Testing locality or noncontextuality with lowest moments

Adam Bednorz,[1,*] Witold Bednorz,[2,†] and Wolfgang Belzig[3]
[1]*Faculty of Physics, University of Warsaw, Hoża 69, PL-00681 Warsaw, Poland*
[2]*Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, ul. Banacha 2, 02-097 Warsaw, Poland*
[3]*Fachbereich Physik, Universität Konstanz, D-78457 Konstanz, Germany*
(Dated: August 21, 2013)

The quest for fundamental test of quantum mechanics is an ongoing effort. We are addressing the question of what are the lowest possible moments to prove quantum nonlocality and noncontextuality without any further assumption – 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 cannot violate classical inequalities as well 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 fully scalable continuous variable Bell inequality, which might be useful in Bell tests closing all loopholes simultaneously.

Introduction – Certain quantum correlations cannot be reproduced by any classical local hidden variable theory (LHV), 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) the dichotomy of the measurement outcomes or at least some restricted set of outcomes in some generalizations [2], (ii) the freedom of choice of the measured observables [3], and (iii) the time of the choice and measurement of the observable is shorter 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 assumes already 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 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]. In contrast, 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, noncontextuality may be violated for an arbitrary localized state 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 *et al.* (CFRD) [9] 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, C, and nonnegative integers $l, m, n-l-m$, where in general the higher n is, one has greater chances to violate the corresponding CFRD inequality. On a practical level, measuring higher moments or dichotomous variables is not a problem in ideal measurements of bounded variables (or for unbounded variable one can make binning). However, in the vast majority of experiments, especially in condensed matter [10], the interesting information is masked by large classical noise. This noise makes the usual binning unable to retrieve the underlying quantum statistics, which is accessible only by measuring moments and subsequent deconvolution.

In this Letter we ask which are the lowest possible moments to non-classicality 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 [11] – a simple classical construction of a probability reproducing all second order correlations. Note that the standard Bell inequalities [1] require dichotomy $A^2 = 1$, which requires to measure a fourth order correlator satisfying $\langle (A^2-1)^2 \rangle = 0$. Hence, the Bell inequalities are of at least fourth order – not second, as it may appear. The proposed Bell-type tests in condensed matter based on second order correlations [12–14] require an additional assumption of dichoto-

Noncontextuality	Yes	Yes	No
State independent	No	Yes	No
Maximal moments	LHV excluded?		
2nd	No	No	No
3rd	Yes	No	No
4th	Yes	Yes	Yes

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

mous interpretation of the measurement results, which is in general experimentally unverified and does not allow to identify entanglement 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 semi-definite 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 fully scalable Bell-type inequality. State-independent noncontextuality can be violated by a fourth-moment generalization of the Mermin-Peres square [15]. Our results for the gradual possibilities to exclude LHV models under different conditions are summarized in Table I.

Comparing to the previous research, note that the CFRD inequalities are the only known Bell-type inequalities scalable with $A \to \lambda A$, $B \to \mu B$ and so on for more observers. Unfortunately, the original example for a violation involved 20th order correlators and 10 observers [9], but was later reduced to 6th order and 3 observers [16, 17] for Greenberger-Horne-Zeilinger states [18]. On the other hand, the CFRD inequality with 4th moments cannot be violated at all, which has been shown for spins [19], quadratures [20], generalized to 8 settings and proved for separable states [21] and finally proved for all states [17] (we show an alternative proof in the Supplemental Material [26]).

Test of LHV – 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 their own detector ($\alpha = 1, 2, \dots$). For each setting, one can measure multiple real-valued observables (numbered $i = 1, 2, 3, \dots$) so that the measurement of $\hat{X}_{\alpha i}$ gives a real number $X_{\alpha i}$ The projection postulate [22] gives the quantum prediction for correlations, $\langle O_1 \cdots O_n \rangle = \mathrm{Tr}\hat{\rho}\hat{O}_1 \cdots \hat{O}_n$ for commuting observables $\hat{O}_k$. The observables measured by different observers and by one observer $\hat{X}_{\alpha i}$ for a given set-

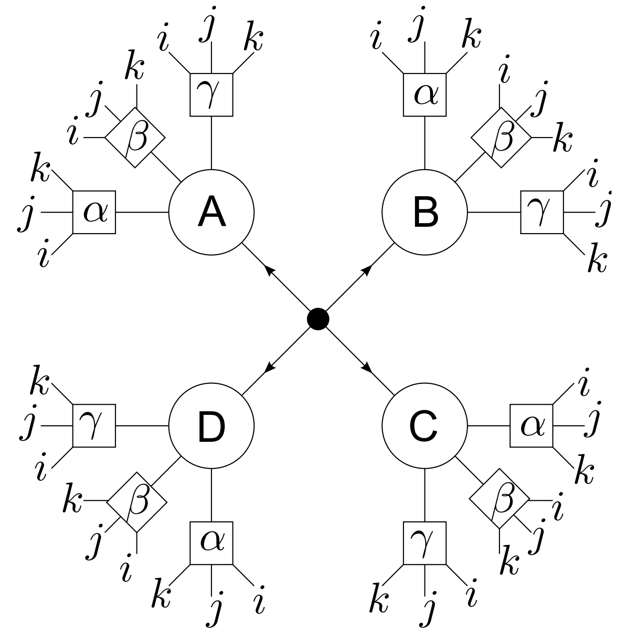

FIG. 1: The general test of local realism. Here we have four observers, Alice, Bob, Charlie and David. Everybody is free to choose between three different settings, α, β and γ and finally they can measure three real, continuous outcomes, e.g. $A_{\alpha i}$. The picture can be generalized to arbitrary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

ting 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 ρ, the LHV model is called *noncontextual*. A single observer suffices to test such LHV as noncontextuality is anyway an experimentally unverifiable assumption – the observer cannot measure simultaneously at two different settings. 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 ρ cannot be altered by the choice of the observable. Noncontextual and local LHVs are ruled out by tests with discrete outcomes [1, 8]. In moment-based tests only a finite number of cross correlations are compared with LHV. Our aim is to find the lowest moments showing nonclassical behavior of quantum correlations.

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 correlations functions can be always reproduced classically [11]. To see this, consider a real symmetric correlation matrix

$$\mathcal{C}_{ij} = \langle X_i X_j\rangle = \mathrm{Tr}\hat{\rho}\{\hat{X}_i,\hat{X}_j\}/2 \quad (1)$$

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 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 definition. However, (1) has much in common with *weak measurement*, for which even correlations of noncommuting observables are experimentally accessible [23], but all measurements are blurred with large noise in this case. Instead, here we use only standard projective detection scheme, where only measurements of commuting observables are possible [22]. Our construction includes all possible first-order averages $\langle X_i\rangle$ by setting one observable to identity or subtracting averages ($X_i \to X_i - \langle X_i\rangle$). Since $\mathrm{Tr}\hat{\rho}\hat{W}^2 \geq 0$ for $\hat{W} = \sum_i \lambda_i\hat{X}_i$ with arbitrary real λ_i, we find that the correlation matrix $\mathcal{C}$ is positive definite. Therefore every correlation can be simulated by a classical Gaussian distribution $\rho \propto \exp(-\sum_{ij}\mathcal{C}^{-1}{}_{ij}X_iX_j/2)$, 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 $X_i \to X_i - \langle X_i\rangle$.

There is an interesting connection between weak positivity and Cirelson's bound [24]. Taking observables A_1, A_2, B_1, 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

$$\begin{aligned}&\langle A_1B_1\rangle + \langle A_1B_2\rangle + \langle A_2B_1\rangle - \langle A_2B_2\rangle \leq\\ &(\langle A_1^2\rangle + \langle A_2^2\rangle + \langle B_1^2\rangle + \langle B_2^2\rangle)/\sqrt{2}.\end{aligned} \quad (3)$$

For $A, B = \pm 1$, the right hand side gives Cirelson'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.

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 (proof in the supplemental Material [26]). The problematic case is semi-positive 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 $\mathrm{Tr}\hat{W}^2 = 0$, which gives $\hat{W} = 0$. We can simply eliminate one of observables by 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 correlations matrix C_{ij} with $i,j \neq m$ is positive definite and the proof in the supplemental material holds [26]. 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 4 observables, nor in any two-dimensional Hilbert space, nor violate local realism (proofs in the supplemental material[26]). There exists, however, an example of violation of state-dependent noncontextuality with five observables in three-dimensional space [26].

Instead, here we show a simple example violating state-dependent noncontextuality, based on the Greenberger-Horne-Zeilinger (GHZ) idea [18]. We consider a three qubit Hilbert space with the 8 basis states are 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 α. 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 $\hat{\rho} = |GHZ\rangle\langle GHZ|$ for the GHZ state

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

Assuming noncontextuality, we have

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

which implies $A_\alpha = -B_\alpha$, so classically $\langle A_1A_2A_3\rangle = -\langle B_1B_2B_3\rangle$. However,

$$\begin{aligned}\langle A_1A_2A_3\rangle &= \mathrm{Tr}\hat{\rho}\hat{A}_1\hat{A}_2\hat{A}_3 = 1,\\ \langle B_1B_2B_3\rangle &= \mathrm{Tr}\hat{\rho}\hat{B}_1\hat{B}_2\hat{B}_3 = 1,\end{aligned} \quad (6)$$

in contradiction with the earlier statement and excluding noncontextual LHV. 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.

Fourth order correlations – noncontextuality – To find a test of noncontextuality we now consider fourth moments. Mermin and Peres [15] 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

$$\begin{array}{|c|c|c|c|}\hline \hat{M}_{ij} & j=1 & j=2 & j=3 \\ \hline i=1 & \hat{\sigma}_1^A & \hat{\sigma}_1^A\hat{\sigma}_1^B & \hat{\sigma}_1^B \\ \hline i=2 & -\hat{\sigma}_1^A\hat{\sigma}_3^B & \hat{\sigma}_2^A\hat{\sigma}_2^B & -\hat{\sigma}_3^A\hat{\sigma}_1^B \\ \hline i=3 & \hat{\sigma}_3^B & \hat{\sigma}_3^A\hat{\sigma}_3^B & \hat{\sigma}_3^A \\ \hline \end{array} \quad (7)$$

where the Pauli observables $\hat{\sigma}_i$ 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 variable M_{ij} then $C_1C_2C_3 = R_1R_2R_3$ in contradiction with the quantum result.

Now we assume that the M are not spin-1/2, but arbitrary operators, which can 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 dichotomy can be replaced by fourth order correlations, without other assumptions on 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 λ_i of $N^T N$ are real and positive. Using the Cauchy inequality we find that $\det(N^T N) = \lambda_1\lambda_2\lambda_3 \leq (\lambda_1 + \lambda_2 + \lambda_3)^3/27 = (\mathrm{Tr} N^T N)^3/27$. We get then

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

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

$$|\langle S\rangle| \leq \sum\nolimits_{ij} \left[\langle M_{ij}^2\rangle\langle M_{ij}^4\rangle/3\right]^{1/2} \quad (9)$$

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 violate are contextual by violating Eq. (9) or other fourth moment based inequalities.

Fourth order correlations – nonlocality – A simple fourth moment-based inequality testing local realism has been considered by CFRD [9]

$$\langle A_1B_1 - A_2B_2\rangle^2 + \langle A_1B_2 + A_2B_1\rangle^2 \leq \langle (A_1^2 + A_2^2)(B_1^2 + B_2^2)\rangle. \quad (10)$$

Note that all averages involve only simultaneously measurable quantities. This constitutes an inequality, which holds classically, involves only 4th order averages and is scalable with respect to A and B. Unfortunately, (10) and its generalizations [21] are not violated at all in quantum mechanics as shown in [17]. We present an alternative proof in [26].

We can ask, if it is possible at all to find a violable inequality involving only fourth order correlation function. The answer is positive, but unfortunately involves a much more complicated inequality, constructed in [11]

$$\begin{aligned} &2|\langle A_1B_1(A_1^2 + B_1^2)\rangle + \langle A_2B_1(A_2^2 + B_1^2)\rangle \\ &+\langle A_1B_2(A_1^2 + B_2^2)\rangle - \langle A_2B_2(A_2^2 + B_2^2)\rangle| \leq \\ &2\left(\langle A_1^4\rangle + \langle A_2^4\rangle + \langle B_1^4\rangle + \langle B_2^4\rangle\right) + \\ &\sum_{X,Y,Z=\{A_1,A_2,B_1,B_2\}}^{Y\neq X;Z\neq X,Y,Y'} (\langle X^4\rangle\langle Y^4\rangle)^{1/4}\langle (Y^2 - Z^2)^2\rangle^{1/2}, \end{aligned} \quad (11)$$

where $Y_i' = Y_{3-i}$. The inequality (11) is jointly invariant under scaling $A \to \lambda A$ and $B \to \lambda B$. For dichotomic outcomes $A_{1,2}^2 = B_{1,2}^2 = 1$ it also reduces to the Bell inequality [1] and can be violated in a standard way.

Note that (11) does not change if A and B register 0 *simultaneously* with large probability. It suits well experiments, where entangled particles are produced rarely but if they appear then both are detected. However, if often only A (or B) registers 0 then the inequality cannot be violated analogously to the detection loophole. We can estimate the detector efficiency using a simple example with outcomes 0 and ± 1. We denote with p_{AB} the coincidence detection probability and by $p_{0B}(p_{A0})$ the probability of Alice(Bob) missing and Bob(Alice) detecting an event. Assuming $p_{A0} = p_{0B}$ and a maximally entangled Bell states Eq. (11) becomes an inequality for the detector efficiency $p_{AB}/(p_{AB} + p_{A0}) \leq 2(2 - 1/2^{1/2})^{1/2} + 1/2^{1/2} - 2 \simeq 98\%$. Hence, a violation is possible only for a detector efficiency greater than $\simeq 98\%$ in contrast to $\simeq 83\%$ for the standard Bell inequality [25]. Nevertheless, the inequality (11) can be made scalable independently for A and B if we substitute $A_j \to \tilde{A}_j = A_j\langle A_j^2\rangle^{-1/2}$ (and analogously with B). Therefore we have shown that a scalable fourth-order Bell-type inequality exists, which might be useful to compensate asymmetries in the detection setups of Alice and Bob.

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 scalable fourth order Bell-type inequality (11) can be violated by reducing the problem to the standard Bell inequality for dichotomic outcomes. It remains an open question if one can find a simpler, preferably scalable fourth order Bell-type inequality, that can be violated in quantum mechanics.

We are grateful for discussions with N. Gisin and M. Reid. A. Bednorz acknowledges financial support by the Polish MNiSW grant IP2011 002371 W. Bednorz acknowledges partial financial support by the Polish MNiSW Grant no. N N201 397437. W. Belzig acknowledges financial support by the DFG via SPP 1285 and SFB 767.

* Electronic address: Adam.Bednorz@fuw.edu.pl
† Electronic address: wbednorz@mimuw.edu.pl

[1] J. S. Bell, Physics **1**, 195 (1964); J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[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] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger , Phys. Rev. Lett. **81**, 5039 (1998); D. N. Matsukevich, P. Maunz, D. L. Moehring, S. Olmschenk, and C. Monroe, Phys. Rev. Lett. **100**, 150404 (2008).
[5] P. M. Pearle, Phys. Rev. D **2**, 1418 (1970); A. Garg and N.D. Mermin, Phys. Rev. D **35**, 3831 (1987); E. Santos, Phys. Rev. A **46**, 3646 (1992).
[6] R. Horodecki, P. Horodecki, M. Horodecki, K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
[7] 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] E. G. Cavalcanti, C. J. Foster, M. D. Reid, and P. D. Drummond, Phys. Rev. Lett. **99**, 210405 (2007).
[10] B. Reulet, J. Senzier, and D. E. Prober, Phys. Rev. Lett. **91**, 196601 (2003); G. Gershon, Y. Bomze, E. V. Sukhorukov, and M. Reznikov **101**, 016803 (2008); J. Gabelli and B. Reulet, J. Stat. Mech. P01049 (2009).
[11] A. Bednorz and W. Belzig, Phys. Rev. B **83**, 125304 (2011).
[12] D. Loss, 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);
[13] C. W. J. Beenakker, Proc. Int. School Phys. E. Fermi, Vol. **162** (IOS Press, Amsterdam, 2006).
[14] I. Neder, N. Ofek, Y. Chung, M. Heiblum, D. Mahalu, and V. Umansky, Nature **448**, 333 (2007).
[15] N.D. Mermin, Phys. Rev. Lett. **65**, 3373 (1990); A. Peres, Phys. Lett. A151, 107 (1990).
[16] Q. Y. He, P. D. Drummond, and M. D. Reid, Phys. Rev. A **83**, 032120 (2011).
[17] A. Salles, D. Cavalcanti, A. Acin, D. Pérez-Garcia, M. M. Wolf, Quant. Inf. Comp. **10** (7-8) 703 (2010).
[18] N. D. Mermin, Phys. Rev. Lett. **65**, 1838 (1990); D. M. Greenberger, M. A. Horne, and A. Zeilinger, in Bells Theorem, Quantum Theory, and Conceptions of the Universe, edited by M. Kafatos (Kluwer Academic, Dordrecht, Holland, 1989), p. 69.
[19] J. Uffink, Phys. Rev. Lett. **88**, 230406 (2002).
[20] A. Salles, D. Cavalcanti, and A. Acin, Phys. Rev. Lett. **101**, 040404 (2008).
[21] E. Shchukin and W. Vogel, Phys. Rev. A **78**, 032104 (2008).
[22] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1932)
[23] A. Bednorz and W. Belzig, Phys. Rev. Lett. **101**, 206803 (2008); **105**, 106803 (2010); Phys. Rev. B **81**, 125112 (2010).
[24] B. S. Cirel'son, Lett. Math. Phys. **4**, 93 (1980).
[25] A. Garg and N.D. Mermin, Phys. Rev. D **35**, 3831 (1987).
[26] See Supplemental Material for additional proofs and technical details.

Supplemental Material

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 correlations can explained also by a positive probability. We also shift all first order averages to zero, $X_i \to X_i - \langle X_i \rangle$. Now, let us define additional discrete events $\{ijkq\}$, $i \neq j \neq k \neq i$ (one such event corresponds to all possible permutations of ijk), $\{ijq\pm\}$, $i \neq j$, $\{ijq\pm\} \neq \{jiq\pm\}$ (here order matters), and $\{iq\}$ with an auxiliary parameter $q \in \{3, -1, -2\}$. Now, suppose that we can measure $\langle X_i X_j X_k \rangle$ (the argument below holds even for noncommuting observables). We assign probability $p(\{ijkq\}) = p(\{ijq\pm\}) = p(\{iq\}) = 1/\lambda^3 > 0$ and the values

$$
\begin{aligned}
&X_{i,j,k}(\{ijkq\}) = q\lambda \langle X_i X_j X_k \rangle^{1/3}/\sqrt[3]{18}, \\
&X_i(\{ijq\pm\}) = \pm\sqrt{2} q Q_{ij}/\sqrt[3]{18},\ X_j(\{ijq\pm\}) = q Q_{ij}/\sqrt[3]{18}, \\
&Q_{ij} = \frac{\lambda}{\sqrt[3]{4}} \left[\langle X_i^2 X_j \rangle - \sum_{k \neq ij} \langle X_i X_j X_k \rangle \right]^{1/3}, \\
&X_i(\{iq\}) = \frac{q\lambda}{\sqrt[3]{18}} \left[\langle X_i^3 \rangle - \sum_{j \neq i} \langle X_j^2 X_i \rangle / 2 \right]^{1/3} \\
&X_l(\{ijkq\}) = X_l(\{ijq\pm\}) = X_l(\{iq\}) = 0,\ l \neq ijk.
\end{aligned}
\tag{A.1}
$$

The cubic root is real-defined for real negative arguments. Note that the special choice of q results in unchanged averages as $3 - 1 - 2 = 0$ but nonzero third order averages as $3^3 - 1^3 - 2^3 = 18$. The remaining Gaussian distribution is rescaled by $1 - c/\lambda^3$, where c is a number of all added events, to restore normalization. Unfortunately, it will modify the correlation matrix $\mathcal{C}$ into a new matrix $\mathcal{C}'$. However, for sufficiently large λ, $\mathcal{C}'$ is arbitrarily close to $\mathcal{C}$, so it must be positive definite and we can find the new Gaussian part in the form $\rho \propto \exp(-\epsilon \sum_{ij} \mathcal{C}'^{-1}{}_{ij} X_i X_j/2)$ where ϵ takes into account that ρ is not normalized to 1 (because of the remaining discrete events). The modified Gaussian part will give the correlation matrix $\mathcal{C}'$, which, by adding the discrete events, turns back into $\mathcal{C}$.

The assignment (A.1) 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 positive signature. If some eigenvalues of $\mathcal{C}$ are 0 then $\mathcal{C}'$ may have a negative eigenvalue for arbitrary λ and we cannot find any Gaussian distribution.

B. NONCONTEXTUALITY IN SIMPLE CASES

Let us examine state-dependent noncontextuality with up to 4 observables. We look for a positive probability $\rho(\{A_i\})$. We have the freedom to set values of correlations of noncommuting products. It is obvious for a single observable, $\rho(A) = \mathrm{Tr}\delta(A - \hat{A})\hat{\rho}$. For two observables $\hat{A}$ and $\hat{B}$, if they commute, then $\rho(A, B) = \mathrm{Tr}\delta(A - \hat{A})\delta(B - \hat{B})\hat{\rho}$. If they do not commute then $\rho(A, B) = \rho(A)\rho(B)$. For three observables $\hat{A}$, $\hat{B}$, $\hat{C}$, if they all commute then $\rho(A, B, C) = \mathrm{Tr}\delta(A - \hat{A})\delta(B - \hat{B})\delta(C - \hat{C})\hat{\rho}$. If they all do not commute, then $\rho(A, B, C) = \rho(A)\rho(B)\rho(C)$. If $\hat{A}\hat{B} = \hat{B}\hat{A}$ and $\hat{B}\hat{C} = \hat{C}\hat{B}$ but $\hat{A}\hat{C} \neq \hat{C}\hat{A}$ then $\rho(A, B, C) = \rho(A, B)\rho(B, C)/\rho(B)$. If $\hat{A}\hat{B} \neq \hat{B}\hat{A}$ and $\hat{B}\hat{C} \neq \hat{C}\hat{B}$ but $\hat{A}\hat{C} = \hat{C}\hat{A}$ then $\rho(A, B, C) = \rho(C, A)\rho(B)$. The case of four observables $\hat{A}$, $\hat{B}$, $\hat{C}$, $\hat{D}$ is a bit longer. If all observables commute or all do not commute then we solve it analogously to the earlier cases. If one observable, say $\hat{D}$, commutes with all three other then $\rho(A, B, C, D) = \rho_D(A, B, C)$ with ρ_D defined as for the three observables but with $\hat{\rho}_D = \hat{\rho}\delta(D - \hat{D})$. If $\hat{D}$ does not commute with all three other then $\rho(A, B, C, D) = \rho(D)\rho(A, B, C)$. If the only commuting pairs are $(\hat{A}, \hat{B})$, $(\hat{B}, \hat{C})$, $(\hat{C}, \hat{D})$ then $\rho(A, B, C, D) = \rho(A, B)\rho(B, C)\rho(C, D)/\rho(B)\rho(C)$. If the only commuting pairs are $(\hat{A}, \hat{B})$ and $(\hat{C}, \hat{D})$ then $\rho(A, B, C, D) = \rho(A, B)\rho(C, D)$. The only remaining case is with noncommuting pairs $(\hat{A}, \hat{B})$ and $(\hat{C}, \hat{D})$ but this is equivalent to the test of local realism with noncommuting $(\hat{A}_1, \hat{A}_2)$

and $(\hat{B}_1, \hat{B}_2)$. We will show in the general proof that this case can be always (if we do not use fourth moments) explained by a LHV model in Section C.

In two-dimensional Hilbert space, observables have the structure $\hat{A} = a_0\hat{1} + \vec{a}\cdot\hat{\vec{\sigma}}$, where $\hat{\vec{\sigma}} = (\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3)$ 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 parallel to the same direction, so that $\vec{a}_\alpha \parallel \vec{a}$, $\vec{b}_\beta \parallel \vec{b}$, $\vec{c}_\gamma \parallel \vec{c}$, etc., where $\vec{a} \nparallel \vec{b}, \vec{c}, \ldots$, $\vec{b} \nparallel \vec{c}, \ldots$, etc. Then we can construct a LHV model defined by $\rho(\{A_\alpha\}, \{B_\beta\}, \{C_\gamma\}, \ldots) = \rho(\{A_\alpha\})\rho(\{B_\beta\})\rho(\{C_\gamma\})\cdots$, where $\rho(\{A_\alpha\}) = \mathrm{Tr}\hat{\rho}\prod_j \delta(A_\alpha - \hat{A}_\alpha)$, etc.

C. THIRD MOMENTS – CONTEXTUAL LHV

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 Section II, we denote $\mathcal{C}_{X\alpha j, Y\beta k} = \langle X_{\alpha j} Y_{\beta k}\rangle$ for $X, Y = A, B, C, \ldots$ and $\alpha, \beta, j, k = 1, 2, \ldots$. For a valid LHV, $\mathcal{C}$ must be positive (semi)definite. The proof is based on two facts

- $\mathcal{C}_{X\alpha j, X\beta k} = \langle X_{\alpha j} X_{\beta k}\rangle$ is not measurable for $\alpha \neq \beta$ (even if accidentally $\hat{X}_{\alpha j}\hat{X}_{\beta k} = \hat{X}_{\beta k}\hat{X}_{\alpha j}$) so it is a free parameter in a LHV model.
- we can always redefine every observable within one observer's setting by real linear transform $\hat{X}_{\alpha m} \to \sum_k \lambda_{\alpha k}\hat{X}_{\alpha k}$ as long as the 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 [1].

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 *measurable* correlations. Suppose that the correlation matrix $\mathcal{C}$ has $\mathcal{N}$ zero eigenvalues with linearly independent eigenvectors

$$W_m = \sum_{\alpha,k}^{X=A,B,\ldots} \lambda^m_{X\alpha k} X_{\alpha k},\ m = 1..\mathcal{N} \tag{C.1}$$

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

$$W_m = \hat{W}_m\hat{\rho} = 0,\ m = 1..\mathcal{N}. \tag{C.2}$$

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

$$X_{\alpha k} + \sum_{\beta j}^{Y\neq X} \lambda^{X\alpha k}_{Y\beta j} Y_{\beta j} = 0 \tag{C.3}$$

plus equations not containing X. If this were not possible then we could reduce the kernel by at least one vector by modifying nonmeasurable correlations in the correlation matrix, keeping its positivity. By such successive reduction we end up with (C.3). Without loss of generality let us take $X = A$. We write (C.2) in the form

$$\sum_{\alpha k} \lambda^m_{\alpha k} A_{\alpha k} + \not{A} = 0. \tag{C.4}$$

where $\not{A}$ replaces all linear combinations of quantities measured by the other observers (B, C, D, ...). By linear eliminations and transformations within setting 1, there exists a form of (C.4) consisting of

$$A_{1k} + \not{1} + \not{A} = 0,\ k = 1, 2, \ldots, \tag{C.5}$$

with $\not{1}$ not containing A_{1j} terms, and other equations that do not contain A_{1j} at all. Suppose that at least one of (C.5) contains an A_{2j} term, so in general (C.5) has the form

$$A_{1k} + \sum_m \lambda_{km} A_{2m} + \not{1}\not{2} + \not{A} = 0,\ k = 1, 2, \ldots \tag{C.6}$$

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

$$\begin{aligned} &A_{1k} + A_{2k} + \not{1}\not{2} + \not{A} = 0,\ k = 1, 2, \ldots, l \\ &A_{1k} + \not{1}\not{2} + \not{A} = 0,\ k = l+1, l+2, \ldots, \\ &A_{2k} + \not{1}\not{2} + \not{A} = 0,\ k = l+1, l+2, \ldots, \end{aligned} \tag{C.7}$$

and other equations that do not contain A_{1j} nor A_{2j} at all. If $l > 0$ then we change $\langle A_{11}A_{21}\rangle \to \langle A_{11}A_{21}\rangle + \epsilon$ with $\epsilon > 0$ in the correlation matrix $\mathcal{C}$. Then $\langle W^2\rangle = 2\epsilon > 0$, where W is the left hand side of the first line in (C.7) 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 ϵ the new correlation matrix $\mathcal{C}$ will be strictly positive for in the space spun by the old non-kernel 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 (C.7). Once we are left with only two last lines of (C.7) we proceed by induction.

Let us assume that, at some stage with a fixed α and l, the kernel equations have the form

$$A_{\xi k} + \sum_{m \leq l} \lambda^{\xi}_{km} A_{\alpha m} + \not{1} \cdots \not{\alpha} + \not{A} = 0 \tag{C.8}$$

for all $\xi < \alpha$ plus other equations not containing A_ξ and $A_{\alpha m}$ with $m \leq l$. Note that the set of possible k can be different for different ξ. If all $\lambda = 0$ then we can proceed to the next induction step, taking next setting. Otherwise, let us denote by Ξ the set of all ξ with $\lambda^{\xi}_{k1} \neq 0$ 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 $\lambda^{\xi}_{k1} = \delta_{k1}$. Now, we make a shift of nonmeasurable correlations $\langle A_{\xi 1}A_{\alpha 1}\rangle \to \langle A_{\xi 1}A_{\alpha 1}\rangle + \epsilon$ and $\langle A_{\xi 1}A_{\eta 1}\rangle \to \langle A_{\xi 1}A_{\eta 1}\rangle - 2\epsilon$ for $\xi, \eta \in \Xi$ with $\epsilon > 0$. Denoting by W_ξ, $\xi \in \Xi$, subsequent left hand sides of (C.8) 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 do not contain $A_{\xi 1}$ nor $A_{\alpha 1}$. For sufficiently small ϵ (every new ϵ is much smaller than all previous ones), the correlation matrix $\mathcal{C}$ on old non-kernel vectors plus W_ξ is strictly positive, similarly as in (C.7). Hence, we kick W_ξ out of the kernel. 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 (C.3).

Let us summarize what we have done. The original correlation matrix (1) may lead us into troubles (violation of noncontextuality). 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 to measure everything in one run of the experiment, leaving more freedom for contextual LHV models.

Now, we *define* all third order correlations, including noncommuting observables. We divide all observables into two families: V_j – appearing in (C.3) and Y_m – the rest. Now,

$$\begin{aligned} &\langle Y_m Y_n Y_p\rangle = \sum_{\sigma(mnp)} \mathrm{Tr}\hat\rho \hat Y_m \hat Y_n \hat Y_p/6, \\ &\langle V_j Y_m Y_n\rangle = \mathrm{Tr}\hat\rho\{\hat V_j, \{\hat Y_m, \hat Y_n\}\}/4, \\ &\langle V_k V_l Y_n\rangle = \mathrm{Tr}\hat\rho(\hat V_j \hat Y_n \hat V_k + \hat V_k \hat Y_n \hat V_j)/2, \\ &\langle V_j V_k V_l\rangle = \sum_{\sigma(jmn)} \mathrm{Tr}\hat\rho \hat V_j \hat V_k \hat V_l/6, \end{aligned} \tag{C.9}$$

where σ denotes all 6 permutations. The above definition is consistent with projective measurement 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 (C.3) and $Z, Z' = V_j, Y_m$. If $Z, Z' = Y_m, Y_n$ it is clear because

$$\hat X \hat\rho = 0. \tag{C.10}$$

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

$$2\langle W Y_m V_j\rangle = \mathrm{Tr}\hat\rho(\hat W \hat Y_m \hat V_j + \hat V_j \hat Y_m \hat W) = 0 \tag{C.11}$$

again because of (C.10). Finally, we need to consider $Z = V_j$, $Z' = V_k$. Because of (C.10), we get

$$6\langle W V_j V_k\rangle = \mathrm{Tr}\hat\rho(\hat V_j \hat W \hat V_k + \hat V_k \hat W \hat V_j) \tag{C.12}$$

Without loss of generality we only need to consider 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 (C.12) vanishes due (C.10). Now W contains A_m. By virtue of (C.3) we can write

$$W = A_m + \sum_n \lambda_n B_n + \not{A}\not{B}, \tag{C.13}$$

where $\not{A}\not{B}$ denotes all terms not containing A and B. Moving A_m and $\sum_n \lambda_n B_n + \not{A}\not{B}$ in opposite direction in (C.12), it can be transformed into

$$\begin{aligned}&\mathrm{Tr}\hat{\rho}(\hat{A}_j\hat{W}\hat{B}_k + \hat{B}_k\hat{W}\hat{A}_j) = \mathrm{Tr}\hat{\rho}(\hat{A}_j\hat{B}_k\hat{A}_m + \hat{A}_m\hat{B}_k\hat{A}_j)\\ &+\mathrm{Tr}\hat{\rho}\left(\left(\sum_n \lambda_n\hat{B}_n + \hat{\not{A}\not{B}}\right)\hat{A}_j\hat{B}_k + \hat{B}_k\hat{A}_j\left(\sum_n \lambda_n\hat{B}_n + \hat{\not{A}\not{B}}\right)\right) = \mathrm{Tr}\hat{\rho}(\hat{A}_j\hat{B}_k\hat{W} + \hat{W}\hat{B}_k\hat{A}_j),\end{aligned}$$

where we used commutation $\hat{A}_j\hat{B}_k = \hat{B}_k\hat{A}_j$. The last expression vanishes due to (C.10). If X 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 (C.12) vanishes due to (C.10). The remaining cases, due to (C.3), have the form $W = A_m + \not{A}$ and (C.12) reads

$$\mathrm{Tr}\hat{\rho}(\hat{A}_j\hat{W}\hat{A}_k + \hat{A}_k\hat{W}\hat{A}_j) = \mathrm{Tr}\hat{\rho}(\hat{A}_j\hat{A}_m\hat{A}_k + \hat{A}_k\hat{A}_m\hat{A}_j) + \mathrm{Tr}\hat{\rho}(\hat{\not{A}}\hat{A}_j\hat{A}_k + \hat{A}_k\hat{A}_j\hat{\not{A}}). \tag{C.14}$$

Now we remember that (C.3) must contain also $W' = A_k - \not{A}'$ so $\hat{A}_k\hat{\rho} = \hat{\not{A}}'\hat{\rho}$ which gives

$$\begin{aligned}&\mathrm{Tr}\hat{\rho}(\hat{A}_j\hat{A}_m\hat{A}_k + \hat{A}_k\hat{A}_m\hat{A}_j) = \mathrm{Tr}\hat{\rho}(\hat{A}_j\hat{A}_m\hat{\not{A}}' + \hat{\not{A}}'\hat{A}_m\hat{A}_j)\\ &= \mathrm{Tr}\hat{\rho}(\hat{\not{A}}'\hat{A}_j\hat{A}_m + \hat{A}_m\hat{A}_j\hat{\not{A}}') = \mathrm{Tr}\hat{\rho}(\hat{A}_k\hat{A}_j\hat{A}_m + \hat{A}_m\hat{A}_j\hat{A}_k),\end{aligned} \tag{C.15}$$

so (C.14) reads $\mathrm{Tr}\hat{\rho}(\hat{W}\hat{A}_j\hat{A}_k + \hat{A}_k\hat{A}_j\hat{W})$ which vanishes due to (C.10). We see that correlations containing arbitrary combinations of left hand sides of (C.3) vanish. Now, we can simply eliminate one observable from each kernel equation (C.3), $\sum_k \lambda_k Z_k = 0$, by substitution $Z_m = -\sum_{k\neq m}\lambda_k Z_k/\lambda_m$ so that only Z_k, $k = 1..l$ remain as independent observables. Now, the correlation matrix $\mathcal{C}$ is strictly positive (kernel is null) and we construct the final LHV model reproducing all measurable quantum first, second and third order correlations as in Section A. The third order correlations involving substituted observables are reproduced by virtue of the just-shown property of (C.9). This completes the proof.

D.VIOLATION OF STATE-DEPENDENT NONCONTEXTUAL LHV WITH THIRD MOMENTS

There exists a third-moment-based state-dependent example violating noncontextuality with 5 observables in three-dimensional Hilbert space. 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 μ. 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 5 commuting pairs and 5 noncommuting pairs. Suppose that an experimentalist measures

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

Let us denote Fourier operators $\hat{A}(q) = \sum_\alpha \hat{A}_\alpha e^{2\pi i\alpha q/5}$. Since $\hat{A}_\alpha = \hat{A}_\alpha^\dagger$, we have $\hat{A}(0) = \hat{A}^\dagger(0)$, $\hat{A}(-1) = \hat{A}(4) = \hat{A}^\dagger(1)$, $\hat{A}(-2) = \hat{A}(3) = \hat{A}^\dagger(2)$. Similarly, for outcomes $A(0) = A^*(0)$, $A(-1) = A(4) = A^*(1)$ and $A(-2) = A(3) = A^*(2)$ (there are either 5 real random variables or 1 real and 2 complex). We can write (D.1) in the equivalent form

$$\mathcal{S} = \langle|A(2)|^2\rangle + \langle(A(0))^2\rangle\cos(\pi/5). \tag{D.2}$$

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

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

Each term of the expansion of the right hand side must contain $A(\pm 2)$ or $A(0)$ because $\pm 1 \pm 1 \pm 1 \not\equiv 0$ so $\mathcal{S} = 0$ implies $\mathcal{Q} = 0$.

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

$$0 = 5\sum_\alpha [\hat{A}_\alpha, \hat{A}_{\alpha+2}]e^{2\pi i\alpha q/5} = \sum_p [\hat{A}(q-p), \hat{A}(p)]e^{-4\pi ip/5} = \sum_p [\hat{A}(p+q), \hat{A}^\dagger(p)]e^{4\pi ip/5} \tag{D.4}$$

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

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

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

$$\hat{A}(0) = a\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ \hat{A}(2) = b\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & i \\ 0 & i & -1 \end{pmatrix}, \hat{A}(1) = c\begin{pmatrix} 0 & 1 & i \\ 1 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \tag{D.6}$$

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

$$\hat{B} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix},\ \hat{C} = \begin{pmatrix} 0 & 1 & i \\ -1 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \tag{D.7}$$

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

Assuming noncontextuality, the quantum mechanical expectation for (D.1) reads,

$$\begin{aligned} \mathcal{S} &= \sum_\alpha \mathrm{Tr}\hat{\rho}\hat{A}_\alpha^2(1 + \cos(\pi/5)) + \sum_\alpha 2\mathrm{Tr}\hat{\rho}\hat{A}_\alpha\hat{A}_{\alpha+2}(\cos(\pi/5) + \cos(2\pi/5)) \\ &= \mathrm{Tr}\hat{\rho}(\hat{A}^\dagger(2)\hat{A}(2) + \hat{A}(2)\hat{A}^\dagger(2) + 2\hat{A}^2(0)\cos(\pi/5))/2 \end{aligned} \tag{D.8}$$

and for (D.3),

$$\mathcal{Q} = 25\sum_\alpha \mathrm{Tr}\hat{\rho}\hat{A}_\alpha^3 = \sum_{q,p,r}^{q+p+r\equiv 0} \mathrm{Tr}\hat{\rho}\hat{A}(q)\hat{A}(p)\hat{A}(r) \tag{D.9}$$

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,

$$\begin{aligned} \mathcal{Q} &= \langle 0|\hat{A}(1)\hat{A}(0)\hat{A}^\dagger(1)|0\rangle + \langle 0|\hat{A}^\dagger(1)\hat{A}(0)\hat{A}(1)|0\rangle + \langle 0|\hat{A}(1)\hat{A}^\dagger(2)\hat{A}(1)|0\rangle + \langle 0|\hat{A}^\dagger(1)\hat{A}(2)\hat{A}^\dagger(1)|0\rangle \\ &= 4a|c|^2 + 8\mathrm{Re}(b^*c^2) = 8(\sqrt{5} - 1) \simeq 9.9, \end{aligned} \tag{D.10}$$

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

E. NO-GO THEOREM ON CFRD INEQUALITIES

Simple fourth order CFRD-type inequalities can be constructed for two observers A and B, with up to 8 settings (and a single real outcome for each setting) [2, 3], $A^r_{0,1,2,3}$, $A^i_{0,1,2,3}$, $B^r_{0,1,2,3}$, $B^i_{0,1,2,3}$, and read

$$\begin{aligned} &|\langle A_0B_0^\dagger + A_1B_1^\dagger + A_2B_2^\dagger + A_3B_3^\dagger\rangle|^2 + |\langle A_0B_1 - A_1B_0 + A_2^\dagger B_3^\dagger - A_3^\dagger B_2^\dagger\rangle|^2 + \\ &|\langle A_0B_2 - A_2B_0 + A_3^\dagger B_1^\dagger - A_1^\dagger B_3^\dagger\rangle|^2 + |\langle A_0B_3 - A_3B_0 + A_1^\dagger B_2^\dagger - A_2^\dagger B_1^\dagger\rangle|^2 \le \\ &\sum_{\alpha\beta}\langle (A_\alpha^\dagger A_\alpha + A_\alpha^\dagger A_\alpha)(B_\beta^\dagger B_\beta + B_\beta^\dagger B_\beta)\rangle/4, \end{aligned} \tag{E.1}$$

where we have denoted $X = X^r + iX^i$, $C = A_\alpha, B_\alpha$. The notation is the same in classical and quantum case except $\hat{}$ and $^\dagger \to {}^*$. We use a complex form only to save space but all the inequality can be expanded into purely real terms [3]. The inequality reduces to (10) if we leave only A_1^r, A_2^r, B_1^r, B_2^r, while other observables are zero. Classically, (E.1) follows from inequality $|\langle z\rangle|^2 \le \langle |z|^2\rangle$ 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 [4]. Below we present an alternative proof.

It suffices to prove (E.1) for pure states, $\hat\rho = |\psi\rangle\langle\psi|$. For mixed states $\hat\rho = \sum_k p_k|\psi\rangle\langle\psi|$, $p_k \ge 0$, $\sum_k p_k = 1$. we apply triangle inequality $|\sum_k p_k z_k| \le \sum_k p_k |z_k|$ and Jensen inequality $(\sum_k p_k|z_k|)^2 \le \sum_k p_k |z_k|^2$, where z_k is the complex correlator in each of the four terms on the left hand side of (E.1) taken for a pure state $|\psi_k\rangle$. If (E.1) 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 (E.1) can be written as

$$\sum_{\alpha\beta}(\langle A_\alpha B_\beta\rangle\langle A^\dagger_\alpha B^\dagger_\beta\rangle - \langle A_\alpha B_\beta\rangle\langle A^\dagger_\beta B^\dagger_\alpha\rangle) + \sum_{\alpha\beta\gamma\delta}\epsilon_{\alpha\beta\gamma\delta}(\langle A_\alpha B_\beta\rangle\langle A_\gamma B_\delta\rangle + \langle A^\dagger_\alpha B^\dagger_\beta\rangle\langle A^\dagger_\gamma B^\dagger_\delta\rangle)/2, \tag{E.2}$$

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

Let us number the four complex correlators inside moduli on the left hand side of (E.1) by $0, 1, 2, 3$, respectively (e.g. 0 is the correlator $\sum_\alpha\langle A_\alpha B^\dagger_\alpha\rangle$). We want to transform (E.1) to a form with a signle real correlator 0 while $1, 2, 3$ vanish. Let us begin with a transformation $C_\alpha \to e^{i\phi_\alpha}C_\alpha$, $C = A, B$, with $\sum_\alpha \phi_\alpha = 0$. Note that $A_0B_1 - A_1B_0 + A^\dagger_2 B^\dagger_3 - A^\dagger_3 B^\dagger_2$ takes just the phase factor $e^{i(\phi_0+\phi_1)}$, so tuning ϕ_α we can always make correlators $1, 2, 3$ real. Making now real rotation in 123 space we can leave only real correlator 3 while 1 and 2 vanish. Still, correlator 0 can have also an unwanted imaginary component, because 0 is invariant under SU(4) transformations. To get rid of it, we have to apply a different transformation $A_0 \to A_0$, $A_1 \to A_1$, $A_2 \to A^\dagger_2$, $A_3 \to A^\dagger_3$, $B_0 \to -B^\dagger_1$, $B_1 \to B^\dagger_0$, $B_2 \to -B_3$, $B_3 \to B_2$, which gives

$$\begin{aligned}
&A_0B_1 - A_1B_0 + A^\dagger_2B^\dagger_3 - A^\dagger_3B^\dagger_2 \to A_0B^\dagger_0 + A_1B^\dagger_1 + A_2B^\dagger_2 + A_3B^\dagger_3,\\
&A_0B_2 - A_2B_0 + A^\dagger_3B^\dagger_1 - A^\dagger_1B^\dagger_3 \to -A_0B_3 + A_3B_0 - A^\dagger_1B^\dagger_2 + A_2B^\dagger_1\\
&A_0B_3 - A_3B_0 + A^\dagger_1B^\dagger_2 - A^\dagger_2B^\dagger_1 \to A_0B_2 - A_2B_0 + A^\dagger_3B^\dagger_1 - A^\dagger_1B^\dagger_3,\\
&A_0B^\dagger_0 + A_1B^\dagger_1 + A_2B^\dagger_2 + A_3B^\dagger_3 \to -A_0B_1 + A_1B_0 - A^\dagger_2B^\dagger_3 + A^\dagger_3B^\dagger_2.
\end{aligned} \tag{E.3}$$

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

$$\mathrm{Re}^2\sum_\alpha\langle A_\alpha B^\dagger_\alpha\rangle \tag{E.4}$$

We apply the triangle inequality

$$\left|\sum_\alpha^{q=r,i}\langle A^q_\alpha B^q_\alpha\rangle\right| \le \sum_\alpha^{q=r,i}|\langle A^q_\alpha B^q_\alpha\rangle|. \tag{E.5}$$

Note that $|\langle A^q_\alpha B^q_\alpha\rangle| \le \langle |A^q_\alpha||B^q_\alpha|\rangle$ where $|X|$ is obtained by reversing signs of all negative eigenvalues of X (in the eigenbasis). To prove (10) we have to show that

$$\left(\sum_\alpha^{q=r,i}\langle |A^q_\alpha||B^q_\alpha|\rangle\right)^2 \le \sum_{\alpha\beta}^{q,p=r,i}\langle |A^q_\alpha|^2|B^p_\beta|^2\rangle \tag{E.6}$$

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

$$|\psi\rangle = \sum_{ki}\psi_{ki}|k_Ai_B\rangle,\ \hat A^x = \sum_{kli}A^x_{kl}|k_Ai_B\rangle\langle l_Ai_B|,\ \hat B^x = \sum_{kij}B^x_{ij}|k_Ai_B\rangle\langle k_Aj_B|. \tag{E.7}$$

The normalization reads $\sum_{ki}|\psi_{ki}|^2=1$. Let us define $\hat\Psi=\sum_{ki}\psi_{ki}|k\rangle\langle i|$, $\hat a^x=\sum_{kl}A_{kl}|k\rangle\langle l|$, $\hat b^x=\sum_{ij}B_{ij}|j\rangle\langle i|$. Now the normalization reads $\mathrm{tr}\,\hat\Psi^\dagger\hat\Psi=1$. One can check the identity $\langle\psi|\hat A^x\hat B^x|\psi\rangle=\mathrm{tr}\,\hat\Psi^\dagger\hat a^x\hat\Psi\hat b^x$. 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 nor even a square matrix. 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 Schmidt decomposition, but it is not necessary. Now (E.6) reads

$$\left(\sum_\alpha^{q=r,i}\mathrm{tr}\,\hat\Psi^\dagger|\hat a^q_\alpha|\hat\Psi|\hat b^q_\alpha|\right)^2\leq\sum_{\alpha\beta}^{q,p=r,i}\mathrm{tr}\,\hat\Psi^\dagger|\hat a^q_\alpha|^2\hat\Psi|\hat b^p_\beta|^2 \tag{E.8}$$

To prove (E.8) we need Lieb concavity theorem [5] which states that for a fixed but arbitrary $\hat\Psi$ and $s\in[0,1]$ the trace class function $f(\hat F,\hat G)=\mathrm{tr}\,\hat\Psi^\dagger\hat F^s\hat\Psi\hat G^{1-s}$ is *jointly concave*, which means that

$$\lambda f(\hat F,\hat G)+(1-\lambda)f(\hat F',\hat G')\leq f(\lambda\hat F+(1-\lambda)\hat F',\lambda\hat G+(1-\lambda)\hat G') \tag{E.9}$$

for $\lambda\in[0,1]$ and arbitrary Hermitian semipositive operators $\hat F,\hat F'$, $\hat G,\hat G'$. By induction (E.9) generalizes straightforward to

$$\sum_\alpha\lambda_\alpha f(\hat F_\alpha,\hat G_\alpha)\leq f\left(\sum_\alpha\lambda_\alpha\hat F_\alpha,\sum_\beta\lambda_\beta\hat G_\alpha\right) \tag{E.10}$$

for $\lambda_\alpha\geq 0$ and $\sum_\alpha\lambda_\alpha=1$ and arbitrary semipositive operators $\hat F_\alpha$, $\hat G_\alpha$. We apply (E.10) to $s=1/2$, $\lambda^q_\alpha=1/8$, $F^q_\alpha=|a^q_\alpha|^2$ and $G^q_\alpha=|b^q_\alpha|^2$ to get

$$\sum_\alpha^{q=r,i}\mathrm{tr}\,\hat\Psi^\dagger|\hat a^q_\alpha|\hat\Psi|\hat b^q_\alpha\leq\mathrm{tr}\,\hat\Psi^\dagger\left(\sum_\alpha^{q=r,i}|\hat a^q_\alpha|^2\right)^{1/2}\hat\Psi\left(\sum_\beta^{p=r,i}|\hat b^p_\beta|^2\right)^{1/2} \tag{E.11}$$

and finally use operator Cauchy-Bunyakovsky-Schwarz inequality $|\,\mathrm{tr}\,\hat c\hat d|^2\leq\mathrm{tr}\,\hat c\hat c^\dagger\,\mathrm{tr}\,\hat d\hat d^\dagger$ to $\hat c=\hat\Psi^\dagger$ and

$$\hat d=\left(\sum_\alpha^{q=r,i}|\hat a^q_\alpha|^2\right)^{1/2}\hat\Psi\left(\sum_\beta^{p=r,i}|\hat b^p_\beta|^2\right)^{1/2} \tag{E.12}$$

which completes the proof. It is impossible to generalize CFRD inequalities to more observables [3].

* Electronic address: Adam.Bednorz@fuw.edu.pl
† Electronic address: wbednorz@mimuw.edu.pl

[1] C.W. Norman, *Undergraduate algebra* (Oxford University Press, 1986).
[2] E. G. Cavalcanti, C. J. Foster, M. D. Reid, and P. D. Drummond, Phys. Rev. Lett. **99**, 210405 (2007).
[3] E. Shchukin and W. Vogel, Phys. Rev. A **78**, 032104 (2008).
[4] A. Salles, D. Cavalcanti, A. Acin, D. Pérez-Garcia, M. M. Wolf, Quant. Inf. Comp. **10** (7-8) 703 (2010).
[5] 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).