by setting all the $\beta_i = 0$. Without losing generality, we can assume that $|\mathfrak{a}_1\rangle$ points in the x direction, i.e., $\alpha_1 = 0$ and $\sin(\alpha_1) = 0$. Then, only $N - 2$ terms of the type $\sin(\alpha_i)\sin(\alpha_{i+1})$ remain and all of them have a positive prefactor. For given values of β_i we can choose the signs of $\alpha_2, \ldots, \alpha_{N-1}$ such that all these terms are negative, while the other parts of the expression are not affected by this. Then, it is clearly optimal to choose $\beta_2 = \beta_3 = \ldots = \beta_N = 0$. This means that all the vectors lie in the x-y plane.

Having set all $\beta_i = 0$, the expression is simplified to $\chi_N = \sum_{i=1}^{N}[\pm]\cos(\alpha_i - \alpha_{i+1})$. We use the notation $\delta_i = \alpha_i - \alpha_{i+1}$ and minimize $\sum_{i=1}^{N}[\pm]\cos(\delta_i)$ under the constraint

$\sum_{i=1}^{N}\delta_i = 0$. Using Lagrange multipliers, it follows that $[\pm]\sin(\delta_i) = \lambda$ for all i.

If N is odd, this means that we can express any δ_i as $\delta_i = \pi/2 \pm \vartheta + 2\pi k_i$ with $\vartheta \geqslant 0$. From $\cos(\pi/2 + \vartheta + 2\pi k_i) = -\cos(\pi/2 - \vartheta + 2\pi k_i)$, it follows that the sign in front of the ϑ should be identical for all δ_i, otherwise, the expression is not minimized. Let us first consider the case that all signs are positive. From the condition $\sum_{i=1}^{N}\delta_i = 0$, it follows that $N(\pi/2) + N\vartheta + 2\pi K = 0$, with $K = \sum_{i=1}^{N}k_i$. Since we wish to minimize χ_N, the angles δ_i should be as close as possible to π, which means that $|\vartheta - \pi/2|$ should be minimal. This leads to the result that one has to choose $K = -(N \pm 1)/2$. Computing the corresponding ϑ leads to $\vartheta = \pi/2 \pm \pi/N$, which results in Eq. (A6). If the signs in front of all ϑ are negative, one can make a similar argument, but this time has to minimize $|\delta_i + \pi|$ or $|\vartheta - 3\pi/2|$. This leads to the same solutions.

If N is even, one has for $i = 2, \ldots, N$ again $\delta_i = \pi/2 \pm \vartheta + 2\pi k_i$ and the first δ_1 can be written as $\delta_1 = -\pi/2 \pm \vartheta + 2\pi k_1$. One can directly see that if the signs in front of ϑ are positive (negative) for all $i = 2, \ldots, N$, it has to be positive (negative) also for $i = 1$. A direct calculation as before leads to $\vartheta = \pi/2 \pm \pi/N$ and, again, to the same bound of Eq. (A6). ∎

Proof of Observation 2. Let us first assume that none of the observables are proportional to the identity, and consider a single sequential measurement $\langle AB\rangle$ of two dichotomic non-commuting observables $A = |A^+\rangle\langle A^+| - |A^-\rangle\langle A^-| = P_+^A - P_-^A$ and $B = |B^+\rangle\langle B^+| - |B^-\rangle\langle B^-| = P_+^B - P_-^B$. We can also express $|A^+\rangle\langle A^+|$ and $|B^+\rangle\langle B^+|$ in terms of their Bloch vectors $|\mathfrak{a}\rangle$ and $|\mathfrak{b}\rangle$. Then, we have that

$$\langle AB\rangle = 2|\langle A^+|B^+\rangle|^2 - 1 = \langle\mathfrak{a}|\mathfrak{b}\rangle. \tag{A8}$$

Note that this means that the mean value $\langle AB\rangle$ is independent of the initial quantum state. To see this relation, we write $\langle AB\rangle = \text{tr}(P_+^B P_+^A \varrho P_+^A P_+^B) - \text{tr}(P_-^B P_+^A \varrho P_+^A P_-^B) - \text{tr}(P_+^B P_-^A \varrho P_-^A P_+^B) + \text{tr}(P_-^B P_-^A \varrho P_-^A P_-^B)$. Using the fact that in a two-dimensional system $|\langle A^+|B^+\rangle|^2 = |\langle A^-|B^-\rangle|^2$ and $|\langle A^-|B^+\rangle|^2 = |\langle A^+|B^-\rangle|^2$ holds, and $\text{tr}(\varrho) = 1$, this can directly be simplified to the above expression. Using the above expression, we can write $\langle\chi_{\text{KCBS}}\rangle = \sum_{i=1}^{5}\langle\mathfrak{a}_i|\mathfrak{a}_{i+1}\rangle$. Then, Lemma 7 proves the desired bound.

It remains to discuss the case where one or more observables in the KCBS inequality are proportional to the identity. Let us first assume that only one observable, say A_1, is proportional to the identity. Then, if the Bloch vector of ϱ is denoted by $|\mathfrak{r}\rangle$, a direct calculation shows that the KCBS operator reads as

$$\langle\chi_{\text{KCBS}}\rangle = \langle\mathfrak{r}|\mathfrak{a}_2\rangle + \sum_{i=2}^{4}\langle\mathfrak{a}_i|\mathfrak{a}_{i+1}\rangle + \langle\mathfrak{a}_5|\mathfrak{r}\rangle, \tag{A9}$$

and Lemma 7 proves again the claim. If two observables A_i and A_j are proportional to the identity, the same rewriting can be applied, if A_i and A_j do not occur jointly in one correlation term. This is the case if $j \neq i \pm 1$. In the other case (say, $A_1 = \mathbb{1}$ and $A_2 = -\mathbb{1}$), one has $\langle A_1 A_2\rangle = -1$ and can rewrite

$$\langle\chi_{\text{KCBS}}\rangle = -1 - \langle\mathfrak{r}|\mathfrak{a}_2\rangle + \sum_{i=3}^{4}\langle\mathfrak{a}_i|\mathfrak{a}_{i+1}\rangle + \langle\mathfrak{a}_4|\mathfrak{r}\rangle, \tag{A10}$$

and Lemma 7 implies that $\langle \chi_{\mathrm{KCBS}} \rangle \geqslant -4\cos(\pi/4) - 1 = -2\sqrt{2} - 1 > -5\cos(\pi/5) = -5(1 + \sqrt{5})/4$. If more than two observables are proportional to the identity, the bound can be proven similarly. ∎

4. Proof of Observation 3

We need a whole sequence of lemmas:

Lemma 8. If one has three dichotomic measurements $A_i, i = 1, 2, 3$ on a three-dimensional quantum system which commute pairwise $[A_i, A_j] = 0$, then either

(a) one of the observables is proportional to the identity $A_i = \pm\mathbb{1}$ for some i or

(b) the product of two observables of the three observables is proportional to the identity $A_i A_j = \pm\mathbb{1}$ for some pair i, j or

(c) the product of all three observables is proportional to the identity $A_1 A_1 A_3 = \pm\mathbb{1}$.

Note that these cases are not exclusive and that for a triple of observables, several of these cases may apply at the same time.

Proof of Lemma 8. This can be proven in the same way as Lemma 5 since all A_i are diagonal in the same basis. ∎

Lemma 9. For sequences of dichotomic measurements, the following inequalities hold:

$$\eta_N \equiv \langle A_1 \rangle + \sum_{i=1}^{N-1} \langle A_i A_{i+1} \rangle - \langle A_N \rangle \leqslant N - 1. \quad \text{(A11)}$$

Here, it is always assumed that two observables which occur in the same sequence commute. Moreover, if we define

$$\zeta_N \equiv \sum_{i=1}^{N} \langle A_i A_{i+1} \rangle - \langle A_N A_1 \rangle, \quad \text{(A12)}$$

then we have

$$\zeta_N \leqslant N - 2 \quad \text{(A13)}$$

in two-dimensional systems, while for three-dimensional systems,

$$\zeta_3 \leqslant 1; \zeta_4 \leqslant 2, \quad \zeta_5 \leqslant \sqrt{5}(4 - \sqrt{5}),$$
$$\zeta_6 \leqslant 1 + \sqrt{5}(4 - \sqrt{5}) = 4(\sqrt{5} - 1) \quad \text{(A14)}$$

holds.

Proof of Lemma 9. If we consider η_N for $N = 2$ both observables commute and the claim $\langle A_1 \rangle + \langle A_1 A_2 \rangle - \langle A_2 \rangle \leqslant 1$ is clear, as it holds for any eigenvector. The bounds for general η_N follow by induction, where in each step of the induction $\langle A_N A_{N+1} \rangle - \langle A_{N+1} \rangle \leqslant 1 - \langle A_N \rangle$ is used, but this is nothing but the bound for $N = 2$.

The bounds for ζ_N are just the ones derived for the generalized KCBS inequalities [see Eq. (6) in the main text and Appendix A 2]. ∎

Lemma 10. Consider the PM square with dichotomic observables on a three-dimensional system, where for one column and one row only the case (c) in Lemma 8 applies. Then, one can not violate the classical bound and one has $\langle \chi_{\mathrm{PM}} \rangle \leqslant 4$.

Proof of Lemma 10. Let us consider the case that the condition holds for the first column and the first row; the other cases are analogous. Then, none of the observables A, B, C, a, α

are proportional to the identity since, otherwise, case (a) in Lemma 8 would apply. These observables can all be written as

$$A = \pm(\mathbb{1} - 2|A\rangle\langle A|), \quad \text{(A15)}$$

with some vector $|A\rangle$, and the vector $|A\rangle$ characterizes the observable A up to the total sign uniquely. In this notation, two observables X and Y commute if and only if the corresponding vectors $|X\rangle$ and $|Y\rangle$ are the same or orthogonal. For our situation, it follows that the vectors $|A\rangle, |B\rangle$, and $|C\rangle$ form an orthonormal basis of the three-dimensional space since if two of them were the same, then for the first row also the case (b) in Lemma 8 would apply. Similarly, the vectors $|A\rangle, |a\rangle$, and $|\alpha\rangle$ form another orthonormal basis of the three-dimensional space. We can distinguish two cases:

Case 1: The vector $|B\rangle$ is neither orthogonal nor parallel to $|a\rangle$. From this, it follows that $|B\rangle$ is also neither orthogonal nor parallel to $|\alpha\rangle$ and, similarly, $|C\rangle$ is neither orthogonal nor parallel to $|a\rangle$ and $|\alpha\rangle$ and vice versa.

Let us consider the observable b in the PM square. This observable can be proportional to the identity, but if this is not the case, the corresponding vector $|b\rangle$ has to be parallel or orthogonal to $|B\rangle$ and $|a\rangle$. Since $|B\rangle$ and $|a\rangle$ are neither orthogonal nor parallel, it has to be orthogonal to both, which means that it is parallel to $|A\rangle$. Consequently, the observable b is either proportional to the identity or proportional to A. Similarly, all the other observables β, c, and γ are either proportional to the identity or proportional to A.

Let us now consider the expectation value of the PM operator $\langle \chi_{\mathrm{PM}} \rangle$ for some quantum state ϱ. We denote this expectation value as $\langle \chi_{\mathrm{PM}} \rangle_\varrho$ in order to stress the dependence on ϱ. The observable A can be written as $A = P_+ - P_-$, where P_+ and P_- are the projectors onto the positive or negative eigenspace. One of these projectors is one dimensional and equals $|A\rangle\langle A|$, and the other other one is two dimensional. For definiteness, let us take $P_+ = |A\rangle\langle A|$ and $P_- = \mathbb{1} - |A\rangle\langle A|$.

Instead of ϱ, we may consider the depolarized state $\sigma = p_+ \varrho_+ + p_- \varrho_-$, with $\varrho_\pm = P_\pm \varrho P_\pm / p_\pm$ and $p_\pm = \mathrm{tr}(P_\pm \varrho P_\pm)$. Our first claim is that, in our situation,

$$\langle \chi_{\mathrm{PM}} \rangle_\varrho = \langle \chi_{\mathrm{PM}} \rangle_\sigma = p_+ \langle \chi_{\mathrm{PM}} \rangle_{\varrho_+} + p_- \langle \chi_{\mathrm{PM}} \rangle_{\varrho_-}. \quad \text{(A16)}$$

It suffices to prove this for all rows and columns separately. Since the observables in each column or row commute, we can first measure observables which might be proportional to A. For the first column and the first row, the statement is clear: We first measure A and the result is the same for ϱ and σ. After the measurement of A, however, the state ϱ is projected either onto ϱ_+ or ϱ_-. Therefore, for the following measurements, it does not matter whether the initial state was ϱ or σ. As an example for the other rows and columns, we consider the second column. Here, we can first measure β and then b and finally B. If β or b are proportional to A, then the statement is again clear. If both β and b are proportional to the identity, then the measurement of $\langle \beta b B \rangle_\varrho$ equals $\pm\langle B \rangle_\varrho$. Then, however, one can directly calculate that $\langle B \rangle_\varrho = \langle B \rangle_\sigma$ since B and A commute.

Having established the validity of Eq. (A16), we proceed by showing that for for each term $\langle \chi_{\mathrm{PM}} \rangle_{\varrho_+}$ and $\langle \chi_{\mathrm{PM}} \rangle_{\varrho_-}$ separately the classical bound holds. For $\langle \chi_{\mathrm{PM}} \rangle_{\varrho_+}$ this is clear: Since $P_+ = |A\rangle\langle A|$, we have that $\varrho_+ = |A\rangle\langle A|$ and $|A\rangle$ is an eigenvector of all observables occurring in the PM square.

Therefore, the results obtained in $\langle\chi_{\mathrm{PM}}\rangle_{\varrho_+}$ correspond to a classical assignment of ± 1 to all observables, and $\langle\chi_{\mathrm{PM}}\rangle_{\varrho_+} \leqslant 4$ follows. For the other term $\langle\chi_{\mathrm{PM}}\rangle_{\varrho_-}$, the problem is effectively a two-dimensional one, and we can consider the restriction of the observables to the two-dimensional space, e.g., $\bar{A} = P_- A P_-$, etc. In this restricted space, we have that $\bar{A}, \bar{b}, \bar{\beta}, \bar{c}$, and $\bar{\gamma}$ are all of them proportional to the identity and, therefore, result in a classical assignment ± 1 independent of ϱ_-. Let us denote these assignments by $\hat{A}, \hat{b}, \hat{\beta}, \hat{c}$, and $\hat{\gamma}$. Then, it remains to be shown that

$$\mathcal{Z} = \hat{A}[\langle\bar{B}\bar{C}\rangle_{\varrho_-} + \langle\bar{\alpha}\bar{a}\rangle_{\varrho_-}] + \hat{b}\hat{c}\langle\bar{a}\rangle_{\varrho_-}$$
$$+ \hat{\beta}\hat{\gamma}\langle\bar{\alpha}\rangle_{\varrho_-} + \hat{b}\hat{\beta}\langle\bar{B}\rangle_{\varrho_-} - \hat{c}\hat{\gamma}\langle\bar{C}\rangle_{\varrho_-} \leqslant 4 \quad \text{(A17)}$$

for all classical assignments and for all states ϱ_-. For observables $\bar{B}$ and $\bar{C}$ we have furthermore that $\bar{B}\bar{C} = \pm\mathbb{1}$ (see Lemma 5), hence, $\bar{B} = \pm\bar{C}$ and similarly $\bar{a} = \pm\bar{\alpha}$. If one wishes to maximize $\mathcal{Z}$ for the case $\hat{A} = +1$, one has to choose $\bar{B} = \bar{C}$ and $\bar{a} = \bar{\alpha}$. Then, the product of the four last terms in $\mathcal{Z}$ equals -1, and $\mathcal{Z} \leqslant 4$ holds. For the case $\hat{A} = -1$, one chooses $\bar{B} = -\bar{C}$ and $\bar{a} = -\bar{\alpha}$, but still the product of the four last terms in $\mathcal{Z}$ equals -1, and $\mathcal{Z} \leqslant 4$. This finishes the proof of the first case.

Case 2: The bases $|A\rangle, |B\rangle, |C\rangle$ *and* $|A\rangle, |a\rangle, |\alpha\rangle$ *are (up to some permutations or signs) the same.* For instance, we can have the case in which $|B\rangle = |a\rangle$ and $|C\rangle = |\alpha\rangle$; the other possibilities can be treated similarly.

In this case, since $|B\rangle$ and $|\alpha\rangle$ are orthogonal, the observable β has to be either proportional to the identity or proportional to A. For the same reason, c has to be either proportional to the identity or to A.

Let us first consider the case in which one of the observables β and c is proportional to A, say $\beta = \pm A$ for definiteness. Then, since $|\beta\rangle = |A\rangle$ and $|B\rangle$ are orthogonal, b can only be the identity or proportional to C. Similarly, γ can only be the identity or proportional to C. It follows that *all* nine observables in the PM square are diagonal in the basis $|A\rangle, |B\rangle, |C\rangle$, and all observables commute. Then, $\langle\chi_{\mathrm{PM}}\rangle \leqslant 4$ follows, as this inequality holds in any eigenspace.

Second, let us consider the case in which β and c are both proportional to the identity. This results in fixed assignments $\hat{\beta}$ and $\hat{c}$ for them. Moreover, B and a differ only by a sign $\hat{\mu}$ (that is, $a = \hat{\mu}B$) and C and α differ only by a sign $\hat{\nu}$ (i.e., $\alpha = \hat{\nu}C$). So, we have to consider

$$\mathcal{X} = \langle ABC\rangle + \hat{\mu}\hat{\nu}\langle ABC\rangle + \hat{\beta}\langle Bb\rangle$$
$$+ \hat{\mu}\hat{c}\langle Bb\rangle + \hat{\nu}\hat{\beta}\langle C\gamma\rangle - \hat{c}\langle C\gamma\rangle. \quad \text{(A18)}$$

In order to achieve $\mathcal{X} > 4$, one has to choose $\hat{\mu} = \hat{\nu}, \hat{\beta} = \hat{\mu}\hat{c}$, and $\hat{c} = -\hat{\nu}\hat{\beta}$. However, the latter is equivalent to $\hat{\beta} = -\hat{\nu}\hat{c}$, showing that this assignment is not possible. Therefore, $\mathcal{X} \leqslant 4$ has to hold. This finishes the proof of the second case. ∎

Lemma 11. Consider the PM square with dichotomic observables on a three-dimensional system, where for one column (or one row) only the case (c) in Lemma 8 applies. Then, one can not violate the classical bound and one has $\langle\chi_{\mathrm{PM}}\rangle \leqslant 4$.

Proof of Lemma 11. We assume that the condition holds for the first column. Then, none of the observables A, a, and α are proportional to the identity, and the corresponding vectors $|A\rangle, |a\rangle$, and $|\alpha\rangle$ form an orthonormal basis of the three-dimensional space.

The idea of our proof is to consider possible other observables in the PM square, which are not proportional to the identity, but also not proportional to A, a, or α. We will see that there are not many possibilities for the observables, and in all cases the bound $\langle\chi_{\mathrm{PM}}\rangle \leqslant 4$ can be proved explicitly.

First, consider the case that there all nontrivial observables in the PM square are proportional to A, a, or α. This means that all observables in the PM square are diagonal in the basis defined by $|A\rangle, |a\rangle$, and $|\alpha\rangle$, and all observables commute. But, then the bound $\langle\chi_{\mathrm{PM}}\rangle \leqslant 4$ is clear.

Second, consider the case that there are several nontrivial observables, which are *not* proportional to A, a, or α. Without losing generality, we can assume that the first of these observables is B. This implies that $|B\rangle$ is orthogonal to $|A\rangle$ and lies in the plane spanned by $|a\rangle$ and $|\alpha\rangle$, but $|a\rangle \neq |B\rangle \neq |\alpha\rangle$. It follows for the observables b and β that they can only be proportional to the identity or to A (see Case 1 in Lemma 10). We denote this as $b = \hat{b}[A]$, where $[A] = A$ or $\mathbb{1}$, and $\hat{b}$ denotes the proper sign, i.e., $b = \hat{b}A$ or $b = \hat{b}\mathbb{1}$. Similarly, we write $\beta = \hat{\beta}[A]$.

Let us assume that there is a second nontrivial observable which is not proportional to A, a, or α (but it might be proportional to B). We can distinguish three cases:

(i) First, this observable can be given by C and C is not proportional to B. Then, this is exactly the situation of Case 1 in Lemma 10, and $\langle\chi_{\mathrm{PM}}\rangle \leqslant 4$ follows.

(ii) Second, this observable can be given by C. However, C is proportional to B. Then, $c = \hat{c}[A]$ and $\gamma = \hat{\gamma}[A]$ follows. Now the proof can proceed as in Case 1 of Lemma 10. One arrives to the same Eq. (A17), with the extra condition that $\bar{B} = \pm\bar{C}$, which was deduced after Eq. (A17) anyway. Therefore, $\langle\chi_{\mathrm{PM}}\rangle \leqslant 4$ has to hold.

(iii) Third, this observable can be given by c. Then, it can not be proportional to B since $|B\rangle$ is not orthogonal to $|a\rangle$. It first follows that $C = \hat{C}[a]$ and $\gamma = \hat{\gamma}[a]$. Combined with the properties of B, one finds that $C = \hat{C}\mathbb{1}$ and $b = \hat{b}\mathbb{1}$ has to hold. Then, the PM inequality reads as

$$\mathcal{Y} = \langle A\alpha a\rangle + \langle B(A\hat{C} + \hat{b}\hat{\beta}[A])\rangle$$
$$+ \hat{\beta}\hat{\gamma}\langle\alpha[A][a]\rangle + \langle c(\hat{b}a - \hat{C}\hat{\gamma}[a])\rangle. \quad \text{(A19)}$$

In this expression, the observables B and c occur only in a single term and a single context. Therefore, for any quantum state, we can obtain an upper bound on $\mathcal{Y}$ by replacing $B \mapsto \pm\mathbb{1}$ and $c \mapsto \pm\mathbb{1}$ with appropriately chosen signs. However, with this replacement, all observables occurring in $\mathcal{Y}$ are diagonal in the basis defined by $|A\rangle, |a\rangle$, and $|\alpha\rangle$, and $\mathcal{Y} = \langle\chi_{\mathrm{PM}}\rangle \leqslant 4$ follows.

In summary, the discussion of the cases (i), (ii), and (iii) has shown the following: It is not possible to have three nontrivial observables in the PM square, which are all of them not proportional to A, a, or α. If one has two of such observables, then the classical bound has been proven.

It remains to be discussed what happen if one has only one observable (say, B), which is not proportional to A, a, or α. However, then the PM inequality can be written similarly as in Eq. (A19), and B occurs in a single context. We can set again $B \mapsto \pm\mathbb{1}$ and the claim follows. ∎

Finally, we can prove our Observation 3:

Proof of Observation 3. Lemmas 10 and 11 solve the problem, if case (c) in one column or row happens. Therefore, we can assume that in all columns and all rows only the cases (a) or (b) from Lemma 8 apply. However, in these cases, we obtain a simple replacement rule: For case (a), one of the observables has to be replaced with a classical value ± 1 and, for case (b), one of the observables can be replaced by a different one from the same row or column. In both cases, the PM inequality is simplified.

For case (a), there are six possible replacement rules, as one of the three observables must be replaced by ± 1. Similarly, for case (b), there are six replacement rules. Therefore, one obtains a finite number, namely, $(6 + 6)^6$ possible replacements. As in the case of the KCBS inequality (see the alternative proof of Observation 1 in Appendix A 1), some of them lead to contradictions (e.g., one may try to set $A = +\mathbb{1}$ from the first column, but $A = -\mathbb{1}$ holds due to the rule from the first row). Taking this into account, one can perform an exhaustive search of all possibilities, preferably by computer. For all cases, either the classical bound holds trivially (e.g., because the assignments require already that one row is -1) or the PM inequality can be reduced, up to some constant, to one of the inequalities in Lemma 8. In most cases, one obtains the classical bound. However, in some cases, the PM inequality is reduced to $\langle \chi_{\mathrm{PM}} \rangle = \zeta_5 + 1$ or $\langle \chi_{\mathrm{PM}} \rangle = \zeta_6$. To give an example, one may consider the square

$$\begin{bmatrix} A & B & C \\ a & b & c \\ \alpha & \beta & \gamma \end{bmatrix} = \begin{bmatrix} A & \mathbb{1} & C \\ a & b & \mathbb{1} \\ \mathbb{1} & \beta & \gamma \end{bmatrix}, \qquad (A20)$$

which results in $\langle \chi_{\mathrm{PM}} \rangle = \zeta_6$ for appropriately chosen A_i. Therefore, from Lemma 9 follows that in three dimensions $\langle \chi_{\mathrm{PM}} \rangle = 4(\sqrt{5} - 1) \approx 4.94$ holds and can indeed be reached. ∎

5. Imperfect measurements

In this section, we discuss the noise robustness of Observation 4. In the first subsection, we prove that Observation 4 also holds for the model of noisy measurements explained in the main text. In the second subsection, we discuss a noise model that reproduces the probabilities of the most general POVM.

a. Noisy measurements

In order to explain the probabilities from a noisy measurement, we first consider the following measurement model: Instead of performing the projective measurement A, one of two possible actions are taken:

(a) with a probability p_A the projective measurement is performed, or

(b) with a probability $1 - p_A$ a completely random outcome ± 1 is assigned independently of the initial state. Here, the results $+1$ and -1 occur with equal probability.

In case (b), after the assignment the physical system is left in one of two possible states ϱ^+ or ϱ^-, depending on the assignment. We will not make any assumptions on $\varrho^\pm$.

Before formulating and proving a bound on $\langle \chi_{\mathrm{PM}} \rangle$ in this scenario, it is useful to discuss the structure of $\langle \chi_{\mathrm{PM}} \rangle$ for the measurement model. A single measurement sequence $\langle ABC \rangle$ is split into eight terms: With a prefactor $p_A p_B p_C$ one has the value, which is obtained, if all measurements are projective; with a prefactor $p_A p_B (1 - p_C)$ one has the value, where A and B are projective, and C is a random assignment, etc. It follows that the total mean value $\langle \chi_{\mathrm{PM}} \rangle$ is an affine function in the probability p_A (if all other parameters are fixed) and also in all other probabilities p_X for the other measurements. Consequently, the maximum of $\langle \chi_{\mathrm{PM}} \rangle$ is attained either at $p_A = 1$ or 0, and similarly for all the measurements. Therefore, for maximizing $\langle \chi_{\mathrm{PM}} \rangle$ it suffices to consider the finite set of cases where, for each observable, either always possibility (a) or always possibility (b) is taken. We can formulate the following:

Proposition 12. Consider noisy measurements as described above. Then, the bound from Observation 4

$$\langle \chi_{\mathrm{PM}} \rangle \leqslant 3\sqrt{3} \qquad (A21)$$

holds.

Proof of Proposition 12. As discussed above, we only have to discuss a finite number of cases. Let us consider a single term $\langle ABC \rangle$. If C is a random assignment, then $\langle ABC \rangle = 0$, independently how A and B are realized. It follows that if C, β, or a are random assignments, then $\langle \chi_{\mathrm{PM}} \rangle \leqslant 4$.

On the other hand, if A is a random assignment, then $\langle ABC \rangle = 0$ as well: (i) If B and C are projective, then the measurement of B and C results in the state-independent mean value $\langle BC \rangle$ [see Eq. (8) in the main text]. This value is independent of the state $\varrho^\pm$ remaining after the assignment of A, hence, $\langle ABC \rangle = \langle AB \rangle - \langle AB \rangle = 0$. (ii) If B is a random assignment, one can also directly calculate that $\langle ABC \rangle = 0$ and the case that (iii) C is a random assignment has been discussed already. Consequently, if A, b, or γ are random assignments, then $\langle \chi_{\mathrm{PM}} \rangle \leqslant 4$.

It remains to discuss the case that B, c, or α are random assignments while all other measurements are projective. First, one can directly calculate that if A, C are projective, and B is a random assignment, then

$$\langle ABC \rangle = \mathrm{tr}(\varrho A)\mathrm{tr}(CX), \qquad (A22)$$

with $X = (\varrho^+ - \varrho^-)/2$. If X is expressed in terms of Pauli matrices, then the length of its Bloch vector does not exceed one since the Bloch vectors of $\varrho^\pm$ are subnormalized.

The estimate of $\langle \chi_{\mathrm{PM}} \rangle$ can now proceed as in the proof of Observation 4, and one arrives at the situation of Lemma 7 in Appendix A 3, where now the vectors are subnormalized, and not necessarily normalized. But, still the bound from Lemma 7 is valid: If the smallest vector in χ_6 has a length ω, one can directly see that $\chi_6 \geqslant \omega[-N \cos(\pi/N)] - (1 - \omega)4$. This proves Proposition 12. ∎

b. More general POVMs

Now, we consider a general dichotomic positive-operator-valued measure (POVM) on a qubit system. This is characterized by two effects E^+ and E^-, where $E^+ + E^- = \mathbb{1}$ and the probabilities of the measurement results are $p^+ = \mathrm{tr}(\varrho E^+)$ and $p^- = \mathrm{tr}(\varrho E^-)$.

These effects have to commute and one can write $E^+ = \alpha|0\rangle\langle 0| + \beta|1\rangle\langle 1|$ and $E^- = \gamma|0\rangle\langle 0| + \delta|1\rangle\langle 1|$ in an appropriate basis. We can assume that $\alpha \geqslant \beta$ and consequently $\delta \geqslant \gamma$. Furthermore, it is no restriction to choose $\beta \leqslant \gamma$. Then,

the effects can be written as $E^+ = \beta\mathbb{1} + (\alpha - \beta)|0\rangle\langle0|$ and $E^- = \beta\mathbb{1} + (\gamma - \beta)\mathbb{1} + (\alpha - \beta)|1\rangle\langle1|$. This means that one can interpret the probabilities of the POVM as coming from the following procedure: With a probability of 2β one assigns a random outcome, with a probability of $\gamma - \beta$ one assigns the fixed value -1, and with a probability of $(\alpha - \beta)$ one performs the projective measurement.

This motivates the following measurement model: Instead of performing the projective measurement A, one of three possible actions is taken:

(i) with a probability p_1^A the projective measurement is performed or

(ii) with a probability p_2^A a fixed outcome ± 1 is assigned independently of the initial state; after this announcement, the state is left in the corresponding eigenstate of A or

(iii) with a probability p_3^A a completely random outcome ± 1 is assigned independently of the initial state.

As above, in case (iii), the physical system is left in one of two possible states ϱ^+ or ϱ^-, but we will not make any assumptions on $\varrho^\pm$. For this measurement model, we have the following:

Proposition 13. In the noise model described above, the PM operator is bounded by

$$\langle\chi_{\mathrm{PM}}\rangle \leqslant 1 + \sqrt{9 + 6\sqrt{3}} \approx 5.404. \quad (A23)$$

Proof of Proposition 13. As in the proof of Proposition 12, we only have to consider a finite set of cases. Let us first discuss the situation, where for each measurement only the possibilities (i) and (ii) are taken.

First, we have to derive some formulas for sequential measurements. The reason is that, if the option (ii) is chosen, then the original formula for sequential measurements, Eq. (15) in the main text, is not appropriate anymore and different formulas have to be used.

In the following, we write $A = (\pm)_A$ if A is a fixed assignment as described in possibility (ii) above. If not explicitly stated otherwise, the observables are measured as projective measurements. Then, one can directly calculate that

$$\langle ABC\rangle = (\pm)_A\langle BC\rangle \quad \text{if} \quad A = (\pm)_A, \quad (A24a)$$

$$\langle ABC\rangle = \mathrm{tr}(\varrho A)\langle BC\rangle \quad \text{if} \quad B = (\pm)_B, \quad (A24b)$$

$$\langle ABC\rangle = (\pm)_C\langle AB\rangle \quad \text{if} \quad C = (\pm)_C. \quad (A24c)$$

Note that in Eq. (A24b) there is no deviation from the usual formula (15) in the main text. Furthermore, we have

$$\langle ABC\rangle = (\pm)_A(\pm)_B\mathrm{tr}(C|B^\pm\rangle\langle B^\pm|) = (\pm)_A\langle BC\rangle$$
$$\text{if} \quad A = (\pm)_A \quad \text{and} \quad B = (\pm)_B, \quad (A25a)$$

$$\langle ABC\rangle = (\pm)_A(\pm)_C\mathrm{tr}(B|A^\pm\rangle\langle A^\pm|) = (\pm)_C\langle AB\rangle$$
$$\text{if} \quad A = (\pm)_A \quad \text{and} \quad C = (\pm)_C, \quad (A25b)$$

$$\langle ABC\rangle = (\pm)_B(\pm)_C\mathrm{tr}(\varrho A)$$
$$\text{if} \quad B = (\pm)_B \quad \text{and} \quad C = (\pm)_C. \quad (A25c)$$

In Eqs. (A25a) and (A25b), $|B^\pm\rangle$ and $|A^\pm\rangle$ denote the eigenstates of B and A, which are left after the fixed assignment.

Equipped with these rules, we can discuss the different cases. First, from Eqs. (A24a), (A24b), and (A25a) it follows

that the proof of Observation 4 does not change, if fixed assignments are made only on the observables which are measured at first or second position of a sequence (i.e., the observables $A, b, \gamma, B, c,$ and α).

However, the structure of the inequality changes if one of the last measurements is a fixed assignment. To give an example, consider the case that the measurement β is a fixed assignment [case (ii) above], while all other measurements are projective [case (i) above]. Using Eq. (A24c) we have to estimate

$$\mathcal{X} = \langle A\rangle\langle BC\rangle + \langle A\rangle\langle\alpha a\rangle + \langle b\rangle\langle ca\rangle$$
$$+ \langle bB\rangle(\pm)_\beta + \langle\gamma\alpha\rangle(\pm)_\beta - \langle\gamma\rangle\langle cC\rangle. \quad (A26)$$

On can directly see that it suffices to estimate

$$\mathcal{X}' = \langle B|C\rangle + \langle\alpha|a\rangle + \langle\varrho|b\rangle\langle c|a\rangle$$
$$+ \langle b|B\rangle + \langle\gamma|\alpha\rangle - \langle\varrho|\gamma\rangle\langle c|C\rangle, \quad (A27)$$

where all expressions should be understood as scalar products of the corresponding Bloch vectors. Then, a direct optimization over the three-dimensional Bloch vectors proves that here

$$\mathcal{X}' \leqslant 1 + \sqrt{9 + 6\sqrt{3}} \approx 5.404 \quad (A28)$$

holds. In general, the observables $\beta, C,$ or a are the possible third measurements in a sequence. One can directly check that, if one or several of them are fixed assignments, then an expression analog to Eq. (A26) arises and the bound of Eq. (A28) holds. Finally, if some of the $\beta, C,$ or a are fixed assignments and, in addition, some of the $A, b, \gamma, B, c,$ and α are fixed assignments, then the comparison between Eq. (A24c) and Eqs. (A25b) and (A25c) shows that no novel types of expressions occur.

It remains to discuss the case where not only the possibilities (i) and (ii) occur, but for one or more measurements also a random assignment [possibility (iii)] is realized. As in the proof of Proposition 12, one finds that only the cases where the second measurements ($B, c,$ and α) are random are interesting. In addition to Eq. (A22), one finds that $\langle ABC\rangle = (\pm_A)\mathrm{tr}(CX)$ if B is random and A is a fixed assignment, and $\langle ABC\rangle = 0$ if B is random and C is a fixed assignment. This shows that no new expressions occur, and proves the claim. ∎

Finally, we would like to add two remarks. First, it should be stressed that the presented noise model still makes assumptions about the measurement, especially about the post-measurement state. Therefore, it is not the most general measurement, and we do not claim that the resulting dimension witnesses are device independent.

Second, we would like to emphasize that the chosen order of the measurements in the definition in Eq. (9) in the main text is important for the proof of the bounds for noisy measurements: For other orders, it is not clear whether the dimension witnesses are robust against imperfections. In fact, for some choices one finds that the resulting inequalities are *not* robust against imperfections: Consider, for instance, a measurement order, where one observable (say, γ for definiteness) is the second observable in one context and the third observable in the other context. Furthermore, assume that γ is an assignment [case (iii) above], while all other measurements are projective. Then, we have to use Eq. (A24b) for the first context of γ, and Eq. (A24c)

for the second context. In Eq. (A24b) there is no difference to the usual formula, especially the formula does not depend on the value assigned to γ. Equation (A24c), however, depends on this value. This means that, for one term in the PM inequality, the sign can be changed arbitrarily and so $\langle \chi_{\text{PM}} \rangle = 6$ can be reached.

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).

[2] O. Gühne and G. Tóth, Phys. Rep. **474**, 1 (2009).

[3] G. Molina-Terriza, J. P. Torres, and L. Torner, Nat. Phys. **3**, 305 (2007).

[4] A. C. Dada, J. Leach, G. S. Buller, M. J. Padgett, and E. Andersson, Nat. Phys. **7**, 677 (2011).

[5] P. Neumann, N. Mizuochi, F. Rempp, P. Hemmer, H. Watanabe, S. Yamasaki, V. Jacques, T. Gaebel, F. Jelezko, and J. Wrachtrup, Science **320**, 1326 (2008).

[6] N. Brunner, S. Pironio, A. Acín, N. Gisin, A. A. Méthot, and V. Scarani, Phys. Rev. Lett. **100**, 210503 (2008).

[7] T. Vértesi and K. F. Pál, Phys. Rev. A **79**, 042106 (2009).

[8] S. Wehner, M. Christandl, and A. C. Doherty, Phys. Rev. A **78**, 062112 (2008).

[9] R. Gallego, N. Brunner, C. Hadley, and A. Acín, Phys. Rev. Lett. **105**, 230501 (2010).

[10] N. Brunner, M. Navascués, and T. Vértesi, Phys. Rev. Lett. **110**, 150501 (2013).

[11] M. Hendrych, R. Gallego, M. Mičuda, N. Brunner, A. Acín, and J. P. Torres, Nat. Phys. **8**, 588 (2012).

[12] J. Ahrens, P. Badziąg, A. Cabello, and M. Bourennane, Nat. Phys. **8**, 592 (2012).

[13] M. M. Wolf and D. Pérez-García, Phys. Rev. Lett. **102**, 190504 (2009).

[14] E. P. Specker, Dialectica **14**, 239 (1960).

[15] A. M. Gleason, J. Math. Mech. **6**, 885 (1957).

[16] J. S. Bell, Rev. Mod. Phys. **38**, 447 (1966).

[17] S. Kochen and E. P. Specker, J. Math. Mech. **17**, 59 (1967).

[18] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman, Phys. Rep. **506**, 1 (2011).

[19] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Phys. Rev. Lett. **101**, 020403 (2008).

[20] A. Cabello, Phys. Rev. Lett. **101**, 210401 (2008).

[21] A. Peres, Phys. Lett. A **151**, 107 (1990).

[22] N. D. Mermin, Phys. Rev. Lett. **65**, 3373 (1990).

[23] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1995).

[24] M. Araújo, M. T. Quintino, C. Budroni, M. T. Cunha, and A. Cabello, Phys. Rev. A **88**, 022118 (2013).

[25] R. Łapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger, Nature (London) **474**, 490 (2011).

[26] J. Ahrens, E. Amselem, A. Cabello, and M. Bourennane, Sci. Rep. **3**, 2170 (2013).

[27] O. Gühne, M. Kleinmann, A. Cabello, J.-Å. Larsson, G. Kirchmair, F. Zähringer, R. Gerritsma, and C. F. Roos, Phys. Rev. A **81**, 022121 (2010).

[28] T. Fritz, New J. Phys. **12**, 083055 (2010).

[29] C. Budroni, T. Moroder, M. Kleinmann, and O. Gühne, Phys. Rev. Lett. **111**, 020403 (2013).

[30] G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, and C. F. Roos, Nature (London) **460**, 494 (2009).

[31] E. Amselem, M. Rådmark, M. Bourennane, and A. Cabello, Phys. Rev. Lett. **103**, 160405 (2009).

[32] O. Moussa, C. A. Ryan, D. G. Cory, and R. Laflamme, Phys. Rev. Lett. **104**, 160501 (2010).

[33] A. Cabello, Phys. Rev. A **82**, 032110 (2010).

[34] S. Yu and C. H. Oh, Phys. Rev. Lett. **108**, 030402 (2012).

[35] M. Kleinmann, C. Budroni, J.-Å. Larsson, O. Gühne, and A. Cabello, Phys. Rev. Lett. **109**, 250402 (2012).

[36] M. Barbieri, Phys. Rev. A **80**, 034102 (2009).