

## Testing locality and noncontextuality with the lowest moments

Adam Bednorz,<sup>1,\*</sup> Witold Bednorz,<sup>2,†</sup> and Wolfgang Belzig<sup>3</sup><sup>1</sup>*Faculty of Physics, University of Warsaw, Hoża 69, PL-00681 Warsaw, Poland*<sup>2</sup>*Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland*<sup>3</sup>*Fachbereich Physik, Universität Konstanz, D-78457 Konstanz, Germany*

(Received 20 August 2013; revised manuscript received 27 January 2014; published 27 February 2014)

The quest for fundamental tests of quantum mechanics is an ongoing effort. We here address the question of what are the lowest possible moments needed to prove quantum nonlocality and noncontextuality without any further assumptions—in particular, without the often assumed dichotomy. We first show that second-order correlations can always be explained by a classical noncontextual local-hidden-variable theory. Similar third-order correlations also cannot violate classical inequalities in general, except for a special state-dependent noncontextuality. However, we show that fourth-order correlations can violate locality and state-independent noncontextuality. Finally we obtain a fourth-order continuous-variable Bell inequality for position and momentum, which can be violated and might be useful in Bell tests, closing all loopholes simultaneously.

DOI: [10.1103/PhysRevA.89.022125](https://doi.org/10.1103/PhysRevA.89.022125)

PACS number(s): 03.65.Ud, 03.67.Mn

### I. INTRODUCTION

Certain quantum correlations cannot be reproduced by any classical local-hidden-variable (LHV) theory, as they violate, e.g., the Bell inequalities for correlations of results of measurements by separate observers [1]. The Bell test must be performed under the following conditions: (i) dichotomy of the measurement outcomes or at least some restricted set of outcomes in some generalizations [2], (ii) freedom of choice of the measured observables [3], and (iii) a shorter time of choice and measurement of the observable than the communication time between the observers. Despite considerable experimental effort [4], the violation has not yet been confirmed conclusively, due to several loopholes [5]. The loopholes reflect the fact that the experiments have not fully satisfied all the conditions (i)–(iii) simultaneously. In fact, the Bell test is stronger than the entanglement criterion, viz., the nonseparability of states [6]. The latter already assumes a quantum mechanical framework (e.g., an appropriate Hilbert space), while the former is formulated classically. The loophole-free violation of a Bell inequality—not just the existence of entanglement—is also necessary to prove the absolute security of quantum cryptography [7].

Nonclassical behavior of quantum correlations can appear also as a violation of noncontextuality. Noncontextuality means that the outcomes of experiments do not depend on the detectors' settings so that there is a common underlying probability for the results of all possible settings while the accessible correlations correspond to commuting sets of observables. The Kochen-Specker theorem ingeniously shows that noncontextuality contradicts quantum mechanics [8]. Noncontextuality is testable in realistic setups [9]. In contrast to noncontextuality, Bell-type tests of nonlocality without further assumptions must exclude also *contextual* LHV models, as correlations of outcomes for different settings are not simultaneously experimentally accessible for a single observer, even if they accidentally commute. Moreover, non-

contextuality may be violated for an arbitrary localized state (state-independent noncontextuality [10]) while Bell-type tests make sense only for nonlocally entangled states. If a Bell-type inequality is violated then state-dependent noncontextuality is violated, too, but not vice versa.

As the Bell and noncontextual inequalities are often restricted to dichotomic outcomes, e.g.,  $A = \pm 1$ , generalizations have been investigated, including the many-outcome case [2]. Recently, Cavalcanti, Foster, Reid, and Drummond, (CFRD) [11] proposed a way to relax the constraint of dichotomy, allowing any unconstrained real value. CFRD constructed a particularly simple class of inequalities holding classically, while seemingly vulnerable by quantum mechanics. The inequalities involve  $n$ th moments  $\langle A^{n-l-m} B^l C^m \rangle$  of observables  $A$ ,  $B$ , and  $C$ , and nonnegative integers  $l$ ,  $m$ , and  $n - l - m$ , where in general the higher  $n$  is, the greater the chances to violate the corresponding CFRD inequality. On a practical level, measuring higher moments or binning is not a problem if the statistics consists of isolated peaks. However, in many experiments, especially in condensed matter [12], the interesting information is masked by large classical noise. This noise then dominates the signal and makes the binning unable to retrieve the underlying quantum statistics, which is accessible only by measuring moments and subsequent deconvolution.

In this paper we ask which are the lowest possible moments to show nonclassicality and systematically investigate whether second-, third-, or fourth-order correlations are sufficient to exclude LHV theories. We first show that second-order inequalities cannot be violated at all because of the so-called weak positivity [13]—a simple classical construction of a probability reproducing all second-order correlations. Note that the standard Bell inequalities [1] require *experimental* verification of the dichotomy  $A^2 = 1$ , which means, e.g., showing that  $\langle (A^2 - 1)^2 \rangle = 0$  by measuring the corresponding fourth-order correlator or applying binning (in which case the correlator is obviously zero). Hence, operationally a standard Bell test is of at least fourth order—not second, as it may appear from the Bell inequalities [1] alone. We emphasize that binning is useless, if the signal is masked by classical noise. The proposed Bell-type tests in condensed matter based on second-order correlations [14–16] require an additional assumption

\*adam.bednorz@fuw.edu.pl

†wbednorz@mimuw.edu.pl

TABLE I. Summary of the feasibility of moment-based tests of LHV theories depending on the conditions: (a) contextuality or noncontextuality and (b) a special or arbitrary input state. The entries answer the questions: Are correlations with moments up to the given order not explicable by a joint positive probability?

	LHV excluded?		
Noncontextuality	Yes	Yes	No
State independent	No	Yes	No
Maximal moments	LHV excluded?		
Second	No	No	No
Third	Yes	No	No
Fourth	Yes	Yes	Yes

of a dichotomous interpretation of the measurement results, which is in general experimentally unverified and does not allow entanglement to be identified unambiguously. Next we will show that Bell-type tests for third moments with standard projective measurements are not possible. Nevertheless, third moments can violate noncontextuality but only for a positive semidefinite correlation matrix and special states. Our main result is to show that generally fourth-order correlators are sufficient to violate state-independent noncontextuality and a Bell-type inequality which can be violated by correlation of position and momentum in a special entangled state. State-independent noncontextuality can be violated by a fourth-moment generalization of the Mermin-Peres square [17]. Our results for the gradual possibilities of excluding LHV models under different conditions are summarized in Table I.

Comparing to previous research, note that the CFRD inequalities are the only known Bell-type inequalities scalable with  $A \rightarrow \lambda A$ ,  $B \rightarrow \mu B$ , and so on for more observers. Unfortunately, the original example for a violation involved 20th-order correlators and ten observers [11], but was later reduced to sixth order and three observers [18,19] for Greenberger-Horne-Zeilinger states [20]. On the other hand, the CFRD inequality with fourth moments cannot be violated at all, which has been shown for spins [21], quadratures [22], generalized to eight settings and proved for separable states [23], and finally proved for all states [19] (we show an alternative proof in Appendix E).

The paper is organized as follows. We start with a general description of tests of contextuality and locality. Then we show that second moments are insufficient to violate locality and noncontextuality. Next, we show that third moments are enough only to show state-dependent contextuality. In the last part we discuss fourth moments, which allow violation of state-independent noncontextuality and locality. The violation of locality is possible with moments of positions and momenta (quadratures).

## II. TEST OF LOCAL-HIDDEN-VARIABLE MODELS

Let us adopt the Bell framework, depicted in Fig. 1. Suppose Alice, Bob, Charlie, etc. are separate observers that can perform measurements on a possibly entangled state, which is described by an initial density matrix  $\hat{\rho}$ . Every observer  $X = A, B, C, \dots$  is free to prepare one of several settings of its own detector ( $\alpha = 1, 2, \dots$ ). For each setting, one can measure multiple real-valued observables (numbered  $i = 1, 2, 3, \dots$ )

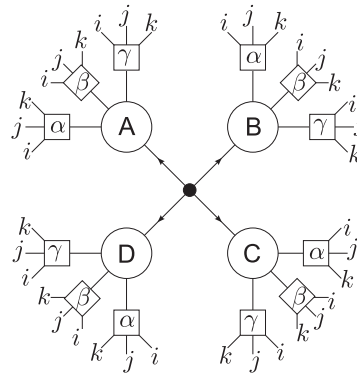


FIG. 1. The general test of local realism. Here we have four observers Alice, Bob, Charlie, and David. Everybody is free to choose between three different settings  $\alpha$ ,  $\beta$ , and  $\gamma$  and finally they can measure three real, continuous outcomes, e.g.,  $A_{\alpha i}$ . The picture can be generalized to arbitrary numbers of observers, settings, and outcomes.

so that the measurement of  $\hat{X}_{\alpha i}$  gives a real number  $X_{\alpha i}$ . The projection postulate gives the quantum prediction for correlations,  $\langle O_1 \dots O_n \rangle = \text{Tr} \hat{\rho} \hat{O}_1 \dots \hat{O}_n$  for commuting observables  $\hat{O}_k$ . The observables measured by different observers and by one observer  $\hat{X}_{\alpha i}$  for a given setting have to commute, viz.,  $[\hat{X}_{\alpha i}, \hat{Y}_{\beta j}] = [\hat{X}_{\alpha i}, \hat{X}_{\alpha j}] = 0$ . The observables for one observer but different settings,  $\hat{X}_{\alpha i}$  and  $\hat{X}_{\beta j}$  for  $\alpha \neq \beta$ , may be noncommuting but may also accidentally commute or even be equal. A LHV model assumes the existence of a joint positive-definite probability distribution of all possible outcomes  $\rho(\{X_{\alpha i}\})$  that reproduces quantum correlations for a given setting. If the accidental equality between observables for different settings,  $\hat{X}_{\alpha i} = \hat{X}_{\beta j}$ , imposes the constraint  $X_{\alpha i} \equiv X_{\beta j}$  in  $\rho$ , the LHV model is called *noncontextual*. A single observer suffices to test such LHVs as noncontextuality is anyway an experimentally unverifiable assumption—the observer cannot measure simultaneously at two different settings. In contrast to noncontextuality, the *locality* test must allow *contextuality*: that even if  $\hat{X}_{\alpha i} = \hat{X}_{\beta j}$  ( $\alpha \neq \beta$ ) then  $X_{\alpha i} \neq X_{\beta j}$  is still possible. The choices of the settings and measurements are required to be fast enough to prevent any communication between observers. Then  $\rho$  cannot be altered by the choice of the observable. Noncontextual and local LHVs can be ruled out by tests with discrete outcomes [1,8]. In moment-based tests only a finite number of cross correlations are compared with LHVs. Our aim is to find the lowest moments showing nonclassical behavior of quantum correlations.

## III. WEAK POSITIVITY

For a moment all observables, commuting or not, will be denoted by  $\hat{X}_i$ . Let us recall the simple proof that first- and second-order correlation functions can always be reproduced classically [13]. To see this, consider a real symmetric correlation matrix

$$C_{ij} = \langle X_i X_j \rangle = \text{Tr} \hat{\rho} \{\hat{X}_i, \hat{X}_j\} / 2 \quad (1)$$

with  $\{\hat{X}, \hat{Y}\} = \hat{X}\hat{Y} + \hat{Y}\hat{X}$  for arbitrary observables  $\hat{X}_i$  and density matrix  $\hat{\rho}$ . Such a relation is consistent with simultaneously measurable correlations. More generally, it holds even in the noncontextual case, when observables from different settings

commute. Only these elements of the matrix  $\mathcal{C}$  are measurable; for the rest (1) is only a definition. Our construction includes all possible first-order averages  $\langle X_i \rangle$  by setting one observable to the identity or subtracting averages ( $X_i \rightarrow X_i - \langle X_i \rangle$ ). Since  $\text{Tr} \hat{\rho} \hat{W}^2 \geq 0$  for  $\hat{W} = \sum_i \lambda_i \hat{X}_i$  with arbitrary real  $\lambda_i$ , we find that the correlation matrix  $\mathcal{C}$  is positive definite. Therefore every correlation can be simulated by a classical Gaussian distribution  $\varrho \propto \exp(-\sum_{ij} \mathcal{C}^{-1}_{ij} X_i X_j / 2)$ , with  $\mathcal{C}^{-1}$  being the matrix inverse of  $\mathcal{C}$ . This is a LHV model reproducing all measurable correlations. We recall that we do not assume dichotomy  $X = \pm 1$ , which is equivalent to  $\langle (X^2 - 1)^2 \rangle = 0$  and requires  $\langle X^4 \rangle$ . For simplicity, from now on we shall fix  $\langle X_i \rangle = 0$ , redefining all quantities  $X_i \rightarrow X_i - \langle X_i \rangle$ .

It is interesting to note that Tsirelson's bound [24] can be seen as a consequence of weak positivity. Taking observables  $A_1, A_2, B_1$ , and  $B_2$ , we have

$$\langle (\sqrt{2}A_1 - B_1 - B_2)^2 \rangle + \langle (\sqrt{2}A_2 - B_1 + B_2)^2 \rangle \geq 0 \quad (2)$$

for the Gaussian distribution with the correlation matrix (1). It is equivalent to

$$\begin{aligned} & \langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \\ & \leq (\langle A_1^2 \rangle + \langle A_2^2 \rangle + \langle B_1^2 \rangle + \langle B_2^2 \rangle) / \sqrt{2}. \end{aligned} \quad (3)$$

For  $A, B = \pm 1$ , the right hand side gives Tsirelson's bound  $2\sqrt{2}$  which is at the same time the maximal quantum value of the left-hand side. On the other hand, the upper classical bound in this case is 2 [1], but it requires assuming dichotomy or equivalently knowledge of higher moments.

#### IV. THIRD MOMENTS

Having learned that second moments do not show nonclassicality at all, we turn to third moments. If the matrix  $\mathcal{C}$  is strictly positive definite, all third-order correlations can be explained by a positive probability as well (the proof is in Appendix A). The problematic case is a semipositive-definite  $\mathcal{C}$ , with at least one 0 eigenvalue. One cannot violate noncontextuality with an arbitrary state and third-order correlations. To see this, let us take the completely random state  $\hat{\rho} \propto \hat{1}$  and suppose that the correlation matrix (1) has a zero eigenvalue for  $\hat{W} = \sum_k \lambda_k \hat{X}_k$ . Then  $\langle W^2 \rangle = 0$  and  $\text{Tr} \hat{W}^2 = 0$ , which gives  $\hat{W} = 0$ . We can simply eliminate one of the observables by the substitution  $\hat{X}_m = -\sum_{k \neq m} \lambda_k \hat{X}_k / \lambda_m$  using the symmetrized order of the operators when noncommuting products appear. Now the remaining correlation matrix  $C_{ij}$  with  $i, j \neq m$  is positive definite and the proof in Appendix A holds. If the correlation matrix has more zero eigenvalues, we repeat the reasoning, until only nonzero eigenvalues remain. Furthermore, third-order correlations alone cannot show noncontextuality in a state-dependent way for up to four observables, nor in any two-dimensional Hilbert space, nor can they violate local realism (the proofs are in Appendixes B and C). There exists, however, an example of violation of state-dependent noncontextuality with five observables in three-dimensional space (Appendix D).

Instead, here we show a simple example violating state-dependent noncontextuality, based on the Greenberger-Horne-Zeilinger (GHZ) idea [20]. We consider a three-qubit Hilbert space with the eight basis states denoted  $|\epsilon_1 \epsilon_2 \epsilon_3\rangle$  with

$\epsilon_\alpha = \pm$ . We have three sets of Pauli matrices  $\hat{\sigma}_j^{(\alpha)}$ , with  $\hat{\sigma}_1 = |- \rangle \langle +| + | + \rangle \langle -|$  and  $\hat{\sigma}_2 = i|- \rangle \langle +| - i| + \rangle \langle -|$ , acting only in the respective Hilbert space of qubit  $\alpha$ . Now let us take the six observables  $\hat{A}_\alpha = \hat{\sigma}_1^{(\alpha)}$ ,  $\hat{B}_\alpha = \hat{C} \hat{\sigma}_2^{(\alpha)}$  for  $\alpha = 1, 2, 3$  and  $\hat{C} = \hat{\sigma}_2^{(1)} \hat{\sigma}_2^{(2)} \hat{\sigma}_2^{(3)}$ . All  $\hat{A}$ 's commute with each other; similarly all  $\hat{B}$ 's commute, and  $\hat{A}_\alpha$  commutes with  $\hat{B}_\alpha$ . We take  $\hat{\rho} = |\text{GHZ}\rangle \langle \text{GHZ}|$  for the GHZ state

$$\sqrt{2}|\text{GHZ}\rangle = |+++ \rangle + |-- \rangle. \quad (4)$$

Assuming noncontextuality, we have

$$\langle (A_\alpha + B_\alpha)^2 \rangle = \text{Tr} \hat{\rho} (\hat{A}_\alpha + \hat{B}_\alpha)^2 = 0, \quad (5)$$

which implies  $A_\alpha = -B_\alpha$ , so classically  $\langle A_1 A_2 A_3 \rangle = -\langle B_1 B_2 B_3 \rangle$ . However,

$$\begin{aligned} \langle A_1 A_2 A_3 \rangle &= \text{Tr} \hat{\rho} \hat{A}_1 \hat{A}_2 \hat{A}_3 = 1, \\ \langle B_1 B_2 B_3 \rangle &= \text{Tr} \hat{\rho} \hat{B}_1 \hat{B}_2 \hat{B}_3 = 1, \end{aligned} \quad (6)$$

in contradiction with the earlier statement and excluding noncontextual LHVs. Hence, we have seen that the third-order correlations may violate noncontextuality for specific states. It should not be surprising that the test is based on violating an equality, instead of an inequality, because third moments can have arbitrary signs.

#### V. FOURTH-ORDER CORRELATIONS: NONCONTEXTUALITY

To find a test of noncontextuality we now consider fourth moments. Mermin and Peres [17] have shown a beautiful example of state-independent violation of noncontextuality using observables on the tensor product of two two-dimensional Hilbert spaces  $\mathcal{H}_A \otimes \mathcal{H}_B$  arranged in a square

$\hat{M}_{ij}$	$j = 1$	$j = 2$	$j = 3$
$i = 1$	$\hat{\sigma}_1^A$	$\hat{\sigma}_1^A \hat{\sigma}_1^B$	$\hat{\sigma}_1^B$
$i = 2$	$-\hat{\sigma}_1^A \hat{\sigma}_3^B$	$\hat{\sigma}_2^A \hat{\sigma}_2^B$	$-\hat{\sigma}_3^A \hat{\sigma}_1^B$
$i = 3$	$\hat{\sigma}_3^B$	$\hat{\sigma}_3^A \hat{\sigma}_3^B$	$\hat{\sigma}_3^A$

(7)

where the Pauli observables  $\hat{\sigma}_i$  are in each Hilbert space ( $\{\hat{\sigma}_i, \hat{\sigma}_j\} = 2\delta_{ij} \hat{1}$ ). Observables in each row and each column commute. We denote products in each column  $\hat{C}_i = \hat{M}_{1i} \hat{M}_{2i} \hat{M}_{3i}$  and row  $\hat{R}_i = \hat{M}_{i1} \hat{M}_{i2} \hat{M}_{i3}$ . We get  $\hat{C}_i = -\hat{1}$  and  $\hat{R}_i = \hat{1}$ . If  $\hat{M}_{ij}$  are replaced by classical variables  $M_{ij}$  then  $C_1 C_2 C_3 = R_1 R_2 R_3$ , in contradiction with the quantum result.

Now we assume that the  $M$  are not spin 1/2, but arbitrary operators, which can be grouped into a Mermin-Peres square fulfilling the corresponding commutation relations  $[\hat{M}_{ij}, \hat{M}_{ik}] = [\hat{M}_{ij}, \hat{M}_{kj}] = 0$  (operators in the same column or row commute). We will show that in this example the dichotomy test can be avoided by fourth-order correlations, without other assumptions on the values  $M_{ij}$ . To see this, note that  $S \equiv \sum_i (C_i - R_i) = \det N$  where  $N_{ij} = M_{i+j, i-j}$  (counting modulo 3). Now, we note that  $(\det N)^2 = \det(N^T N)$  and the eigenvalues  $\lambda_i$  of  $N^T N$  are real and positive. Using the Cauchy inequality we find that  $\det(N^T N) = \lambda_1 \lambda_2 \lambda_3 \leq$

$(\lambda_1 + \lambda_2 + \lambda_3)^3/27 = (\text{Tr} N^T N)^3/27$ . We get then

$$3\sqrt{3}|S| \leq \left( \sum_{ij} M_{ij}^2 \right)^{3/2} \leq 3 \sum_{ij} |M_{ij}|^3, \quad (8)$$

where we used the Hölder inequality in the last step. Now, we take the average of the above equation, use  $|\langle S \rangle| \leq \langle |S| \rangle$ , and apply the Cauchy-Bunyakovsky-Schwarz inequality  $\langle |xy| \rangle \leq (\langle x^2 \rangle \langle y^2 \rangle)^{1/2}$  to  $x = M_{ij}$  and  $y = M_{ij}^2$ . We obtain finally an inequality obeyed by all noncontextual theories,

$$|\langle S \rangle| \leq \sum_{ij} [(M_{ij}^2 \langle M_{ij}^4 \rangle) / 3]^{1/2}. \quad (9)$$

The inequality involves maximally fourth-order correlations and every correlation is measurable (corresponds to commuting observables). One can check that if  $M_{ij}$  correspond to (7) then the left-hand side of (9) is 6 while the right-hand side of (9) is  $3\sqrt{3}$ , giving a contradiction. Hence, a violation of (9) is possible, but it remains to be shown that systems with naturally continuous variables are contextual by violating Eq. (9) or other fourth-moment-based inequalities.

## VI. FOURTH-ORDER CORRELATIONS: NONLOCALITY

A simple fourth-moment-based inequality testing local realism has been considered by CFRD [11]:

$$\begin{aligned} & \langle A_1 B_1 - A_2 B_2 \rangle^2 + \langle A_1 B_2 + A_2 B_1 \rangle^2 \\ & \leq \langle (A_1^2 + A_2^2)(B_1^2 + B_2^2) \rangle. \end{aligned} \quad (10)$$

Note that all averages involve only simultaneously measurable quantities. This constitutes an inequality, which holds classically, involves only fourth-order averages, and is scalable with respect to  $A$  and  $B$ . Unfortunately, (10) and its generalizations [23] are not violated at all in quantum mechanics as shown in [19]. We present an alternative proof in Appendix E.

Unfortunately a violable two-party fourth-order inequality is much more complicated [13]. A different, but quadripartite, inequality can be obtained by a slight modification of CFRD inequalities [11]. It reads

$$|\langle ABCD \rangle|^2 \leq \langle |AB|^2 \rangle \langle |CD|^2 \rangle \quad (11)$$

where  $A = A_1 + iA_2$ , etc., so that both sides, when expanded, contain only simultaneously measurable correlations (because  $|\langle ABCD \rangle|^2 = \langle \text{Re} ABCD \rangle^2 + \langle \text{Im} ABCD \rangle^2$  is free from products  $\langle A_1 A_2 \dots \rangle$  and  $|A|^2 = A_1^2 + A_2^2$  on the right-hand side) It follows from the generalized triangle inequality  $|\langle Z \rangle| \leq \langle |Z| \rangle$  for  $Z = ABCD$  and the Cauchy-Bunyakovsky-Schwarz inequality  $\langle XY \rangle^2 \leq \langle X^2 \rangle \langle Y^2 \rangle$  for  $X = |AB|$  and  $Y = |CD|$ . See more details in Appendix F.

Interestingly, the inequality (11) can be violated by correlations of positions and momenta. Let us take standard harmonic oscillator operators  $\sqrt{2}\hat{A} = \hat{X}_A + i\hat{P}_A$  with  $[\hat{X}_A, \hat{P}_A] = i$  ( $\hbar = 1$ ) so  $A_1 \rightarrow \hat{X}_A/\sqrt{2}$ ,  $A_2 \rightarrow \hat{P}_A/\sqrt{2}$ , and  $[\hat{A}, \hat{A}^\dagger] = 1$  and analogously for  $B, C$ , and  $D$ . In the Fock basis  $\hat{A}|n\rangle_A = \sqrt{n}|n-1\rangle_A$ , etc. Now take a specific entangled state in the product space of  $A, B, C$ , and  $D$ ,  $|\psi\rangle = \sum_{n \geq 0} z_n |nnnn\rangle$  with real  $z_n$  (for simplicity) and check if (11) holds also quantum mechanically. We find that  $\langle \psi | \hat{A} \hat{B} \hat{C} \hat{D} | \psi \rangle = \sum_n n^2 z_n z_{n-1}$  while  $\langle \psi | (\hat{A}^\dagger \hat{A} + \hat{A} \hat{A}^\dagger)(\hat{B}^\dagger \hat{B} + \hat{B} \hat{B}^\dagger) | \psi \rangle =$

$\sum_n z_n^2 (2n+1)^2$ , and similarly for  $C$  and  $D$ . Due to symmetry between the oscillators, the inequality (11) is equivalent to  $\langle ABCD \rangle \leq \langle |AB|^2 \rangle$ , and the quantum mechanical prediction reads  $\sum_{n=0}^N n^2 z_n z_{n-1} \leq \sum_{n=0}^N z_n^2 (n+1/2)^2$ . This is equivalent to the positivity of the  $(N+1) \times (N+1)$  matrix  $M$  with entries  $M_{nn} = (n+1/2)^2$  for  $n = 0, 1, \dots, N$  and  $M_{n,n+1} = M_{n+1,n} = -(n+1)^2/2$  for  $n = 0, 1, \dots, N-1$ , and 0 otherwise. However, for  $N \geq 10$  we get  $\det M < 0$  so it must have a negative eigenvalue. A numerical check for  $N = 10$  shows that, e.g., the state with  $\{z_n\} = \{0.83, 0.42, 0.27, 0.18, 0.13, 0.09, 0.07, 0.05, 0.03, 0.02, 0.01\}$  violates (11). The generation of the highly entangled state violation (11) will be difficult but possible because techniques of generation of multipartite entangled optical states already exist [25].

## VII. CONCLUSIONS

We have proved that one cannot show nonclassicality by violating inequalities containing only up to third-order correlations, except state-dependent contextuality. Fourth-order correlations are sufficient to violate locality and state-independent noncontextuality but the corresponding inequalities are quite complicated. A fourth-order quadripartite Bell-type inequality (11) can be violated by fourth-order correlations of position and momentum or quadratures for special entangled states.

## ACKNOWLEDGMENTS

We are grateful for discussions with N. Gisin and M. Reid. A.B. acknowledges financial support by the Polish MNiSW Grant No. IP2011 002371. W. Bednorz acknowledges partial financial support by the Polish MNiSW Grant No. N N201 608740. W. Belzig acknowledges financial support by the DFG via SPP 1285 and SFB 767.

## APPENDIX A: POSITIVE DEFINITE CORRELATIONS

Let us assume that the correlation matrix  $\mathcal{C}$  from (1) is strictly positive definite, having all eigenvalues positive. We will prove that every third-order correlation can be explained also by a positive probability. We also shift all first-order averages to zero,  $X_i \rightarrow X_i - \langle X_i \rangle$ . So far the distribution of  $X$  was Gaussian and  $X$  were continuous, but in this case all central third moments are zero. To allow for nonzero third moments we have to change the probability. The simplest (but not the only) way is to change the probability at particular values of  $X$  to get a non-Gaussian distribution. We define additional labels  $\{ijkq\}$ ,  $i \neq j \neq k \neq i$  (in this case 1 for all possible permutations of  $ijk$ ),  $\{ijq\pm\}$ ,  $i \neq j$ ,  $\{ijq\pm\} \neq \{jiq\pm\}$  (here order matters), and  $\{iq\}$  with an auxiliary parameter  $q \in \{3, -1, -2\}$ . The modified distribution reads

$$\begin{aligned} \varrho(X) &= \varrho_G(X) + \lambda^{-3} \sum_L \prod_j \delta(X_j - W_j(L)), \\ \varrho_G(X) &= \frac{1 - c/\lambda^3}{(2\pi)^{n/2} (\det \mathcal{C})^{1/2}} e^{-\sum_{ij} c_{ij}^{-1} X_i X_j / 2}, \end{aligned} \quad (A1)$$

where  $\varrho_G$  is the ‘‘old’’ Gaussian (renormalized) while the second part is the sum over  $\delta$  peaks at particular points depending on the label  $L$ . Here  $c$  is the number of all labels  $L$  and  $\lambda > 0$  is some very large real parameter such that

$c/\lambda^3 < 1$ . The positions of the peaks are

$$W_{i,j,k}(\{ijkq\}) = q\lambda \langle X_i X_j X_k \rangle^{1/3} / \sqrt[3]{18},$$

$$W_i(\{ijq\pm\}) = \pm \sqrt{2} q Q_{ij} / \sqrt[3]{18},$$

$$W_j(\{ijq\pm\}) = q Q_{ij} / \sqrt[3]{18},$$

$$Q_{ij} = \frac{\lambda}{\sqrt[3]{4}} \left[ \langle X_i^2 X_j \rangle - \sum_{k \neq ij} \langle X_i X_j X_k \rangle \right]^{1/3}, \quad (\text{A2})$$

$$W_i(\{iq\}) = \frac{q\lambda}{\sqrt[3]{18}} \left[ \langle X_i^3 \rangle - \sum_{j \neq i} \langle X_j^2 X_i \rangle / 2 \right]^{1/3},$$

$$W_l(\{ijkq\}) = W_l(\{ijq\pm\}) = W_l(\{iq\}) = 0, \quad l \neq ijk.$$

The cubic root is defined real for real negative arguments. Here  $\langle X_i X_j X_k \rangle$  are the desired third moments (the argument holds even for noncommuting observables). Note that the special choice of  $q$  results in unchanged averages  $\langle X_i \rangle$  as  $3 - 1 - 2 = 0$  but nonzero third-order averages as  $3^3 - 1^3 - 2^3 = 18$ . The calculation of the third moments gives exactly the desired values. Unfortunately, it will modify the correlation matrix  $\mathcal{C}$ . However, the correction is  $\sim 1/\lambda$ . The modified correlation matrix is then arbitrarily close to  $\mathcal{C}$  at  $\lambda \rightarrow \infty$ , so it must be positive definite and we can find the new Gaussian part in the form  $\varrho_G \propto \exp(-\sum_{ij} \mathcal{C}'^{-1}_{ij} X_i X_j / 2)$ , where the matrix  $\mathcal{C}'$  gives the correct total second-order correlations.

The assignment (A2) is certainly not unique; one could easily find a lot of different ones also reproducing correctly third-order correlations. However, the bottom line is that the proof works only if  $\mathcal{C}$  has a positive signature. If some eigenvalues of  $\mathcal{C}$  are 0 (which occurs when a particular  $X_i$  is in fact linearly dependent on the others) then  $\mathcal{C}'$  may have a negative eigenvalue for arbitrary  $\lambda$  and we cannot find any Gaussian distribution, as shown in the example in Sec. IV.

## APPENDIX B: NONCONTEXTUALITY IN SIMPLE CASES

Let us examine state-dependent noncontextuality with up to four observables  $\hat{A}_i$ ,  $i = 1, 2, 3, 4$ , with the outcomes  $A_i$  or  $A, B, C, D$ . We look for a positive probability  $\varrho(\{A_i\})$  that reproduces correctly all first, second, and third moments calculated by quantum rules. We have the freedom to set values of correlations of noncommuting products of observables because they are not measurable simultaneously. The construction of the probability depends on the commutation properties of the set  $\{\hat{A}_i\}$  and is shown for various cases in Table II. We denote  $\rho(\{A_i\}) = \text{Tr} \hat{\rho} \Pi_i \delta(A_i - \hat{A}_i)$  for every subset of commuting  $\hat{A}_i$ .

The only difficult case is with noncommuting pairs  $(\hat{A}_1, \hat{A}_2)$  and  $(\hat{B}_1, \hat{B}_2)$  but this is equivalent to the test of local realism. We will show in the general proof in Appendix C that this case can be always (if we do not use fourth moments) explained by a LHV model. Thus, we have shown that it is possible to define positive probability distributions  $\varrho$  that reproduces all quantum first, second, and third moments of measurable (commuting) combinations of up to four observables.

In two-dimensional Hilbert space the situation is somewhat simpler and we can find a classical construction for an arbitrary number of observables (not limited to 4).

TABLE II. Construction of positive probabilities  $\varrho$  for all cases of up to four observables. Here the link  $-$  means the observables commute (those not linked do not commute). Exceptions:  $\varrho = 0$  when the denominator is zero.

Observables	$\varrho(A, B, C, D)$
$A B C D$	$\rho(A)\rho(B)\rho(C)\rho(D)$
$A - B C D$	$\rho(A, B)\rho(C)\rho(D)$
$A D$	$\rho(A, B, C)\rho(D)$
$/ \setminus$	
$B - C$	$\rho(A)\rho(B, C)\rho(D)$
$A D$	
$B - C$	$\rho(A, B)\rho(A, C)\rho(D)/\rho(A)$
$B - A - C D$	$\rho(A, B)\rho(A, C)\rho(A, D)/\rho^2(A)$
$B - A - D$	
$ $	$\rho(A, B, C)\rho(A, D)/\rho(A)$
$C$	
$A - D$	$\rho(A, B, C)\rho(A, D)/\rho(A)$
$/ \setminus$	
$B - C$	$\rho(A, B, C, D)$
$A - D$	
$  \times  $	$\rho(A, B, C, D)$
$B - C$	
$A - B - C - D$	$\rho(A, B)\rho(B, C)\rho(C, D)/\rho(B)\rho(C)$
$A - B C - D$	$\rho(A, B)\rho(C, D)$
$A$	$\rho(A, B, C)\rho(B, C, D)/\rho(B, C)$
$/ \setminus$	
$B - C$	$\rho(A, B, C)\rho(B, C, D)/\rho(B, C)$
$\setminus /$	
$D$	$\rho(A, B, C)\rho(B, C, D)/\rho(B, C)$
$A_1 - B_1$	
$ $	Appendix C
$B_2 - A_2$	

The observables have the structure  $\hat{A} = a_0 \hat{1} + \vec{a} \cdot \hat{\sigma}$ , where  $\hat{\sigma} = (\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3)$  with standard Pauli matrices  $\hat{\sigma}_j$ , satisfying  $\{\hat{\sigma}_j, \hat{\sigma}_m\} = 2\delta_{jm} \hat{1}$ . Observables  $\hat{A}$  and  $\hat{B}$  commute if and only if  $\vec{a} \parallel \vec{b}$ . We can group all observables (their number is arbitrary) parallel to the same direction, so that  $\vec{a}_\alpha \parallel \vec{a}$ ,  $\vec{b}_\beta \parallel \vec{b}$ ,  $\vec{c}_\gamma \parallel \vec{c}$ ,  $\dots$ , where  $\vec{a} \not\parallel \vec{b}, \vec{c}, \dots$ ,  $\vec{b} \not\parallel \vec{c}, \dots$ , etc. Then we construct a LHV model defined by  $\varrho(\{A_\alpha\}, \{B_\beta\}, \{C_\gamma\}, \dots) = \rho(\{A_\alpha\})\rho(\{B_\beta\})\rho(\{C_\gamma\}) \dots$ , where  $\rho(\{A_\alpha\}) = \text{Tr} \hat{\rho} \Pi_j \delta(A_\alpha - \hat{A}_\alpha)$  and similarly for the other sets. This means that all (noncontextual) third moments for a two-level system are reproduced by a classical probability.

On the other hand we will see in Appendix D an example of the violation of state-dependent noncontextuality involving a three-dimensional Hilbert space and five observables.

## APPENDIX C: THIRD MOMENTS: CONTEXTUAL LHV MODELS

We will present a general proof that third-order correlations can be explained by a LHV model, if contextuality is allowed and no assumption on higher-order moments or dichotomy is made. As in Sec. II, we denote  $\mathcal{C}_{X_{\alpha j}, Y_{\beta k}} = \langle X_{\alpha j} Y_{\beta k} \rangle$  for  $X, Y = A, B, C, \dots$  and  $\alpha, \beta, j, k = 1, 2, \dots$ . For a valid LHV theory,  $\mathcal{C}$  must be positive (semi)definite.

### 1. Assumptions

The proof is based on two facts:

(a)  $\mathcal{C}_{X\alpha_j, X\beta_k} = \langle X_{\alpha_j} X_{\beta_k} \rangle$  is not measurable for  $\alpha \neq \beta$  (even if accidentally  $\hat{X}_{\alpha_j} \hat{X}_{\beta_k} = \hat{X}_{\beta_k} \hat{X}_{\alpha_j}$ ) because  $\alpha$  and  $\beta$  correspond to two different settings of the same observer which cannot be realized simultaneously. So it is a free parameter in a LHV model.

(b) We can always redefine every observable within one observer's setting by a real linear transformation  $\hat{X}_{\alpha m} \rightarrow \sum_k \lambda_{\alpha k} \hat{X}_{\alpha k}$  as long as linear independence is preserved, because all such observables commute with each other.

The proof involves a kind of Gauss elimination on a set of linear equations [26].

### 2. Problem of zero eigenvalues

The first choice for  $\mathcal{C}$  will be (1), which is positive semidefinite. We shall see that this choice must be sometimes modified, without affecting the *measurable* correlations. Suppose that the correlation matrix  $\mathcal{C}$  has  $\mathcal{N}$  zero eigenvalues with linearly independent zero eigenvectors

$$W_m = \sum_{\alpha, k}^{X=A, B, \dots} \lambda_{X\alpha k}^m X_{\alpha k}, \quad m = 1, \dots, \mathcal{N} \quad (\text{C1})$$

with the property  $\langle W_m^2 \rangle = 0$ . This implies  $\text{Tr} \hat{\rho} \hat{W}_m^2 = 0$ , which gives

$$W_m = \hat{W}_m \hat{\rho} = 0, \quad m = 1, \dots, \mathcal{N}. \quad (\text{C2})$$

The above set of linear equations can be modified as in the usual algebra; we can multiply equations by nonzero numbers and add up, as long as linear independence holds. The vectors  $W_m$  span the kernel of the correlation matrix. We shall prove that for a given observer  $X$  the above set of equations can be written in the form

$$X_{\alpha k} + \sum_{\beta j}^{Y \neq X} \lambda_{Y\beta j}^{X\alpha k} Y_{\beta j} = 0, \quad (\text{C3})$$

where we sum over all observers different from  $X$  and all their settings and observables plus equations not containing  $X$ . If this is not possible then we shall prove that we can reduce the kernel by at least one vector by modifying nonmeasurable correlations in the correlation matrix, keeping its positivity. By such successive reduction we will end up with (C3). For the Bell case ( $A$  and  $B$ ,  $\alpha = 1, 2$ ) (C3) reduces either to trivial single vectors  $A_\alpha - \lambda B_\alpha$  or a set

$$\begin{aligned} A_1 &= \lambda_{11} B_1 + \lambda_{12} B_2, \\ A_2 &= \lambda_{21} B_1 + \lambda_{22} B_2 \end{aligned} \quad (\text{C4})$$

with invertible matrix  $\lambda$ . The original correlation matrix (1) may lead us into trouble for some correlations (violation of noncontextuality), which are anyway unobservable so we do not need to bother in contextual LHV models. Therefore, sometimes we have to modify it slightly to relax dangerous constraints. The resulting LHV correlation matrix can be different from (1) but only for nonmeasurable correlations. We make use of the fact that quantum mechanics does not permit measuring everything in one run of the experiment, leaving more freedom for contextual LHV models.

### 3. Reduction of zero eigenvectors

We shall prove that all zero eigenvectors can be eliminated except those in the form of (C3). Without loss of generality let us take  $X = A$ . We write (C2) in the form

$$\sum_{\alpha k} \lambda_{\alpha k}^m A_{\alpha k} + \mathcal{A} = 0, \quad (\text{C5})$$

where  $\mathcal{A}$  replaces all linear combinations of quantities measured by the other observers ( $B, C, D, \dots$ ), e.g.,  $\mathcal{A}$  can be  $2B_{11} - 3B_{11} + B_{21} - 5C_{13}$ . By linear eliminations and transformations within setting 1, there exists a form of (C5) consisting of

$$A_{1k} + \mathcal{V} + \mathcal{A} = 0, \quad k = 1, 2, \dots, \quad (\text{C6})$$

with  $\mathcal{V}$  not containing  $A_{1j}$  terms, and other equations that do not contain  $A_{1j}$  at all. Suppose that at least one of (C6) contains an  $A_{2j}$  term, so in general (C6) has the form

$$A_{1k} + \sum_m \lambda_{km} A_{2m} + \mathcal{V} + \mathcal{A} = 0, \quad k = 1, 2, \dots \quad (\text{C7})$$

with at least one  $\lambda_{km} \neq 0$  and  $\mathcal{V}$  denoting all terms not containing  $A_{1j}$  and  $A_{2j}$ . By linear eliminations and transformations within settings 1 and 2 we arrive at

$$\begin{aligned} A_{1k} + A_{2k} + \mathcal{V} + \mathcal{A} &= 0, \quad k = 1, 2, \dots, l, \\ A_{1k} + \mathcal{V} + \mathcal{A} &= 0, \quad k = l+1, l+2, \dots, \\ A_{2k} + \mathcal{V} + \mathcal{A} &= 0, \quad k = l+1, l+2, \dots, \end{aligned} \quad (\text{C8})$$

and other equations that do not contain  $A_{1j}$  nor  $A_{2j}$  at all (if we have a single observable for each setting then we can omit the index  $k$ ). If  $l > 0$  then we change  $\langle A_{11} A_{21} \rangle \rightarrow \langle A_{11} A_{21} \rangle + \epsilon$  with  $\epsilon > 0$  in the correlation matrix  $\mathcal{C}$  (or  $\langle A_1 A_2 \rangle$  for single observables). Then  $\langle W^2 \rangle = 2\epsilon > 0$ , where  $W$  is the left-hand side of the first line in (C8) for  $k = 1$ . Correlations involving other kernel vectors remain unaffected as none of them contains  $A_{11}$  nor  $A_{21}$ . For sufficiently small but positive  $\epsilon$  the new correlation matrix  $\mathcal{C}$  will be strictly positive in the space spanned by the old nonkernel vectors plus  $W$ . In this way we reduce by 1 the dimension of the kernel. By repeating this reasoning we kick out of the kernel all vectors on the left-hand side of the first line of (C8). Once we are left with only the two last lines of (C8) we proceed by induction.

Let us assume that, at some stage with a fixed  $\alpha$ , the kernel equations have the form

$$A_{\xi k} + \sum_m \lambda_{km}^{\xi} A_{\alpha m} + \mathcal{V} \cdots \mathcal{Q} + \mathcal{A} = 0 \quad (\text{C9})$$

for all  $\xi < \alpha$ , plus other equations not containing  $A_{\xi}$  and  $A_{\alpha m}$ . Note that the set of possible  $k$  can be different for different  $\xi$ . If all  $\lambda = 0$  then we can proceed to the next induction step, taking the next setting. Otherwise, let us denote by  $\Xi$  the set of all  $\xi$  with  $\lambda_{k1}^{\xi} \neq 0$  for some  $k$  (we fix the other index to 1 without loss of generality). By linear eliminations we find only one such  $k$  for each  $\xi \in \Xi$  so that  $\lambda_{k1}^{\xi} = \delta_{k1}$ . Now, we make a shift of the nonmeasurable correlations  $\langle A_{\xi 1} A_{\alpha 1} \rangle \rightarrow \langle A_{\xi 1} A_{\alpha 1} \rangle + \epsilon$  and  $\langle A_{\xi 1} A_{\eta 1} \rangle \rightarrow \langle A_{\xi 1} A_{\eta 1} \rangle - 2\epsilon$  for  $\xi, \eta \in \Xi$  with  $\epsilon > 0$ . Denoting by  $W_{\xi}$ ,  $\xi \in \Xi$ , the subsequent left-hand sides of (C9) for  $k = 1$ , we have  $\langle W_{\xi} W_{\eta} \rangle = 2\epsilon \delta_{\xi\eta}$ . Correlations with other kernel vectors remain zero as they contain neither  $A_{\xi 1}$  nor

$A_{\alpha 1}$ . For sufficiently small  $\epsilon$  (every new  $\epsilon$  is much smaller than all previous ones), the correlation matrix  $\mathcal{C}$  on old nonkernel vectors plus  $W_\xi$  is strictly positive, as in (C8). Hence, we kick  $W_\xi$  out of the kernel. By repeating this step for subsequent  $m$  we get rid of all unwanted kernel vectors and can proceed with the induction step. Then we repeat it for each observer, to finally arrive at the desired form (C3).

#### 4. Construction of third moments

Now, we *define* all third-order correlations, including noncommuting observables. We divide all observables into two families:  $V_j$ —appearing in (C3) and  $Y_m$ —the rest. Now,

$$\begin{aligned}\langle Y_m Y_n Y_p \rangle &= \sum_{\sigma(mnp)} \text{Tr} \hat{\rho} \hat{Y}_m \hat{Y}_n \hat{Y}_p / 6, \\ \langle V_j Y_m Y_n \rangle &= \text{Tr} \hat{\rho} \{ \hat{V}_j, \{ \hat{Y}_m, \hat{Y}_n \} \} / 4, \\ \langle V_k V_l Y_n \rangle &= \text{Tr} \hat{\rho} (\hat{V}_j \hat{Y}_n \hat{V}_k + \hat{V}_k \hat{Y}_n \hat{V}_j) / 2, \\ \langle V_j V_k V_l \rangle &= \sum_{\sigma(jkm)} \text{Tr} \hat{\rho} \hat{V}_j \hat{V}_k \hat{V}_l / 6,\end{aligned}\quad (\text{C10})$$

where  $\sigma$  denotes all six permutations. The above definition is consistent with projective measurements for all measurable correlations.

We have to check if  $\langle WZZ' \rangle = 0$  for  $W$  given by an arbitrary linear combination of left-hand sides of (C3) and  $Z, Z' = V_j, Y_m$ . If  $Z, Z' = Y_m, Y_n$  it is clear because

$$\hat{W} \hat{\rho} = 0. \quad (\text{C11})$$

If  $Z = Y_m, Z' = V_j$ , then

$$2 \langle W Y_m V_j \rangle = \text{Tr} \hat{\rho} (\hat{W} \hat{Y}_m \hat{V}_j + \hat{V}_j \hat{Y}_m \hat{W}) = 0 \quad (\text{C12})$$

again because of (C11). Finally, we need to consider  $Z = V_j, Z' = V_k$ . Because of (C11), we get

$$6 \langle W V_j V_k \rangle = \text{Tr} \hat{\rho} (\hat{V}_j \hat{W} \hat{V}_k + \hat{V}_k \hat{W} \hat{V}_j). \quad (\text{C13})$$

Without loss of generality we need to consider only two cases. The first one is  $V_j = A_j, V_k = B_k$ . If  $W$  does not contain  $A$  or  $B$  then we can move it to the left or right and (C13) vanishes due to (C11). Now suppose  $W$  contains  $A_m$ . By virtue of (C3) we can write

$$W = A_m + \sum_n \lambda_n B_n + \mathcal{A}\mathcal{B}, \quad (\text{C14})$$

where  $\mathcal{A}\mathcal{B}$  denotes all terms not containing  $A$  and  $B$ . Moving  $A_m$  and  $\sum_n \lambda_n B_n + \mathcal{A}\mathcal{B}$  in opposite directions in (C13), it can be transformed into

$$\begin{aligned}\text{Tr} \hat{\rho} (\hat{A}_j \hat{W} \hat{B}_k + \hat{B}_k \hat{W} \hat{A}_j) &= \text{Tr} \hat{\rho} (\hat{A}_j \hat{B}_k \hat{A}_m + \hat{A}_m \hat{B}_k \hat{A}_j) \\ &+ \text{Tr} \hat{\rho} \left[ \left( \sum_n \lambda_n \hat{B}_n + \mathcal{A}\mathcal{B} \right) \hat{A}_j \hat{B}_k \right. \\ &+ \left. \hat{B}_k \hat{A}_j \left( \sum_n \lambda_n \hat{B}_n + \mathcal{A}\mathcal{B} \right) \right] \\ &= \text{Tr} \hat{\rho} (\hat{A}_j \hat{B}_k \hat{W} + \hat{W} \hat{B}_k \hat{A}_j),\end{aligned}$$

where we used the commutation rule  $\hat{A}_j \hat{B}_k = \hat{B}_k \hat{A}_j$ . The last expression vanishes due to (C11). If  $W$  contains  $B_m$ , we proceed analogously.

The last case is  $V_j = A_j, V_k = A_k$ . If  $W$  does not contain any  $A$  terms then we can move  $W$  to the left or right and (C13) vanishes due to (C11). The remaining cases, due to (C3), have the form  $W = A_m + \mathcal{A}$ , and (C13) reads

$$\begin{aligned}\text{Tr} \hat{\rho} (\hat{A}_j \hat{W} \hat{A}_k + \hat{A}_k \hat{W} \hat{A}_j) &= \text{Tr} \hat{\rho} (\hat{A}_j \hat{A}_m \hat{A}_k + \hat{A}_k \hat{A}_m \hat{A}_j) \\ &+ \text{Tr} \hat{\rho} (\hat{\mathcal{A}} \hat{A}_j \hat{A}_k + \hat{A}_k \hat{A}_j \hat{\mathcal{A}}).\end{aligned}\quad (\text{C15})$$

Now we remember that (C3) must contain also  $W' = A_k - \mathcal{A}'$  so  $\hat{A}_k \hat{\rho} = \hat{\mathcal{A}}' \hat{\rho}$ , which gives

$$\begin{aligned}\text{Tr} \hat{\rho} (\hat{A}_j \hat{A}_m \hat{A}_k + \hat{A}_k \hat{A}_m \hat{A}_j) &= \text{Tr} \hat{\rho} (\hat{A}_j \hat{A}_m \hat{\mathcal{A}}' + \hat{\mathcal{A}}' \hat{A}_m \hat{A}_j) \\ &= \text{Tr} \hat{\rho} (\hat{\mathcal{A}}' \hat{A}_j \hat{A}_m + \hat{A}_m \hat{A}_j \hat{\mathcal{A}}') \\ &= \text{Tr} \hat{\rho} (\hat{A}_k \hat{A}_j \hat{A}_m + \hat{A}_m \hat{A}_j \hat{A}_k),\end{aligned}\quad (\text{C16})$$

so (C15) reads  $\text{Tr} \hat{\rho} (\hat{W} \hat{A}_j \hat{A}_k + \hat{A}_k \hat{A}_j \hat{W})$ , which vanishes due to (C11). We see that correlations containing arbitrary combinations of left-hand sides of (C3) vanish. Now, we can simply eliminate one observable from each kernel equation (C3),  $\sum_k \lambda_k Z_k = 0$ , by the substitution  $Z_m = -\sum_{k \neq m} \lambda_k Z_k / \lambda_m$ , so that only  $Z_k, k = 1, \dots, l$  remain as independent observables. Hence, the correlation matrix  $\mathcal{C}$  is strictly positive (the kernel is null) and we construct the final LHV model reproducing all measurable quantum first-, second-, and third-order correlations as in Appendix A. The third-order correlations involving substituted observables are reproduced by virtue of the just shown property of (C10). This completes the proof.

#### APPENDIX D: VIOLATION OF STATE-DEPENDENT NONCONTEXTUALITY WITH THIRD MOMENTS

There exists a third-moment-based state-dependent example violating noncontextuality with five observables in a three-dimensional Hilbert space, which we will construct now. Let us take observables  $\hat{A}_\alpha$ , for  $\alpha = 1, 2, 3, 4, 5$ . Below all summations are over the set  $\{1, 2, 3, 4, 5\}$  and indices are counted modulo 5,  $\alpha + 5\mu \equiv \alpha$  with integer  $\mu$ . We assume that  $\hat{A}_\alpha \hat{A}_{\alpha+2} = \hat{A}_{\alpha+2} \hat{A}_\alpha$  but  $\hat{A}_\alpha \hat{A}_{\alpha+1} \neq \hat{A}_{\alpha+1} \hat{A}_\alpha$ , so there are five commuting pairs and five noncommuting pairs. Suppose that an experimentalist measures

$$\begin{aligned}\mathcal{S} &= \left\langle \left( \sum_\alpha A_\alpha \cos \frac{4\pi\alpha}{5} \right)^2 \right\rangle + \left\langle \left( \sum_\alpha A_\alpha \sin \frac{4\pi\alpha}{5} \right)^2 \right\rangle \\ &+ \left\langle \left( \sum_\alpha A_\alpha \right)^2 \right\rangle \cos \frac{\pi}{5} = \sum_\alpha \langle A_\alpha^2 \rangle [1 + \cos(\pi/5)] \\ &+ \sum_\alpha 2 \langle A_\alpha A_{\alpha+2} \rangle [\cos(\pi/5) + \cos(2\pi/5)].\end{aligned}\quad (\text{D1})$$

Let us denote Fourier operators by  $\hat{A}(q) = \sum_\alpha \hat{A}_\alpha e^{2\pi i \alpha q / 5}$ . Since  $\hat{A}_\alpha = \hat{A}_\alpha^\dagger$ , we have  $\hat{A}(0) = \hat{A}^\dagger(0)$ ,  $\hat{A}(-1) = \hat{A}(4) = \hat{A}^\dagger(1)$ , and  $\hat{A}(-2) = \hat{A}(3) = \hat{A}^\dagger(2)$ . Similarly, for outcomes  $A(0) = A^*(0)$ ,  $A(-1) = A(4) = A^*(1)$ , and  $A(-2) = A(3) =$

$A^*(2)$  (there are either five real random variables or one real and two complex). We can write (D1) in the equivalent form

$$\mathcal{S} = \langle |A(2)|^2 \rangle + \langle [A(0)]^2 \rangle \cos(\pi/5). \quad (\text{D2})$$

If  $\mathcal{S} = 0$  then  $A(0) = A(2) = 0$ . Let us further take

$$\mathcal{Q} = 25 \sum_{\alpha} \langle A_{\alpha}^3 \rangle = \sum_{q,p,r}^{q+p+r=0} \langle A(q)A(p)A(r) \rangle. \quad (\text{D3})$$

Each term of the expansion of the right-hand side must contain  $A(\pm 2)$  or  $A(0)$  because  $\pm 1 \pm 1 \pm 1 \neq 0$  so  $\mathcal{S} = 0$  implies  $\mathcal{Q} = 0$ .

Denoting the commutator by  $[\hat{X}, \hat{Y}] = \hat{X}\hat{Y} - \hat{Y}\hat{X}$ , we have

$$\begin{aligned} 0 &= 5 \sum_{\alpha} [\hat{A}_{\alpha}, \hat{A}_{\alpha+2}] e^{2\pi i \alpha q/5} \\ &= \sum_p [\hat{A}(q-p), \hat{A}(p)] e^{-4\pi i p/5} \\ &= \sum_p [\hat{A}(p+q), \hat{A}^{\dagger}(p)] e^{4\pi i p/5}. \end{aligned} \quad (\text{D4})$$

By inverse Fourier transform, satisfying the above relations for  $q = 1, \dots, 5$  is equivalent to  $[\hat{A}_{\alpha}, \hat{A}_{\alpha+2}] = 0$ . In fact, there are only three independent equations in (D4) for  $q = 0, 1, 2$  because  $q = 3, 4$  can be obtained from Hermitian conjugation of  $q = 2, 1$  with some factor. We obtain

$$\begin{aligned} [\hat{A}(1), \hat{A}^{\dagger}(1)] \sin \frac{\pi}{5} - [\hat{A}(2), \hat{A}^{\dagger}(2)] \sin \frac{2\pi}{5} &= 0, \\ [\hat{A}(1), \hat{A}(0)] \sin \frac{2\pi}{5} - [\hat{A}(2), \hat{A}^{\dagger}(1)] \sin \frac{\pi}{5} &= 0, \\ [\hat{A}(2), \hat{A}(0)] \sin \frac{\pi}{5} - [\hat{A}^{\dagger}(2), \hat{A}^{\dagger}(1)] \sin \frac{2\pi}{5} &= 0. \end{aligned} \quad (\text{D5})$$

In the basis  $|0\rangle, |1\rangle, |2\rangle$ , we take

$$\begin{aligned} \hat{A}(0) &= a \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \hat{A}(2) = b \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & i \\ 0 & i & -1 \end{pmatrix}, \\ \hat{A}(1) &= c \begin{pmatrix} 0 & 1 & i \\ 1 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \end{aligned} \quad (\text{D6})$$

with real  $a$  and complex  $b, c$ . We have  $[\hat{A}(0), \hat{A}(2)] = \hat{A}(1)\hat{A}(2) = \hat{A}(2)\hat{A}(1) = 0$ ,  $[\hat{A}(1), \hat{A}^{\dagger}(1)] = 2|c|^2\hat{B}$ ,  $[\hat{A}(2), \hat{A}^{\dagger}(2)] = 4|b|^2\hat{B}$ ,  $[\hat{A}(1), \hat{A}(0)] = ac\hat{C}$ , and  $[\hat{A}(2), \hat{A}^{\dagger}(1)] = -2bc^*\hat{C}$ , where

$$\hat{B} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \hat{C} = \begin{pmatrix} 0 & 1 & i \\ -1 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}. \quad (\text{D7})$$

To satisfy (D5), we need  $|c|^2 = 4|b|^2 \cos(\pi/5)$  and  $bc^* = -ac \cos(\pi/5)$ , satisfied by  $b = 1$ ,  $c = 2\sqrt{\cos(\pi/5)}$ , and  $a = -1/\cos(\pi/5)$ .

Assuming noncontextuality, the quantum mechanical expectation for (D1) reads

$$\begin{aligned} \mathcal{S} &= \sum_{\alpha} \text{Tr} \hat{\rho} \hat{A}_{\alpha}^2 [1 + \cos(\pi/5)] \\ &+ \sum_{\alpha} 2 \text{Tr} \hat{\rho} \hat{A}_{\alpha} \hat{A}_{\alpha+2} [\cos(\pi/5) + \cos(2\pi/5)] \\ &= \text{Tr} \hat{\rho} [\hat{A}^{\dagger}(2)\hat{A}(2) + \hat{A}(2)\hat{A}^{\dagger}(2) + 2\hat{A}^2(0) \cos(\pi/5)]/2, \end{aligned} \quad (\text{D8})$$

and for (D3),

$$\mathcal{Q} = 25 \sum_{\alpha} \text{Tr} \hat{\rho} \hat{A}_{\alpha}^3 = \sum_{q,p,r}^{q+p+r=0} \text{Tr} \hat{\rho} \hat{A}(q)\hat{A}(p)\hat{A}(r). \quad (\text{D9})$$

For  $\hat{\rho} = |0\rangle\langle 0|$ , we have  $\hat{A}(0, \pm 2)\hat{\rho} = \hat{\rho}\hat{A}(0, \pm 2) = 0$ , so  $\mathcal{S} = 0$ . By explicit calculation we find

$$\begin{aligned} \mathcal{Q} &= \langle 0|\hat{A}(1)\hat{A}(0)\hat{A}^{\dagger}(1)|0\rangle + \langle 0|\hat{A}^{\dagger}(1)\hat{A}(0)\hat{A}(1)|0\rangle \\ &+ \langle 0|\hat{A}(1)\hat{A}^{\dagger}(2)\hat{A}(1)|0\rangle + \langle 0|\hat{A}^{\dagger}(1)\hat{A}(2)\hat{A}^{\dagger}(1)|0\rangle \\ &= 4a|c|^2 + 8\text{Re}(b^*c^2) = 8(\sqrt{5} - 1) \simeq 9.9, \end{aligned} \quad (\text{D10})$$

in clear contradiction to the classical prediction  $\mathcal{Q} = 0$ .

#### APPENDIX E: NO-GO THEOREM ON TWO-PARTY CFRD INEQUALITIES

Simple fourth-order CFRD-type inequalities can be constructed for two observers  $A$  and  $B$ , with up to eight settings (and a single real outcome for each setting) [11,23],  $A_{\alpha}^{r/i}, B_{\alpha}^{r/i}$  with  $\alpha = 0, 1, 2, 3$ , and read

$$\begin{aligned} &|\langle A_0 B_0^{\dagger} + A_1 B_1^{\dagger} + A_2 B_2^{\dagger} + A_3 B_3^{\dagger} \rangle|^2 \\ &+ |\langle A_0 B_1 - A_1 B_0 + A_2 B_3^{\dagger} - A_3 B_2^{\dagger} \rangle|^2 \\ &+ |\langle A_0 B_2 - A_2 B_0 + A_3 B_1^{\dagger} - A_1 B_3^{\dagger} \rangle|^2 \\ &+ |\langle A_0 B_3 - A_3 B_0 + A_1 B_2^{\dagger} - A_2 B_1^{\dagger} \rangle|^2 \\ &\leq \sum_{\alpha\beta} \langle (A_{\alpha}^{\dagger} A_{\alpha} + A_{\alpha} A_{\alpha}^{\dagger})(B_{\beta}^{\dagger} B_{\beta} + B_{\beta} B_{\beta}^{\dagger}) \rangle / 4, \end{aligned} \quad (\text{E1})$$

where we have denoted  $C = C^r + iC^i$ ,  $C = A_{\alpha}, B_{\alpha}$ . The notation is the same in the classical and quantum cases except  $\dagger \rightarrow *$ . We use the complex form only to save space but all the inequalities can be expanded into purely real terms [23]. The inequality reduces to (10) if we leave only  $A_1^r, A_2^r, B_1^r$ , and  $B_2^r$ , while other observables are zero. Classically, (E1) follows from the inequality  $|\langle z \rangle|^2 \leq \langle |z|^2 \rangle$  applied to each term on the left-hand side and summed up. Surprisingly, the inequality is not violated at all in quantum mechanics, which has been proved in [19]. Below we present an alternative proof.

It suffices to prove (E1) for pure states  $\hat{\rho} = |\psi\rangle\langle\psi|$ . For mixed states  $\hat{\rho} = \sum_k p_k |\psi_k\rangle\langle\psi_k|$ ,  $p_k \geq 0$ , and  $\sum_k p_k = 1$ . We apply the triangle inequality  $|\sum_k p_k z_k| \leq \sum_k p_k |z_k|$  and the Jensen inequality  $(\sum_k p_k |z_k|)^2 \leq \sum_k p_k |z_k|^2$ , where  $z_k$  is the complex correlator in each of the four terms on the left-hand side of (E1) taken for a pure state  $|\psi_k\rangle$ . If (E1) is valid for each  $|\psi_k\rangle$  then it holds for the mixture, too.



Let us focus then on pure states. Note that the sum of the last three terms on the left-hand side of (E1) can be written as

$$\begin{aligned} & \sum_{\alpha\beta} (\langle A_\alpha B_\beta \rangle \langle A_\alpha^\dagger B_\beta^\dagger \rangle - \langle A_\alpha B_\beta \rangle \langle A_\beta^\dagger B_\alpha^\dagger \rangle) \\ & + \sum_{\alpha\beta\gamma\delta} \epsilon_{\alpha\beta\gamma\delta} (\langle A_\alpha B_\beta \rangle \langle A_\gamma B_\delta \rangle + \langle A_\alpha^\dagger B_\beta^\dagger \rangle \langle A_\gamma^\dagger B_\delta^\dagger \rangle) / 2, \end{aligned} \quad (\text{E2})$$

using the completely antisymmetric tensor  $\epsilon$  with  $\epsilon_{0123} = 1$ . Therefore the whole inequality is invariant under SU(4) transformations of  $A_\alpha$  and  $B_\beta$  treated as components of four-dimensional vectors (it is straightforward to verify the invariance of other parts of the inequality). We recall that these external transformations do not interfere with the internal Hilbert spaces  $\mathcal{H}_{A,B}$ .

Let us number the four complex correlators inside the moduli on the left-hand side of (E1) by 0, 1, 2, 3, respectively (e.g., 0 is the correlator  $\sum_\alpha \langle A_\alpha B_\alpha^\dagger \rangle$ ). We want to transform (E1) to a form with a single real correlator 0 while 1, 2, and 3 vanish. Let us begin with a transformation  $C_\alpha \rightarrow e^{i\phi_\alpha} C_\alpha$ ,  $C = A, B$ , with  $\sum_\alpha \phi_\alpha = 0$ . Note that  $A_0 B_1 - A_1 B_0 + A_2^\dagger B_3^\dagger - A_3^\dagger B_2^\dagger$  just gets the phase factor  $e^{i(\phi_0 + \phi_1)}$ , so by tuning  $\phi_\alpha$  we can always make the correlators 1, 2, and 3 real. Making now a real rotation in 123 space we can leave only the real correlator 3 while 1 and 2 vanish. Still, the correlator 0 can also have an unwanted imaginary component, because 0 is invariant under SU(4) transformations. To get rid of it, we have to apply a different transformation  $A_0 \rightarrow A_0$ ,  $A_1 \rightarrow A_1$ ,  $A_2 \rightarrow A_2^\dagger$ ,  $A_3 \rightarrow A_3^\dagger$ ,  $B_0 \rightarrow -B_1^\dagger$ ,  $B_1 \rightarrow B_0^\dagger$ ,  $B_2 \rightarrow -B_3$ ,  $B_3 \rightarrow B_2$ , which gives

$$\begin{aligned} & A_0 B_1 - A_1 B_0 + A_2^\dagger B_3^\dagger - A_3^\dagger B_2^\dagger \\ & \rightarrow A_0 B_0^\dagger + A_1 B_1^\dagger + A_2 B_2^\dagger + A_3 B_3^\dagger, \\ & A_0 B_2 - A_2 B_0 + A_3^\dagger B_1^\dagger - A_1^\dagger B_3^\dagger \\ & \rightarrow -A_0 B_3 + A_3 B_0 - A_1^\dagger B_2^\dagger + A_2^\dagger B_1^\dagger, \\ & A_0 B_3 - A_3 B_0 + A_1^\dagger B_2^\dagger - A_2^\dagger B_1^\dagger \\ & \rightarrow A_0 B_2 - A_2 B_0 + A_3^\dagger B_1^\dagger - A_1^\dagger B_3^\dagger, \\ & A_0 B_0^\dagger + A_1 B_1^\dagger + A_2 B_2^\dagger + A_3 B_3^\dagger \\ & \rightarrow -A_0 B_1 + A_1 B_0 - A_2^\dagger B_3^\dagger + A_3^\dagger B_2^\dagger. \end{aligned} \quad (\text{E3})$$

It is clear that the inequality (E1) remains unchanged [we can change signs in the second and fourth parts of (E3)]. Now the correlator 0 vanishes because it is moved to  $-1$  and  $1$  is moved to  $0$  ( $2 \rightarrow -3$ ,  $3 \rightarrow 2$ ). Applying again an SU(4) transformation, we can get correlator 1 real while 2 and 3 vanish and 0 remains null because it is invariant under SU(4). Applying again (E3) we get only a single real term in 0. In this way, the left-hand side of (E1) reads

$$\left( \text{Re} \sum_\alpha \langle A_\alpha B_\alpha^\dagger \rangle \right)^2. \quad (\text{E4})$$

We apply the triangle inequality

$$\left| \sum_\alpha^{q=r,i} \langle A_\alpha^q B_\alpha^q \rangle \right| \leq \sum_\alpha^{q=r,i} |\langle A_\alpha^q B_\alpha^q \rangle|. \quad (\text{E5})$$

Note that  $|\langle A_\alpha^q B_\alpha^q \rangle| \leq \langle |A_\alpha^q| |B_\alpha^q| \rangle$ , where  $|X|$  is obtained by reversing the signs of all negative eigenvalues of  $X$  (in the eigenbasis). To prove (10) we have to show that

$$\left( \sum_\alpha^{q=r,i} \langle |A_\alpha^q| |B_\alpha^q| \rangle \right)^2 \leq \sum_{\alpha\beta}^{q,p=r,i} \langle |A_\alpha^q|^2 |B_\beta^p|^2 \rangle. \quad (\text{E6})$$

We decompose  $|\psi\rangle$  and arbitrary operators  $\hat{A}^x$  and  $\hat{B}^x$  in the basis  $|k_A i_B\rangle$  of the joint Hilbert space  $\mathcal{H}_A \otimes \mathcal{H}_B$ ,

$$\begin{aligned} |\psi\rangle &= \sum_{ki} \psi_{ki} |k_A i_B\rangle, \quad \hat{A}^x = \sum_{kli} A_{kl}^x |k_A i_B\rangle \langle l_A i_B|, \\ \hat{B}^x &= \sum_{kij} B_{ij}^x |k_A i_B\rangle \langle k_A j_B|. \end{aligned} \quad (\text{E7})$$

The normalization reads  $\sum_{ki} |\psi_{ki}|^2 = 1$ . Let us define  $\hat{\Psi} = \sum_{ki} \psi_{ki} |k\rangle \langle i|$ ,  $\hat{a}^x = \sum_{kl} A_{kl} |k\rangle \langle l|$ , and  $\hat{b}^x = \sum_{ij} B_{ij} |j\rangle \langle i|$ . Now the normalization reads  $\text{tr} \hat{\Psi}^\dagger \hat{\Psi} = 1$ . One can check the identity  $\langle \psi | \hat{A}^x \hat{B}^x | \psi \rangle = \text{tr} \hat{\Psi}^\dagger \hat{a}^x \hat{\Psi} \hat{b}^x$ . We stress that  $\hat{a}^x$  and  $\hat{b}^x$  are no longer operators in  $\mathcal{H}_A \otimes \mathcal{H}_B$ , but in  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , respectively, while  $\hat{\Psi}$  is a linear transformation from  $\mathcal{H}_B$  to  $\mathcal{H}_A$ , which need not be represented by a Hermitian or even a square matrix. We note that such a manipulation is possible only for two observers. By taking suitable bases, we could even make  $\hat{\Psi}$  diagonal, real, and positive, analogously to a Schmidt decomposition, but it is not necessary. Now (E6) reads

$$\left( \sum_\alpha^{q=r,i} \text{tr} \hat{\Psi}^\dagger |\hat{a}_\alpha^q| \hat{\Psi} |\hat{b}_\alpha^q| \right)^2 \leq \sum_{\alpha\beta}^{q,p=r,i} \text{tr} \hat{\Psi}^\dagger |\hat{a}_\alpha^q|^2 \hat{\Psi} |\hat{b}_\beta^p|^2. \quad (\text{E8})$$

To prove (E8) we need the Lieb concavity theorem [27] which states that for a fixed but arbitrary  $\hat{\Psi}$  and  $s \in [0, 1]$  the trace class function  $f(\hat{F}, \hat{G}) = \text{tr} \hat{\Psi}^\dagger \hat{F}^s \hat{\Psi} \hat{G}^{1-s}$  is *jointly concave*, which means that

$$\begin{aligned} & \lambda f(\hat{F}, \hat{G}) + (1 - \lambda) f(\hat{F}', \hat{G}') \\ & \leq f(\lambda \hat{F} + (1 - \lambda) \hat{F}', \lambda \hat{G} + (1 - \lambda) \hat{G}') \end{aligned} \quad (\text{E9})$$

for  $\lambda \in [0, 1]$  and arbitrary Hermitian semipositive operators  $\hat{F}, \hat{F}', \hat{G}, \hat{G}'$ . By induction (E9) generalizes straightforwardly to

$$\sum_\alpha \lambda_\alpha f(\hat{F}_\alpha, \hat{G}_\alpha) \leq f \left( \sum_\alpha \lambda_\alpha \hat{F}_\alpha, \sum_\beta \lambda_\beta \hat{G}_\beta \right) \quad (\text{E10})$$

for  $\lambda_\alpha \geq 0$  and  $\sum_\alpha \lambda_\alpha = 1$  and arbitrary semipositive operators  $\hat{F}_\alpha$  and  $\hat{G}_\alpha$ . We apply (E10) for  $s = 1/2$ ,  $\lambda_\alpha^q = 1/8$ ,  $F_\alpha^q = |a_\alpha^q|^2$ , and  $G_\alpha^q = |b_\alpha^q|^2$  to get

$$\begin{aligned} & \sum_\alpha^{q=r,i} \text{tr} \hat{\Psi}^\dagger |\hat{a}_\alpha^q| \hat{\Psi} |\hat{b}_\alpha^q| \\ & \leq \text{tr} \hat{\Psi}^\dagger \left( \sum_\alpha^{q=r,i} |a_\alpha^q|^2 \right)^{1/2} \hat{\Psi} \left( \sum_\beta^{p=r,i} |b_\beta^p|^2 \right)^{1/2}. \end{aligned} \quad (\text{E11})$$

Finally we use the operator Cauchy-Bunyakovsky-Schwarz inequality  $|\text{tr } \hat{c} \hat{d}|^2 \leq \text{tr } \hat{c} \hat{c}^\dagger \text{tr } \hat{d} \hat{d}^\dagger$  for  $\hat{c} = \hat{\Psi}^\dagger$  and

$$\hat{d} = \left( \sum_{\alpha}^{q=r,i} |\hat{a}_{\alpha}^q|^2 \right)^{1/2} \hat{\Psi} \left( \sum_{\beta}^{p=r,i} |\hat{b}_{\beta}^p|^2 \right)^{1/2}, \quad (\text{E12})$$

which completes the proof. It is impossible to generalize CFRD inequalities to more observables [23].

#### APPENDIX F: FOUR-PARTY CFRD INEQUALITIES

For a complex random variable  $Z$  we have a generalized triangle (in the complex plane) inequality  $|\langle Z \rangle| \leq \langle |Z| \rangle$ . Now, for complex random variables  $A, B, C, D$  we have

$$|\langle ABC \rangle|^2 \leq \langle |ABC|^2 \rangle \leq \langle |AB|^2 \rangle \langle |C|^2 \rangle \quad (\text{F1})$$

and

$$|\langle ABCD \rangle|^2 \leq \langle |ABCD|^2 \rangle \leq \langle |AB|^2 \rangle \langle |CD|^2 \rangle, \quad (\text{F2})$$

where we use the Cauchy-Bunyakovsky-Schwarz inequality in the last step. Complex variables can be constructed out of real ones,  $A = A_1 + iA_2$ , etc., where  $A_1$  and  $A_2$  are real. Both sides of the inequalities can be expanded in real variables, in such a way that no average contains simultaneously  $A_1$  and  $A_2$ . In particular,

$$\begin{aligned} \langle |AB|^2 \rangle &= \langle A_1^2 B_1^2 \rangle + \langle A_1^2 B_2^2 \rangle + \langle A_2^2 B_1^2 \rangle + \langle A_2^2 B_2^2 \rangle, \\ \langle |C|^2 \rangle &= \langle C_1^2 \rangle + \langle C_2^2 \rangle, \end{aligned} \quad (\text{F3})$$

while

$$\begin{aligned} |\langle ABCD \rangle|^2 &= \langle ABCD \rangle^* \langle ABCD \rangle \\ &= \langle \text{Re } ABCD \rangle^2 + \langle \text{Im } ABCD \rangle^2, \end{aligned} \quad (\text{F4})$$

where

$$\begin{aligned} \langle \text{Re } ABCD \rangle &= \langle A_1 B_1 C_1 D_1 \rangle - \langle A_1 B_1 C_2 D_2 \rangle \\ &\quad - \langle A_2 B_2 C_1 D_1 \rangle + \langle A_2 B_2 C_2 D_2 \rangle \\ &\quad - \langle A_1 B_2 C_1 D_2 \rangle - \langle A_1 B_2 C_2 D_1 \rangle \\ &\quad - \langle A_2 B_1 C_1 D_2 \rangle - \langle A_2 B_1 C_2 D_1 \rangle \end{aligned} \quad (\text{F5})$$

and

$$\begin{aligned} \langle \text{Im } ABCD \rangle &= \langle A_1 B_1 C_1 D_2 \rangle + \langle A_1 B_1 C_2 D_1 \rangle \\ &\quad + \langle A_1 B_2 C_1 D_1 \rangle + \langle A_2 B_1 C_1 D_1 \rangle \\ &\quad - \langle A_1 B_2 C_2 D_2 \rangle - \langle A_2 B_1 C_2 D_2 \rangle \\ &\quad - \langle A_2 B_2 C_1 D_2 \rangle - \langle A_2 B_2 C_2 D_1 \rangle. \end{aligned} \quad (\text{F6})$$

As quantum counterexamples, let us take spin observables  $\sigma_1 = |+\rangle\langle -| + |-\rangle\langle +|$  and  $\sigma_2 = i|-\rangle\langle +| - i|+\rangle\langle -|$ . Now  $A_1 = \sigma_1^A$  and  $A_2 = \sigma_2^A$  so that  $A = A_1 + iA_2 = \sigma_+^A = 2|+\rangle\langle -|$ , etc. Taking Greenberger-Horne-Zeilinger states

$$\begin{aligned} \sqrt{2}|\psi\rangle &= |+++\rangle + |--\rangle, \\ \sqrt{2}|\psi\rangle &= |++++\rangle + |----\rangle, \end{aligned}$$

we get on the left-hand side of (F1) 16 while the right-hand side is equal to 8, and on the left-hand side of (F2) 64 while the right-hand side is equal to 16. So in both cases they are violated.

We can test the inequalities also by position and momentum measurement. Let us take  $\sqrt{2}A = X_A + iP_A$  with  $[X_A, P_A] = i$  ( $\hbar = 1$ ) so  $A_1 = X_A/\sqrt{2}$ ,  $A_2 = P_A/\sqrt{2}$ , and  $[A, A^\dagger] = 1$ , and analogously for  $B, C$ , and  $D$ . In the Fock basis  $A|n\rangle_A = \sqrt{n}|n-1\rangle_A$  and so on. Now we take a generic entangled state

$$|\psi\rangle = \sum_{n \geq 0} z_n |nnnn\rangle$$

with real  $z_n$  (for simplicity) and check if (F2) holds. Note that  $\langle ABCD \rangle = \sum_n z_n^2 z_{n-1}$  while

$$\begin{aligned} \langle |AB|^2 \rangle &= \langle |CD|^2 \rangle = \langle (AA^\dagger + A^\dagger A)(BB^\dagger + B^\dagger B) \rangle / 4 \\ &= \sum_n z_n^2 (n+1/2)^2. \end{aligned} \quad (\text{F7})$$

In this case, if (F2) holds then also  $\langle ABCD \rangle \leq \langle |AB|^2 \rangle$  holds, which yields

$$\sum_n n^2 z_n z_{n-1} \leq \sum_n z_n^2 (n+1/2)^2. \quad (\text{F8})$$

This is equivalent to the positivity of the  $(N+1) \times (N+1)$  matrix  $M$  with entries  $M_{nn} = (n+1/2)^2$  for  $n = 0, 1, \dots, N$  and  $M_{n,n+1} = M_{n+1,n} = -(n+1)^2/2$  for  $n = 0, 1, \dots, N-1$ , and 0 otherwise. However, for  $N = 10$  we get  $2^{22} \det M = -21\,772\,303\,951\,061\,875$  so it must have a negative eigenvalue. The numerical check shows that the minimal eigenvalue of  $M$  is  $\lambda_{\min} = -0.002\,879\,31$  while the normalized coefficients  $z_n$  read  $(z_0, z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, z_9, z_{10}) = (0.828\,979, 0.419\,264, 0.265\,03, 0.181\,928, 0.129\,563, 0.093\,487\,9, 0.067\,152\,3, 0.047\,126\,4, 0.031\,430\,2, 0.018\,836\,4, 0.008\,542\,37)$  which violates (F2). Note that for larger  $N$  one can get a smaller  $\lambda_{\min}$ , e.g.  $-0.093$  for  $N = 3000$ .

Taking an analogous state

$$|\psi\rangle = \sum_n z_n |nnn\rangle, \quad (\text{F9})$$

unfortunately one cannot violate (F1) which reads in this case

$$\left( \sum_n z_n z_{n-1} n^{3/2} \right)^2 \leq \sum_n z_n^2 (n+1/2)^2 \sum_n z_n^2 (n+1/2). \quad (\text{F10})$$

One can see it from the Minkowski inequality

$$\left( \sum_n x_n y_n \right) \leq \sum_n x_n^2 \sum_n y_n^2, \quad (\text{F11})$$

taking  $x_n = z_{n-1} \sqrt{n-1/2}$  and  $y_n = \sqrt{n^3/(n-1/2)}$  for  $n = 1, 2, \dots$ . Note also that  $n^3/(n-1/2) \leq (n+1/2)^2$  because  $n^3 \leq (n-1/2)(n+1/2)^2 = (n^2 - 1/4)(n+1/2) = n^2 - n/4 + n^2/2 - 1/8$ , which is true due to the fact that  $n^2/2 - n/4 - 1/8 \geq 0$  for  $n \geq 1$ .

Interestingly, in the case of the tri- and quadripartite CFRD inequalities, the Lieb theorem, used in Appendix E for two parties, does not prevent the violation of a classical inequality, even in the fourth-moment version. However, the violating state in the position-momentum space is quite complicated, so an open question remains whether any simpler fourth-order inequality or simpler violating state exists.

- [1] J. S. Bell, *Physics* (Long Island City, NY) **1**, 195 (1964); J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969); A. Shimony, [plato.stanford.edu/entries/bell-theorem/](http://plato.stanford.edu/entries/bell-theorem/)
- [2] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, *Phys. Rev. Lett.* **88**, 040404 (2002).
- [3] J. H. Conway and S. Kochen, *Found. Phys.* **36**, 1441 (2006).
- [4] A. Aspect, J. Dalibard, and G. Roger, *Phys. Rev. Lett.* **49**, 1804 (1982); G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, *ibid.* **81**, 5039 (1998); W. Tittel, J. Brendel, H. Zbinden, and N. Gisin, *ibid.* **81**, 3563 (1998); M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, *Nature (London)* **409**, 791 (2001); D. N. Matsukevich, P. Maunz, D. L. Moehring, S. Olmschenk, and C. Monroe, *Phys. Rev. Lett.* **100**, 150404 (2008); M. Ansmann *et al.*, *Nature (London)* **461**, 504 (2009); M. Giustina *et al.*, *ibid.* **497**, 227 (2013); B. G. Christensen *et al.*, *Phys. Rev. Lett.* **111**, 130406 (2013).
- [5] P. M. Pearle, *Phys. Rev. D* **2**, 1418 (1970); A. Garg and N. D. Mermin, *ibid.* **35**, 3831 (1987); E. Santos, *Phys. Rev. A* **46**, 3646 (1992); M. Genovese, *Phys. Rep.* **413**, 319 (2005).
- [6] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
- [7] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, *Rev. Mod. Phys.* **74**, 145 (2002); A. Acin, N. Gisin, and L. Masanes, *Phys. Rev. Lett.* **97**, 120405 (2006).
- [8] S. Kochen and E. P. Specker, *J. Math. Mech.* **17**, 59 (1967).
- [9] Y.-F. Huang, C.-F. Li, Y.-S. Zhang, J.-W. Pan, and G.-C. Guo, *Phys. Rev. Lett.* **90**, 250401 (2003); Y. Hasegawa, R. Loidl, G. Badurek, M. Baron, and H. Rauch, *ibid.* **97**, 230401 (2006); A. Cabello, *ibid.* **104**, 220401 (2010).
- [10] A. Cabello, *Phys. Rev. Lett.* **101**, 210401 (2008); E. Amselem, M. Radmark, M. Bourennane, and A. Cabello, *ibid.* **103**, 160405 (2009); A. R. Plastino and A. Cabello, *Phys. Rev. A* **82**, 022114 (2010); M. Kleinmann, C. Budroni, J.-A. Larsson, O. Gühne, and A. Cabello, *Phys. Rev. Lett.* **109**, 250402 (2012).
- [11] E. G. Cavalcanti, C. J. Foster, M. D. Reid, and P. D. Drummond, *Phys. Rev. Lett.* **99**, 210405 (2007).
- [12] B. Reulet, J. Senzier, and D. E. Prober, *Phys. Rev. Lett.* **91**, 196601 (2003); G. Gershon, Y. Bomze, E. V. Sukhorukov, and M. Reznikov, *ibid.* **101**, 016803 (2008); J. Gabelli and B. Reulet, *J. Stat. Mech.* (2009) P01049.
- [13] A. Bednorz and W. Belzig, *Phys. Rev. B* **83**, 125304 (2011).
- [14] D. Loss and E. V. Sukhorukov, *Phys. Rev. Lett.* **84**, 1035 (2000); G. Burkard, D. Loss, and E. V. Sukhorukov, *Phys. Rev. B* **61**, R16303 (2000); S. Kawabata, *J. Phys. Soc. Jpn.* **70**, 1210 (2001); G. B. Lesovik, T. Martin, G. Blatter *et al.*, *Eur. Phys. J. B* **24**, 287 (2001).
- [15] C. W. J. Beenakker, in *Proceedings of the International School of Physics "Enrico Fermi," Quantum Computers, Algorithms and Chaos*, edited by G. Casati, D. L. Shepelyansky, P. Zoller, and G. Benenti, Vol. 162 (IOS Press, Amsterdam, 2006), pp. 307–347.
- [16] I. Neder, N. Ofek, Y. Chung, M. Heiblum, D. Mahalu, and V. Umansky, *Nature (London)* **448**, 333 (2007).
- [17] N. D. Mermin, *Phys. Rev. Lett.* **65**, 3373 (1990); A. Peres, *Phys. Lett. A* **151**, 107 (1990).
- [18] Q. Y. He, P. D. Drummond, and M. D. Reid, *Phys. Rev. A* **83**, 032120 (2011).
- [19] A. Salles, D. Cavalcanti, A. Acin, D. Pérez-García, and M. M. Wolf, *Quantum Inf. Comput.* **10**, 703 (2010).
- [20] N. D. Mermin, *Phys. Rev. Lett.* **65**, 1838 (1990); D. M. Greenberger, M. A. Horne, and A. Zeilinger, in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, edited by M. Kafatos (Kluwer Academic, Dordrecht, 1989), p. 69.
- [21] J. Uffink, *Phys. Rev. Lett.* **88**, 230406 (2002).
- [22] A. Salles, D. Cavalcanti, and A. Acin, *Phys. Rev. Lett.* **101**, 040404 (2008).
- [23] E. Shchukin and W. Vogel, *Phys. Rev. A* **78**, 032104 (2008).
- [24] B. S. Cirel'son, *Lett. Math. Phys.* **4**, 93 (1980).
- [25] Z. Zhao, Y.-A. Chen, A.-N. Zhang, T. Yang, H. J. Briegel, and J.-W. Pan, *Nature (London)* **430**, 54 (2004); P. P. Munhoz and F. L. Semião, *Eur. Phys. J. D* **59**, 509 (2010); T. Sh. Iskhakov, I. N. Agafonov, M. V. Chekhova, and G. Leuchs, *Phys. Rev. Lett.* **109**, 150502 (2012); D. Pagel, H. Fehske, J. Sperling, and W. Vogel, *Phys. Rev. A* **88**, 042310 (2013); M. Krenn, M. Huber, R. Fickler, R. Lapkiewicz, S. Ramelow, and A. Zeilinger, [arXiv:1306.0096](https://arxiv.org/abs/1306.0096).
- [26] C. W. Norman, *Undergraduate Algebra* (Oxford University Press, Oxford, 1986).
- [27] E. H. Lieb, *Adv. Math.* **11**, 267 (1973); M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000), Appendix 6; R. Bhatia, *Matrix Analysis* (Springer, New York, 1997).