Bound on Bell inequalities by fraction of determinism and reverse triangle inequality

P. Joshi,[1,2] K. Horodecki,[1,3] M. Horodecki,[1,2] P. Horodecki,[4] R. Horodecki,[1] Ben Li,[5] S. J. Szarek,[5,6] and T. Szarek[7]

[1]*National Quantum Information Center of Gdańsk, PL-81–824 Sopot, Poland*
[2]*Institute of Theoretical Physics and Astrophysics, University of Gdańsk, PL-80–952 Gdańsk, Poland*
[3]*Institute of Informatics, University of Gdańsk, PL-80–952 Gdańsk, Poland*
[4]*Faculty of Applied Physics and Mathematics, Technical University of Gdańsk, PL-80–233 Gdańsk, Poland*
[5]*Department of Mathematics, Case Western Reserve University, Cleveland, Ohio 44106-7058, USA*
[6]*Institut de Mathématiques de Jussieu-PRG, Université Pierre et Marie Curie-Paris 6, F-75252 Paris, France*
[7]*Institute of Mathematics, University of Gdańsk, PL-80–952 Gdańsk, Poland*

(Received 18 February 2015; revised manuscript received 14 July 2015; published 29 September 2015)

It is an established fact that entanglement is a resource. Sharing an entangled state leads to nonlocal correlations and to violations of Bell inequalities. Such nonlocal correlations illustrate the advantage of quantum resources over classical resources. In this paper, we quantitatively study Bell inequalities with $2 \times n$ inputs. As found in Gisin *et al.* [Int. J. Quantum. Inform. **05**, 525 (2007)], quantum mechanical correlations cannot reach the algebraic bound for such inequalities. Here we uncover the heart of this effect, which we call the *fraction of determinism*. We show that any quantum statistics with two parties and $2 \times n$ inputs exhibit a nonzero fraction of determinism, and we supply a quantitative bound for it. We then apply it to provide an explicit *universal upper bound* for Bell inequalities with $2 \times n$ inputs. As our main mathematical tool, we introduce and prove a *reverse triangle inequality*, stating in a quantitative way that if some states are far away from a given state, then their mixture is also. The inequality is crucial in deriving the lower bound for the fraction of determinism, but is also of interest on its own.

DOI: 10.1103/PhysRevA.92.032329 PACS number(s): 03.67.Bg, 03.67.Mn

I. INTRODUCTION

Since Bell's paper [1], entanglement has been studied and explored in depth. Entanglement has been used in many information-processing applications in which it either yields an advantage over the classical setting, e.g., in communication complexity [2], or where a classical counterpart simply doesn't exist, e.g., in quantum key distribution (QKD) [3], its device independent variant (DIQKD) [4], teleportation, superdense coding [5], or pseudotelepathy (PT) [6,7].

Although quantum theory allows for violations of Bell inequalities, in certain cases the violations cannot reach their maximum algebraic value. Tsirelson was the first to find upper bounds on the Bell values for quantum theory [8] and to relate them to Grothendieck's inequality. A significant amount of work has been done to explain why quantum mechanics does not lead to "algebraic" violations of Bell inequalities [9,10]. In Ref. [11], Wehner and Oppenheim argued that the trade-off between the so-called steerability and uncertainty determines how nonlocal a theory is. In Ref. [12], Cleve *et al.* gave an upper bound for the winning probability for XOR games in the quantum setting; their bound depends on the classical winning probability and Grothendieck's constant. (Note that the XOR game is a so-called nonlocal game, and Bell inequalities can be alternatively formulated as classical bounds on winning probabilities of nonlocal games [13,14].)

The approach to bound quantum violations via a Grothendieck-type constant K_G is now quite common and reasonably well understood. It leads to estimates for the Bell values β that are of the form $\beta_{\mathrm{qm}} \leqslant K_G \beta_{\mathrm{loc}}$ [15]. In this work we develop a different strategy, where the quantum Bell value of a given inequality depends on the difference between its optimal algebraic value β_{alg} and its optimal classical value β_{loc}.

Specifically, we study quantitatively Bell inequalities with $2 \times n$ inputs and give a *universal bound* on quantum Bell values of these inequalities. To find such a bound for $2 \times n$ Bell inequalities, we introduce the notion of the *fraction of determinism* (FOD) and show that the FOD is nonzero and depends only on the number of outcomes of each party (but not on n nor on the dimension of the underlying Hilbert space). Next, we ascertain that the presence of the FOD in quantum theory limits quantum Bell values. Our paper is inspired by Gisin *et al.* [16], where it is shown that there are some Bell inequalities (two-party pseudotelepathy games), for which quantum resources achieve the algebraic violation. They show that to achieve such violations, each party needs to have at least three input settings. In other words, there is no $2 \times n$ Bell inequality for which quantum theory attains the algebraic violation. Here we uncover the heart of this effect—the fraction of determinism—and are able to give a quantitative bound for it.

While looking for a lower bound for the FOD, we proved a fundamental property of quantum states which is interesting on its own. Namely, if ρ_1 and ρ_2 are far from σ, then any convex mixture of them is also far from σ. More precisely, if $\Delta_1 = \|\rho_1 - \sigma\| \geqslant 2 - \epsilon$ and $\Delta_2 = \|\rho_2 - \sigma\| \geqslant 2 - \epsilon$ for some $\epsilon \geqslant 0$, then, for all $p \in [0,1]$,

$$\Delta = \|p\rho_1 + (1-p)\rho_2 - \sigma\|_1 \geqslant 2 - O(\sqrt{\epsilon}), \qquad (1)$$

where $\|\rho\| \stackrel{\text{def}}{=} \mathrm{Tr}\sqrt{\rho^\dagger \rho}$ is the trace norm. Since the inequality (1) bounds the trace distance between σ and the mixture $\rho = p\rho_1 + (1-p)\rho_2$ from below (see Fig. 1), we refer to it as a "reverse triangle inequality" (RTI). Interestingly, it turns out that for classical states (i.e., probability densities, which can be represented as commuting density matrices) one can find the lower bound of Δ with the defect term linear in ϵ, while

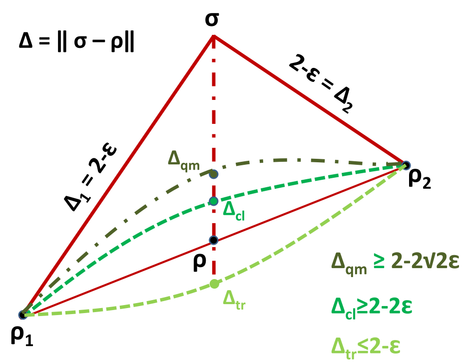

FIG. 1. (Color online) Pictorial representation of different bounds of $\|\sigma - \rho\|$. The triangle inequality gives an upper bound $2-\epsilon$, whereas *reverse triangle inequalities* give lower bounds $2-2\sqrt{2\epsilon}$ for general quantum states and $2-2\epsilon$ for classical (or commuting) states.

for noncommuting quantum states one cannot, in general, have dependence better than $O(\sqrt{\epsilon})$.

The second fundamental property which we use here follows from nonsignalling—impossibility of instantaneous communication. Namely, by performing a measurement on one site of the entangled state, one can create only those ensembles which give rise to the same density matrix—the reduced state of the entangled state. This implies that if we consider two such ensembles, there must be at least two elements (one from one ensemble, and the other from the second ensemble) that are not perfectly distinguishable. It has been apparently not studied to what extent they have to be indistinguishable. Here, by using the reverse triangle inequality, we are able to give a robust quantitative bound (Corollary 1), which is independent of the dimension of the underlying Hilbert space. We shall use it further to give a lower bound for the FOD, which in turn will allow us to upper-bound quantum violations for all $2 \times n$ Bell inequalities.

The paper is organized as follows. In Sec. II, we introduce necessary definitions and the role of the FOD. In Secs. III and IV, we present, respectively, a summary of our main results and sketches of their derivations. The special case when Bob has two inputs with binary outcomes is analyzed in Sec. V. For this case, we have explicitly calculated bounds for the FOD and for the classical fraction. Finally, we conclude our work in Sec. VII. Details of most proofs are relegated to the Appendixes.

II. PRELIMINARIES

A. Definitions

Box. Consider two distant parties, Alice and Bob (A and B), sharing a physical system. Each of them perform measurements labeled as $x \in \{x_1,\dots x_{n_A}\}$ and $y \in \{y_1,\dots y_{n_B}\}$, respectively; we will refer to such a setup as having "$n_A \times n_B$ input settings" or "measurement settings," or simply "$n_A \times n_B$ inputs." The outcomes of Alice and Bob are labeled as a and b, respectively. A box is defined as a family of joint probability distributions $p(a,b|x,y)$, i.e., $P = \{p(a,b|x,y)\}$. By a *nonsignalling box* (NS box) we mean a box which satisfies the following conditions:

$$\begin{aligned} p(b|y) &= \sum_{a=1}^{|x|} p(a,b|x,y) \\ &= \sum_{a=1}^{|x'|} p(a,b|x',y)\ \forall b,x,x' \quad \text{and} \quad y, \\ p(a|x) &= \sum_{b=1}^{|y|} p(a,b|x,y) \\ &= \sum_{b=1}^{|y'|} p(a,b|x,y') \quad \forall a,x,y' \quad \text{and} \quad y, \end{aligned} \tag{2}$$

where by $|z|$ we denote number of outcomes an observable z takes. A *local box* (or *classical box*) is defined as a box where joint probabilities can be expressed as

$$p(a,b|x,y) = \int_\Lambda q(\lambda)p(a|x,\lambda)p(b|y,\lambda)d\lambda, \tag{3}$$

where the hidden variable λ is distributed according to some probability density $q(\lambda)$. Such boxes satisfy, by definition (see below), every Bell inequality. We say that a box P is a *quantum box* (QM box) if the conditional probabilities can be realized as $p(a,b|x,y) = \mathrm{Tr}(M_a^x \otimes N_b^y \rho_{AB})$, where ρ_{AB} is a shared quantum state between parties A and B, and M_a^x and N_b^y are measurements for A and B, respectively (that is, for each inputs x and y, $\{M_a^x\}_{a=1}^{|x|}$ and $\{N_b^y\}_{b=1}^{|y|}$ are POVMs, i.e., families of positive operators satisfying $\sum_a M_a^x = I$ and $\sum_b N_b^y = I$). In this work, we only consider NS boxes. Notice that since the nonsignalling conditions are linear constraints, they define a polytope (NS polytope). Local boxes and QM boxes belong to this polytope.

Bell inequalities. Let $\mathcal{S} = \{s_{a,b}^{x,y}\}$ be a real vector and $P = \{p(a,b|x,y)\}$ be a box. The condition,

$$\mathcal{S}\cdot P := \sum_{x,y,a,b} s_{a,b}^{x,y}\ p(a,b|x,y) \leqslant \beta, \tag{4}$$

is called a *Bell inequality* if it is satisfied by any local box P [13].

Fraction of determinism (FOD). Let P be a nonsignalling box. Consider representations of P as a convex combination $P = (1-c)X + cD$, where X is an NS box and D is a deterministic box (i.e., a box, for which all the conditional probabilities are either 0 or 1). The *fraction of determinism* of P is then defined as the maximal possible weight of D in such representations, i.e.,

$$\mathrm{FOD} := \max_{D,X}\{c \mid P = (1-c)X + cD\}. \tag{5}$$

Classical fraction (CF). For a nonsignalling box P, we similarly consider representations of P as a convex combination $P = (1-\sum_i c_i)X + \sum_i c_i D_i$, where X is an NS box and D_i's are deterministic boxes. The classical fraction of P is then defined as the maximal combined weight of deterministic boxes in decompositions of the above form, i.e.,

$$\mathrm{CF} := \max_{\{D_i\},X}\left\{\sum_i c_i \,|\, P = \left(1-\sum_i c_i\right)X + \sum_i c_i D_i\right\}. \tag{6}$$

Notice that the set of local boxes is a polytope (*local polytope*), whose vertices are the deterministic boxes. Equivalently, local boxes are exactly convex combinations of deterministic boxes. Consequently, an alternative definition of CF is via a formula analogous to (5), with the admissible decompositions being now $P = (1 - c)X + cL$, where L is any local (or classical) box. A similar quantity was introduced in Ref. [17], where they call it "local fraction."

Note that the FOD, the CF, and the cost of nonlocality $c_{\rm nl}$ [18] satisfy the following relations:

$$\mathrm{FOD} \leqslant \mathrm{CF} = 1 - c_{\rm nl}. \tag{7}$$

While the FOD may be strictly smaller than the CF, it is easy to see that if the CF is nonzero, so is the FOD.

B. The role of the fraction of determinism

In the classical theory, the FOD is never zero. This is because—as noted above—every local box is a convex combination of deterministic boxes and some coefficients in that representation must be clearly nonzero. On the other hand, PR boxes [9] are completely noiseless and have zero fraction of determinism. This follows from the fact that they are vertices of the NS polytope and hence cannot be expressed as a nontrivial mixture of NS boxes; since they are not local themselves, their classical fraction is zero and, *a fortiori*, their FOD is also zero.

In quantum theory, the set of boxes is larger than the local polytope. In particular, there exist quantum boxes with zero fraction of determinism [16]. However, as will be explained in the next section, this phenomenon cannot materialize for quantum boxes with $2 \times n$ measurement settings. In turn, a nonzero FOD for a particular box, or for a class of boxes, limits the corresponding Bell values β and, in particular, prevents an algebraic violation. Indeed, for a given $\mathcal{S}$, let

$$\beta_{\rm det} = \beta_{\rm det}^{\mathcal{S}} := \max\{\mathcal{S}\cdot D \mid D \text{ is a deterministic box}\}, \tag{8}$$

and

$$\beta_{\rm alg} = \beta_{\rm alg}^{\mathcal{S}} := \max\{\mathcal{S}\cdot X \mid X \text{ is an NS box}\}, \tag{9}$$

be, respectively, the optimal deterministic and algebraic β values for the inequality associated with $\mathcal{S}$. Note that in (8) we would obtain the same value if we optimized over all classical boxes; this is because a linear function on a compact convex set (the local polytope) attains its maximum on extreme points (the deterministic boxes). In other words, $\beta_{\rm det} = \beta_{\rm loc}$; in what follows the two quantities will be used interchangeably.

Having defined the needed concepts, we present next a crucial observation.

Observation 1: Let P be an NS box such that $c = \mathrm{FOD}(P) > 0$. Then any Bell expression $\beta(P) := \mathcal{S}\cdot P$ is upper bounded by

$$\beta(P) = \mathcal{S}\cdot P \leqslant \beta_{\rm alg} - c(\beta_{\rm alg} - \beta_{\rm det}). \tag{10}$$

Proof. The proof is straightforward. We can decompose box P as $P = cD + (1 - c)X$, where D is a deterministic box and X is an NS box. Therefore,

$$\mathcal{S}\cdot P = c\,\mathcal{S}\cdot D + (1 - c)\,\mathcal{S}\cdot X \leqslant c\,\beta_{\rm det} + (1 - c)\,\beta_{\rm alg}. \tag{11}$$

■

This observation implies that if there is *any* violation at all (i.e., if $\beta_{\rm alg} > \beta_{\rm det}$), then the β value corresponding to P is *strictly* smaller than $\beta_{\rm alg}$. Note that essentially the same argument gives an identical bound with $c = \mathrm{CF}(P)$.

III. SUMMARY OF THE RESULTS

In this section we give a *universal* lower bound on the FOD for the $2 \times n$ inputs quantum scenario, depending only on the number of outcomes of both parties. As explained above, such an estimate implies an upper bound for all corresponding Bell inequalities in terms of the classical and the algebraic β values of each inequality. A summary of our main results is as follows.

Theorem 1: For an arbitrary QM box with $n_B = 2$ and $n_A = n$, i.e., $2 \times n$ inputs, the fraction of determinism is bounded as follows:

$$\mathrm{FOD} \geqslant \frac{71 - 17\sqrt{17}}{16k(l-1)l_1 l_2} \approx \frac{0.0567}{k(l-1)l_1 l_2}, \tag{12}$$

where $k = \max\{|x_1|,\ldots|x_n|\}$, $l_1 = |y_1|$, $l_2 = |y_2|$, and $l = \max\{l_1,l_2\}$.

To illustrate an application of Theorem 1, let us consider the (essentially simplest possible) instance of 2×2 input settings with binary outcomes, i.e., $n = k = l = 2$. In that case, the bound (12) becomes

$$\mathrm{FOD} \geqslant 7.08753 \times 10^{-3}. \tag{13}$$

Using the estimate (13), we can find an upper bound on quantum value of the CHSH inequality [13] which is

$$\beta_{\rm qm}^{\rm CHSH} \leqslant 4 - 7.08753 \times 10^{-3}(4-2) \leqslant 3.98583, \tag{14}$$

where we used (10) and the facts that $\beta_{\rm alg} = 4$ (attained on a PR box) and $\beta_{\rm det} = 2$.

In Sec. V, we will derive estimates for the FOD and the CF in the particular case of $2 \times n$ inputs with $2 \times k$ outputs without using our general theorem. When specified to $k = 2$, the resulting bounds are

$$\mathrm{FOD} \geqslant \frac{0.10961}{4} \quad \text{and} \quad \mathrm{CF} \geqslant \frac{0.11226}{4}. \tag{15}$$

Using these estimates, one can deduce from (10) bounds for the CHSH scenario which are slightly better than (14), namely

$$\beta_{\rm qm}^{\rm CHSH} \leqslant 4 - \frac{0.10961 * 2}{4} \leqslant 3.94519, \tag{16}$$

when using the FOD, and

$$\beta_{\rm qm}^{\rm CHSH} \leqslant 4 - \frac{0.11226 * 2}{4} \leqslant 3.94386, \tag{17}$$

when using the CF. We realize that these are still weak bounds, but the importance of this study lies in its generality: These bounds are valid for *any* (appropriately scaled) Bell inequality with $2 \times n$ inputs.

To prove Theorem 1 we need the following fundamental property of quantum states. We state it here as it may be of independent interest.

Theorem 2 (reverse triangle inequality): Let $\epsilon \geqslant 0$ and assume that the states ρ_i,σ satisfy

$$\|\rho_i - \sigma\| \geqslant 2 - \epsilon \tag{18}$$

for $i = 1,\ldots,l$. Then, for any probability distribution $\{p_i\}_{i=1}^{l}$,

(1) for any states ρ_i,σ satisfying (18)

$$\left\|\sum_{i=1}^{l} p_i\rho_i-\sigma\right\| \geqslant 2-2\sqrt{l\epsilon}, \tag{19}$$

(2) for commuting states ρ_i,σ satisfying (18)

$$\left\|\sum_{i=1}^{l} p_i\rho_i-\sigma\right\| \geqslant 2-l\epsilon, \tag{20}$$

(3) there exist three noncommuting states ρ_1,ρ_2, and σ satisfying (18) such that

$$\left\|\frac{\rho_1+\rho_2}{2}-\sigma\right\| \leqslant 2-\sqrt{2\epsilon}. \tag{21}$$

Remark. The third assertion says that when $l=2, 2-\sqrt{2\epsilon}$ is the best possible bound one can hope to achieve for general states. Hence, one cannot have better lower bound in (19) than $2-O(\sqrt{\epsilon})$.

In the following section we shall sketch a derivation of Theorem 1 from Theorem 2. The proof of Theorem 2 and of its generalization, Proposition 2, is given in Appendix A.

IV. FRACTION OF DETERMINISM IN QM

Our main objective in this section is to prove Theorem 1. To do that, i.e., to lower-bound the FOD of a given box, essentially requires one to look for deterministic structures in the box. Among all possible deterministic structures, picking the one with the maximal weight gives the value of the FOD of the box. As a step in that direction, we note first a simple fact following directly from the definition.

Observation 2: Consider a box $P=\{p(a,b|x,y)\}$ with inputs $\{x_1,\ldots,x_{n_A}\}$ on Alice's side and $\{y_1,\ldots,y_{n_B}\}$ on Bob's side. Then the following two conditions are equivalent.

(1) We can find outcomes $a^{(1)},\ldots,a^{(n_A)}$, $b^{(1)},\ldots,b^{(n_B)}$ such that

$$\forall s,t \qquad p(a^{(s)},b^{(t)}|x_s,y_t) \geqslant c. \tag{22}$$

(2) $\mathrm{FOD}(P)\geqslant c$.

This is illustrated pictorially in Fig. 2(a) (when $n_B=|y_1|=|y_2|=2$). Although this is an important observation, it does not easily allow calculating explicit lower bounds for the FOD. Therefore, we reformulate estimating the FOD of a QM box as an optimization problem as follows.

Proposition 1: Consider a QM box with two inputs $\{y_1,y_2\}$ on Bob's side and n inputs $\{x_1,\ldots,x_n\}$ on Alice's side. Then (22) is satisfied with $c=c_0$, where

$$c_0=\inf_{\xi,\xi'}\max_{i,j}\min_{x}\max_{a}\min\left\{p_i\mathrm{Tr}\big(M_a^x\rho_i\big),q_j\mathrm{Tr}\big(M_a^x\sigma_j\big)\right\}. \tag{23}$$

The infimum is taken here over all ensembles $\xi=\{(p_i,\rho_i)\}_{i=1}^{|y_1|}$, $\xi'=\{(q_j,\sigma_j)\}_{j=1}^{|y_2|}$ (where $\{p_i\}$ and $\{q_j\}$ are probability distributions, and ρ_i and σ_j are states) satisfying

$$\sum_i p_i\rho_i=\sum_j q_j\sigma_j, \tag{24}$$

and the first minimum over all inputs $x\in\{x_1,\ldots,x_n\}$ (with $\{M_a^x\}$ the corresponding POVM measurements).

Proof. The argument depends on showing that c_0 defined via (23) works as a bound in (22), which then allows us to appeal to Observation 2.

By definition, a quantum box is realized via POVMs $\{M_a^x\}$ (with $x\in\{x_1,\ldots,x_n\}$) on Alice's side, two POVMs $\{N_b^{y_1},N_{b'}^{y_2}\}$ on Bob's side, and a shared quantum state ρ_{AB}. Depending on Bob's measurement choice (y_1 or y_2), an ensemble $\xi_0=\{p(b_i|y_1),\rho_i\}_{i=1}^{|y_1|}$ or $\xi_0'=\{p(b_j'|y_2),\sigma_j\}_{j=1}^{|y_2|}$ is created at Alice's site, where $p(b_i|y_1)=:p_i$ and $p(b_j'|y_2)=:q_j$ are the marginal conditional probabilities. In terms of these marginals, one can express joint probabilities as follows.

$$\begin{aligned} p(a,b_i|xy_1)&=p_i\mathrm{Tr}\big(M_a^x\rho_i\big),\\ p(a,b_j'|xy_2)&=q_j\mathrm{Tr}\big(M_a^x\sigma_j\big). \end{aligned} \tag{25}$$

Note that these ensembles satisfy

$$\mathrm{Tr}_B(\rho_{AB})=\sum_i p(b_i|y_1)\rho_i=\sum_j p(b_j'|y_2)\sigma_j, \tag{26}$$

so that the condition in (24) holds. Now, for a fixed pair of indices (i,j) and an input x, pick an outcome $a=a_{ij}^{(x)}$ such

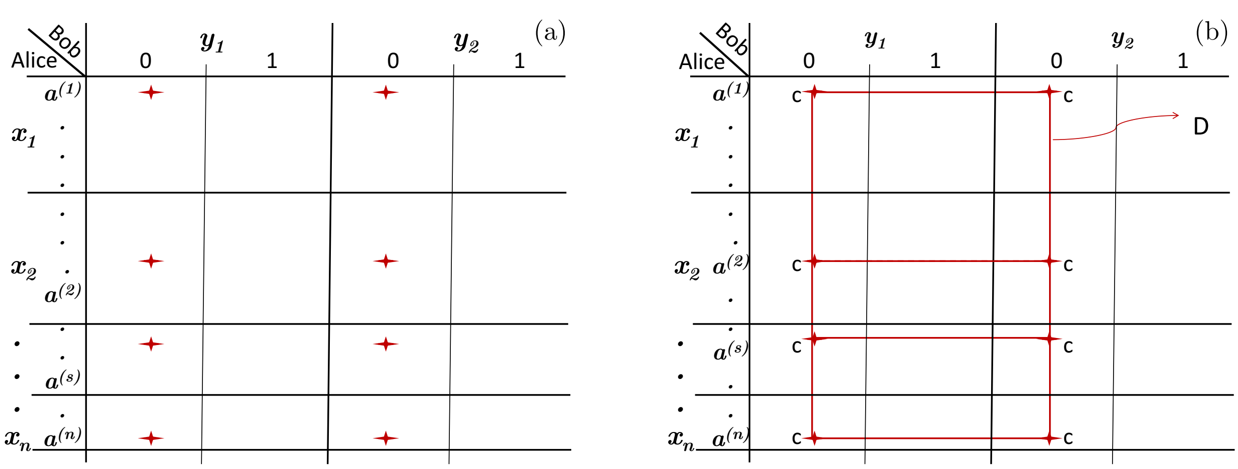

FIG. 2. (Color online) Illustration of FOD. (a) The outcomes $a^{(1)},\ldots,a^{(n)}$ satisfying (22). (These are the optimal outcomes referred to in the proof of Propostion 5.) Note that here $b^{(1)}=0=b^{(2)}$. (b) Deterministic box D: the rectangle is a deterministic box appearing with weight c in the decomposition of P.

that $\min\{p(ab_i|xy_1),p(ab'_j|xy_2)\}$ is maximal. Essentially, by fixing a pair of indices (i,j), we fix "one side" of (the support of) a deterministic box. The optimal outputs $\{a_{ij}^{(x)}\}$ define "the other side" [see Fig. 2(a)] and the smallest probability gives the weight of that box. Optimizing over all pairs (i,j) leads to the choices of $a^{(s)}=a_{i_0j_0}^{(x_s)}$, $b^{(1)}=b_{i_0}$, and $b^{(2)}=b'_{j_0}$ that satisfy (22) with the constant given by (23), as needed. ■

In order to be able to apply Proposition 1, we need to supply an appropriate lower bound on the parameter c_0 defined by (23). To that end, let us analyze the optimization problem implicit in (23). For given $\xi,\xi',\{M_a^x\}$ and the selected i,j, it is *a priori* conceivable that the two densities (25) are disjointly supported (when considered as functions of a), and so any choice of a will lead to the minimum of the two probabilities being 0. Another way to describe such a situation is that the POVM $\{M_a^x\}$ can *perfectly* distinguish ρ_i from σ_j. However, it is easy to see that this can only happen if the states ρ_i and σ_j "live" on orthogonal subspaces of the underlying Hilbert space or, equivalently, if $\|\rho_i-\sigma_j\|=2$. Further, if this unfortunate state of affairs persists for every pair i,j, it follows that the mixtures $\sum_i p_i\rho_i$ and $\sum_j q_j\sigma_j$ themselves "live" on orthogonal subspaces and hence can be perfectly distinguished by $\{M_a^x\}$, which is of course impossible since they coincide by (24). This shows that there is a choice of $i\,(=i_0), j\,(=j_0)$, and then of $a\,(=a_{i_0j_0}^{x_s})$, such that the objective function in the optimization problem (23) is strictly positive. Note, however, that these considerations do not yield yet any explicit lower bound nor any hint of uniformity over ξ,ξ', and x.

A version of the above reasoning is the gist of the argument in Ref. [16]. In what follows we shall provide quantitative statements elaborating on the points made above. The first lemma asserts that if $\|\rho-\sigma\|$ is noticeably smaller than 2, then there is a limit on how well a POVM can distinguish ρ and σ. More precisely, we have the following.

Lemma 1: Let $\epsilon\geqslant 0$ and suppose that $\|\rho-\sigma\|\leqslant 2-\epsilon$. Then, for any POVM $\{M_a\}_{a=1}^k$, there exists an outcome a_0 such that

$$\mathrm{Tr}\big(M_{a_0}\rho\big)\geqslant\frac{\epsilon}{2k}\quad\text{and}\quad\mathrm{Tr}\big(M_{a_0}\sigma\big)\geqslant\frac{\epsilon}{2k}. \tag{27}$$

The proof is—unsurprisingly—based on the Helstrom formula [19], which relates distinguishability of quantum states via POVMs to their trace distance. The details are presented in Appendix B. The importance of the Lemma for our results lies in the fact that it uncovers a deterministic structure in every quantum box with $2\times n$ inputs.

Using Lemma 1, we can replace marginal conditional probabilities $\mathrm{Tr}(M_a^x\rho_i)$ and $\mathrm{Tr}(M_a^x\sigma_j)$ in (23) by ϵ (depending on i and j). Having done this, we can get rid of the optimization over inputs and outputs of Alice, i.e., over a and over all POVMs $\{M_a^x\}$. As a consequence, we have the following.

Observation 3: The quantity c_0 of (23) satisfies $c_0\geqslant c_1$ with

$$c_1=\frac{1}{2k}\inf_{\xi,\xi'}\max_{i,j}\big((2-\|\rho_i-\sigma_j\|)\min\{p_i,q_j\}\big), \tag{28}$$

where the infimum is taken over all ensembles $\xi=\{(p_i,\rho_i)\}_{i=1}^{|y_1|}$, $\xi'=\{(q_j,\sigma_j)\}_{j=1}^{|y_2|}$ satisfying

$$\sum_i p_i\rho_i=\sum_j q_j\sigma_j. \tag{29}$$

Proof. Lemma 1 tells us that, for given ξ,ξ',i,j, and x,

$$\max_a\min\big\{p_i\mathrm{Tr}\big(M_a^x\rho_i\big),q_j\mathrm{Tr}\big(M_a^x\sigma_j\big)\big\}\geqslant\min\{p_i,q_j\}\frac{\epsilon_{ij}}{2k}, \tag{30}$$

where $\epsilon_{ij}=2-\|\rho_i-\sigma_j\|$. To conclude that $c_0\geqslant c_1$, it remains to notice that the left-hand side and the right-hand side of (30) represent the expressions appearing, respectively, in (23) and (28) (the definitions of c_0 and c_1), and that the right-hand side doesn't depend on x. ■

Having simplified the FOD, we are ready for the crucial step, which involves quantifying the limitations on the trace distances $\|\rho_i-\sigma_j\|$. Our aim is to estimate those distances in terms of the number of Bob's outcomes. Interestingly, this leads to an important independent result, which is a purely geometric property of quantum states, and which is stated as Theorem 2. For our purposes, we need a slight generalization of the first statement of the theorem.

Proposition 2: Assume that $\epsilon\geqslant 0$ and that the states ρ_i,σ_j satisfy $\|\rho_i-\sigma_j\|\geqslant 2-\epsilon$ for all $i\in\{1,\ldots,l_1\}$ and all $j\in\{1,\ldots,l_2\}$. Then, for any probability distributions $\{p_i\}_{i=1}^{l_1}$ and $\{q_j\}_{j=1}^{l_2}$,

$$\left\|\sum_{i=1}^{l_1}p_i\rho_i-\sum_{j=1}^{l_2}q_j\sigma_j\right\|\geqslant 2-2\sqrt{l_1l_2\epsilon}. \tag{31}$$

Proposition 2 follows by essentially the same argument as Theorem 2 (see Appendix A). Alternatively, a statement with a little worse dependence on ϵ,l_1,l_2 in (31) can be formally derived from Theorem 2 by applying it twice.

For facility of application, it is more convenient to restate Proposition 2 in the contrapositive form. Note that, by continuity, it doesn't matter whether we state the results with strict or nonstrict inequalities, as long as we are consistent.

Corollary 1: Let $\theta\in[0,2]$ and assume that two ensembles $\{(p_i,\rho_i)\}_{i=1}^{l_1}$, $\{(q_j,\sigma_j)\}_{j=1}^{l_2}$ satisfy

$$\left\|\sum_i p_i\rho_i-\sum_j q_j\sigma_j\right\|\leqslant\theta. \tag{32}$$

Then there exist i_0 and j_0 such that

$$\|\rho_{i_0}-\sigma_{j_0}\|\leqslant 2-\epsilon, \tag{33}$$

where

$$\epsilon=\frac{1}{l_1l_2}\left(\frac{2-\theta}{2}\right)^2. \tag{34}$$

We are now almost done: Choosing $i=i_0$ and $j=j_0$ in (28) yields a nonzero lower bound for c_1, hence for c_0. However, it may still happen that, for the chosen pair of indices, the probabilities p_{i_0}, q_{j_0} are very small. To guard against that risk, we truncate the ensembles so that the minimal probability is bounded away from zero. Such smaller ensembles do not give rise to the same density matrix. However, their density matrices will be still close, provided we did not truncate too much. Here is a precise quantitative statement to that effect.

Lemma 2: Let $\mathcal{E}_1=\{p_i,\rho_i\}_{i=1}^{l_1}$, $\mathcal{E}_2=\{q_j,\sigma_j\}_{j=1}^{l_2}$ be two ensembles which give rise to the same density matrix. Given

$\tilde{l}_1 \leqslant l_1$ and $\tilde{l}_2 \leqslant l_2$, we set

$$\delta_1 = 1 - \sum_{i=1}^{\tilde{l}_1} p_i, \quad \delta_2 = 1 - \sum_{j=1}^{\tilde{l}_2} q_j, \tag{35}$$

and define new ensembles,

$$\tilde{\mathcal{E}}_1 = \{\tilde{p}_i,\rho_i\}_{i=1}^{\tilde{l}_1}, \quad \tilde{\mathcal{E}}_2 = \{\tilde{q}_j,\sigma_j\}_{j=1}^{\tilde{l}_2}, \tag{36}$$

where $\tilde{p}_i = p_i/(1-\delta_1)$ and $\tilde{q}_j = q_j/(1-\delta_2)$. Then the ensembles $\tilde{\mathcal{E}}_1,\tilde{\mathcal{E}}_2$ satisfy

$$\left\| \sum_{i=1}^{\tilde{l}_1} \tilde{p}_i\rho_i - \sum_{j=1}^{\tilde{l}_2} \tilde{q}_j\sigma_j \right\| \leqslant \frac{2\max\{\delta_1,\delta_2\}}{1-\min\{\delta_1,\delta_2\}}. \tag{37}$$

Thus we can use the new ensembles (36) to show that there exist a pair of states ρ_{i_0} and σ_{j_0} with $\|\rho_{i_0} - \sigma_{j_0}\| \leqslant 2-\epsilon$, and that at the same time the weights of the states satisfy $p_{i_0} \geqslant p_{\tilde{l}_1}$, $q_{j_0} \geqslant q_{\tilde{l}_2}$. By adjusting $\tilde{l}_1$ and $\tilde{l}_2$ properly, we can simultaneously secure bounds on both the weights and the norm, and complete the proof of our main result.

Proof of Theorem 1. We start by recalling Observation 2, which states that FOD $\geqslant c$ if c satisfies (22). In Proposition 1, we showed that a particular $c = c_0$ satisfies (22), namely

$$c_0 = \inf_{\xi,\xi'} \max_{i,j} \min_{x} \max_{a} \min\left\{p_i \mathrm{Tr}\big(M_a^x\rho_i\big), q_j \mathrm{Tr}\big(M_a^x\sigma_j\big)\right\}. \tag{38}$$

To simplify the optimization problem implicit in (38), another lower bound for the FOD ($c_1 \leqslant c_0$) was obtained in Observation 3 by using Lemma 1, to wit,

$$c_1 = \frac{1}{2k} \inf_{\xi,\xi'} \max_{i,j} \big((2 - \|\rho_i - \sigma_j\|)\min\{p_i,q_j\}\big), \tag{39}$$

We now want to appeal to Corollary 1. However, the probabilities p_i and q_j may *a priori* be very small for the chosen pair of indices $(i,j) = (i_0,j_0)$. Hence, we first need to truncate the ensembles appropriately following the scheme outlined in Lemma 2. Assume that $\{p_i\}$ and $\{q_j\}$ are arranged in the nonincreasing order and let $\mu \geqslant 2$ be a parameter. Choose the largest $\tilde{l}_1$ and $\tilde{l}_2$ such that $p_{\tilde{l}_1} \geqslant \frac{1}{(l-1)\mu}$ and $q_{\tilde{l}_2} \geqslant \frac{1}{(l-1)\mu}$, where $l = \max\{l_1,l_2\}$. (Note that $\tilde{l}_1$ and $\tilde{l}_2$ exist because of our assumption $\mu \geqslant 2$.) Then

$$\delta_1 = \sum_{i=\tilde{l}_1+1}^{l_1} p_i < (l-1) \times \frac{1}{(l-1)\mu} = \frac{1}{\mu}, \tag{40}$$

and, similarly,

$$\delta_2 < \frac{1}{\mu}. \tag{41}$$

Lemma 2 yields now the following estimate on the truncated ensemble:

$$\left\| \sum_{i=1}^{\tilde{l}_1} \tilde{p}_i\rho_i - \sum_{j=1}^{\tilde{l}_2} \tilde{q}_j\sigma_j \right\| < \frac{2/\mu}{1-1/\mu} = \frac{2}{\mu-1}. \tag{42}$$

Consequently, it follows from Corollary 1 that there are indices i_0, j_0 such that

$$\|\rho_{i_0} - \sigma_{j_0}\| < 2 - \epsilon, \tag{43}$$

where

$$\epsilon = \frac{1}{l_1 l_2}\left(\frac{\mu-2}{\mu-1}\right)^2. \tag{44}$$

Next, taking into account that for the ensembles in question,

$$p_{i_0} \geqslant \frac{1}{(l-1)\mu} \quad \text{and} \quad q_{j_0} \geqslant \frac{1}{(l-1)\mu}, \tag{45}$$

Eq. (39) yields

$$c_0 \geqslant c_1 \geqslant \frac{1}{2k}\frac{\epsilon}{(l-1)\mu}. \tag{46}$$

Finally, substituting the value of ϵ given by (44) into (46) and recalling that $\mu \geqslant 2$ was arbitrary, we are led to

$$\text{FOD} \geqslant \frac{1}{2k(l-1)l_1 l_2} \times \max_{\mu \geqslant 2} \frac{1}{\mu}\left(\frac{\mu-2}{\mu-1}\right)^2. \tag{47}$$

To complete the proof, it remains to observe that the function $f(\mu) = \frac{1}{\mu}(\frac{\mu-2}{\mu-1})^2$ attains its maximum—equal to $\frac{71-17\sqrt{17}}{8}$ – at $\mu_0 = \frac{5+\sqrt{17}}{2}$. ■

In the next section we will present a refinement of the above argument when the party that has two allowed inputs has only binary outcomes. For that case, we not only explicitly calculate a (better) lower bound for the FOD, but we also find a lower bound for the CF which is slightly larger than that for the FOD.

V. FOD AND CF FOR A SIMPLE BOB

We devote this section to finding an upper bound for quantum Bell values in the particular case where Bob has two inputs and two outputs, while Alice has n inputs and k outputs. We will argue along the same lines as in the proof of Theorem 1. However, rather than appealing directly to the conclusion of the theorem, we will use the simplicity of our setting to further optimize some steps in the argument, particularly the one involving truncations. Again, we are looking for structures resembling deterministic boxes within a quantum box. The weights of such deterministic boxes yield lower bounds the FOD and the CF of the box. The details are explained below (and in Appendix C).

Since Bob is the party with two inputs and two outputs, he can create ensembles $\{p_i,\rho_i\}_{i=0}^1$ or $\{q_j,\sigma_j\}_{j=0}^1$ at Alice's site by making, respectively, measurement y_1 or y_2 on his part of the shared quantum state. Lemma 1 asserts that for all pairs of ρ_i and σ_j, and for all POVMs $\{M_a^x\}$ we have

$$\begin{aligned} &\exists \quad a_0,a_1,a_2,a_3 \quad \text{such that} \\ &\mathrm{Tr}\big(M_{a_0}^x\rho_0\big) \geqslant \epsilon_{00}/2k, \quad \text{and} \quad \mathrm{Tr}\big(M_{a_0}^x\sigma_0\big) \geqslant \epsilon_{00}/2k \\ &\mathrm{Tr}\big(M_{a_1}^x\rho_0\big) \geqslant \epsilon_{01}/2k, \quad \text{and} \quad \mathrm{Tr}\big(M_{a_1}^x\sigma_1\big) \geqslant \epsilon_{01}/2k \\ &\mathrm{Tr}\big(M_{a_2}^x\rho_1\big) \geqslant \epsilon_{10}/2k, \quad \text{and} \quad \mathrm{Tr}\big(M_{a_2}^x\sigma_0\big) \geqslant \epsilon_{10}/2k \\ &\mathrm{Tr}\big(M_{a_3}^x\rho_1\big) \geqslant \epsilon_{11}/2k, \quad \text{and} \quad \mathrm{Tr}\big(M_{a_3}^x\sigma_1\big) \geqslant \epsilon_{11}/2k, \end{aligned} \tag{48}$$

where $\epsilon_{ij} = 2 - \|\rho_i - \sigma_j\|$. (Note that the outcomes a_i depend on the input setting x.) This means that if $\epsilon_{ij} > 0$ and when Bob obtains outcomes $(b,b') = (b_i,b'_j)$ for inputs (y_1,y_2), then for any POVM of Alice there exists at least one outcome on her side such that once she obtains it, she cannot tell apart measurement choices of Bob with certainty (i.e., she cannot determine whether Bob chose y_1 or y_2 to create the ensemble). Let us call such outcome of Alice a *confusing outcome*. For example, in the first pair of inequalities in (48), if $\epsilon_{00} > 0$, then the outcome a_0 of the POVM does not differentiate with certainty between ρ_0 and σ_0. There are four pairs (b,b'), hence there are four potential confusing outcomes corresponding to each of these four cases.

Consider now the particular case when Bob's outcomes are 0 for measurement y_1 and also 0 for measurement y_2. Let Alice choose input setting x and let us say that, for this setting (and for the above Bob's pair of outcomes), a_0 is her confusing outcome. Using Eq. (48), we have $\mathrm{Tr}(M^x_{a_0}\rho_0) \geqslant \epsilon_{00}/2k$ and $\mathrm{Tr}(M^x_{a_0}\sigma_0) \geqslant \epsilon_{00}/2k$. Indeed, this is true for any input choice of Alice and a_0 depends on the input. Now, note that the joint probabilities from the quantum box can be expressed as follows:

$$\begin{aligned} p(a_0,0|x,y_1) &= p_0\mathrm{Tr}\big(M^x_{a_0}\rho_0\big),\\ p(a_0,0|x,y_1) &= q_0\mathrm{Tr}\big(M^x_{a_0}\sigma_0\big), \end{aligned} \tag{49}$$

where p_0 and q_0 denote the marginal probabilities of Bob's outcome 0 while measuring, respectively, y_1 and y_2. Thus we obtain that they both can be bounded as follows:

$$p(a_0,0|x,y_1) \geqslant c_{00}, \quad p(a_0,0|x,y_1) \geqslant c_{00}, \tag{50}$$

with

$$c_{00} := \frac{1}{2k}\min\{p_0\epsilon_{00},q_0\epsilon_{00}\}. \tag{51}$$

By this construction, we can create a deterministic box (say D_{00}) with fraction equal to c_{00}. In other words, every quantum box P_Q satisfies the relation $P_Q = (1-c_{00})X + c_{00}D_{00}$. In such a way, we can create four different decompositions of $P_Q = (1-c_{ij})X + c_{ij}D_{ij}$ using deterministic boxes ($\{D_{ij}\}^1_{i,j=0}$) corresponding to each of the four outcome pairs (b,b') of Bob. Using these decompositions and the definition of FOD we get

$$\begin{aligned} \mathrm{FOD} \geqslant \frac{1}{2k}\max\big\{&\min\{p_0\epsilon_{00},q_0\epsilon_{00}\}, \min\{p_1\epsilon_{11},q_1\epsilon_{11}\},\\ &\min\{p_0\epsilon_{01},q_1\epsilon_{01}\}, \min\{p_1\epsilon_{10},q_0\epsilon_{10}\}\big\}. \end{aligned} \tag{52}$$

Notice that the obtained bound still depends on the shared state and on Bob's measurements (through the ensembles $\{p_i,\rho_i\}$ and $\{q_j,\sigma_j\}$). Later on (Appendix C) we will optimize over the ensembles, thus obtaining a *universal bound* for the FOD of any quantum box with $2\times n$ inputs and $2\times k$ outcomes for Bob and Alice, respectively, which is as follows:

$$\mathrm{FOD} \geqslant \frac{0.10961}{2k}. \tag{53}$$

Using this value of FOD in (10), we can find the following upper bound on quantum Bell values for the CHSH inequalities,

$$\beta^{\mathrm{CHSH}}_{\mathrm{qm}} \leqslant 4 - \frac{0.10961 * 2}{4} \approx 3.94519. \tag{54}$$

Next, we will show how this bound can be improved using the classical fraction. We have more flexibility here because we can, instead of a single deterministic box, consider mixtures of boxes corresponding to different pairs of Bob's outcomes. To this end, we notice that there is a possibility that there may exist a POVM for Alice such that she obtains a single confusing outcome for two or more different cases (e.g., when she obtains a confusing outcome a_0, she is unable to distinguish between measurement choices of Bob not only in the case when Bob obtains (0,0), but also in the case when he obtains (1,0)). So, in the worst case, for some measurement choices there may be just one confusing outcome at Alice's side for all the four different cases as shown in Fig. 3 in the last row of the box. In that case, the quantum box does not satisfy $P_Q = (1-\sum c_{ij})X + \sum^1_{i,j=0} c_{ij}D_{ij}$ because this would require us to use some probabilities twice. Thus, to be on the safe side, we will use only orthogonal pairs of deterministic boxes (i.e., either $P_Q = (1-c_{00}-c_{11})X + c_{00}D_{00} + c_{11}D_{11}$ or $P_Q = (1-c_{01}-c_{10})X + c_{01}D_{01} + c_{10}D_{10}$). Using these decompositions and the definition of CF we get

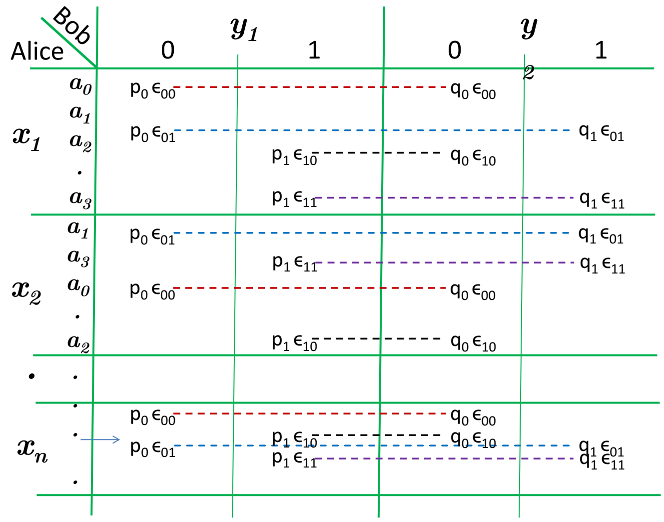

FIG. 3. (Color online) The box $\{p(a,b|x,y)\}$ of Alice and Bob. Dashed lines represent which pairs give rise to confused outcomes (a_i's) and their lower bounds. Note that, for some choice of inputs (x,y), some or all confused outcomes (a_i's) may coincide with each other.

$$\begin{aligned} \mathrm{CF} \geqslant \frac{1}{2k}\max\big\{&\min\{p_0\epsilon_{00},q_0\epsilon_{00}\} + \min\{p_1\epsilon_{11},q_1\epsilon_{11}\},\\ &\min\{p_0\epsilon_{01},q_1\epsilon_{01}\} + \min\{p_1\epsilon_{10},q_0\epsilon_{10}\}\big\}. \end{aligned} \tag{55}$$

After optimizing over p_0,q_0, ρ_i's, and σ_j's, (Appendix C contains detailed calculations) one can obtain the following bound on the CF for $2\times n$ inputs and $2\times k$ outputs for Bob and Alice.

$$\mathrm{CF} \geqslant \frac{0.11226}{2k}. \tag{56}$$

Using this estimate one can infer from (10) the following upper bound on quantum Bell values for CHSH inequalities (i.e., when $n=2$ and $k=2$).

$$\beta^{\mathrm{CHSH}}_{\mathrm{qm}} \leqslant 4 - \frac{0.11226 * 2}{4} \approx 3.94386. \tag{57}$$

Notice that these bounds are slightly better than the one obtained in (14). However, even these slightly better bounds are quite weak, but since they hold for any $2\times n$ Bell inequalities, they presumably cannot be strengthened substantially.

VI. DISCUSSION

While the bounds we obtained are universal, they are rather weak. In addition to universality, there are two sources of this weakness. On one hand, we have applied a series of inequalities and even if all of them were tight, their combination would lose that precision. On the other hand, the very concept of the FOD does not lead in general to optimal bounds (i.e., the Tsirelson bounds). To see this, let us note that even for classical boxes the FOD can be very small, leading to a weak bound.

Let us restrict our discussion to the CHSH case (hence 2×2 inputs with binary outcomes). Consider the classical box P_{noise} that can be called maximally mixed and is given by

$$p(ab|xy) = 1/4 \text{ for all } a,b,x,y . \tag{58}$$

Since all probabilities are equal to $1/4$, it follows from Observation 2 that

$$\text{FOD}(P_{\text{noise}}) = \tfrac{1}{4}. \tag{59}$$

Indeed, any choice of $a^{(1)},a^{(2)},b^{(1)},b^{(2)}$ works in (22) with $c = 1/4$ and clearly no choice allows $c > 1/4$. Our inequality (10) yields, now,

$$\beta^{\text{CHSH}}(P_{\text{noise}}) \leqslant 3.5 . \tag{60}$$

Thus even the bound obtained this way for classical boxes is worse than the Tsirelson bound. (However, as we noted in Sec. II B, we were guaranteed to get a nontrivial bound, that is, strictly smaller than $\beta^{\text{CHSH}}_{\text{alg}} = 4$.)

Note that, in the above example, using the classical fraction of P_{noise} [$\text{CF}(P_{\text{noise}}) = 1$] would clearly give the correct value $\beta^{\text{CHSH}}_{\text{loc}} = 2$. However, even having a general way of calculating the classical fraction would not always lead to tight estimates. Indeed, while—as shown by a simple separation argument—for every box P there is a Bell expression $\mathcal{S}\cdot P$, for which the bound (10) with $c = \text{CF}(P)$ is saturated (cf. the comment following Observation 1), this is no longer true for an arbitrary pair $P,\mathcal{S}$.

VII. CONCLUSION

In this paper we gave an explicit *universal bound* for $2 \times n$ Bell inequalities, which is independent of the number n of inputs and of the dimension of the underlying Hilbert space. Specifically, we showed that this universal bound depends only on the number of outputs of the two parties, and on the difference between the optimal algebraic value and the optimal deterministic value of the inequality. We showed that the presence of the fraction of determinism (FOD) in $2 \times n$ BI prevents quantum Bell values from achieving the maximal algebraic value. Hence this result is a quantitative variant of the theorem shown by Gisin *et al.* in Ref. [16], which states that there exist no $2 \times n$ inputs pseudotelepathy game. Although these bounds are not tight, one can improve them by considering the classical fraction (CF) and by using it to strengthen the estimate. We have analyzed a simple case where the CF gives a better bound than taking into account merely the FOD.

To obtain the above results, we established a *reverse triangle inequality* (RTI), which is an independent result of its own interest. While the triangle inequality gives upper bounds on the trace distance between a state and mixtures of other states, the RTI bounds that trace distance from below. We have determined that this lower bound is different for general (i.e., possibly noncommuting) states than when considering only commuting states. The bound in the commuting case is sharp, and the one in the noncommuting case is close to being sharp.

ACKNOWLEDGMENTS

We thank Aram Harrow for a suggestion that led to a simpler proof of our geometric result and to a better constant. P.J. thanks P. Mazurek for useful discussions. We also thank Andrew Blasius for his helpful comments regarding the manuscript. Finally, we thank the anonymous referee, whose criticism of the earlier version of the paper led to substantial improvement in the presentation. This work was supported by ERC QOLAPS, EC IP QESSENCE, EC Grant RAQUEL, MNiSW Grant No. IdP2011 000361, and NCBiR-CHIST-ERA Project QUASAR. P.J. was also supported by Grant No. MPD/2009-3/4 from Foundation for Polish Science. S.J.S. was partially supported by grants from the National Science Foundation (USA) and by Grant No. 2011-BS01-008-02 from ANR (France). T.S. was partially supported by the National Science Centre of Poland, Grant No. DEC-2012/07/B/ST1/03320. K.H. acknowledges Grant No. BMN 538-5300-0637-1. Part of this work was done at the National Quantum Information Centre of Gdańsk.

APPENDIX A: PROOF OF THEOREM 2 AND A DISCUSSION OF ITS OPTIMALITY

Theorem 2: Let $\epsilon \geqslant 0$ and assume that the states ρ_i,σ satisfy

$$\|\rho_i - \sigma\| \geqslant 2 - \epsilon, \tag{A1}$$

for $i = 1, \dots, l$. Then, for any probability distribution $\{p_i\}_{i=1}^{l}$,

(1) for any states ρ_i, σ satisfying (A1)

$$\left\| \sum_{i=1}^{l} p_i \rho_i - \sigma \right\| \geqslant 2 - 2\sqrt{l\epsilon}, \tag{A2}$$

(2) for commuting states ρ_i, σ satisfying (A1)

$$\left\| \sum_{i=1}^{l} p_i \rho_i - \sigma \right\| \geqslant 2 - l\epsilon. \tag{A3}$$

(3) there exist three noncommuting states ρ_1,ρ_2, and σ satisfying (A1) such that

$$\left\| \frac{\rho_1 + \rho_2}{2} - \sigma \right\| \leqslant 2 - \sqrt{2\epsilon}. \tag{A4}$$

Proof. We start by recalling two well-known facts.

Rotfel'd inequality [20]. Let f be a concave function on $[0,\infty)$ such that $f(0) \geqslant 0$ and let $A_1, \dots, A_l \geqslant 0$. Then,

$$\text{Tr} f\left(\sum_{i=1}^{l} A_i\right) \leqslant \text{Tr} \sum_{i=1}^{l} f(A_i). \tag{A5}$$

Rotfel'd inequality is usually stated for just two matrices (i.e., $l = 2$), but the general case follows easily by induction.

Fuchs–van de Graaf inequalities [21]. These inequalities give two-sided bounds for the trace distance between two quantum states σ and τ in terms of *fidelity* between σ and τ, which is defined as $F(\sigma,\tau) = \mathrm{Tr}\sqrt{\sqrt{\sigma}\,\tau\sqrt{\sigma}}$. We have

$$1 - F(\sigma,\tau) \leqslant \tfrac{1}{2}\|\sigma - \tau\| \leqslant \sqrt{1 - F(\sigma,\tau)^2}. \tag{A6}$$

Rotfel'd inequality applied with $f(t) = \sqrt{t}$ allows us to upper-bound fidelity of the mixture $\sum_{i=1}^{l} p_i\rho_i =: \rho$ in terms of individual fidelities:

$$\begin{aligned} F(\sigma,\rho) &= \mathrm{Tr}\sqrt{\sqrt{\sigma}\left(\sum_{i=1}^{l} p_i\rho_i\right)\sqrt{\sigma}} \\ &= \mathrm{Tr}\sqrt{\sum_{i=1}^{l} p_i\sqrt{\sigma}\rho_i\sqrt{\sigma}} \\ &\leqslant \sum_{i=1}^{l}\sqrt{p_i}\,\mathrm{Tr}\sqrt{\sqrt{\sigma}\rho_i\sqrt{\sigma}} \\ &= \sum_{i=1}^{l}\sqrt{p_i}\,F(\sigma,\rho_i). \end{aligned} \tag{A7}$$

The second inequality in (A6) can be rewritten as

$$F(\sigma,\tau)^2 \leqslant 1 - \tfrac{1}{4}\|\sigma - \tau\|^2, \tag{A8}$$

which combined with the hypothesis $\|\rho_i - \sigma\| \geqslant 2 - \epsilon$ leads to

$$F(\sigma,\rho_i) \leqslant \sqrt{1 - \tfrac{1}{4}(2-\epsilon)^2} = \sqrt{\epsilon - \tfrac{\epsilon^2}{4}} \leqslant \sqrt{\epsilon}. \tag{A9}$$

Inserting this bound into (A7) and using the Cauchy-Schwarz inequality yields

$$F(\sigma,\rho) \leqslant \sum_{i=1}^{l}\sqrt{p_i}\sqrt{\epsilon} \leqslant \sqrt{l\epsilon}. \tag{A10}$$

We are now in a position to appeal to the first of the Fuchs–van de Graaf inequalities (A6) to obtain

$$\tfrac{1}{2}\|\rho - \sigma\| \geqslant 1 - F(\sigma,\rho) \geqslant 1 - \sqrt{l\epsilon}, \tag{A11}$$

or $\|\rho - \sigma\| \geqslant 2 - 2\sqrt{l\epsilon}$, as needed for part 1 of the theorem.

For part 2 of the theorem, let us first reformulate the statement in the language of probability densities (non-negative functions with unit integral) and the usual L_1 norm $\|\cdot\|_1$.

Let g_i,h be probability densities satisfying $\|g_i - h\|_1 = \int|g_i - h| \geqslant 2 - \epsilon$ for $i = 1,\ldots,l$. Then, for any weights $\{p_i\}_{i=1}^{l}$,

$$\left\|\sum_{i=1}^{l} p_i g_i - h\right\|_1 \geqslant 2 - l\epsilon, \tag{A12}$$

and the inequality is sharp.

Now, since for any real u,v we have the identity $|u - v| = u + v - 2\min\{u,v\}$, the condition $\|g_i - h\|_1 = \int g_i + h - 2\min\{g_i,h\} \geqslant 2 - \epsilon$ translates to $\int\min\{g_i,h\} \leqslant \epsilon/2$. Accordingly, if $g = \sum_{i=1}^{l} p_i g_i$, then

$$\min\{g,h\} \leqslant \sum_{i=1}^{l}\min\{g_i,h\}, \tag{A13}$$

and so $\int\min\{g,h\} \leqslant l\epsilon/2$, which is again equivalent to $\|f - g\|_1 \geqslant 2 - l\epsilon$. This proves part 2; moreover, the proof provides a guide for constructing—for any instance of parameters—examples when the inequality (A12) (or (3)) is sharp.

While the "threshold for significance" in the bounds in (A2) and (A12) is roughly the same ($l\epsilon \ll 1$), the dependence on $l\epsilon$ as that quantity goes to 0 is different. What is interesting is that this difference between the classical and quantum settings is genuine and not just an artifact of the argument. This is the content of part 3 of the theorem, which shows that the $O(\sqrt{\epsilon})$ dependence in (A2) is optimal, at least in the case $l = 2$.

To simplify the exposition, let us first reformulate the problem by considering a slightly more general question.

What is the optimal function $\epsilon \mapsto \phi(\epsilon)$ such that whenever ρ_1,ρ_2,σ are positive semidefinite matrices whose trace is at most 1 and such that $\mathrm{Tr}\rho_i + \mathrm{Tr}\sigma - \|\rho_i - \sigma\| \leqslant \epsilon$ for $i = 1,2$, then $\mathrm{Tr}\rho + \mathrm{Tr}\sigma - \|\rho - \sigma\| \leqslant \phi(\epsilon)$ for any convex combination $\rho = p\rho_1 + (1-p)\rho_2$?

The point is that the optimal function ϕ for this relaxed problem is the same as for the original problem, where all the traces are required to be *equal to* 1, at the cost of increasing the dimension by 2. Indeed, if ρ_i,σ are as above, we may define states $\tilde\rho_i,\tilde\sigma$ by

$$\tilde\rho_i = \begin{bmatrix}\rho_i & 0 & 0\\ 0 & 1-\mathrm{Tr}\rho_i & 0\\ 0 & 0 & 0\end{bmatrix}, \quad \tilde\sigma = \begin{bmatrix}\sigma & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 1-\mathrm{Tr}\sigma\end{bmatrix}. \tag{A14}$$

It is then easy to see that $2 - \|\tilde\rho_i - \tilde\sigma\| = \mathrm{Tr}\rho_i + \mathrm{Tr}\sigma - \|\rho_i - \sigma\|$, and similarly for $\tilde\rho = p\tilde\rho_1 + (1-p)\tilde\rho_2$.

With this reformulation, it is enough to look at 2×2 matrices and $p_1 = p_2 = \frac{1}{2}$. Given $r \in [0,1]$, consider

$$\sigma = \begin{bmatrix}0 & 0\\ 0 & r\end{bmatrix}, \quad \rho_i = \begin{bmatrix}1-r & \pm\sqrt{r(1-r)}\\ \pm\sqrt{r(1-r)} & r\end{bmatrix}, \tag{A15}$$

where $i = 1$ corresponds to the plus sign and $i = 2$ to the minus. One directly checks that

$$\mathrm{Tr}\rho_i + \mathrm{Tr}\sigma - \|\rho_i - \sigma\| = 1 + r - \sqrt{1 + 2r - 3r^2}. \tag{A16}$$

On the other hand, if $\rho = \frac{1}{2}(\rho_1 + \rho_2)$, then

$$\mathrm{Tr}\rho + \mathrm{Tr}\sigma - \|\rho - \sigma\| = 2r. \tag{A17}$$

In our setting, this means that if $\epsilon := 1 + r - \sqrt{1 + 2r - 3r^2}$ (which covers all possible values $\epsilon \in [0,2]$ as r varies over $[0,1]$), then $\phi(\epsilon) \geqslant 2r$. Since $(2r)^2 \geqslant 2(1 + r - \sqrt{1 + 2r - 3r^2}) = 2\epsilon$ for $r \in [0,1]$, this shows that $\phi(\epsilon) \geqslant \sqrt{2\epsilon}$. In other words, for $l = 2$ one cannot have a lower bound in (A2) that is better than $2 - \sqrt{2\epsilon}$.

While this example does not directly address the case $l > 2$, we know that—already in the classical setting—one can not have a nontrivial bound if $l\epsilon$ is not small enough, and so the dependence of the bound in (A2) on l cannot be too far from optimal. ∎

Proposition 2: Assume that $\epsilon \geqslant 0$ and that the states ρ_i,σ_j satisfy $\|\rho_i - \sigma_j\| \geqslant 2-\epsilon$ for all $i \in \{1,\ldots,l_1\}$ and $j \in \{1,\ldots,l_2\}$. Then, for any probability distribution $\{p_i\}_{i=1}^{l_1}$ and $\{q_j\}_{j=1}^{l_2}$,

$$\left\| \sum_{i=1}^{l_1} p_i\rho_i - \sum_{j=1}^{l_2} q_j\sigma_j \right\| \geqslant 2 - 2\sqrt{l_1 l_2 \epsilon}. \tag{A18}$$

Proof. The proof is a straightforward extension of the previous argument. Using Rotfel'd inequality one more time for $F(\sigma,\rho_i)$ in (A7) gives us the following,

$$F(\sigma,\rho_i) \leqslant \sum_j^{l_2} \sqrt{q_j} F(\sigma_j,\rho_i). \tag{A19}$$

Since we are assuming $\|\rho_i - \sigma_j\| \geqslant 2-\epsilon$, the same calculation as in (A9) yields

$$F(\sigma_j,\rho_i) \leqslant \sqrt{\epsilon}. \tag{A20}$$

Therefore, combining (A7), (A19), and (A20) we are led to

$$\begin{aligned} F(\sigma,\rho) &\leqslant \sum_i^{l_1} \sqrt{p_i} \sum_j^{l_2} \sqrt{q_j} F(\sigma_j,\rho_i) \\ &\leqslant \sum_{i,j}^{l_1,l_2} \sqrt{p_i q_j}\sqrt{\epsilon} \leqslant \sqrt{l_1 l_2 \epsilon}. \end{aligned} \tag{A21}$$

It remains to use the first Fuchs–van de Graaf relation (A6) to obtain (A18). ∎

APPENDIX B: FRACTION OF DETERMINISM IN QM

Lemma 1: Let $\epsilon \geqslant 0$ and suppose that $\|\rho - \sigma\| \leqslant 2-\epsilon$. Then, for any POVM $\{M_a\}_{a=1}^k$, there exists an outcome a_0 such that

$$\mathrm{Tr}\big(M_{a_0}\rho\big) \geqslant \frac{\epsilon}{2k} \quad \text{and} \quad \mathrm{Tr}\big(M_{a_0}\sigma\big) \geqslant \frac{\epsilon}{2k}. \tag{B1}$$

Proof. We shall show that if, on the contrary, for all a we have either $\mathrm{Tr}(\rho M_a) < \frac{\epsilon}{2k}$ or $\mathrm{Tr}(\sigma M_a) < \frac{\epsilon}{2k}$, then

$$p_e < \frac{\epsilon}{4}, \tag{B2}$$

where p_e is the probability of error in distinguishing ρ versus σ with equal *a priori* probabilities. Since it is known that probability is given by the Helstrom relation [19],

$$p_e(\rho,\sigma) = \tfrac{1}{2} - \tfrac{1}{4}\|\rho - \sigma\|, \tag{B3}$$

this will contradict the hypothesis $\|\rho - \sigma\| \leqslant 2-\epsilon$ and prove the lemma.

To that end, let us define two sets: $I_\rho = \{a : \mathrm{Tr}(\sigma M_a) < \frac{\epsilon}{2k}\}$ and $I_\sigma = I \setminus I_\rho$, where I is the set of all indices a. By the above assumption, for all $a \in I_\sigma$ we have $\mathrm{Tr}(\rho M_a) < \frac{\epsilon}{2k}$. Our decision scheme will be now as follows: If $a \in I_\rho$ then the state is ρ, otherwise it is σ. With this decision scheme we have

$$\begin{aligned} p_e &\leqslant \frac{1}{2}\mathrm{Tr}\left(\sum_{a\in I_\rho} M_a\sigma\right) + \frac{1}{2}\mathrm{Tr}\left(\sum_{a\in I_\sigma} M_a\rho\right) \\ &< \frac{1}{2}|I_\rho|\frac{\epsilon}{2k} + \frac{1}{2}|I_\sigma|\frac{\epsilon}{2k} = \frac{\epsilon}{4}, \end{aligned} \tag{B4}$$

which shows (B2) and completes the proof of Lemma 1. ∎

Lemma 2: Let $\mathcal{E}_1 = \{p_i,\rho_i\}_{i=1}^{l_1}$ and $\mathcal{E}_2 = \{q_j,\sigma_j\}_{j=1}^{l_2}$ be two ensembles which give rise to the same density matrix. Given $\tilde{l}_1 \leqslant l_1$ and $\tilde{l}_2 \leqslant l_2$, we set

$$\delta_1 = 1 - \sum_{i=1}^{\tilde{l}_1} p_i, \quad \delta_2 = 1 - \sum_{j=1}^{\tilde{l}_2} q_j, \tag{B5}$$

and define new ensembles,

$$\tilde{\mathcal{E}}_1 = \{\tilde{p}_i,\rho_i\}_{i=1}^{\tilde{l}_1}, \quad \tilde{\mathcal{E}}_2 = \{\tilde{q}_j,\sigma_j\}_{j=1}^{\tilde{l}_2}, \tag{B6}$$

where $\tilde{p}_i = p_i/(1-\delta_1)$ and $\tilde{q}_j = q_j/(1-\delta_2)$. Then the ensembles $\tilde{\mathcal{E}}_1,\tilde{\mathcal{E}}_2$ satisfy

$$\left\| \sum_{i=1}^{\tilde{l}_1} \tilde{p}_i\rho_i - \sum_{j=1}^{\tilde{l}_2} \tilde{q}_j\sigma_j \right\| \leqslant \frac{2\max\{\delta_1,\delta_2\}}{1-\min\{\delta_1,\delta_2\}}. \tag{B7}$$

Proof. Since $\sum_{i=1}^{l_1} p_i\rho_i = \sum_{j=1}^{l_2} q_j\sigma_j$, it follows from the triangle inequality that

$$\begin{aligned} \left\| \sum_{i=1}^{\tilde{l}_1} p_i\rho_i - \sum_{j=1}^{\tilde{l}_2} q_j\sigma_j \right\| &\leqslant \left\| \sum_{i=\tilde{l}_1+1}^{l_1} p_i\rho_i \right\| + \left\| \sum_{j=\tilde{l}_2+1}^{l_2} q_j\sigma_j \right\| \\ &= \delta_1 + \delta_2. \end{aligned} \tag{B8}$$

Let $\theta_1 = 1-\delta_1, \theta_2 = 1-\delta_2$. Then

$$\begin{aligned} (1-\delta_1)&\left\| \sum_{i=1}^{\tilde{l}_1} \tilde{p}_i\rho_i - \sum_{j=1}^{\tilde{l}_2} \tilde{q}_j\sigma_j \right\| \\ &= \left\| \theta_1 \sum_{i=1}^{\tilde{l}_1} \tilde{p}_i\rho_i - \theta_1 \sum_{j=1}^{\tilde{l}_2} \tilde{q}_j\sigma_j \right\| \\ &= \left\| \theta_1 \sum_{i=1}^{\tilde{l}_1} \tilde{p}_i\rho_i - \theta_2 \sum_{j=1}^{\tilde{l}_2} \tilde{q}_j\sigma_j + (\theta_2-\theta_1)\sum_{j=1}^{\tilde{l}_2} \tilde{q}_j\sigma_j \right\| \\ &\leqslant \left\| \sum_{i=1}^{\tilde{l}_1} p_i\rho_i - \sum_{j=1}^{\tilde{l}_2} q_j\sigma_j \right\| + |\delta_1 - \delta_2|. \end{aligned} \tag{B9}$$

Combining (B8) and (B9) we are led to

$$\left\| \sum_{i=1}^{\tilde{l}_1} \tilde{p}_i\rho_i - \sum_{j=1}^{\tilde{l}_2} \tilde{q}_j\sigma_j \right\| \leqslant \frac{\delta_1+\delta_2+|\delta_1-\delta_2|}{1-\delta_1}. \tag{B10}$$

Finally, using the identity $a+b+|a-b| = 2\max\{a,b\}$ and noticing that the same estimate holds if we replace δ_1 with δ_2, we obtain the required estimate.

APPENDIX C: FOD AND CF FOR $l = 2$

For a quantum box with $2 \times n$ inputs and $2 \times k$ outputs for, respectively, Bob and Alice, the FOD and the CF were estimated in Sec. V as follows:

$$\text{FOD} \geqslant \frac{1}{2k} \max\{\min\{p_0\epsilon_{00},q_0\epsilon_{00}\}, \min\{p_1\epsilon_{11},q_1\epsilon_{11}\}, \min\{p_0\epsilon_{01},q_1\epsilon_{01}\}, \min\{p_1\epsilon_{10},q_0\epsilon_{10}\}\},$$

and

$$\text{CF} \geqslant \frac{1}{2k} \max\{\min\{p_0\epsilon_{00},q_0\epsilon_{00}\} + \min\{p_1\epsilon_{11},q_1\epsilon_{11}\}, \min\{p_0\epsilon_{01},q_1\epsilon_{01}\} + \min\{p_1\epsilon_{10},q_0\epsilon_{10}\}\},$$

where $\epsilon_{ij} = 2 - \|\rho_i - \sigma_j\|$.

The above bounds are independent of the choices of measurements of Alice. We would like to find a bound for the FOD and the CF, which only depends on the number of outcomes of Alice. To achieve this, we have to optimize over p_0,q_0, ρ_i's, and σ_j's.

By renaming the labels, we may assume that

$$p_0 \geqslant q_0 \geqslant q_1 \geqslant p_1, \tag{C1}$$

in which case the above estimates simplify to

$$\text{FOD} \geqslant \frac{1}{2k} \max\{q_0\epsilon_{00}, p_1\epsilon_{11}, q_1\epsilon_{01}, p_1\epsilon_{10}\}, \tag{C2}$$

and

$$\text{CF} \geqslant \frac{1}{2k} \max\{q_0\epsilon_{00} + p_1\epsilon_{11}, q_1\epsilon_{01} + p_1\epsilon_{10}\}. \tag{C3}$$

To come up with explicit bounds, we first note that Corollary 1 implies that $\max_{ij} \epsilon_{ij} \geqslant \frac{1}{4}$. Another bound can be obtained from Lemma 2 by truncating from our ensembles the elements with smaller probabilities (i.e., in view of (C1), $\{p_1,\rho_1\}$ and $\{q_1,\sigma_1\}$). The truncated ensembles are then simply $\{1,\rho_0\}$ and $\{1,\sigma_0\}$, and since $\delta_1 = p_1$ and $\delta_2 = q_1$, the conclusion of Lemma 2 translates to

$$\epsilon_{00} = 2 - \|\rho_0 - \sigma_0\| \geqslant 2\left(1 - \frac{\max\{p_1,q_1\}}{1 - \min\{p_1,q_1\}}\right) = 2\left(1 - \frac{q_1}{p_0}\right). \tag{C4}$$

To produce a lower bound for the FOD we consider now the following two cases.

(1) If $\epsilon_{00} \geqslant 1/4$, then

$$\text{FOD} \geqslant \frac{1}{2k} \max\left\{\frac{q_0}{4}, 2q_0\left(1 - \frac{q_1}{p_0}\right)\right\}. \tag{C5}$$

(2) If $\epsilon_{00} < 1/4$, then some ϵ_{ij} other than ϵ_{00} is greater than or equal to $\frac{1}{4}$ and so

$$\text{FOD} \geqslant \frac{1}{2k} \max\left\{\frac{p_1}{4}, 2q_0\left(1 - \frac{q_1}{p_0}\right)\right\}. \tag{C6}$$

Since $\frac{p_1}{4} \leqslant \frac{q_0}{4}$ in view of our assumption (C1), a lower bound for the FOD is given by (C6), the smaller of the two values obtained in cases 1 and 2.

All that remains is to minimize over all p_i's and q_j's satisfying the constraints, i.e., $0 \leqslant p_1 \leqslant q_1 \leqslant \frac{1}{2}$. We have

$$\min_{0 \leqslant p_1 \leqslant q_1 \leqslant \frac{1}{2}} \max\left\{\frac{p_1}{4}, 2q_0\left(1 - \frac{q_1}{p_0}\right)\right\} \geqslant \frac{5 - \sqrt{17}}{8} \geqslant 0.10961, \tag{C7}$$

and, therefore,

$$\text{FOD} \geqslant \frac{0.10961}{2k}. \tag{C8}$$

We can similarly estimate CF, which in the present setting verifies

$$\text{CF} \geqslant \frac{0.11226}{2k}. \tag{C9}$$

[1] J. S. Bell, Physics (NY) **1**, 195 (1964).

[2] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, Rev. Mod. Phys. **82**, 665 (2010).

[3] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002).

[4] A. Acin, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).

[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[6] G. Brassard, R. Cleve, and A. Tapp, Phys. Rev. Lett. **83**, 1874 (1999).

[7] G. Brassard, A. Broadbent, and A. Tapp, Found. Phys. **35**, 1877 (2005).

[8] B. Tsirelson, Lett. Math. Phys. **4**, 93 (1980).

[9] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).

[10] R. Ramanathan, J. Tuziemski, M. Horodecki, and P. Horodecki, arXiv:1410.0947.

[11] J. Oppenheim and S. Wehner, Science **330**, 1072 (2010).

[12] R. Cleve, P. Hoyer, B. Toner, and J. Watrous, in *Proceedings of 19th IEEE Annual Conference on Computational Complexity (CCC 2004)* (IEEE, New York, 2004), pp. 236–249, arxiv:quant-ph/0404076.

[13] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).

[14] H. Buhrman, O. Regev, G. Scarpa, and R. de Wolf, Theory of Computing **8**, 623 (2012).

[15] M. Junge, C. Palazuelos, D. Pérez-García, I. Villanueva, and M. M. Wolf, Commun. Math. Phys. **300**, 715 (2010).

[16] N. Gisin, A. A. Methot, and V. Scarani, Int. J. Quantum. Inform. **05**, 525 (2007).

[17] J. Barrett, A. Kent, and S. Pironio, Phys. Rev. Lett. **97**, 170409 (2006).

[18] N. Brunner, D. Cavalcanti, A. Salles, and P. Skrzypczyk, Phys. Rev. Lett. **106**, 020402 (2011).

[19] C. W. Helstrom, J. Stat. Phys. **1**, 231 (1969).

[20] J.-C. Bourin, Proc. Am. Math. Soc. **138**, 495 (2010).

[21] C. A. Fuchs and J. van de Graaf, IEEE Trans. Inf. Theory **45**, 1216 (1999).