General properties of nonsignaling theories

Ll. Masanes,[1] A. Acin,[2] and N. Gisin[3]
[1]*Department of Mathematics, University of Bristol, BS8 1TW Bristol, United Kingdom*
[2]*ICFO-Institut de Ciències Fotòniques, 08034 Barcelona, Spain*
[3]*GAP-Optique, University of Geneva, 20 Rue de l'École de Médicine, CH-1211 Geneva 4, Switzerland*
(Received 18 September 2005; published 30 January 2006)

This article identifies a series of properties common to all theories that do not allow for superluminal signaling and predict the violation of Bell inequalities. Intrinsic randomness, uncertainty due to the incompatibility of two observables, monogamy of correlations, impossibility of perfect cloning, privacy of correlations, bounds in the shareability of some states; all these phenomena are solely a consequence of the no-signaling principle and nonlocality. In particular, it is shown that for any distribution, the properties of (i) nonlocal, (ii) no arbitrarily shareable, and (iii) positive secrecy content are equivalent.

DOI: 10.1103/PhysRevA.73.012112 PACS number(s): 03.65.Ud, 03.65.Ta

I. INTRODUCTION

There are two experimental facts that, when considered together, significantly restrict any possible physical theory that aims to account for them. The first one is the constancy of the speed of light in any reference frame. This implies that no signal carrying information can propagate faster than light. More generally, we refer to the impossibility of sending information arbitrarily fast as the no-signaling principle. The second fact is the existence of correlations between spacelike separated events that violate Bell inequalities [1,2]. This means that such correlations cannot be explained by strategies arranged in the past. Models accounting for such correlations can be constructed by assuming some signaling between the correlated events. But this seems to contradict the first experimental fact. This is the reason why such correlations are called nonlocal. Despite this, physical theories exist that predict the violation of Bell inequalities and are nonsignaling, an example being quantum mechanics (QM).

QM is not the unique theory consistent with the two mentioned experimental facts. It is well known that nonsignaling correlations exist that are more nonlocal than the ones predicted by QM. Indeed, Popescu and Rohrlich proved that there are nonsignaling correlations giving a Bell inequality violation larger than the quantum-mechanical prediction [3]. This suggests the possible existence of theories, different from QM, that allow for Bell inequality violation without contradicting the no-signaling principle. Although there is no experimental reason to reject QM, it is highly desirable to know the nature of these alternative theories in order to "study quantum physics from the outside." In this article, we aim at providing a unified picture for the static part (we do not consider dynamics) of all such theories, identifying a series of features common to all of them.

Analyzing these common properties can be very useful in gaining a better understanding of QM. It is often said that the postulates of QM do not have a clear physical meaning, especially when compared with the postulates of other theories, such as relativity or thermodynamics. The postulates of QM imply no signaling (if we assume locality of interactions) and nonlocality. It was proposed by Popescu and Rohrlich to consider no signaling and the existence of nonlocal correlations as proper physical principles. Could these two principles, together with other independent postulates, imply QM? How would these other postulates look? For such an enterprise, it is very important to learn all the consequences that follow from these two principles without any extra assumption.

From an information-theoretical point of view, it is also worth looking at a framework more general than QM, as illustrated by several recent works analyzing the use of nonlocal correlations as an information-theoretical resource [4–6]. This is of particular interest in the case of secret communication: there, the security of a protocol relies on some assumptions on the eavesdropper capabilities. Usually, it is assumed that her computational power is bounded, or that her action is constrained by QM laws. It is then desirable to weaken the strength of these assumptions as much as possible. In this sense, a secret key distribution was recently proposed in Ref. [7] and its security proved solely using the no-signaling principle. In this article, we extend the connection between nonlocality and secrecy at the level of an equivalence. Notice that the fact that a probability distribution contains secrecy does not imply that it can be distilled into a secret key (see below).

Summary and results. The article is organized as follows. In Sec. II nonsignaling correlations are introduced, local and nonlocal ones are distinguished. Special emphasis is made on a particular family of distributions that we call isotropic, which will prove very useful in later reasonings.

In Sec. III, different aspects of monogamy in nonlocal correlations are presented. In particular, the complete equivalence between locality and infinite shareability is proven (Sec. III A). In Sec. III B, through some examples, we survey the complex structure of the monogamy relations.

In Sec. IV we prove that any nonsignaling theory that predicts the violation of a Bell inequality has a no-cloning theorem. Some additional analysis is made for the case of QM.

In Sec. V we prove that nonsignaling correlations contain secrecy (in the sense of cost) if and only if they are nonlocal. In Sec. VI we review the fact that all nonlocal correlations must have nondeterministic outcomes. And, in Sec. VI A we show that the more incompatible two observables are, the more uncertain their outcomes.

Finally, we conclude with some final remarks, exposing some open question. Some additional material and proofs is contained in the appendixes.

II. DEFINITIONS AND GENERAL FRAME

Consider n parties—Alice, Bob, Clare,…—each possessing a physical system, which can be measured with different observables. Denote by x_k the observable chosen by party k, and by a_k the corresponding measurement outcome. The joint probability distribution for the outcomes, conditioned on the observables chosen by the n parties is

$$P(a_1, \dots, a_n | x_1, \dots, x_n). \tag{1}$$

One can formulate this scenario in an equivalent and slightly more abstract way. Imagine that each of the n parties has a physical device with an input and an output. Just after the kth party inputs x_k, the device outputs a_k, and it cannot be used anymore. Throughout this article, we assume that inputs and outputs take values from finite, but arbitrarily large, alphabets $x_k \in \{0,1,\dots,X_k-1\}$ and $a_k \in \{0,1,\dots,A_k-1\}$. Notice that, without loss of generality, we assume that all observables belonging to one party have the same number of outcomes.

It is useful to look at these conditioned probability distributions (1) as points in a large dimensional space. The set of all these points (1) is a convex polytope. Unless no other constraints are imposed, (1) can be any vector of positive numbers, satisfying the normalization conditions

$$\sum_{a_1,\dots a_n} P(a_1, \dots, a_n | x_1, \dots, x_n) = 1 \tag{2}$$

for all input values $x_1, \dots, x_n$.

A. Nonsignaling correlations

The n-partite distribution $P(a_1,\dots,a_n|x_1,\dots,x_n)$ is nonsignaling, when the marginal distribution for each subset of parties $\{a_{k_1},\dots,a_{k_m}\}$ only depends on its corresponding inputs

$$P(a_{k_1}, \dots, a_{k_m} | x_1, \dots, x_n) = P(a_{k_1}, \dots, a_{k_m} | x_{k_1}, \dots, x_{k_m}). \tag{3}$$

It turns out that very few of these conditions are linearly independent. It was proved in Ref. [4] that all conditions of the form (3) can be derived from the following.

Condition. For each $k \in \{1,\dots,n\}$ the marginal distribution obtained when tracing out a_k is independent of x_k:

$$\sum_{a_k} P(a_1, \dots, a_k, \dots, a_n | x_1, \dots, x_k, \dots, x_n) = \sum_{a_k} P(a_1, \dots, a_k, \dots, a_n | x_1, \dots, x'_k, \dots, x_n), \tag{4}$$

for all values of $a_1,\dots,a_{k-1},a_{k+1},\dots,a_m$ and $x_1,\dots,x_{k-1},x_k,x'_k,x_{k+1},\dots,x_n$.

These linear constraints characterize an affine set. The intersection of this set with the polytope of distributions (1) gives another convex polytope. Throughout this article, whenever we refer to distributions, correlations, states or points, we always assume they belong to the nonsignaling polytope.

B. Local correlations

Local correlations are the ones that can be generated if the parties share classical information, or equivalently, the ones that can be written as

$$P(a_1, \dots, a_n | x_1, \dots, x_n) = \sum_e P(e) P(a_1|x_1,e) \cdots P(a_n|x_n,e). \tag{5}$$

This subset of correlations is a convex polytope delimited by two kinds of facets. The first kind warrants that all the components of Eq. (5) are positive, and thus, they are also facets of the nonsignaling polytope. The second kind of facets can be violated by nonlocal correlations, and are called Bell inequalities. For a complete introduction to Bell inequalities, polytopes, and related topics see Ref. [8]. Throughout this article we assume that all Bell inequalities have been normalized (with a transformation of the form $\mathcal{B} \to \alpha\mathcal{B}+\beta$, where α and β are real numbers), such that the local bound is $\mathcal{B}[P_{\text{local}}] \leq 0$, and the maximal violation compatible with no signaling is $\mathcal{B}[P_{\max}]=1$.

As said above, local correlations can be generated with shared randomness and local operations. In expression (5), the random variable e stands for the information shared among the parties, sometimes called local hidden variable. Depending on its value, the kth party locally generates $P(a_k|x_k,e)$. The distributions that cannot be written similar to Eq. (5) are called nonlocal.

C. Quantum correlations

We call quantum those correlations that can be generated if the parties share quantum information [entanglement], or equivalently, those correlations that can be written as

$$P(a_1, \dots, a_n | x_1, \dots, x_n) = \text{tr}[F_{a_1}^{(x_1)} \otimes \cdots \otimes F_{a_n}^{(x_n)} \rho], \tag{6}$$

where ρ is a quantum state, namely, a unit trace, semidefinite positive matrix, and $\{F_0^{(x_k)},\dots,F_{A_k-1}^{(x_k)}\}$ define what is called a positive operator valued measure [9]. That is, a set of positive operators $\{F_{a_k}^{(x_k)}\}$ satisfying $\Sigma_{a_k} F_{a_k}^{(x_k)} = \mathbb{1}, \forall\ x_k$.

D. Isotropic correlations

Let us define a particular family of bipartite distributions with binary input/output. In the case where the marginal distributions for a and b are unbiased, all the information of $P(a,b|x,y)$ is contained in the four correlation functions

$$C_{xy} = +P(0,0|x,y) + P(1,1|x,y) - P(0,1|x,y) - P(1,0|x,y), \tag{7}$$

for $xy=00,01,10,11$. One can always fix $C_{00},C_{01},C_{10} \geq 0$ by performing local reversible transformations. Once we

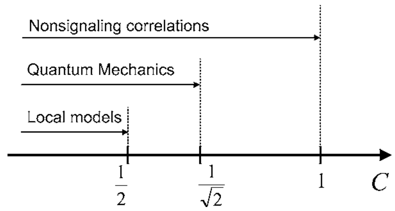

FIG. 1. Value of C for isotropic correlations (12).

have a distribution in this canonical form, its nonlocality is decided by the CHSH inequality [10]

$$\mathcal{B}_{\text{CHSH}} = \frac{1}{2}[C_{00} + C_{01} + C_{10} - C_{11}] - 1, \tag{8}$$

where we have written it in our standard form. We call isotropic, denoted by $P_{\text{iso}}(a,b|x,y)$, those correlations with unbiased marginal distributions for a and b that satisfy

$$C_{00} = C_{01} = C_{10} = -C_{11} \geqslant 0. \tag{9}$$

This family depends on a unique parameter $C = C_{00}$, whose relation to the CHSH violation is

$$\mathcal{B}_{\text{CHSH}}[P_{\text{iso}}] = 2C - 1. \tag{10}$$

In Fig. 1 we can see for which values of C the distribution P_{iso} belongs to the local and quantum set. When $C=1$, this distribution is known as a PR box [3,4], and is usually written as

$$P_{\text{PR}}(a,b|x,y) = \begin{cases} 1/2 & \text{if } a+b \bmod 2 = xy, \\ 0 & \text{otherwise}. \end{cases} \tag{11}$$

This distribution can be considered the paradigm of nonlocal, nonsignaling correlations (see Ref. [11]). With this definition, we can express any P_{iso} as the following mixture:

$$P_{\text{iso}} = CP_{\text{PR}} + (1 - C)P_N^A P_N^B, \tag{12}$$

where P_N^A is the local noise distribution for Alice, independently of the inputs. Thus, one can interpret C as the probability of sharing a PR box instead of local noise.

III. MONOGAMY OF NONLOCAL CORRELATIONS

While classical correlations can be shared among an indefinite number of parties, it is well known that quantum correlations cannot. This fact is often called monogamy of entanglement [12]. In this section we prove that this is a generic feature of all nonsignaling theories.

First, let us recall a result already mentioned in Ref. [4]. All Bell inequalities for which the maximal violation consistent with no signaling is attained by a unique distribution, have monogamy constraints. Suppose that $\mathcal{B}$ is a Bell inequality with unique maximal violator $P_{\max}$. If Alice-Bob maximally violate this inequality $\mathcal{B}[P(a,b|x,y)]=1$, then, Alice and Clare are completely uncorrelated. To prove this, first notice that because all Bell inequalities $\mathcal{B}[P]$ are linear in P, $P_{\max}$ must be an extreme of the Alice-Bob polytope. Otherwise, the maximal violator would not be unique. Second, using the definition of marginal distribution and the no-signaling condition we have

$$\begin{aligned} P_{\max}(a,b|x,y) &= \sum_c P(a,b,c|x,y,z) \\ &= \sum_c P(a,b|x,y,z,c)P(c|x,y,z) \\ &= \sum_c P(a,b|x,y,z,c)P(c|z), \end{aligned} \tag{13}$$

for all z. But, because $P_{\max}(a,b|x,y)$ is extremal, any such decompositions must consist of only one term. This implies that Clare is uncorrelated with Alice and Bob.

Actually, one can prove that all the CGLMP inequalities have a unique nonsignaling probability distribution achieving its algebraic maximum. This well-known set of inequalities was first proposed in Ref. [13] for the case of two inputs of d possible outputs. One can easily see that imposing nosignaling and maximal violation of CGLMP inequality identifies a unique probability distribution $P(a,b|x,y)$. This means that this set of Bell inequalities have the previous monogamy condition.

A. *m*-shareability and locality

Shareability represents a natural property in the analysis of the monogamy of correlations. A bipartite probability distribution $P(a,b|x,y)$ is said to be m shareable with respect to Bob, if there exists an $(m+1)$-partite distribution $P(a,b_1,\ldots,b_m|x,y_1,\ldots,y_m)$ being symmetric with respect to $(b_1,y_1)\cdots(b_m,y_m)$, with marginals $P(a,b_i|x,y_i)$ equal to the original distribution $P(a,b|x,y)$. The following result shows the relation between shareability and nonlocality.

Result 1. If $P(a,b|x,y)$ is m-shareable with respect to Bob, then it satisfies all Bell inequalities with m (*or less*) different values for the input y.

Proof. To prove this statement we construct a local model for $P(a,b|x,y)$ when y is constrained to $y=1,\ldots,m$ (without loss of generality). By assumption $P(a,b_1,\ldots,b_m|x,y_1,\ldots,y_m)$ with the abovementioned properties exists, then so the marginal $P(b_1,\ldots,b_m|y_1,\ldots,y_m)$ and the conditional $P(a|x,b_1,\ldots,b_m,y_1,\ldots,y_m)$ do. In this local model, the information shared by the parties, that is the variable e in Eq. (5), is the string $e=(b_1,\ldots,b_m)$, when the corresponding inputs are fixed to $y_1=1,\ldots,y_m=m$. Thus, using the definition of conditional probabilities, we can decompose $P(a,b|x,y)$ in the following way:

$$\begin{aligned} P(a,b|x,y) = \sum_{b_1,\ldots,b_m} & P(b_1,\ldots,b_m|y_1=1,\ldots,y_m=m) \\ & \times P(a|x,b_1,\ldots,b_m,y_1=1,\ldots,y_m=m) \times \delta_{b,b_y}, \end{aligned} \tag{14}$$

where in δ_{b,b_y} the variables b and y are the ones appearing in $P(a,b|x,y)$, and b_y is the yth component of the shared information $(b_1,\ldots,b_m)$. The three factors in each term of Eq. (14) have to be interpreted as the $P(e)$, $P(a|x,e)$ and $P(b|y,e)$ appearing in the decomposition (5), respectively.

Note that this result represents the extension of theorem 2 in Ref. [14], derived for quantum states, to the more general nonlocal scenario. It also implies that if a state is X and Y shareable with respect to Alice and Bob, then it is local. In particular, two-shareable states do not violate the CHSH or CGMLP inequalities.

A converse of the previous result is also true: if a state is local, then it is ∞ shareable with respect to any party. To show the last statement, we explicitly construct the extension (to m Bobs) for the arbitrary local correlations written in Eq. (5):

$$P(a,b_1,\ldots,b_m|x,y_1,\ldots,y_m) = \sum_e P(e)P(a|x,e)P(b_1|y_1,e)\cdots P(b_m|y_m,e), \quad (15)$$

with each distribution $P(b_i|y_i,e)$ being equal to the $P(b|y,e)$ that appears in Eq. (5). We can merge the previous two statements into the following one.

Result 2. Locality and ∞ shareability are equivalent properties.

This result is analogous to what happens in QM: a bipartite quantum state is ∞ shareable if and only if it is separable [15].

B. Examples

In what follows, we show that the CHSH inequality presents an even stronger kind of monogamy.

Result 3. Consider a binary input and output tripartite distribution $P(a,b,c|x,y,z)$. If Alice and Bob's marginal is nonlocal, then Alice and Clare's marginal must be local:

$$\mathcal{B}_{\text{CHSH}}[P(a,b|x,y)] > 0 \Rightarrow \mathcal{B}_{\text{CHSH}}[P(a,c|x,z)] \leq 0.$$

Proof. We prove this statement by contradiction. Suppose that there exists a tripartite distribution $P(a,b,c|x,y,z)$ such that both $P(a,b|x,y)$ and $P(a,c|x,z)$ are nonlocal. Then Alice-Bob, and simultaneously Alice-Clare, can depolarize their bipartite correlations and transform them into isotropic ones, without decreasing the Bell violation. This procedure is shown in Appendix A. Then, if Alice-Bob have larger C than Alice-Clare, Bob decreases it until both are equal (this procedure is also explained in Appendix A). An analogous thing is done in the opposite situation. After this manipulation, both marginals are isotropic and have the same value of C. This implies that the two marginals are equal, and thus two shareable. In Sec. III A we have seen that a two shareable state with binary inputs is always local. This finishes the construction of the contradiction.

In more general situations strict monogamy no longer holds. Indeed, one can easily design a situation where Alice shares a PR box with Bob, and another with Clare. This corresponds to a case where Alice can choose between four inputs and four outputs, while Bob and Clare are restricted to the simplest case of $Y=Z=B=C=2$. Clearly, the corresponding Alice-Bob and Alice-Clare distribution violate the CHSH inequality. A nicer and more symmetric example, with only two inputs for each party, is given by the following tripartite distribution

$$P^{ABC} = \frac{1}{2}P^{AB}_{\text{PR}\{0,1\}}P^{C}_{N\{0,1\}} + \frac{1}{2}P^{AC}_{\text{PR}\{2,3\}}P^{B}_{N\{2,3\}}, \quad (16)$$

where $P_{\text{PR}\{\alpha,\beta\}}$ is a PR box with outputs restricted to $a,b \in \{\alpha,\beta\}$, $P_{N\{\alpha,\beta\}}$ is a local noise distribution with outputs restricted to $a,b \in \{\alpha,\beta\}$, and the superindices label the parties. In what follows, we prove that the Alice-Bob marginal

$$P^{AB} = \frac{1}{2}P^{AB}_{\text{PR}\{0,1\}} + \frac{1}{2}P^{A}_{N\{2,3\}}P^{B}_{N\{2,3\}}, \quad (17)$$

is nonlocal. Assume the opposite: P^{AB} can be expressed as a mixture of local extreme points (5). Because each local extreme point has determined outcomes, we can split the local mixture into a part with outcomes $\{2,3\}$, and a part with outcomes $\{0,1\}$. The last would correspond to a local expansion of $P^{AB}_{\text{PR}\{0,1\}}$, but we know that such thing does not exist. Now, using the symmetry of Eq. (16), we conclude that its marginals P^{AB} and P^{AC} are both nonlocal.

In the case $X=Y=2$ and A,B arbitrary, there is a situation where strong monogamy still holds: where the reduced states of Alice-Bob and Alice-Clare consist both on isotropic correlations with nonuniform noise (independent of the inputs). First, let us generalize the idea of isotropic distributions for arbitrary output alphabets. The generalization of the PR box is [4]

$$P_{\text{PR}}(a,b|x,y) = \begin{cases} 1/A & \text{if } a-b \bmod A = xy, \\ 0 & \text{otherwise}. \end{cases} \quad (18)$$

In a natural way, we define

$$P^{AB}_{\text{iso}} = CP^{AB}_{\text{PR}} + (1-C)P^{A}_{\text{ind}}P^{B}_{\text{ind}}, \quad (19)$$

where P^{A}_{ind} is an arbitrary local distribution for Alice, independent of the inputs. It is clear that if Alice and Bob add to their outputs a shared random number modulo A:

$$a \to a + r \bmod A, \quad (20)$$

$$b \to b + r \bmod A, \quad (21)$$

their distribution becomes

$$P^{AB}_{\text{iso}} \to CP^{AB}_{\text{PR}} + (1-C)P^{A}_{N}P^{B}_{N}, \quad (22)$$

where $P^{A/B}_{N}$ is the (local) uniform distribution independent of the inputs x/y. As in the case $A=B=2$, if C is positive, one of the parties can decrease its value by performing a local operation. Using the same trick as before, one can prove that all tripartite distributions where the marginals Alice-Bob and Alice-Clare are both isotropic with nonuniform noise (19), show strong monogamy.

IV. NO CLONING

The quantum no-cloning theorem represents one of the cornerstones of quantum information theory. It is usually explained as a consequence of the nonorthogonality of quantum states and the linearity of quantum time evolution. The relation between quantum cloning and no signaling has also been studied by several authors. Indeed, if one assumes that

(i) states are described by vectors in Hilbert spaces, (ii) probabilities are obtained according to the usual trace rule, and (iii) no-signaling, the optimal fidelity of a cloning machine cannot be larger than the one allowed by quantum dynamics [16]. In what follows, we formulate the problem independently of QM and show the following.

Result 4. All nonsignaling theories predicting the violation Bell inequalities have a no-cloning theorem.

A similar result was proved for the case of the CHSH inequality by Werner, in Ref. [17]. Here we prove it for general nonlocal theories, not necessarily violating the CHSH inequality. Suppose that there exists a machine to which we can input a physical system (in an arbitrary state), and it outputs two systems in exactly the same state as the original one. We do not make any restriction on how the output systems are correlated, as long as its reduced state is identical to the input one. We call such an engine a perfect cloning machine. Let us consider the following situation: Alice and Bob share the nonlocal distribution $P(a,b|x,y)$, and perform the following two spacelike separated events. On one site, Alice chooses the input x_0 and obtains the output a_0. On the other site, Bob performs m clones of its original system. For an observer who sees first the event on Alice's site, the description of Bob's input system is $P(b|y,x_0,a_0)$. For this observer, Bob's system is completely uncorrelated with the rest of the universe, and the functioning of the perfect cloning machine is unambiguous:

$$P(b|y,x_0,a_0) \rightarrow P(b_1,\ldots,b_m|y_1,\ldots,y_m,x_0,a_0). \tag{23}$$

As mentioned above, we do not make any restriction to the joint state of all clones $P(b_1,\ldots,b_m|y_1,\ldots,y_m,x_0,a_0)$ as long as when tracing all systems but one the distribution $P(b_i|y_i,x_0,a_0)$ is the same as the original one $P(b|y,x_0,a_0)$. Because we consider a perfect cloning machine there is no distinction between pure and mixed states: all are perfectly cloned. For an observer who first sees Bob's operation, its description of the physical situation is

$$P(a,b_1,\ldots,b_m|x,y_1,\ldots,y_m). \tag{24}$$

But, because all descriptions must give consistent predictions, the descriptions from the point of view of the two mentioned observers (23) and (24) must be the same, up to conditioning on a. This implies that the original distribution $P(a,b|x,y)$ is m shareable. More concretely, because m is arbitrary, we can say that $P(a,b|x,y)$ is ∞ shareable. According to the result of Sec. III A, the original distribution $P(a,b|x,y)$ must be local, in contradiction with the initial assumption.

Phase covariant cloning machine. Once we have ruled out the existence of a perfect cloning machine, it is interesting to look for the optimal imperfect one. Suppose that its action is

$$P(a,b|x,y) \rightarrow P(a,b_1,b_2|x,y_1,y_2), \tag{25}$$

where, without loss of generality we can assume that the final distribution is symmetric with respect to (b_1,y_1) and (b_2,y_2). By definition, the reduced distribution $P(a,b_i|x,y_i)$ is two shareable. This implies that it cannot violate any two-input Bell inequality. In particular, if the initial distribution $P(a,b|x,y)$ has $Y=2$, the resulting clones are correlated with Alice's system in a local way.

Let us consider a particular case in the binary input-output scenario. Consider that Alice and Bob share an isotropic distribution with parameter C. Bob clones his subsystem, and, according to the previous paragraph, the resulting clones are locally correlated with Alice's subsystem. If we suppose that the clones are isotropically correlated with Alice, the maximum value for their parameter is $C_{\rm cln}=1/2$. Thus, the *shrinking factor* associated to this cloning operation is

$$\frac{C_{\rm cln}}{C} = \frac{1}{2C}. \tag{26}$$

Now, consider the isotropic correlations that arise when measuring a singlet with the observables that maximize the CHSH violation, that is $P_{\rm iso}$ with $C=1/\sqrt{2}$. In this case, the shrinking factor (26) coincides with the one of the phase covariant quantum cloning machine $1/\sqrt{2}$ [18], that is QM attains this maximum value for the cloning of nonlocal correlations. In this sense, QM clones the quantum correlations achieving the Cirelson bound in an optimal way.

V. NONLOCALITY AND PRIVACY

The monogamy of correlations and the impossibility of perfect cloning seem immediately to be related to the concept of privacy. If two honest parties know to share correlations with some degree of monogamy, they can estimate and possibly bound their correlations with a third dishonest party, the eavesdropper. In this section we strengthen this intuitive idea, proving that under the no-signaling assumption, a probability distribution contains secrecy if and only if it is nonlocal. Recall that this does not mean that this probability distribution can be transformed into a secret key.

For the sake of simplicity we consider the bipartite case. In a cryptographic scenario, one usually considers two honest parties (Alice and Bob) each possessing a random variable A and B, and an eavesdropper (Eve) having E. The correlations among the three random variables are described by a probability distribution P_{ABE}. On the other hand, it is meant by nonlocal correlations those probability distributions conditioned on some inputs $P(a,b|x,y)$ that cannot be written in the form of Eq. (5). It is in principle not so evident how to relate the two scenarios. For instance, how to add (i) the third party in the nonlocal scenario or (ii) the missing inputs for Alice and Bob in the cryptographic scenario. Therefore, before proving the equivalence between privacy and nonlocality one has to connect the two considered scenarios.

A. Secret correlations

A tripartite probability distribution (without inputs) P_{ABE} among two honest parties and an eavesdropper contains secrecy when it cannot be generated by local operations and public communication (LOPC), i.e., its formation requires the use of a private channel or secret bits [19]. On the other hand, P_{ABE} can be generated by LOPC, if there exists a stochastic map $E \rightarrow E'$ such that

$$P_{AB|E'} = P_{A|E'} P_{B|E'}. \tag{27}$$

We say that P_{ABE} contains secrecy [19] when this is not possible. We stress that this does not mean that many copies of P_{ABE} can later be used to obtain a secret key by LOPC. Indeed, there are probability distributions with positive secrecy content, which cannot be distilled into a secret key by LOPC [20].

Now, suppose Alice and Bob share a distribution $P(a,b|x,y)$. They decide the inputs according to uniform distributions $p(x)=1/X$ and $p(y)=1/Y$ [21]. Then, Alice's and Bob's information is, respectively, $A=(a,x)$ and $B=(b,y)$. The random variables A and B are correlated according to

$$P_{AB} = P(a,b|x,y)\frac{1}{XY}. \tag{28}$$

Can Alice and Bob bound Eve's information on their outcomes from their observed correlations? Can one prove that all possible extension P_{ABE} of P_{AB}, derived from $P(a,b|x,y)$ through Eq. (28), contain secrecy? This is of course impossible if no assumption on the possible extensions are made. In general, Alice and Bob can never exclude that Eve has a perfect copy of their outcomes, unless some constraints are imposed. However, if it is assumed that no faster-than-light communication is possible, not all possible extension of the initial bipartite probability distribution are allowed. Let us only consider extensions $P(a,b,e|x,y)$ compatible with no signaling. Thus, to each $P(a,b|x,y)$ we can associate a family of tripartite distributions

$$P_{ABE} = P(a,b,e|x,y)\frac{1}{XY}, \tag{29}$$

where $E=e$. We say that $P(a,b|x,y)$ contains secrecy if all its associated P_{ABE} contain secrecy.

B. All nonlocal correlations contain secrecy

The aim of this section is to show the link between the nonlocal properties of $P(a,b|x,y)$ and the secrecy content of any possible extension P_{ABE}, defined through Eq. (29). Before proceeding, note that an equivalent way of defining local correlations is as follows: a probability distribution $P(a,b|x,y)$ is local (5) when there exists a (nonsignaling) extension $P(a,b,e|x,y)$ such that

$$P(a,b|x,y,e) = P(a|x,e)P(b|y,e). \tag{30}$$

Now, assume one has a bipartite distribution $P(a,b|x,y)$ for which there exists an extension P_{ABE} with no secrecy content, that is

$$P_{AB|E} = P_{A|E} P_{B|E}. \tag{31}$$

Because processing the outcomes of a nonsignaling distribution gives another nonsignaling distribution, any transformation $E \rightarrow E'$ is included in the arbitrariness of the extension $P(a,b,e|x,y)$. By using the definition of conditional probabilities, one can see that Eq. (31) is equivalent to Eq. (30). That is, P_{ABE} has no secrecy if and only if there exists an extension of $P(a,b|x,y)$ satisfying Eq. (30), which is to say that $P(a,b|x,y)$ is local. This establishes the following equivalence.

Result 5. A distribution contains secrecy if and only if it is nonlocal.

This result is completely analogous to the quantum case, where it is known that a bipartite state contains secrecy if and only if it is entangled [22]. Of course, in this case eavesdroppers must be limited by QM.

It was already proven in Ref. [7] that all local correlations (5) can be distributed by LOPC. The public message that one of the parties, say Alice, should send to the rest in order to create the correlations, is precisely the (hidden) variable e that appears in Eq. (5). Therefore, if Alice and Bob's probability distribution is local, they cannot exclude that the global probability distribution including Eve does not contain any secrecy.

The following natural question is to identify those nonlocal correlations distillable to a secret key and whether they can be distributed using quantum states [23]. This will define those quantum correlations secure against an eavesdropper only limited by the no-signaling principle [7].

VI. NONLOCALITY AND RANDOMNESS

We first start by showing that all nonlocal correlations have random outcomes (see also Ref. [3]). Consider a deterministic bipartite distribution $P_{\mathrm{DET}}(a,b|x,y)$. That is, a and b are deterministic functions of (x,y) ($a=f[x,y]$ and $b=g[x,y]$). Using this and no signaling, we can get the following equalities:

$$\begin{aligned} P_{\mathrm{det}}(a,b|x,y) &= \delta_{(a,b),(f[x,y],g[x,y])} = \delta_{a,f[x,y]}\delta_{b,g[x,y]} \\ &= P(a|x,y)P(b|x,y) = P(a|x)P(b|y). \end{aligned} \tag{32}$$

The last line is a distribution of the form (5). Therefore, all deterministic distributions are local. Or in other words, all nonlocal states have uncertain outcomes. This fact can be straightforwardly extended to the n-party case.

Summarizing, in any nonsignaling theory with nonlocal correlations there are two kinds of randomness. The first one reflects our ignorance and corresponds to those probability distributions that can be written as the convex combination of extreme points. But, as in QM, there is also an intrinsic randomness even for extreme points, or pure states. The PR box (11) is an example of a pure state with uncertain outcomes.

Incompatible observables and uncertainty. Within QM it is said that two observables (O_0,O_1) are compatible if there exists a more complete one O of which both are functions $(O_0,O_1)=f(O)$. Consider $P(a,b|x,y)$, we say that the two observables in Bob's site b_0 and b_1 (corresponding to the inputs $y=0,1$) are compatible, if there exists a joint distribution for both $P'(a,b_0,b_1|x)$. That is

$$\sum_{b_0} P'(a,b_0,b_1|x) = P(a,b_1|x,y=1), \tag{33}$$

$$\sum_{b_1} P'(a,b_0,b_1|x) = P(a,b_0|x,y=0). \tag{34}$$

Or in other words, $P(a,b|x,y)$ is two shareable with respect to Bob if we restrict to $y=0,1$.

When the observables (b_0,b_1) are not compatible, a possible way of quantifying the degree of incompatibility is

$$\begin{aligned}\mathrm{inc}[b_0,b_1] = \min\{\eta > 0&: P(a,b|x,y) \\ &= \eta P_{\mathrm{inc}}(a,b|x,y) + (1-\eta)P_{\mathrm{com}}(a,b|x,y)\},\end{aligned} \tag{35}$$

where $P_{\mathrm{com}}(a,b|x,y)$ is a distribution where b_0 and b_1 are compatible and $P_{\mathrm{inc}}(a,b|x,y)$ is an arbitrary one. It is clear that the range of $\mathrm{inc}[b_0,b_1]$ is $[0,1]$, and $\mathrm{inc}[b_0,b_1]=0$ if and only if b_0 and b_1 are compatible. In Appendix B it is proven that in the binary input-output case, this minimization yields the CHSH violation

$$\mathrm{inc}[b_0,b_1] = \mathcal{B}_{\mathrm{CHSH}}[P(a,b|x,y)]. \tag{36}$$

In the case of binary outputs or inputs, we are able to establish a direct relation between $\mathrm{inc}[b_0,b_1]$ and the uncertainty of b_0 and b_1:

Result 6. In the binary output case $[A=B=2]$ the following constraints hold:

$$H(b_0) \geqslant h\left(\frac{1}{2}\,\mathrm{inc}[b_0,b_1]\right), \tag{37}$$

$$H(b_1) \geqslant h\left(\frac{1}{2}\,\mathrm{inc}[b_0,b_1]\right), \tag{38}$$

where $H(b)$ is the entropy of the output b and $h(x)$ is the binary entropy of x [24]. These inequalities also hold for arbitrary output and binary input $[X=Y=2]$.

The proof of this result is in Appendix B. Although this has the flavor of the Heisenberg uncertainty relations, it differs in the fact that here we do not have a trade off between the uncertainty of each observable. In particular, if b_0 is deterministic, inequality (37) implies $\mathrm{inc}[b_0,b_1]=0$, and hence, nothing prevents b_1 from being deterministic too. It is also remarkable that a deterministic observable is compatible with any other.

VII. CONCLUSIONS

In this work, we have identified a series of features common to all physical theories that do not allow for instantaneous transmission of information, and predict the violation of Bell inequalities. As shown, these two assumptions are sufficient to prove the following.

Constraints on how nonlocality is distributed among the correlations of different pairs of particles in multipartite scenarios.

Impossibility of perfect cloning of states.

Strict equivalence of the following properties: nonlocality, bounded shareability, positive secrecy content.

A relation for the incompatibility of two observables and the uncertainty of their outcomes.

Hence, some properties traditionally attributed to QM are generic within this family of physical theories. For example, the fact that two observables cannot be simultaneously measured on the same system (incompatibility), becomes necessary to explain the correlations observed in some experiments (violation of CHSH [2]), independently of the fact that we use models based on noncommuting operators to explain such experiments (see also Ref. [17]). Moreover, a no-cloning theorem can be derived without invoking any nonorthogonality of states.

This indicates how constraining is the demand that a theory compatible with special relativity predicts the violation of Bell inequalities. One could actually say that there is not much room left out of QM.

From a more fundamental point of view, this work proposes a different approach to the study of quantum properties. In general, QM has been studied in comparison with classical mechanics, that is, starting from a more restrictive theory. Here, the idea is to start from a more general family of theories, and to study "quantum" properties common to all of them. It is then an open research project to identify those additional postulates that allow one to recover the whole quantum structure.

ACKNOWLEDGMENTS

We thank J. Barrett, S. Iblisdir, N. S. Jones, G. Molina-Terriza, S. Popescu, and V. Scarani for discussions. This work is supported by the ESF, an MCYT "Ramón y Cajal" grant, the Generalitat de Catalunya, the Swiss NCCR "Quantum Photonics," OFES within the EU Project No. RESQ (IST-2001-37559) and the U.K. Engineering and Physical Sciences Research Council (IRC QIP).

APPENDIX A: DEPOLARIZATION AND SHRINKING

In this appendix it is shown that, in the case $X=Y=A=B=2$, any distribution can be transformed into an isotropic one maintaining the CHSH violation (8) invariant. We call this process "depolarization." We also show that the parameter C of an isotropic distribution can be decreased with local operations. We call this operation "shrinking."

Depolarization. This transformation can be implemented by using three bits of shared randomness and local operations, in the following two steps.

First step, Alice and Bob perform with probability $1/2$ one of the following two operations: (1) nothing and (2) flip a and b. This makes the correlations locally unbiased.

Second step, with probability $1/4$ both parties perform one of the following four operations. (1) nothing, (2) flip $a_{x=1}$ and y, (3) flip x and $b_{y=1}$, and (4) flip x, $a_{x=0}$, y, and $b_{b=1}$, where flipping $a_{x=1}$ means that a is only flipped when $x=1$, that is $a \rightarrow a+x \bmod 2$. After the second step, the resulting correlations satisfy Eq. (9). It can be seen that both steps keep invariant the violation of the CHSH inequality.

Shrinking. A useful observation is that when $C>0$, the value of $\mathcal{B}_{\mathrm{CHSH}}$ can always be decreased by performing an operation in one site. This is accomplished when one party, say Bob, outputs b with probability $1-\epsilon$, and an unbiased random bit with probability ϵ. This operation implements the transformation $C \rightarrow (1-\epsilon)C$.

APPENDIX B: PROOFS OF SEC. VI

Result. In the case $A=B=X=Y=2$ the degree of incompatibility of two observables is

$$\mathrm{inc}[b_0,b_1] = \mathcal{B}_{\mathrm{CHSH}}[P]. \tag{B1}$$

Proof. In the binary input case, the fact that P_{com} in Eq. (35) is two shareable implies that it is local. Hence, if $P=\eta P_{\mathrm{inc}}+(1-\eta)P_{\mathrm{com}}$ we have that

$$\mathcal{B}_{\mathrm{CHSH}}[P] \leqslant \eta \mathcal{B}_{\mathrm{CHSH}}[P_{\mathrm{inc}}] \leqslant \eta. \tag{B2}$$

One can always express an arbitrary nonlocal distribution P (with binary input and output) as a mixture of a PR box and a local distribution saturating the CHSH inequality. If we do so, and assign $P_{\mathrm{inc}}=P_{\mathrm{PR}}$ and $\mathcal{B}_{\mathrm{CHSH}}[P_{\mathrm{com}}]=0$, the minimum value $\eta=\mathcal{B}_{\mathrm{CHSH}}[P]$ is actually achieved.

Result 6. In the binary output case $[A=B=2]$ the following constraints hold:

$$H(b_0) \geqslant h\left(\frac{1}{2}\,\mathrm{inc}[b_0,b_1]\right), \tag{B3}$$

$$H(b_1) \geqslant h\left(\frac{1}{2}\,\mathrm{inc}[b_0,b_1]\right), \tag{B4}$$

where $H(b)$ is the entropy of the output b and $h(x)$ is the binary entropy of x [24]. These inequalities also hold for arbitrary output and binary input $[X=Y=2]$.

Proof. Let us prove the above inequalities (B3), (B4) for the binary output case. It is shown in this case [11] that, for all extreme points, the one party marginals are deterministic or unbiased: $[P(b=0|y),P(b=1|y)] \in \{[0,1],[1,0],[1/2,1/2]\}$. In the next we see that, if one observable, say $y=0$, is deterministic $[P(b_0|0)=0,1]$ then it is compatible with all the rest. To see this suppose that the outcome of b_0 is always $b_0=\beta$, then, for any y, the joint distribution $P(a,b_0,b_y|x,y)=P(a,b_y|x,y)\,\delta_{b_0,\beta}$ exists. Then, b_0 and b_y are compatible by definition. Now, let us decompose P_{inc} as a mixture of extreme points. This mixture must not contain extreme points having the marginal of b_0 or the marginal of b_1 deterministic. Otherwise, one could move this extreme point to the mixture of compatible ones P_{com}, decreasing the value of η. Thus, the marginals for b_0 and b_1 taken from P_{inc} are always unbiased. Therefore, $\mathrm{inc}[b_0,b_1]$ is the probability of getting with certainty an unbiased outcome. The situation where b_0 and b_1 have minimal entropy is when P_{com} is deterministic. Suppose that $P_{\mathrm{com}}(b=0|y=0)=1$, then recalling Eq. (35)

$$P(b=1|y=0) = \mathrm{inc}[b_0,b_1]P_{\mathrm{inc}}(b=1|y=0) = \frac{1}{2}\,\mathrm{inc}[b_0,b_1], \tag{B5}$$

and thus the entropy of b_0 is $H(b_0)=h(\mathrm{inc}[b_0,b_1]/2)$. The same holds for b_1. In general, when P_{com} is not deterministic, the entropies will be larger than the bounds (B3), (B4).

Let us prove that the bounds (B3), (B4) also hold in the case where inputs are binary, and the outputs belong to larger alphabets. In that case, all extreme points have been classified in Ref. [4]. There, it is shown that all extreme points have local marginals where all outcomes with nonzero probability are equiprobable. As discussed before, if we write P_{inc} as a mixture of extreme points, the marginals for b_0 and b_1 given by these extreme points must have at least two outcomes with nonzero probability. Otherwise the two observables are compatible and we can attach the extreme point to P_{inc}, decreasing η. The situation where b_0 and b_1 have minimal entropy is when P_{com} is deterministic, and P_{inc} has only two outcomes with nonzero probability for b_0 and b_1. In such case, the inequalities (B3), (B4) are saturated. When P_{inc} has more than two outcomes with nonzero probability for b_0 and b_1, the entropies will be larger.

[1] J. S. Bell, Physics (Long Island City, N.Y.) **1**, 195 (1964).
[2] A. Aspect, Nature (London) **398**, 189 (1999).
[3] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[4] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, Phys. Rev. A **71**, 022101 (2005).
[5] W. van Dam, e-print quant-ph/0501159; S. Wolf and J. Wullschleger, e-print quant-ph/0502030; H. Buhrman, M. Christandl, F. Unger, S. Wehner, and A. Winter, e-print quant-ph/0504133; T. Short, N. Gisin, and S. Popescu, e-print quant-ph/0504134.
[6] J. Barrett, e-print quant-ph/0508211.
[7] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).
[8] R. Werner and M. M. Wolf, e-print quant-ph/0107093.
[9] See, for instance, A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1995).
[10] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[11] N. S. Jones and Ll. Masanes, e-print quant-ph/0506182; J. Barrett and S. Pironio, Phys. Rev. Lett. **95**, 140401 (2005).
[12] V. Coffman, J. Kundu, and W. K. Wootters, Phys. Rev. A **61**, 052306 (2000); T. J. Osborne, e-print quant-ph/0502176.
[13] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, Phys. Rev. Lett. **88**, 040404 (2002).
[14] B. M. Terhal, A. C. Doherty, and D. Schwab, Phys. Rev. Lett. **90**, 157903 (2003).
[15] R. F. Werner, Lett. Math. Phys. **17**, 359 (1989).
[16] N. Gisin, Phys. Lett. A **242**, 1 (1998).
[17] R. F. Werner, e-print quant-ph/0101061.
[18] C. -S. Niu and R. B. Griffiths, Phys. Rev. A **60**, 2764 (1999); N. J. Cerf, J. Mod. Opt. **47**, 187 (2000).
[19] R. Renner and S. Wolf, *Advances in Cryptology—EUROCRYPT 2003*, Vol. 2656 of Lecture Notes in Computer Science (Springer-Verlag, Berlin, 2003), pp. 562–577.

[20] N. Gisin and S. Wolf, *Proceedings of CRYPTO 2000*, Vol. 1880 of Lecture Notes in Computer Science (Springer-Verlag, Berlin, 2000), p. 482; A. Acin, J. I. Cirac, and Ll. Masanes, Phys. Rev. Lett. **92**, 107903 (2004).

[21] Actually all the results are independent of the choices $p(x)$ and $p(y)$, if all the terms in these distributions are different from zero.

[22] A. Acin and N. Gisin, Phys. Rev. Lett. **94**, 020501 (2005).

[23] A. Acin, N. Gisin, and Ll. Masanes, e-print quant-ph/0510094.

[24] The entropy of a probability distribution $P(a)$ is

$$H(a) = -\sum_a P(a)\log_2 P(a).$$

The binary entropy function is

$$h(x) = -x\log_2 x - (1-x)\log_2(1-x).$$