REVIEWS OF MODERN PHYSICS, VOLUME 86, APRIL–JUNE 2014

Bell nonlocality

Nicolas Brunner

*Département de Physique Théorique, Université de Genève, 1211 Genève, Switzerland
and H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol,
BS8 1TL, United Kingdom*

Daniel Cavalcanti

*Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3,
Singapore 117543
and Ciències Fotòniques, Mediterranean Technology Park,
08860 Castelldefels (Barcelona), Spain*

Stefano Pironio

Laboratoire d'Information Quantique, Université Libre de Bruxelles (ULB), Belgium

Valerio Scarani

*Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3,
Singapore 117543
and Department of Physics, National University of Singapore, 3 Science Drive 2,
Singapore 117542*

Stephanie Wehner

*Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3,
Singapore 117543
and School of Computing, National University of Singapore, 13 Computing Drive,
Singapore 117417*

(published 18 April 2014; corrected 23 May 2014)

Bell's 1964 theorem, which states that the predictions of quantum theory cannot be accounted for by any local theory, represents one of the most profound developments in the foundations of physics. In the last two decades, Bell's theorem has been a central theme of research from a variety of perspectives, mainly motivated by quantum information science, where the nonlocality of quantum theory underpins many of the advantages afforded by a quantum processing of information. The focus of this review is to a large extent oriented by these later developments. The main concepts and tools which have been developed to describe and study the nonlocality of quantum theory and which have raised this topic to the status of a full subfield of quantum information science are reviewed.

DOI: 10.1103/RevModPhys.86.419 PACS numbers: 03.65.Ud

CONTENTS

I. Introduction 420
A. Nonlocality in a nutshell 421
B. The limitations of noncommunicating Ph.D. students 422
C. Scope of this review 423
D. Outline 423
II. Mathematical Characterization of Nonlocal Correlations 423
A. General definitions 424
1. No-signaling correlations 424
2. Local correlations 424
3. Quantum correlations 425
B. Bell inequalities 425
1. The local polytope 426
2. Facet Bell inequalities 426
3. Examples 427
4. Nonlocal games 428
a. XOR games 429
b. An example: CHSH as a game 429
c. Projection and unique games 429
d. Other special classes of games 429
C. Bell inequality violations 429
1. Quantum bounds 430
a. Properties of quantum correlations 430
b. Correlation inequalities 430
c. State and measurement dependent bounds 431
d. General bounds 431
2. No-signaling bounds 432
D. Multipartite correlations 433
E. Nonlocality without inequalities 434
F. Quantifying nonlocality 435
G. Multiple rounds and parallel repetition 435

III. Nonlocality and Quantum Theory 437
A. Nonlocality versus entanglement 437
1. Single-copy nonlocality 438
2. Hidden nonlocality 440
3. Multicopy nonlocality 441
4. More general scenarios 441
5. Entanglement distillation and nonlocality 441
6. Nonlocality and teleportation 441
7. More nonlocality with less entanglement 442
B. Nonlocality versus Hilbert space dimension 442
1. Minimal Hilbert space dimension and dimension witnesses 442
2. Grothendieck's constant and Bell inequalities with unbounded violation 443
C. Simulation of quantum correlations 444
1. Simulating the singlet state 444
2. Other quantum states 444
3. Elitzur-Popescu-Rohrlich decomposition 444
IV. Applications of Quantum Nonlocality 445
A. Communication complexity 445
B. Information theory 445
C. Quantum cryptography 446
1. Initial developments 446
2. Randomness versus nonlocality 447
a. Quantitative measures of randomness 447
b. Randomness and Bell violations 447
3. Device-independent randomness generation 448
4. Device-independent quantum key distribution 449
D. Other device-independent protocols 451
V. Information-theoretic Perspective on Nonlocality 451
A. Properties of no-signaling correlations 452
B. Nonlocality measures, interconversion, and distillation 452
C. Consequences of superstrong nonlocality 453
1. Information-theoretic consequences 453
a. Communication complexity and nonlocal computation 453
b. Information causality 454
c. Limitations on multipartite correlations 454
2. Physical consequences 454
a. Macroscopic locality 454
b. Uncertainty and information 454
c. Local quantum mechanics 455
D. Nonlocality in generalized probabilistic theories 455
VI. Multipartite Nonlocality 455
A. Defining multipartite nonlocality 455
1. Genuine multipartite nonlocality *à la* Svetlichny 456
2. Beyond Svetlichny's model 456
B. Detecting genuine multipartite nonlocality 457
1. Svetlichny's inequality 457
2. Generalizations to more parties, measurements, and dimensions 458
C. Monogamy 458
D. Nonlocality of multipartite quantum states 458
1. Multipartite nonlocality versus multipartite entanglement 458
2. Greenberger-Horne-Zeilinger states 459
3. Graph states 459
4. Nonlocality of other multipartite quantum states 460
VII. Experimental Aspects 461
A. Bell experiments 461
1. Photons 461
2. Atoms 462
3. Hybrid schemes and other systems 463
B. Loopholes 463
1. Detection loophole 463
a. Faking Bell inequality violations with postselection 463
b. Taking into account no-click events 463
c. Threshold efficiencies 464
2. Locality loophole 465
3. Finite statistics 466
C. Toward a loophole-free Bell test 466
1. Photons 466
2. Continuous variable systems 467
3. Atom-atom and atom-photon entanglement 467
D. Bell tests without alignment 467
VIII. Related Concepts 468
A. Bilocality and more general correlation scenarios 468
B. No-go theorems for nonlocal models 469
1. Models *à la* Leggett 469
2. Superluminal signaling models 469
C. Steering 470
D. Semiquantum games 470
IX. Conclusion 470
Acknowledgments 470
Appendix: Guide to Bell inequalities 471
1. Bipartite Bell inequalities 471
a. Binary outputs: $(2, m, 2)$ 471
b. Arbitrary number of outputs: $(2, m, \Delta)$ 471
2. Multipartite Bell inequalities 471
a. Binary outputs: $(n, m, 2)$ 471
b. Arbitrary number of outputs: (n, m, Δ) 471
c. Bell inequalities detecting genuine multipartite nonlocality 471
References 471

I. INTRODUCTION

In 1964, Bell proved that the predictions of quantum theory are incompatible with those of any physical theory satisfying a natural notion of locality[1] (Bell, 1964). Bell's theorem has deeply influenced our perception and understanding of physics, and arguably ranks among the most profound scientific discoveries ever made. With the advent of quantum information science, considerable interest has been devoted to Bell's theorem. In particular, a wide range of concepts and technical tools have been developed for describing and studying the nonlocality of quantum theory. These represent the main focus of this review. Hence we will not discuss, at least not directly, the extensive literature dealing with the conceptual implications of Bell's theorem from a traditional foundational perspective. Skipping many important contributions before and after Bell's ground-breaking discovery, the most notable one being the famous Einstein-Podolosky-Rosen paper (Einstein, Podolsky, and Rosen, 1935), we start straightaway with the mathematical formulation of a locality constraint in the context of certain experiments involving separate systems and its violation by the predictions of quantum theory.

[1] To avoid any misunderstanding from the start, by “locality” we do not mean the notion used within quantum mechanics and quantum field theory that operators defined in spacelike separated regions commute. Bell's notion of locality is different and is clarified below.

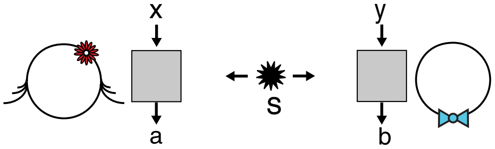

FIG. 1 (color online). Sketch of a Bell experiment. A source (S) distributes two physical systems to distant observers, Alice and Bob. Upon receiving their systems, each observer performs a measurement on it. The measurement chosen by Alice is labeled x and its outcome a. Similarly, Bob chooses measurement y and gets outcome b. The experiment is characterized by the joint probability distribution $p(ab|xy)$ of obtaining outcomes a and b when Alice and Bob choose measurements x and y.

A. Nonlocality in a nutshell

In a typical “Bell experiment,” two systems which may have previously interacted—for instance they may have been produced by a common source—are now spatially separated and are each measured by one of two distant observers, Alice and Bob (see Fig. 1). Alice may choose one out of several possible measurements to perform on her system and we let x denote her measurement choice. For instance, x may refer to the position of a knob on her measurement apparatus. Similarly, we let y denote Bob's measurement choice. Once the measurements are performed, they yield outcomes a and b on the two systems. The actual values assigned to the measurement choices x, y and outcomes a, b are purely conventional; they are mere macroscopic labels distinguishing the different possibilities.

From one run of the experiment to the other, the outcomes a and b that are obtained may vary, even when the same choices of measurements x and y are made. These outcomes are thus in general governed by a probability distribution $p(ab|xy)$, which can of course depend on the particular experiment being performed. By repeating the experiment a sufficient number of times and collecting the observed data, one gets a fair estimate of such probabilities.

When such an experiment is actually performed—say, by generating pairs of spin-$1/2$ particles and measuring the spin of each particle in different directions—it will in general be found that

$$p(ab|xy) \neq p(a|x)p(b|y), \tag{1}$$

implying that the outcomes on both sides are not statistically independent from each other. Even though the two systems may be separated by a large distance, and may even be spacelike separated, the existence of such correlations is nothing mysterious. In particular, it does not necessarily imply some kind of direct influence of one system on the other, for these correlations may simply reveal some dependence relation between the two systems which was established when they interacted in the past. This is at least what one would expect in a local theory.

We formalized the idea of a *local* theory more precisely. The assumption of locality implies that we should be able to identify a set of past factors, described by some variables λ, having a joint causal influence on both outcomes, and which fully account for the dependence between a and b. Once all such factors have been taken into account, the residual indeterminacies about the outcomes must now be decoupled; that is, the probabilities for a and b should factorize

$$p(ab|xy,\lambda) = p(a|x,\lambda)p(b|y,\lambda). \tag{2}$$

This factorability condition simply expresses the fact that we have found an explanation according to which the probability for a depends only on the past variables λ and on the local measurement x, but not on the distant measurement and outcome, and analogously for the probability to obtain b. The variable λ will not necessarily be constant for all runs of the experiment, even if the procedure which prepares the particles to be measured is held fixed, because λ may involve physical quantities that are not fully controllable. The different values of λ across the runs should thus be characterized by a probability distribution $q(\lambda)$. Combined with the above factorability condition, we can thus write

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

where we also implicitly assumed that the measurements x and y can be freely chosen in a way that is independent of λ, i.e., that $q(\lambda|x,y) = q(\lambda)$. This decomposition now represents a precise condition for locality in the context of Bell experiments.[2] Note that no assumptions of determinism or of a “classical behavior” are being involved in Eq. (3): we assumed that a (and similarly b) is only *probabilistically* determined by the measurement x and the variable λ, with no restrictions on the physical laws governing this causal relation. Locality is the crucial assumption behind Eq. (3). In relativistic terms, it is the requirement that events in one region of space-time should not influence events in spacelike separated regions.

It is now a straightforward mathematical theorem[3] that the predictions of quantum theory for certain experiments involving entangled particles do not admit a decomposition of the form (3). To establish this result, we consider for simplicity an experiment where there are only two measurement choices per observer $x, y \in \{0, 1\}$ and where the possible outcomes take also two values labeled $a, b \in \{-1, +1\}$. Let $\langle a_x b_y \rangle = \sum_{a,b} ab\, p(ab|xy)$ be the expectation value of the product ab for given measurement choices (x, y) and consider the expression $S = \langle a_0 b_0 \rangle + \langle a_0 b_1 \rangle + \langle a_1 b_0 \rangle - \langle a_1 b_1 \rangle$, which is

[2] Bell also used the term local *causality* instead of locality. Local *hidden-variable* or local *realistic* models are also frequently used to refer to the existence of a decomposition of Eq. (3); see Goldstein *et al.* (2011) and Norsen (2009) for a critical discussion of these terminologies.

[3] It is relatively frequent to see a paper claiming to “disprove” Bell's theorem or that a mistake in the derivation of Bell inequalities has been found. However, once one accepts the definition (3), it is a quite trivial *mathematical theorem* that this definition is incompatible with certain quantum predictions. Such papers are thus either using (possibly unaware) a different definition of locality or they are erroneous. Quantum Randi challenges have been proposed to confront Bell deniers in a pedagogical way (Gill, 2012; Vongehr, 2012).

a function of the probabilities $p(ab|xy)$. If these probabilities satisfy the locality decomposition (3), we necessarily have that

$$S = \langle a_0 b_0 \rangle + \langle a_0 b_1 \rangle + \langle a_1 b_0 \rangle - \langle a_1 b_1 \rangle \leq 2, \tag{4}$$

which is known as the Clauser-Horne-Shimony-Holt (CHSH) inequality (Clauser *et al.*, 1969). To derive this inequality, we can use Eq. (3) in the definition of $\langle a_x b_y \rangle$, which allows us to express this expectation value as an average $\langle a_x b_y \rangle = \int d\lambda q(\lambda) \langle a_x \rangle_\lambda \langle b_y \rangle_\lambda$ of a product of local expectations $\langle a_x \rangle_\lambda = \sum_a a p(a|x,\lambda)$ and $\langle b_y \rangle_\lambda = \sum_b b p(b|y,\lambda)$ taking values in $[-1, 1]$. Inserting this expression into Eq. (4), we can write $S = \int d\lambda q(\lambda) S_\lambda$, with $S_\lambda = \langle a_0 \rangle_\lambda \langle b_0 \rangle_\lambda + \langle a_0 \rangle_\lambda \langle b_1 \rangle_\lambda + \langle a_1 \rangle_\lambda \langle b_0 \rangle_\lambda - \langle a_1 \rangle_\lambda \langle b_1 \rangle_\lambda$. Since $\langle a_0 \rangle_\lambda, \langle a_1 \rangle_\lambda \in [-1, 1]$, this last expression is smaller than $S_\lambda \leq |\langle b_0 \rangle_\lambda + \langle b_1 \rangle_\lambda| + |\langle b_0 \rangle_\lambda - \langle b_1 \rangle_\lambda|$. Without loss of generality, we can assume that $\langle b_0 \rangle_\lambda \geq \langle b_1 \rangle_\lambda \geq 0$ which yields $S_\lambda = 2\langle b_0 \rangle_\lambda \leq 2$, and thus $S = \int d\lambda q(\lambda) S_\lambda \leq 2$.

Consider now the quantum predictions for an experiment in which the two systems measured by Alice and Bob are two qubits in the singlet state $|\Psi^-\rangle = (1/\sqrt{2})(|01\rangle - |10\rangle)$, where we have used the shortcut notation $|ab\rangle \equiv |a\rangle \otimes |b\rangle$, and where $|0\rangle$ and $|1\rangle$ are conventionally the eigenstates of σ_z for the eigenvalues $+1$ and -1, respectively. Let the measurement choices x and y be associated with vectors $\vec{x}$ and $\vec{y}$ corresponding to measurements of $\vec{x} \cdot \vec{\sigma}$ on the first qubit and of $\vec{y} \cdot \vec{\sigma}$ on the second qubit, where $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ denotes the Pauli vector. According to quantum theory, we then have the expectations $\langle a_x b_y \rangle = -\vec{x} \cdot \vec{y}$. Let the two settings $x \in \{0, 1\}$ correspond to measurements in the orthogonal directions $\hat{e}_1$ and $\hat{e}_2$, respectively, and the settings $y \in \{0, 1\}$ to measurements in the directions $-(\hat{e}_1 + \hat{e}_2)/\sqrt{2}$ and $(-\hat{e}_1 + \hat{e}_2)/\sqrt{2}$. We then have $\langle a_0 b_0 \rangle = \langle a_0 b_1 \rangle = \langle a_1 b_0 \rangle = 1/\sqrt{2}$ and $\langle a_1 b_1 \rangle = -1/\sqrt{2}$, hence

$$S = 2\sqrt{2} > 2, \tag{5}$$

in contradiction with Eq. (4) and thus with the locality constraint (3). This is the content of Bell's theorem, establishing the nonlocal character of quantum theory and of any model reproducing its predictions.

The CHSH inequality (4) is an example of a *Bell inequality*, a linear inequality for the probabilities $p(ab|xy)$ that is necessarily verified by any model satisfying the locality condition (3), but which can be *violated* by suitable measurements on a pair of quantum particles in an entangled state. The violation of these inequalities and the predictions of quantum theory were first confirmed experimentally by Freedman and Clauser (1972), then more convincingly by Aspect, Grangier, and Roger (1982b), and in many other experiments since.

Before outlining in more detail the content of this review, we first reconsider Bell's locality condition from a more operational perspective, which illustrates the spirit underlying this review.

B. The limitations of noncommunicating Ph.D. students

Consider a quantum apparatus which can perform a measurement on a quantum system in a state ρ_A. If measurement x is chosen, an output a is obtained. Quantum theory predicts the statistics $p(a|x)$ for the outcomes given the measurements. Suppose that a Ph.D. student, who cannot realize such a quantum experiment, is instead provided with unlimited classical computational power and a source of random numbers. If the student is competent, he can simulate the same statistics as in the quantum experiment based only on the description of the state ρ_A and of the measurement x to be performed on it. This is not a particularly deep remark: it is the daily job of physicists all over the world and an obvious consequence of the fact that the theory allows one to predict the results.

Now consider two quantum devices in two distant locations A and B performing measurements x and y on two systems in a joint state ρ_{AB}. Quantum theory allows one to compute the joint probabilities $p(ab|xy)$, so certainly the above student can simulate the experiment if he is given all the relevant information. However, in the quantum experiment the two locations can be sufficiently separated so that no information on y is available at the location A before a result is obtained, and similarly no information on x is available at B. Can two students, one at A and the other at B, simulate the quantum statistics in the same circumstances? As before, the students cannot manipulate any quantum systems, but they have unlimited computational power, access to a source of random numbers, and a perfect description of the joint state ρ_{AB}. Although they cannot communicate once the measurements are specified, they may have set up in advance a common strategy and have shared some common classical data λ, which can vary across different simulation runs according to a probability distribution $q(\lambda)$. In full generality, the output of the first student will thus be characterized by a probability distribution $p(a|x,\lambda)$, which is fixed by their common strategy and the joint state ρ_{AB}, but which may depend on the specific measurement x chosen and of the data λ shared with the second student. Similarly the output of the second student is given by a probabilistic function $p(b|y,\lambda)$. The joint statistics simulated by the two students are thus characterized by the probabilities

$$p(ab|xy) = \int d\lambda q(\lambda) p(a|x,\lambda) p(b|y,\lambda), \tag{6}$$

which is nothing but the locality condition (3). This condition thus admits a very simple and operational interpretation: it characterizes the correlations that can be reproduced with classical resources by our two noncommunicating students. The fact that certain experiments involving entangled quantum states violate Bell inequalities then imply that the two students cannot simulate such experiments. The violations of Bell inequalities can thus be interpreted as establishing a gap between what noncommunicating observers having access to classical or to quantum entangled resources can achieve. Note that locality, i.e., the constraint that the two observers cannot communicate, is the important limitation here. As we said previously, if all the information about x and y is available to one of the students, it is always possible to reproduce the quantum statistics using only classical resources.

There is another point worth noting here. The fact that entangled quantum systems are able to do things completely different from classical systems is well known. Indeed, for more than a century physicists have discovered that classical physics does not explain everything. However, if given only the statistics of a real quantum experiment and of a classical simulation of it, there is no way to tell the difference. The brute measurement data produced, for instance, by a Stern-Gerlach experiment can be simulated classically; its “nonclassicality” becomes evident only when one takes into account the fact that a magnetic moment is being measured and that the measurements are associated with the direction of the gradient of a magnetic field. In the case of Bell nonlocality, however, the real quantum experiment and its (attempted) simulation can be distinguished solely from the measurement data, without having to specify which physical degree of freedom is involved or which measurements are performed. This property is referred to as *device independence*. Interpreted in this way, the violation of Bell inequalities can be seen as a detector of entanglement that is robust to any experimental imperfection: as long as a violation is observed, we have the guarantee, independently of any implementation details, that the two systems are entangled. This remark is important: since entanglement is at the basis of many protocols in quantum information, and, in particular, quantum cryptographic protocols, it opens the way to device-independent tests of their performance.

C. Scope of this review

We have given here only a succinct and intuitive presentation of the locality condition from which Bell's theorem follows. This naturally raises a series of questions: What are the precise physical assumptions on which this condition is based? Can we rigorously justify, in particular, on relativistic grounds, the notion of locality captured by this condition? To what extent does nonlocality, i.e., the violation of Eq. (3), conflict with relativity? What do the various interpretations of quantum theory have to say about this issue? We do not address here such questions that have been the subject of extensive analysis and discussion by both physicists and philosophers of science since Bell's discovery. A recent concise review has been written from this perspective (Goldstein *et al.*, 2011). Bell's collection of papers on the subject (Bell, 2004) is a must read, in which he explains and develops his main result from a variety of perspectives. In particular, the two articles (Bell, 1975, 1990) introduce the principle of *local causality*—a precise formulation of the notion of relativistic locality—from which the condition (3) can be derived; see also Norsen (2007). For a discussion of the implications of nonlocality for relativity, see Maudlin (2002).

This review has a more technical flavor: How can one show that the measurement statistics of a given experiment do not satisfy the condition (3)? How can one derive Bell inequalities in a systematic way? Which entangled states violate these inequalities, and which ones do not? Can quantum nonlocality be exploited for information processing, and if yes how? How should one design the best experimental test of quantum nonlocality, etc.? Although they may have foundational motivations or implications, the works discussed here have an original technical component. Many of them also follow a recent trend in which nonlocality is considered from an operational perspective and where its relations with other topics in quantum information science, such as the theory of entanglement or cryptography, are investigated. Finally, we focus on progress reported in the last 15 years or so. For works on Bell nonlocality before this period or for aspects not covered here, see Clauser and Shimony (1978), Home and Selleri (1991), Khalfin and Tsirelson (1992), Mermin (1993), Tsirelson (1993), Zeilinger (1999), Werner and Wolf (2001a), Genovese (2005), and Buhrman *et al.* (2010), and references therein.

D. Outline

The outline of this review is as follows. Section II is devoted to setting up some general definitions and presenting a mathematical characterization of nonlocal correlations. In particular, we study the general properties of correlations that can arise between local, quantum, and no-signaling systems. We address the problem of deriving Bell inequalities from the locality condition (3) and determining their maximal quantum violations. Section III addresses nonlocality in quantum theory. The main question is to understand how quantum nonlocality relates to certain properties of quantum resources, such as entanglement and Hilbert space dimension. The relation between nonlocality and information is discussed in Secs. IV and V. We first present in Sec. IV various applications of quantum nonlocality, such as communication complexity, quantum cryptography, and device-independent quantum information processing. Section V provides an information-theoretic perspective on nonlocality, in which nonlocal correlations are viewed as a fundamental resource. Notably, these ideas stimulated a series of works trying to recover the structure of quantum correlations (and more generally of quantum theory itself) from information-theoretic principles. Section VI is devoted to the nonlocality of multipartite systems. The notions of genuine multipartite nonlocality and monogamy of correlations are discussed, as well as their relevance for quantum multipartite systems. In Sec. VII we review the experimental work that has been achieved on quantum nonlocality, where Bell inequality violations have been demonstrated using a variety of different physical systems and experimental configurations. We also discuss the loopholes that may affect Bell experiments and report recent progress made toward a loophole-free Bell experiment. Finally, Sec. VIII deals with variations around Bell's theorem, in which different notions of nonlocality, stronger or weaker than Bell's, are considered. Section IX gives our conclusion. And finally, the Appendix provides a guide referencing Bell inequalities for a wide range of Bell scenarios.

II. MATHEMATICAL CHARACTERIZATION OF NONLOCAL CORRELATIONS

This section presents the main concepts and tools for characterizing nonlocal correlations. The notations introduced here will be used throughout this review. For clarity, the discussion focuses mainly on the case of two observers, generalizations to the multipartite case being usually straightforward (see also Secs. II.D and VI for results specific to the multipartite case).

A. General definitions

As in the Introduction, we consider two distant observers, Alice and Bob, performing measurements on a shared physical system, for instance, a pair of entangled particles. Each observer has a choice of m different measurements to perform on his system. Each measurement can yield Δ possible outcomes. Abstractly we describe the situation by saying that Alice and Bob have access to a "black box." Each party locally selects an input (a measurement setting) and the box produces an output (a measurement outcome). We refer to this scenario as a *Bell scenario*.

We label the inputs of Alice and Bob $x, y \in \{1, \ldots\ldots, m\}$ and their outputs $a, b \in \{1, \ldots\ldots, \Delta\}$, respectively. The labels attributed to the inputs and outputs are purely conventional, and the results presented here are independent of this choice. Some parts of this review might use other notations for convenience. In particular, when the outputs are binary, it is customary to write $a, b \in \{-1, 1\}$ or $a, b \in \{0, 1\}$.

Let $p(ab|xy)$ denote the joint probability to obtain the output pair (a, b) given the input pair (x, y). A Bell scenario is then completely characterized by $\Delta^2 m^2$ such joint probabilities, one for each possible pair of inputs and outputs. Following the terminology introduced by Tsirelson (1993), we refer to the set $\mathbf{p} = \{p(ab|xy)\}$ of all these probabilities as a *behavior*. Informally, we simply refer to them as the *correlations* characterizing the black box shared by Alice and Bob. A behavior can be viewed as a point $\mathbf{p} \in \mathbb{R}^{\Delta^2 m^2}$ belonging to the probability space $\mathcal{P} \subset \mathbb{R}^{\Delta^2 m^2}$ defined by the positivity constraints $p(ab|xy) \geq 0$ and the normalization constraints $\sum_{a,b=1}^{\Delta} p(ab|xy) = 1$. Due to the normalization constraints $\mathcal{P}$ is a subspace of $\mathbb{R}^{\Delta^2 m^2}$ of dimension $\dim \mathcal{P} = (\Delta^2 - 1)m^2$.

The existence of a given physical model behind the correlations obtained in a Bell scenario translates into additional constraints on the behaviors $\mathbf{p}$. Three main types of correlations can be distinguished.

1. No-signaling correlations

The first natural limitation on behaviors $\mathbf{p}$ are the *no-signaling* constraints (Cirel'son, 1980; Popescu and Rohrlich, 1994), formally expressed as

$$\sum_{b=1}^{\Delta} p(ab|xy) = \sum_{b=1}^{\Delta} p(ab|xy'), \quad \text{for all } a, x, y, y',$$
$$\sum_{a=1}^{\Delta} p(ab|xy) = \sum_{a=1}^{\Delta} p(ab|x'y), \quad \text{for all } b, y, x, x'. \tag{7}$$

These constraints have a clear physical interpretation: they imply that the local marginal probabilities of Alice $p(a|x) \equiv p(a|xy) = \sum_{b=1}^{\Delta} p(ab|xy)$ are independent of Bob's measurement setting y, and thus Bob cannot signal to Alice by his choice of input (and the other way around). In particular, if Alice and Bob are spacelike separated, the no-signaling constraints (7) guarantee that Alice and Bob cannot use their black box for instantaneous signaling, preventing a direct conflict with relativity.

Let $\mathcal{NS}$ denote the set of behaviors satisfying the no-signaling constraints (7). It is not difficult to see that $\mathcal{NS}$ is an affine subspace of $\mathbb{R}^{\Delta^2 m^2}$ of dimension

$$\dim \mathcal{NS} = 2(\Delta - 1)m + (\Delta - 1)^2 m^2 =: t, \tag{8}$$

see, e.g., Pironio (2005). One can thus parametrize points in $\mathcal{NS}$ using t numbers rather than the $\Delta^2 m^2$ numbers [or $(\Delta^2 - 1)m^2$ taking into account normalization] necessary to specify a point in the general probability space $\mathcal{P}$. A possible parametrization is given by the set of probabilities $\{p(a|x), p(b|y), p(ab|xy)\}$, where $a, b = 1, \ldots, \Delta - 1$ and $x, y = 1, \ldots, m$. There are indeed t such probabilities and their knowledge is sufficient to reconstruct the full list of $p(ab|xy)$ for any a, b, x, and y. Seen as a subset of $\mathbb{R}^t$, the no-signaling set is thus uniquely constrained by the $\Delta^2 m^2$ positivity constraints $p(ab|xy) \geq 0$ (which have to be reexpressed in terms of the chosen parametrization).

In the case of binary outcome $(\Delta = 2)$, an alternative parametrization is provided by the $2m + m^2$ correlators $\{\langle A_x \rangle, \langle B_y \rangle, \langle A_x B_y \rangle\}$, where

$$\langle A_x \rangle = \sum_{a \in \{\pm 1\}} a p(a|x), \qquad \langle B_y \rangle = \sum_{b \in \{\pm 1\}} b p(b|y), \tag{9}$$

$$\langle A_x B_y \rangle = \sum_{a,b \in \{\pm 1\}} ab p(ab|xy), \tag{10}$$

and we assumed $a, b \in \{-1, 1\}$. Joint probabilities and correlators are related as $p(ab|xy) = [1 + a\langle A_x \rangle + b\langle B_y \rangle + ab\langle A_x B_y \rangle]/4$. Thus an arbitrary no-signaling behavior must satisfy $1 + a\langle A_x \rangle + b\langle B_y \rangle + ab\langle A_x B_y \rangle \geq 0$ for all a, b, x, and y. See Bancal, Gisin, and Pironio (2010) for a more general definition of correlators for the $\Delta > 2$ case.

A particular subset of interest of $\mathcal{NS}$ in the $\Delta = 2$ case is the one for which $\langle A_x \rangle = \langle B_y \rangle = 0$. We refer to this set as the *correlation space* $\mathcal{C}$. In terms of the m^2 correlators (10), an arbitrary point in $\mathcal{C}$ is constrained only by the inequalities $-1 \leq \langle A_x B_y \rangle \leq 1$. Bell inequalities that involve only the quantities $\langle A_x B_y \rangle$, such as the CHSH inequality, are called correlation inequalities.

2. Local correlations

A more restrictive constraint than the no-signaling condition is the locality condition discussed in the Introduction. Formally, the set $\mathcal{L}$ of *local behaviors* is defined by the elements of $\mathcal{P}$ that can be written in the form

$$p(ab|xy) = \int_{\Lambda} d\lambda q(\lambda) p(a|x, \lambda) p(b|y, \lambda), \tag{11}$$

where the (hidden) variables λ are arbitrary variables taking value in a space Λ and distributed according to the probability density $q(\lambda)$ and where $p(a|x, \lambda)$ and $p(b|y, \lambda)$ are local probability response functions for Alice and Bob, respectively. Operationally, one can also think about λ as shared randomness; that is, some shared classical random bits, where Alice will choose an outcome a depending on both her measurement setting x and λ and similarly for Bob.

Whereas any local behavior satisfies the no-signaling constraint, the converse does not hold. There exist no-signaling behaviors which do not satisfy the locality conditions. Hence the set of local correlations is strictly smaller than the set of no-signaling correlations; that is, $\mathcal{L} \subset \mathcal{NS}$.

Correlations that cannot be written in the above form are said to be *nonlocal*. Note that this can happen only if $\Delta \geq 2$ and $m \geq 2$ (otherwise it is always possible to build a local model for any behavior in $\mathcal{P}$). In the following, we thus always assume $\Delta \geq 2$, $m \geq 2$.

3. Quantum correlations

Finally, we consider the set of behaviors achievable in quantum mechanics. Formally, the set $\mathcal{Q}$ of *quantum behaviors* corresponds to the elements of $\mathcal{P}$ that can be written as

$$p(ab|xy) = \mathrm{tr}\,(\rho_{AB} M_{a|x} \otimes M_{b|y}), \tag{12}$$

where ρ_{AB} is a quantum state in a joint Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ of arbitrary dimension, $M_{a|x}$ are measurement operators [positive operator valued measure (POVM) elements] on $\mathcal{H}_A$ characterizing Alice's measurements (thus $M_{a|x} \geq 0$ and $\sum_{a=1}^{\Delta} M_{a|x} = \mathbb{1}$), and similarly $M_{b|y}$ are operators on $\mathcal{H}_B$ characterizing Bob's measurements.

Note that, without loss of generality, we can always assume the state to be pure and the measurement operators to be orthogonal projectors, if necessary by increasing the dimension of the Hilbert space. That is, we can equivalently write a quantum behavior as

$$p(ab|xy) = \langle\psi| M_{a|x} \otimes M_{b|y} |\psi\rangle, \tag{13}$$

where $M_{a|x} M_{a'|x} = \delta_{aa'} M_{a|x}$, $\sum_a M_{a|x} = \mathbb{1}_A$ and similarly for the operators $M_{b|y}$.

A different definition of quantum behaviors is also possible, where instead of imposing a tensor product structure between Alice's and Bob's systems, we merely require that their local operators commute (Tsirelson, 1993). We call the corresponding set $\mathcal{Q}'$, i.e., a behavior $\mathbf{p}$ belongs to $\mathcal{Q}'$ if

$$p(ab|xy) = \langle\psi| M_{a|x} M_{b|y} |\psi\rangle, \tag{14}$$

where $|\psi\rangle$ is a state in a Hilbert space H, and $M_{a|x}$ and $M_{b|y}$ are orthogonal projectors on H defining proper measurements and satisfying $[M_{a|x}, M_{b|y}] = 0$. The former definition (13) is standard in nonrelativistic quantum theory, while the second one (14) is natural in relativistic quantum field theory. Since $[M_{a|x} \otimes \mathbb{1}_B, \mathbb{1}_A \otimes M_{b|y}] = 0$ it is immediate that $\mathcal{Q} \subseteq \mathcal{Q}'$. It is an open question, known as Tsirelson's problem, whether the inclusion is strict, i.e., $\mathcal{Q} \neq \mathcal{Q}'$ (Scholz and Werner, 2008; Tsirelson, 1993; Junge *et al.*, 2011; Fritz, 2012a). In the case where the Hilbert spaces $\mathcal{H}$, $\mathcal{H}_A$, and $\mathcal{H}_B$ are finite, it is known that Eqs. (13) and (14) coincide (Tsirelson, 1993; Doherty *et al.*, 2008; Navascues *et al.*, 2011). It is also known that $\mathcal{Q} = \mathcal{Q}'$ if Alice has a binary choice of inputs with two outputs each, independently of Bob's number of inputs and outputs (Navascues *et al.*, 2011). More precisely, in this case any element of $\mathcal{Q}'$ can be approximated arbitrarily well by an element of $\mathcal{Q}$. For many applications and results, it does not matter whether we consider the quantum sets $\mathcal{Q}$ or $\mathcal{Q}'$. In the following, we drop the distinction and use the notation $\mathcal{Q}$ to refer to both sets, except when results are specific to only one definition.

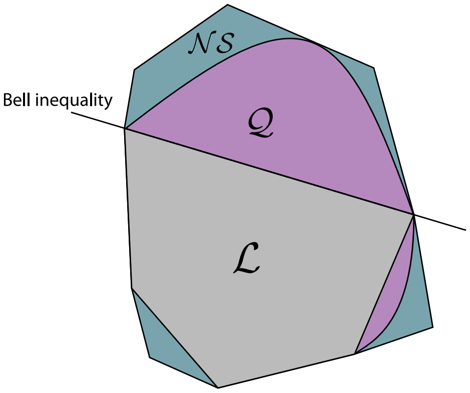

FIG. 2 (color online). Sketch of the no-signaling ($\mathcal{NS}$), quantum ($\mathcal{Q}$), and local ($\mathcal{L}$) sets. Notice the strict inclusions $\mathcal{L} \subset \mathcal{Q} \subset \mathcal{NS}$. Moreover, $\mathcal{NS}$ and $\mathcal{L}$ are polytopes, i.e., they can be defined as the convex combination of a finite number of extremal points. The set $\mathcal{Q}$ is convex, but not a polytope. The hyperplanes delimiting the set $\mathcal{L}$ correspond to Bell inequalities.

It can easily be shown that any local behavior admits a description of Eq. (12) and thus belongs to $\mathcal{Q}$ (Pitowsky, 1986). Moreover, any quantum behavior satisfies the no-signaling constraints. However, there are quantum correlations that do not belong to the local set (this follows from the violation of Bell inequalities) and, as we will see, there are no-signaling correlations that do not belong to the quantum set (Khalfin and Tsirelson, 1985; Rastall, 1985; Popescu and Rohrlich, 1994). In general, we thus have the strict inclusions $\mathcal{L} \subset \mathcal{Q} \subset \mathcal{NS}$ (see Fig. 2). Furthermore, it can be shown that $\dim \mathcal{L} = \dim \mathcal{Q} = \dim \mathcal{NS} = t$ (Pironio, 2005), where t is defined in Eq. (8).

In the following sections, we discuss the properties of $\mathcal{L}$, $\mathcal{Q}$, and $\mathcal{NS}$ in more detail. In particular, we see how it is possible to decide if a given behavior belongs or not to one of these sets. We show how each set can be characterized in terms of Bell-type inequalities and discuss how to compute bounds for Bell-type expression for behaviors in $\mathcal{L}$, $\mathcal{Q}$, and $\mathcal{NS}$.

B. Bell inequalities

The sets $\mathcal{L}$, $\mathcal{Q}$, and $\mathcal{NS}$ are closed, bounded, and convex. That is, if $\mathbf{p}_1$ and $\mathbf{p}_2$ belong to one of these sets, then the mixture $\mu\mathbf{p}_1 + (1-\mu)\mathbf{p}_2$ with $0 \leq \mu \leq 1$ also belongs to this set. The convexity of $\mathcal{Q}$ can be established for instance by following the argument in Pitowsky (1986). By the hyperplane separation theorem, it follows that for each behavior $\hat{\mathbf{p}} \in \mathbb{R}^t$ that does not belong to one of the sets $\mathcal{K} = \mathcal{L}$, $\mathcal{Q}$, or $\mathcal{NS}$ there exists a hyperplane that separates this $\hat{p}$ from the corresponding set (see Fig. 2). That is, if $\hat{\mathbf{p}} \notin \mathcal{K}$, then there exists an inequality of the form

$$\mathbf{s} \cdot \mathbf{p} = \sum_{abxy} s_{xy}^{ab} p(ab|xy) \leq S_k \tag{15}$$

that is satisfied by all $\mathbf{p} \in \mathcal{K}$ but which is violated by $\hat{\mathbf{p}}$: $\mathbf{s} \cdot \hat{\mathbf{p}} > S_k$. In the case of the local set $\mathcal{L}$, such inequalities are simply Bell inequalities. Thus any nonlocal behavior violates a Bell inequality. An example of such an inequality is the CHSH inequality (4) that we introduced in Sec. I.A. The inequalities associated with the quantum set, which characterize the limits of $\mathcal{Q}$, are often called quantum Bell inequalities or Tsirelson inequalities.

In the following, we refer to an arbitrary $\mathbf{s} \in \mathbb{R}^t$ as a Bell expression and to the minimal value S_l such that $\mathbf{s} \cdot \mathbf{p} \le S_l$ holds for all $\mathbf{p} \in \mathcal{L}$ as the local bound of this Bell expression. Similarly, we define the quantum bound S_q and the no-signaling bound S_{ns} as the analog quantities for the sets $\mathcal{Q}$ and $\mathcal{NS}$. If $S_q > S_l$ we also say that quantum mechanics violates the Bell inequality $\mathbf{s} \cdot \mathbf{p} \le S_l$. When such a behavior is observed one speaks of a Bell inequality violation.

1. The local polytope

We now investigate how Bell inequalities, i.e., the hyperplanes characterizing the set $\mathcal{L}$, can be found. To this end, it is useful to note that we can express local correlations in a simpler form. The first step is to realize that local correlations can, equivalent to Eq. (11), be defined in terms of *deterministic* local hidden-variable models. In a deterministic model, the local response functions $p(a|x,\lambda)$ and $p(b|y,\lambda)$ only take the value 0 or 1, that is, the hidden variable λ fully specifies the outcome that is obtained for each measurement. No such requirement is imposed on the general stochastic model (11). That both definitions are equivalent follows from the fact that any local randomness present in the response functions $p(a|x,\lambda)$ and $p(b|y,\lambda)$ can always be incorporated in the shared random variable λ. To see this, introduce two parameters $\mu_1, \mu_2 \in [0,1]$ in order to define a new hidden variable $\lambda' = (\lambda, \mu_1, \mu_2)$. Let

$$p'(a|x,\lambda') = \begin{cases} 1, & \text{if } F(a-1|x,\lambda) \le \mu_1 < F(a|x,\lambda), \\ 0, & \text{otherwise}, \end{cases} \tag{16}$$

where $F(a|x,\lambda) = \sum_{\tilde{a} \le a} p(\tilde{a}|x,\lambda)$, be a new response function for Alice and define a similar one for Bob. If we choose $q'(\lambda') = q'(\lambda, \mu_1, \mu_2) = q(\lambda)$ for the new hidden variable distribution, that is, if we uniformly randomize over μ_1 and μ_2, we clearly recover the predictions of the general, stochastic model (11). The newly defined model, however, is deterministic. This equivalence between the two models was first noted by Fine (1982).

We can further simplify the definition by noting that we need to consider only a finite number of hidden variables. Indeed, in a deterministic model, each hidden variable λ defines an assignment of one of the possible outputs to each input. The model as a whole is a probabilistic mixture of these deterministic assignments of outputs to inputs, with the hidden variable specifying which particular assignment is chosen in each run of the experiment. Since the total number of inputs and outputs is finite, there can be only a finite number of such assignments, and hence a finite number of hidden variables.

More precisely, we can rephrase the local model (11) as follows. Let $\lambda = (a_1, \ldots, a_m; b_0, \ldots, b_m)$ define an assignment of outputs a_x and b_y for each of the inputs $x = 1, \ldots, m$ and $y = 1, \ldots, m$. Let $\mathbf{d}_\lambda \in \mathcal{L}$ denote the corresponding deterministic behavior

$$d_\lambda(ab|xy) = \begin{cases} 1, & \text{if } a = a_x \text{ and } b = b_y, \\ 0, & \text{otherwise}. \end{cases} \tag{17}$$

There are Δ^{2m} possible output assignments and therefore Δ^{2m} such local deterministic behaviors. A behavior $\mathbf{p}$ is local if and only if it can be written as a convex combination of these deterministic points, that is, if

$$\mathbf{p} = \sum_\lambda q_\lambda \mathbf{d}_\lambda, \qquad \text{with } q_\lambda \ge 0, \quad \sum_\lambda q_\lambda = 1. \tag{18}$$

This last representation is particularly useful as it provides an algorithm for determining if a given behavior $\mathbf{p}$ is local (Zukowski et al., 1999; Kaszlikowski et al., 2000). Indeed, determining whether there exist weights q_λ satisfying the linear constraints in Eq. (18) is a typical instance of a *linear programming* problem (Boyd and Vandenberghe, 2004) for which there exist algorithms that run in time that is polynomial in the number of variables. Note, however, that since there are Δ^{2m} possible λ the size of this particular linear program is extremely large and hence the algorithm is not efficient by itself. Every linear program comes in a primal and a dual form. The dual form of the linear program associated with Eq. (18) has an interesting physical interpretation. Indeed, it can be formulated as

$$\begin{aligned} \max_{(\mathbf{s},S_l)} \quad & \mathbf{s} \cdot \mathbf{p} - S_l, \\ \text{s.t.} \quad & \mathbf{s} \cdot \mathbf{d}_\lambda - S_l \le 0, \quad \lambda = 1, \ldots, \Delta^{2m}, \\ & \mathbf{s} \cdot \mathbf{p} - S_l \le 1. \end{aligned} \tag{19}$$

If $\mathbf{p}$ is local, the maximum S of the above program is $S \le 0$. If $\mathbf{p}$ is nonlocal, the maximum is $S = 1$, i.e., the program returns an inequality $\mathbf{s} \cdot \mathbf{d}_\lambda \le S_l$ satisfied by all deterministic points (and hence, by convexity, by all local points), but violated by $\mathbf{p}$: $\mathbf{s} \cdot \mathbf{p} = S_l + 1 > S_l$. That is, Eq. (19) provides a procedure for finding, for any $\mathbf{p}$, a Bell inequality that detects its nonlocality.

Since the set $\mathcal{L}$ is the convex hull of a finite number of points, it is a *polytope*. The local deterministic behaviors $\mathbf{d}_\lambda$ correspond to the vertices, or extreme points, of the polytope. It is a basic result in polyhedral theory, known as Minkowski's theorem, that a polytope can, equivalently to Eq. (18) as the convex hull of its vertices, be represented as the intersection of finitely many half-spaces. Hence, we have that $\mathbf{p} \in \mathcal{L}$ if and only if

$$\mathbf{s}^i \cdot \mathbf{p} \le S_l^i \quad \forall \; i \in I, \tag{20}$$

where I indexes a finite set of linear inequalities. If, on the other hand, p is nonlocal, it necessarily violates one of the inequalities in Eq. (20). Thus the local set $\mathcal{L}$ can be characterized by a finite set of Bell inequalities.

2. Facet Bell inequalities

If $\mathbf{s} \cdot \mathbf{p} \le S_l$ is a valid inequality for the polytope $\mathcal{L}$, then $F = \{\mathbf{p} \in \mathcal{L} | \mathbf{s} \cdot \mathbf{p} = S_l\}$ is called a face of $\mathcal{L}$. Faces of dimension $\dim F = \dim \mathcal{L} - 1 = t - 1$ are called facets of $\mathcal{L}$ and the corresponding inequalities are called facet Bell inequalities. The terminology "tight Bell inequalities" is also used.[4] Facet inequalities are important because they provide a minimal representation of the set $\mathcal{L}$ in Eq. (20): minimal as they are necessarily required in the description (20), and since any other Bell inequality can be written as a non-negative combination of the facet inequalities. These notions are easily understood and visualized in two or three dimensions (note, however, that our low-dimensional intuition is often unreliable in higher dimensions). A more general discussion of polytope theory (not applied to Bell inequalities) can be found in, e.g., Schrijver (1989) and Ziegler (1995). The connection between optimal Bell inequalities and polytope theory was realized early by Froissard (1981) and later by Garg and Mermin (1984), Pitowsky (1989), Peres (1999), and Werner and Wolf (2001b).

Facet Bell inequalities provide a practical description of the local polytope $\mathcal{L}$. Usually, however, we start from the vertices of $\mathcal{L}$, which are the local deterministic behaviors $\mathbf{d}_\lambda$. The task of determining the facets of a polytope, given its vertices, is known as the facet enumeration or convex hull problem. For sufficiently simple cases, it is possible to obtain all the facets with the help of computer codes, such as CDD (Fukuda, 2003) or PORTA (Christof and Lobel, 1997), which are specifically designed for convex hull computations. However, such programs become prohibitively time consuming as the number of parties, inputs, or outputs grow. Note also that the simpler problem of determining whether a behavior is local using the linear program associated with Eq. (19) also becomes rapidly impractical for a large number of inputs m since the number of deterministic points scales exponentially with m. Results in computer science tell us that this problem is in general extremely difficult (Babai, Fortnow, and Lund, 1991). Evidence in this direction was first given by Pitowsky (1989). Then it was proven that deciding whether a behavior is local for the class of Bell scenarios with binary outputs ($\Delta = 2$) and m inputs is NP complete (Avis *et al.*, 2004). It is therefore highly unlikely that the problem of characterizing the local polytope admits a simple solution in full generality.

In the following, we list some facet Bell inequalities of interest. Note that the positivity conditions [corresponding to $p(ab|xy) \geq 0$] are always facets of the local polytope, but obviously are never violated by any physical theory. All other facet inequalities are violated by some no-signaling behaviors and possibly by some quantum behaviors. It is in fact an open question whether there exist facet inequalities in the bipartite case (different than the positivity ones) that are not violated by any quantum behaviors [such inequalities are known in Bell scenarios with more parties (Almeida, Bancal *et al.*, 2010)]. Note also that if an inequality defines a facet of the local polytope then it is obviously also the case for all the inequalities obtained from it by relabeling the outputs, inputs, or parties. What we mean thus in the following by an "inequality" is the whole class of inequalities obtained by such operations. Finally, it was shown by Pironio (2005) that there exists a hierarchical structure in the facial structure of local polytopes, in the sense that a facet Bell inequality of a given polytope with Δ outputs and m inputs can always be extended (or lifted) to any polytope with $\Delta' \geq \Delta$ and/or $m' \geq m$ (and also to polytopes corresponding to more parties) in such a way as to define a facet of the new polytope.

3. Examples

The simplest nontrivial Bell scenario corresponds to the case $\Delta = 2$, $m = 2$. The corresponding local polytope was completely characterized by Froissard (1981) and independently by Fine (1982). In this case, there is only one (nontrivial) facet inequality: the CHSH inequality introduced in Eq. (4). It was shown by Pironio (2004) that the CHSH inequality is also the only facet inequality for all polytopes with two inputs and two outputs for Alice and an arbitrary number of inputs and outputs for Bob.

The case $\Delta = 2$, $m = 3$ was computationally solved by Froissard (1981) who found that, together with the CHSH inequality, the inequality

$$\begin{aligned} p_1^A + p_1^B - p_{11} - p_{12} - p_{13} - p_{21} - p_{31} - p_{22} \\ + p_{23} + p_{32} \geq -1 \end{aligned} \tag{21}$$

is facet defining, where $p_x^A = p(a = 1|x)$, $p_y^B = p(b = 1|y)$, and $p_{xy} = p(a = 1, b = 1|xy)$. This result was later on rederived by Sliwa (2003) and Collins and Gisin (2004). The Froissard inequality is also referred to as the I_{3322} inequality, following the terminology of Collins and Gisin (2004). Note that this inequality could be generalized for the case of an arbitrary number of measurements m with binary outcomes, a family known as the I_{mm22} inequalities (Collins and Gisin, 2004), proven to be facets by Avis and Ito (2007).

For Δ arbitrary and $m = 2$, Collins, Gisin, Linden *et al.* (2002) introduced the following inequality [we used the notation of Acín, Gill, and Gisin (2005)]:

$$[a_1 - b_1] + [b_1 - a_2] + [a_2 - b_2] + [b_2 - a_1 - 1] \geq d - 1, \tag{22}$$

where $[a_x - b_y] = \sum_{j=0}^{\Delta-1} j p(a - b = j \bmod \Delta|xy)$ and similarly for the other terms. Note that for convenience the measurement outcomes are now denoted as $a, b \in \{0, 1, ..., \Delta - 1\}$. This inequality is known as the Collins-Gisin-Linden-Massar-Popescu (CGLMP) inequality. For $\Delta = 2$, it reduces to the CHSH inequality. It has been shown to be facet defining for all Δ by Masanes (2003).

The above inequality can be extended to an arbitrary number of inputs m in the following way (Barrett, Kent, and Pironio, 2006):

$$\begin{aligned} [a_1 - b_1] + [b_1 - a_2] + [a_2 - b_2] + \cdots + [a_m - b_m] \\ + [b_m - a_1 - 1] \geq d - 1. \end{aligned} \tag{23}$$

Although this Bell inequality is not a facet inequality, it is useful in several contexts. In the case $\Delta = 2$, it reduces to the

[4] Note, however, that in polytope theory a tight inequality refers merely to an inequality that "touches" the polytope, i.e., such that $F \neq \emptyset$.

chained inequality introduced by Pearle (1970) and Braunstein and Caves (1990).

Beyond these simple cases, a large zoology of Bell inequalities has been derived and it would be impossible to discuss them all here in detail, in particular, given the increase of complexity with larger values of Δ and m. For instance, in the case $\Delta = 2$, there is only one (nontrivial) facet Bell inequality for $m = 2$, two inequalities for $m = 3$, but already for $m = 4$ their number is not known (Brunner and Gisin, 2008). For $m = 10$, there are at least 44 368 793 inequalities (Avis *et al.*, 2004) (and this value is probably a gross underestimate). To complete the simple examples given above, we mention some recent papers where new Bell inequalities have been derived. Collins and Gisin (2004) and Brunner and Gisin (2008) obtained several facet Bell inequalities by numerically solving the convex hull problem for small values of Δ and m. Avis *et al.* (2004, 2005) and Avis and Ito (2007) obtained large families of Bell inequalities by establishing a relation between the local polytope for $\Delta = 2$ and a high-dimensional convex polytope called the cut polytope in polyhedral combinatorics. Vértesi (2008), Vértesi and Pal (2008), and Pal and Vértesi (2009) proposed new algorithms to construct families of facet and nonfacet Bell inequalities in the $\Delta = 2$ case. Methods exploiting symmetries to generate Bell inequalities for arbitrary Δ and m (and an arbitrary number n of parties) have been investigated by Bancal, Gisin, and Pironio (2010) and Bancal, Branciard *et al.* (2012). While we focused here on the case where Δ and m are finite, it is also possible to define Bell inequalities taking a continuous set of values for the outputs (Cavalcanti *et al.*, 2007; Salles *et al.*, 2010) or the inputs (Kaszlikowski and Zukowski, 2000; Aharon *et al.*, 2013).

Finally note that nonlinear Bell inequalities have also been considered. Quadratic inequalities were discussed by Uffink (2002), while Cavalcanti *et al.* (2007) and Salles *et al.* (2010) considered Bell inequalities based on moments of the probability distribution. Another approach, based on entropic quantities, was introduced by Braunstein and Caves (1988) and further developed by Cerf and Adami (1997) and Chaves and Fritz (2012).

4. Nonlocal games

Bell inequalities are also referred to as nonlocal games or sometimes simply as games. Looking at Bell inequalities through the lens of games often provides an intuitive understanding of their meaning. Such games enjoy a long history in computer science where they are known as interactive proof systems; see Condon (1989) for an early survey. More recently, they have also been studied in the quantum setting, under the name of interactive proof systems with entanglement (Cleve *et al.*, 2004). In order to make such literature accessible, we see how the two concepts can be translated into each other.[5]

When talking about a game, we imagine that there is an outside party, the referee that plays the game against Alice and Bob. In this context, parties or systems are referred to as *players*. Papers dealing with interactive proof systems also refer to the referee as the verifier and to the players as *provers*. The referee chooses a question $x \in X$ for Alice and $y \in Y$ for Bob according to some probability distribution $\pi: X \times Y \to [0,1]$ from some set of possible questions X and Y. Upon receiving x from the referee, Alice (Bob) returns an answer $a \in R_A$ $(b \in R_B)$ from some set of possible answers R_A (R_B). The referee then decides whether these answers are winning answers for the questions he posed according to the rules of the game. These rules are typically expressed by means of a predicate $V: R_A \times R_B \times X \times Y \to \{0,1\}$, where $V(a,b,x,y) = 1$ if and only if Alice and Bob win against the referee by giving answers a and b for questions x and y. To emphasize the idea that the correct answers depend on the questions given, one often writes the predicate as $V(a,b|x,y)$.

Alice and Bob are fully aware of the rules, that is, they know the predicate V and the distribution π. Before the game starts, they can agree on any strategy that may help them thwart the referee. However, once the game starts they can no longer communicate. In particular, this means that Alice does not know which question is given to Bob and vice versa. In the classical setting, such a strategy consists of shared randomness, which is the computer science name for local hidden variables. In the quantum case, Alice and Bob's strategy consists of a choice of shared quantum state and measurements.

The relation between games and Bell inequalities becomes apparent by noting that the questions are simply labels for measurement settings. That is, using our earlier notation we can take $X = Y = \{1, \ldots, m\}$. Note that we can without loss of generality assume that the number of settings $|X|$ and $|Y|$ are the same; otherwise, we can simply extend the number of settings for each party but never employ them. Similarly, the answers correspond to measurement outcomes. That is, we can take $R_A = R_B = \{1, \ldots, \Delta\}$.

Any strategy leads to some particular probabilities $p(a,b|x,y)$ that Alice and Bob give answers a, b for questions x, y, respectively. In the language of Bell inequalities, this is simply the probability that Alice and Bob obtain measurement outcomes a and b when performing the measurements labeled x and y. The probability that Alice and Bob win against the referee for some particular strategy can thus be written as

$$p_{\text{win}} = \sum_{x,y} \pi(x,y) \sum_{a,b} V(a,b|x,y) p(a,b|x,y). \tag{24}$$

In the classical or quantum setting, one can consider the maximum winning probability that Alice and Bob can achieve. For instance, considering classical resources, we have

$$\max p_{\text{win}} = S_l, \tag{25}$$

where the maximization is taken over all deterministic strategies of Alice and Bob. Note that this leads to the familiar form of a Bell inequality

$$p_{\text{win}} = \mathbf{s} \cdot \mathbf{p} \leq S_l, \tag{26}$$

where the coefficients are given by

[5] For the purpose of illustration, we will here restrict ourselves to the case of only two parties, Alice and Bob. However, the relation holds for an arbitrary amount of parties.

$$s_{x,y}^{a,b} = \pi(x,y)V(a,b|x,y). \tag{27}$$

Hence games form a subset of general Bell inequalities. In complexity theory, the winning probability is also often referred to as the value of the game.

a. XOR games

A class of games that is very well understood are so-called XOR games (Cleve *et al.*, 2004). In an XOR game, each player has only two possible answers $a, b \in \{0,1\}$. To decide whether Alice and Bob win, the referee computes the XOR $c = a \oplus b := a + b \bmod 2$ and then bases his decision solely on c. For such games the predicate is generally written as $V(c|x,y) := \sum_a V(a, b = c \oplus a|x,y)$. We see later that it is easy to find the optimal quantum strategy for XOR games, and indeed the structure of their optimal measurements is entirely understood. Also, multiplayer XOR games are reasonably well understood and bounds relating the classical and quantum winning probabilities are known (Briet and Vidick, 2013).

We simply note for the moment that XOR games are equivalent to correlation Bell inequalities with binary outcomes. Indeed, from Eq. (10) it follows that we can write $p(a \oplus b = 0|x,y) = \frac{1}{2}(1 + \langle A_x B_y \rangle)$ and $p(a \oplus b = 1|x,y) = \frac{1}{2}(1 - \langle A_x B_y \rangle)$. The winning probability for an XOR game can thus be written as

$$p_{\rm win} = \frac{1}{2}\sum_{x,y}\pi(x,y)\sum_{c\in\{0,1\}} V(c|x,y)[1 + (-1)^c \langle A_x B_y \rangle], \tag{28}$$

which is the general form of a correlation Bell inequalities. XOR games can thus be recast as correlation inequalities and vice versa.

b. An example: CHSH as a game

An illustrative example of how correlation Bell inequalities transform into games and vice versa is provided by the CHSH inequality. For convenience, we take here $X = Y = \{0,1\}$ (instead of $\{1,2\}$), as well as $R_A = R_B = \{0,1\}$. Viewing CHSH as a game, the rules state that Alice and Bob win if and only if $x \cdot y = a \oplus b$. Plugging this into Eq. (28) one obtains

$$p_{\rm win}^{\rm CHSH} = \frac{1}{2}\left(1 + \frac{S}{4}\right), \tag{29}$$

where S is the CHSH expression as given in Eq. (4). Indeed one has $S \leq 2$ for any classical strategy. Hence, the probability for Alice and Bob to win the game using classical resources is at most $3/4$. Using quantum resources, the winning probability is at most $(1 + 1/\sqrt{2})/2 \approx 0.85$, as given by Tsirelson's bound $S \leq 2\sqrt{2}$.

c. Projection and unique games

A projection game is a game in which for every pair of questions x and y to Alice and Bob, and for every answer a of Alice, there exists a unique winning answer for Bob. In the quantum information literature, these are also often simply called unique games. However, in the classical computer science literature and also some of the quantum information literature the term unique game can also refer to a game for which for any pair of questions (x,y) there exists a permutation $\pi_{x,y}$ over the set $\{1,\ldots,\Delta\}$ of possible answers such that Alice and Bob win if and only if their answers obey $a = \pi_{x,y}(b)$. In terms of the predicate this means that $V(a,b|x,y) = 1$ if and only if $b = \pi_{x,y}(a)$. Note that in this language, every unique game is a projection game because there is only one correct answer for Bob for each x, y, and a. However, not every projection game forms a unique game.

A more general notion which imposes a limit on the number of winning answers are k-to-k' games (Kempe, Regev, and Toner, 2010). More precisely, a game is k to k' if for all questions x and y the following two conditions hold: for all answers a of Alice there exist at most k winning answers for Bob, and for all answers b of Bob there exist at most k' winning answers for Alice. A projection game is thus a k-to-k' game for $k = k' = 1$.

d. Other special classes of games

Several other special classes of games have been studied on occasion. A linear game is a game for which one can associate the set of possible answers $\{1,\ldots,\Delta\}$ with an Abelian group G of size Δ and find a function $W: \{1,\ldots,m\}^{\times 2} \to G$ such that $V(a,b|x,y) = 1$ if and only if $a - b = W(x,y)$. Any linear game is a unique game and has been shown to have the special property to be a uniform game, that is, a game in which there exists an optimal quantum strategy such that the marginal distributions $p(a|x)$ and $p(b|y)$ are the uniform distributions over R_A and R_B, respectively (Kempe, Regev, and Toner, 2010). Furthermore, a game may be called *free* if the questions are drawn from a product distribution, that is, $\pi(x,y) = \pi_A(x) \times \pi_B(y)$ for some distributions π_A and π_B (Kempe and Vidick, 2011). A game is called *symmetric* if for all questions x, y and all answers a, b we have $V(a,b|x,y) = V(b,a|y,x)$ (Dinur and Reingold, 2006; Kempe and Vidick, 2011). An example of a game that is both free and symmetric is given by the CHSH game above. Another class of games that has drawn attention in the computer science literature is characterized merely by the fact that there exists a quantum strategy that wins the game with probability $p_{\rm win} = 1$. Such games are sometimes also called Kochen-Specker games (or pseudotelepathy or Greenberger-Horne-Zeilinger games) due to the fact that the optimal quantum strategy yields a so-called Kochen-Specker set (Renner and Wolf, 2004), a concept in contextuality which is outside the scope of this review [see Brassard, Broadbent, and Tapp (2005) for a survey on such games; see also the related discussion in Sec. II.E].

C. Bell inequality violations

In the above discussion, we saw that it is in principle possible to decide (albeit very inefficiently) whether a given behavior is local and to compute the local bound S_l of an arbitrary Bell expression. In this section, we look at the analogous problem in the quantum and no-signaling cases. We review the existing methods for computing the quantum and no-signaling bounds, i.e., the maximal quantum and no-signaling violations, of an arbitrary Bell expression $\mathbf{s}$. Such methods can also be used to determine if a given behavior admits a quantum or no-signaling representation and thus this section is more generally concerned with the problem of practical characterizations of the quantum and no-signaling sets beyond the formal definitions (7) and (12).

1. Quantum bounds

a. Properties of quantum correlations

Before discussing in more detail how one can compute the quantum bound S_q of a Bell expression, we briefly discuss the general structure of the quantum set $\mathcal{Q}$. Recall that a behavior $\mathbf{p}$ is quantum if, as defined in Eq. (14), it can be written as $p(ab|xy) = \langle\psi|M_{a|x}M_{b|y}|\psi\rangle$, where $|\psi\rangle$ is a state in a Hilbert space $\mathcal{H}$, and $M_{a|x}$ and $M_{b|y}$ are orthogonal projectors on $\mathcal{H}$ defining proper measurements and satisfying $[M_{a|x}, M_{b|y}] = 0$. (For characterizing the quantum set it is convenient to assume we impose commutation relations rather than a tensor product structure and we follow this approach in the remainder of this section.)

As mentioned, the local set $\mathcal{L}$ is strictly contained in the quantum set $\mathcal{Q}$, i.e., there are quantum behaviors that are nonlocal, and thus in general $S_q > S_l$. There are two basic requirements that any quantum behavior must satisfy to be nonlocal. First, Alice's different measurements must be noncommuting as well as Bob's (Fine, 1982). Second, the state ρ must be entangled. Without surprise, quantum nonlocality can thus be traced back to the two features usually seen as distinguishing quantum from classical physics: noncommutativity and entanglement.

Contrary to the local set, the set $\mathcal{Q}$ of quantum correlations is generally not a polytope. It cannot therefore be described by a finite number of extreme points or a finite number of linear inequalities. It is not difficult to see though that all extremal points of $\mathcal{L}$, i.e., the local deterministic behaviors, are also extremal points of $\mathcal{Q}$. Furthermore, certain faces of $\mathcal{L}$ are also faces of $\mathcal{Q}$. An example is provided by the $(\Delta - 1)$-dimensional face associated with the hyperplanes $p(ab|xy) = 0$ [note, however, that the corresponding Bell inequalities $p(ab|xy) \geq 0$ cannot be violated by any physical behavior]. Thus, while Q is not a polytope, its boundary contains some flat regions. Linden *et al.* (2007)showed that the local and quantum sets have common faces which correspond to Bell inequalities that can be violated by certain no-signaling behaviors. As mentioned earlier, it is an open question whether there exist such examples of maximal dimension, i.e., whether there exist facets of $\mathcal{L}$ corresponding to Bell inequalities that are not violated by $\mathcal{Q}$ but which can be violated by $\mathcal{NS}$ [such examples are known in the tripartite case (Almeida, Bancal *et al.*, 2010)].

The boundary of the nonlocal part of $\mathcal{Q}$ may also contain flat regions, i.e., the maximal violation of a Bell inequality may sometimes be realized with two or more different nonlocal quantum behaviors. The question of when an extremal quantum behavior can be realized by a unique quantum representation (up to unitary equivalence) was considered by Franz, Furrer, and Werner (2011), where it was, in particular, shown that in the correlation space $\mathcal{C}$ all nonlocal extremal behaviors are uniquely realizable in the cases $m = 2$ and $m = 3$. Examples of noncorrelation Bell inequalities that are maximally violated by unique quantum behaviors have been given by Acín, Massar, and Pironio (2012). These inequalities are maximally violated by partially entangled states, thus showing that these state are necessary to characterize the boundary of the quantum region; see also Liang, Vértesi, and Brunner (2011) and Vidick and Wehner (2011). Note, however, that the so-called embezzling state (van Dam and Hayden, 2003) is universal in the sense that any two-party Bell inequality can be maximally violated using an embezzling state (de Oliveira Oliveira, 2010) up to a small error term.

We now focus more specifically on the problem of computing the quantum bound of a Bell expression. Recall that $\mathcal{Q}$ as any convex compact set can be described by an infinite system of linear inequalities of the form $\mathbf{s} \cdot \mathbf{p} \leq S_q$, here the quantum Bell inequalities. Given an arbitrary Bell expression s, its corresponding quantum bound is given by

$$S_q = \max_{p \in \mathcal{Q}} \mathbf{s} \cdot \mathbf{p} = \max_{\mathcal{S}} ||\mathcal{S}||, \tag{30}$$

where

$$\mathcal{S} = \sum_{abxy} s_{xy}^{ab} M_{a|x} M_{b|y} \tag{31}$$

is the *Bell operator* associated with s, $||\mathcal{S}||$ denotes the spectral norm (largest eigenvalue) of $\mathcal{S}$, and the above optimization is performed over all possible Bell operators $\mathcal{S}$ associated with $\mathbf{s}$. That is, over all possible measurements $\{M_{a|x}\}_a$ and $\{M_{b|x}\}_b$, where the coefficients s_{xy}^{ab} are given by the choice of $\mathbf{s}$. In the case of the CHSH expression, the Bell operator takes the form $\mathcal{S} = \hat{A}_1(\hat{B}_1 + \hat{B}_2) + \hat{A}_2(\hat{B}_1 - \hat{B}_2)$, where $\hat{A}_x$, $\hat{B}_y$ are arbitrary ± 1-eigenvalued observables. Following Landau (1987), we can derive the quantum bound of the CHSH inequality (Cirel'son, 1980) by noting that $\mathcal{S}^2 = 4 + [\hat{A}_1, \hat{A}_2][\hat{B}_1, \hat{B}_2]$, from which it follows that $||\mathcal{S}^2|| \leq 8$ and hence $||\mathcal{S}|| \leq 2\sqrt{2}$. Computing the quantum bound of other Bell expressions is a more complicated business. It has been shown to be an NP-hard problem in the tripartite case (Kempe *et al.*, 2011).

b. Correlation inequalities

The case of quantum correlation inequalities defined in the correlation space $\mathcal{C}$ is particularly well understood thanks to Tsirelson (Cirel'son, 1980; Tsirelson, 1987, 1993). Recall that, in the correlation space, a behavior is defined by the m^2 correlators $\langle A_x B_y \rangle$. It is easy to see that such a behavior is quantum if we write $\langle A_x B_y \rangle = \langle\psi|\hat{A}_x \otimes \hat{B}_y|\psi\rangle$ for some quantum state $|\psi\rangle$ in $\mathcal{H}_A \otimes \mathcal{H}_B$ and some ± 1-eigenvalued quantum observables A_x on $\mathcal{H}_A$ and B_y on $\mathcal{H}_B$. Tsirelson showed that it is sufficient to consider $\dim \mathcal{H}_A = \dim \mathcal{H}_B = 2^m$ if m is even and $\dim \mathcal{H}_A = \dim \mathcal{H}_B = 2^{m+1}$ if m is odd, and $|\psi\rangle$ is a maximally entangled state in $\mathcal{H}_A \otimes \mathcal{H}_B$. Furthermore, he showed that the m^2 correlators $\langle A_x B_y \rangle$ are quantum if and only if there exist $2m$ unit vectors $\hat{v}_x$ and $\hat{w}_y$ in $\mathbb{R}^{2m}$ such that

$$\langle A_x B_y \rangle = \hat{v}_x \cdot \hat{w}_y \tag{32}$$

for all $x, y \in \{1, \ldots, m\}$. This last representation is particularly useful, as deciding if a behavior can be written in the form (32) can be cast as a semidefinite program (SDP) for which efficient algorithms are available (Cleve *et al.*, 2004; Wehner, 2006b). This means that the problem of computing the winning probability of the game $p_{\rm win}$ is in the complexity class EXP (exponential time), since SDPs can be solved in polynomial time but the input is of exponential size. However, combining Jain *et al.* (2010) and Wehner (2006a) one now knows that the problem of computing $p_{\rm win}$ for XOR games lies in the complexity class PSPACE (the set of all decision problems that can be solved by a Turing machine using a polynomial amount of space).

This technique can be used to compute tight bounds for two-outcome correlation inequalities, i.e., XOR games. In particular, the quantum bounds for the CHSH inequality and the chained inequalities [Eq. (23) in the case $\Delta = 2$] can easily be obtained in this way (Wehner, 2006b). It should be noted that this SDP technique can be seen as a special case of the general SDP method discussed in Sec. II.C.1.d.

In the $\Delta = 2$, $m = 2$, this SDP approach can be used to yield a complete description of $\mathcal{Q}\cap\mathcal{C}$ (i.e., the quantum part of the correlation space $\mathcal{C}$) in terms of a finite set of nonlinear inequalities: a behavior is quantum if and only if it satisfies

$$|\mathrm{asin}\langle A_1B_1\rangle + \mathrm{asin}\langle A_1B_2\rangle + \mathrm{asin}\langle A_2B_1\rangle - \mathrm{asin}\langle A_2B_2\rangle| \le \pi \tag{33}$$

together with the inequalities obtained by permuting the $\langle A_xB_y\rangle$ in Eq. (33) (Cirel'son, 1980; Tsirelson, 1987; Landau, 1988; Masanes, 2003). For further results and a more detailed discussion of the characterization of $\mathcal{Q}$ in the correlation space $\mathcal{C}$, see Tsirelson (1987, 1993) and Avis, Moriyama, and Owari (2009).

It is interesting to note that it is much harder to determine the optimal local bound S_l for a correlation Bell inequality than it is to compute the quantum one unless $P = \mathrm{NP}$ (Cleve *et al.*, 2004). That is, the quantum problem is actually easier than the classical one.

c. State and measurement dependent bounds

We now return to the general case of quantum correlations in the probability space $\mathcal{P}$. To compute the quantum bound (30) of a Bell expression, the first simple approach is to introduce an explicit parametrization of a family of Bell operators $\mathcal{S}$ in a Hilbert space $H = H_A \otimes H_B$ of fixed dimension $\dim H = d_H$, and to maximize $||\mathcal{S}||$ over all operators in this family. In general, however, we have no *a priori* guarantee that the optimal quantum bound can be realized using a Bell operator from this particular family. Furthermore, most optimization methods cannot guarantee convergence to the global extremum. This approaches therefore typically yields only lower bounds on S_q. It is nevertheless very useful when looking for an explicit quantum violation of a Bell inequality $\mathbf{s}\cdot\mathbf{p} \le S_l$ (although we have no guarantee that this is the optimal quantum violation).

Rather than directly trying to obtain a state-independent bound by maximizing the norm of the Bell operator, it is often easier to compute the quantum bound for a fixed quantum state $|\psi\rangle$, i.e., maximize $\langle\psi|\mathcal{S}|\psi\rangle$ over all Bell operators $\mathcal{S}$. This optimization can be dealt with as previously by introducing an explicit parametrization of a family of Bell operators. Another possibility, introduced by Liang and Doherty (2007), is to exploit the fact that, for a given quantum state, a Bell expression is bilinear in the measurement operators, that is, it is linear in the operators $\{M_{a|x}\}$ for fixed $\{M_{b|y}\}$ and linear in the $\{M_{b|y}\}$ for fixed $\{M_{a|x}\}$. When the measurements on one system are fixed, the problem of finding the optimal measurements for the other system can therefore be cast as a SDP. This SDP can then be used as the basis for an iterative algorithm: fix Bob's measurements and find Alice's optimal ones; with these optimized measurements for Alice now fixed, find the optimal ones for Bob; then optimize again over Alice's measurements and so on, until the quantum value converges within the desired numerical precision. A similar iterative algorithm was introduced by Werner and Wolf (2001a) for correlation inequalities. In this case, once the measurements for one party are fixed, optimization of the other party's measurements can be carried out explicitly. This turns out to be true not only for correlation inequalities but for any Bell expression with binary outcomes (Liang and Doherty, 2007). Finally, we note that a method for finding an optimal Bell operator for a fixed quantum state can again be used in an iterative algorithm to find a state-independent bound (Pal and Vértesi, 2009): starting with an initial quantum state (e.g., a maximally entangled state), find the corresponding optimal Bell operator; then find the optimal quantum state associated with this Bell operator (i.e., the eigenvector associated with the largest eigenvalue); and repeat these steps starting from this new state.

d. General bounds

The techniques just described provide lower bounds on S_q. Looking at Eq. (30) it becomes clear that finding S_q can be understood as an instance of polynomial optimization. More specifically, we want to optimize Eq. (31) over noncommutative variables $M_{a|x}$, $M_{b|y}$ subject to certain constraints, namely, that such variables form quantum measurements and Alice measurement operators commute with those of Bob. It is known that in principle any polynomial optimization problem in commutative variables can be solved using a hierarchy of SDPs—two general methods that are dual to each other were introduced by Lasserre (2001) and Parrilo (2003), respectively.

It turns out that these techniques can be extended to the quantum setting (Navascues, Pironio, and Acín, 2007, 2008; Doherty *et al.*, 2008), yielding a powerful approach to obtaining upper bounds on S_q, i.e., of deriving constraints satisfied by the entire quantum set. This method was originally introduced by Navascues, Pironio, and Acín (2007), which follows the ideas of Lasserre (2001). The idea is basically the following. Let $|\psi\rangle$ and $\{M_{a|x}\}$, $\{M_{b|y}\}$ define a quantum realization of a behavior $p \in \mathcal{Q}'$, i.e., $p(ab|xy) = \langle\psi|M_{a|x}M_{b|y}|\psi\rangle$. Let $\mathcal{O}$ be a set of k operators consisting of all operators $M_{a|x}$ and $M_{b|y}$ together with some finite products of them. For instance, $\mathcal{O}$ may consist of all operators of the form $M_{a|x}$, $M_{b|y}$, $M_{a|x}M_{a'|x'}$, $M_{a|x}M_{b|y}$, and $M_{b|y}M_{b'|y'}$. Denote by O_i ($i = 1,\ldots,k$) the elements of $\mathcal{O}$ and introduce the $k\times k$ matrix Γ with entries $\Gamma_{ij} = \langle\psi|O_i^\dagger O_j|\psi\rangle$, called the *moment matrix* associated with $\mathcal{O}$. Then the following properties are easily established (independently of the particular quantum realization considered): (i) $\Gamma \succeq 0$ is semidefinite positive, (ii) the entries of Γ satisfy a series of linear inequalities, and (iii) the probabilities $p(ab|xy)$ defining the behavior $\mathbf{p}$ correspond to a subset of the entries of Γ. A necessary condition for a behavior $\mathbf{p}$ to be quantum is therefore that there exists a moment matrix Γ with the above properties, a problem that can be determined using SDP. For any $\mathcal{O}$, the set of behaviors $\mathbf{p} \in \mathcal{P}$ for which there exists such a moment matrix thus define a set $\mathcal{Q}_{\mathcal{O}}$ that contains the quantum set $\mathcal{Q}'$ (and thus also $\mathcal{Q}$). Optimizing a Bell expression (which is linear in $\mathbf{p}$) over this set $\mathcal{Q}_{\mathcal{O}}$ is also a SDP and yields an upper bound on S_q. Consider, in particular, the case where $\mathcal{O}$ is the set of all operators consisting of a product of at most ν of the operators $M_{a|x}$ and $M_{b|y}$ and denote the corresponding set of behaviors Q_ν. Then the associated SDP defines for $\nu = 1, 2, \ldots,$ a hierarchy $\mathcal{Q}_1 \supseteq \mathcal{Q}_2 \supseteq \cdots \supseteq \mathcal{Q}$ of relaxations approximating better and better the quantum region from the outside (see Fig. 3). Or, equivalently, they define a decreasing series of upper bounds on the quantum bound S_q of any Bell expression.

Subsequently, following the ideas of Parrilo (2003), Doherty *et al.* (2008) constructed the SDP hierarchy that is dual to Navascues, Pironio, and Acín (2007). It relies on the fact that for any Bell operator $\mathcal{S}$ we have $\xi = \hat{b}_q \mathbb{1} - \mathcal{S} \geq 0$ (i.e., $\|\mathcal{S}\| \leq \hat{b}_q$) if and only if the polynomial ξ can be written as a (weighted) sum of squares of other polynomials. We can thus think of minimizing $\hat{b}_q$ such that ξ is a sum of squares of polynomials in order to find $\|\mathcal{S}\|$. If we limit the degree of these polynomials, the problem can be cast as an SDP. Very roughly, at level ℓ of the SDP hierarchy we then limit the degree to be at most 2ℓ, leading to better and better bounds for increasing values of ℓ.

Doherty *et al.* (2008) and Navascues, Pironio, and Acín (2008) showed that this hierarchy of SDP relaxations converges in the asymptotic limit to the set $\mathcal{Q}'$ [see also Pironio, Navascues, and Acín (2010) for a more general approach not limited to quantum correlations]. It is also possible to certify that a behavior $\mathbf{p}$ belongs to the quantum set $\mathcal{Q}$ or to obtain the optimal bound S_q of a Bell expression at a finite step in the hierarchy [see, e.g., (Doherty *et al.* (2008) and Navascues, Pironio, and Acín (2008) for a number of examples]. A criterion has been introduced by Navascues, Pironio, and Acín (2008) to determine when this happens and to reconstruct from the moment matrix Γ a quantum realization of this optimal solution in terms of an explicit state $|\psi\rangle$ and operators $M_{a|x}$ and $M_{b|y}$. Optimality at a finite step in the hierarchy can also be determined by comparing the SDP upper bound with a lower bound obtained by searching over explicit families of quantum Bell operators. Pal and Vértesi (2009), for instance, determined the optimal quantum value S_q of 221 Bell expressions in this way at the third step of the hierarchy. Even if they do not always provide an optimal bound, numerical examples show that low-order steps of the hierarchy usually already approximate very well the quantum bound. Kempe, Regev, and Toner (2010) proved that for a certain particular family of Bell scenarios, known as *unique games* (see Sec. II.B.4.c), the first step of the hierarchy always provides a good approximation of the quantum set. We also note that, for correlation inequalities, the first step of the hierarchy always provides the optimal solution as it is equivalent to the SDP approaches based on Tsirelson results mentioned earlier.

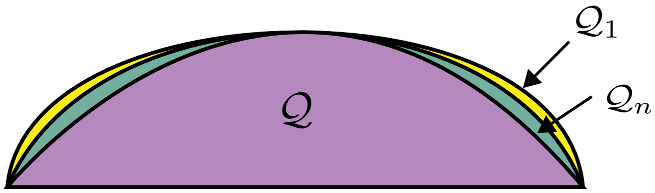

FIG. 3 (color online). Hierarchy of sets $\mathcal{Q}_\nu$ generated by the hierarchy of SDPs defined by Navascues, Pironio, and Acín (2007) (see Sec. II.C.1.d). Each set in the hierarchy better approximates the set of quantum correlations $\mathcal{Q}$. In the CHSH scenario, the set $\mathcal{Q}_1$ already achieves the maximum quantum value of the CHSH inequality, i.e., Tsirelson's bound.

In the $\Delta = 2$, $m = 2$ case, the set $\mathcal{Q}_1$ corresponding to the first step of the hierarchy was analytically characterized by Navascues, Pironio, and Acín (2007). A behavior $\mathbf{p}$ belongs to $\mathcal{Q}_1$ if and only if $\langle A_i \rangle^2 = 1$ or $\langle B_j \rangle^2 = 1$ for some $i, j = 1, 2$ or if it satisfies the inequality

$$|\mathrm{asin}\langle D_{11}\rangle + \mathrm{asin}\langle D_{12}\rangle + \mathrm{asin}\langle D_{21}\rangle - \mathrm{asin}\langle D_{22}\rangle| \leq \pi \qquad (34)$$

together with the inequalities obtained from this one by permuting the D_{ij}, where

$$D_{ij} = \frac{\langle A_i B_j \rangle - \langle A_i \rangle \langle B_j \rangle}{\sqrt{(1 - \langle A_i \rangle^2)(1 - \langle B_j \rangle^2)}}. \qquad (35)$$

The nonlinear inequalities (34) thus form a necessary condition for a behavior to be quantum. They strengthen the inequalities (33) to which they reduce when $\langle A_i \rangle = \langle B_j \rangle = 0$.

2. No-signaling bounds

We now consider the problem of computing bounds on Bell expressions for no-signaling correlations. Contrary to the case of local and quantum correlations, this turns out to be a rather easy task. To understand why note that, as mentioned, once the no-signaling constraints (7) are taken into account, e.g., by introducing a parametrization of the relevant affine subspace $\mathbb{R}^t$, the set $\mathcal{NS}$ of no-signaling behaviors is uniquely determined by the set of $\Delta^2 m^2$ positivity inequalities $p(ab|xy) \geq 0$. Deciding whether a behavior belongs to $\mathcal{NS}$ can thus easily be verified by checking that all positivity inequalities are satisfied. Since there are $\Delta^2 m^2$ such inequalities, this is a problem whose complexity scales polynomially with the number of inputs and outputs. More generally, linear programming can be used to efficiently determine the no-signaling bound S_{ns} of an arbitrary Bell expression $\mathbf{s}$, as used, e.g., by Toner (2009). Especially in the case of multipartite correlations it is sometimes convenient to compute S_{ns} to obtain a (crude) bound for S_q.

Finally, we remark that since $\mathcal{NS}$ is defined by a finite number of linear inequalities, it is, as the local set, a polytope and can also be described as the convex hull of a finite set of vertices. These can be obtained from the list of facets [the inequalities $p(ab|xy) \geq 0$] using the same polytope algorithms that allow one to list the facets of $\mathcal{L}$ given its vertices. The vertices of $\mathcal{L}$, the local deterministic points $\mathbf{d}_\lambda$, are clearly also vertices of $\mathcal{NS}$ (since they cannot be written as a convex combination of any other behavior). All other vertices of $\mathcal{NS}$ are nonlocal.

The geometry of the no-signaling set and its relation to $\mathcal{L}$ is particularly simple for the $\Delta = 2$, $m = 2$ Bell scenario. In this case, the no-signaling behaviors form an 8-dimensional subspace of the full probability space $\mathcal{P}$. The local polytope consists of 16 vertices, the local deterministic points, and 24 facets. Sixteen of these facets are positivity inequalities and eight are different versions, up to relabeling of the inputs and outputs, of the CHSH inequality. The no-signaling polytope, on the other hand, consists of 16 facets, the positivity inequalities, and 24 vertices. Sixteen of these vertices are the local deterministic ones and eight are nonlocal vertices, all equivalent up to relabeling of inputs and outputs to the behavior

$$p(ab|xy) = \begin{cases} 1/2, & \text{if } a \oplus b = xy, \\ 0, & \text{otherwise}, \end{cases} \tag{36}$$

which is usually referred to as a PR box. It is easily verified that the PR box violates the CHSH inequality (4) up to the value $\mathbf{s} \cdot \mathbf{p} = 4 > 2$, the algebraic maximum. In the language of games, this means that the CHSH game can be won with probability $p_{\text{win}}^{\text{CHSH}} = 1$. There exists a one-to-one correspondence between each version of the PR box and of the CHSH inequality, in the sense that each PR box violates only one of the CHSH inequalities. The PR box was introduced by Khalfin and Tsirelson (1985), Rastall (1985), and Popescu and Rohrlich (1994). Since the maximal quantum violation of the CHSH inequality is $2\sqrt{2}$, it provides an example of a no-signaling behavior that is not quantum, implying that in general $Q \neq NS$. The relation between $\mathcal{L}$, $\mathcal{Q}$, and $\mathcal{NS}$ in the $\Delta = 2$, $m = 2$ case is represented in Fig. 4.

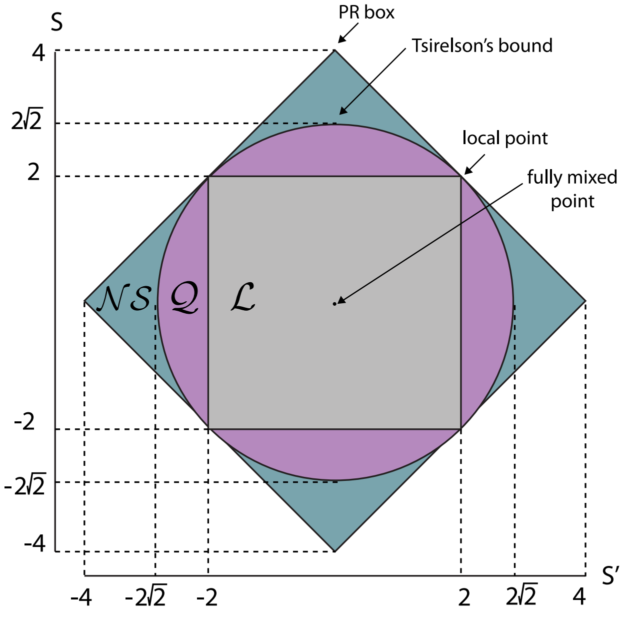

FIG. 4 (color online). A two-dimensional section of the no-signaling polytope in the CHSH scenario ($m = \Delta = 2$). The vertical axis represents the CHSH value S, while the horizontal axis represents the value of a symmetry of the CHSH expression S' (where inputs have been relabeled). Local correlations satisfy $|S| \leq 2$ and $|S'| \leq 2$. The PR box is the no-signaling behavior achieving the maximum CHSH value $S = 4$. Tsirelson's bound corresponds to the point where $S = 2\sqrt{2}$, i.e., the maximum CHSH value that a quantum behavior can achieve.

The complete list of all no-signaling vertices is also known in the case of two inputs ($m = 2$) and an arbitrary number of outputs (Barrett, Linden *et al.*, 2005) and in the case of two outputs ($\Delta = 2$) and an arbitrary number of inputs (Jones and Masanes, 2005; Barrett and Pironio, 2005). In both cases, the corresponding nonlocal vertices can be seen as straightforward generalizations of the PR box.

D. Multipartite correlations

Although we focused for simplicity in the preceding sections on Bell scenarios involving $n = 2$ systems, most of the above definitions and basic results extend straightforwardly to the case of an arbitrary number $n > 2$ of systems. For instance, in the tripartite case a behavior $p(abc|xyz)$ is no signaling when

$$\sum_c p(abc|xyz) = \sum_{c'} p(abc'|xyz') \quad \forall \; a, b, x, y, z, z' \tag{37}$$

and similar relations obtained from permutations of the parties; a behavior is local if it can be written as a convex combination of a finite number of deterministic behaviors $d_\lambda(abc|xyz)$; Bell inequalities correspond to faces of the corresponding polytope, and so on. Next we discuss a few notable results obtained in the multipartite case. Note that many references cited in the previous sections also contain results for more than two parties.

As in the bipartite case, one can consider Bell inequalities that involve only full correlators in the case where all measurements have binary outcomes. In the $n = 3$ case, for instance, such an inequality would involve only terms of the form $\langle A_x B_y C_z \rangle = \sum_{a,b,c=\pm 1} abc\, p(abc|xyz)$, and similarly for more parties. All correlation Bell inequalities with $m = 2$ inputs have been derived by Werner and Wolf (2001b) and Zukowski and Brukner (2002) for an arbitrary number n of parties. There are 2^{2^n} such inequalities (with redundancies under relabeling) which can be summarized in a single, but nonlinear inequality. Notable inequalities that are part of this family are the inequalities introduced by Mermin (1990a) and further developed by Ardehali (1992) and Belinskii and Klyshko (1993). In the case $n = 3$, the Mermin inequality takes the form

$$|\langle A_1 B_2 C_2 \rangle + \langle A_2 B_1 C_2 \rangle + \langle A_2 B_2 C_1 \rangle - \langle A_1 B_1 C_1 \rangle| \leq 2. \tag{38}$$

It is associated with the Greenberger-Horne-Zeilinger (GHZ) paradox (see Sec. II.E) in the sense that correlations that exhibit the GHZ paradox violate it up to the algebraic bound of 4. Werner and Wolf (2001b) also investigated the structure of the quantum region in the full correlation space. In particular, it was shown that the quantum bound of all inequalities introduced by Werner and Wolf (2001b) and Zukowski and Brukner (2002) is achieved by the n-partite GHZ state $(|00\cdots 0\rangle + |11\cdots 1\rangle)/\sqrt{2}$.

Sliwa (2003) derived all facet Bell inequalities (in the full probability space) for three parties in the case $\Delta = 2$, $m = 2$. There are 46 inequivalent such inequalities. All of these are violated in quantum mechanics, except for the inequality considered by Almeida, Bancal *et al.* (2010). Pironio, Bancal, and Scarani (2011) listed all vertices of the no-signaling polytope corresponding to the same Bell scenario. Interestingly, they are also 46 inequivalent classes of no-signaling vertices. Fritz (2012b) showed that there actually exists a bijection between facet Bell inequalities and no-signaling vertices for every Bell scenario with two inputs and outputs, independent of the number of parties.

Evidently, the structure of nonlocal correlations is much richer (and less understood) in the multipartite case than in the bipartite one. In particular, there exist different definitions of nonlocality that refine the straightforward extension of the bipartite definition. This question and others that are more specific to the multipartite scenario are discussed in Sec. VI.

E. Nonlocality without inequalities

To demonstrate that some quantum correlations $\mathbf{p}$ are nonlocal it is sufficient, as discussed in Sec. II.B, to exhibit a Bell inequality that is violated by $\mathbf{p}$. In certain cases, however, it is possible to directly show that quantum predictions are incompatible with those of any local model via a simple logical contradiction that does not involve any inequality (although such arguments can obviously always be rephrased as the violation of a Bell inequality). Here we present two examples of such “Bell's theorem without inequalities,” namely, the Greenberger-Horne-Zeilinger paradox and a construction due to Hardy.

The situation considered by Greenberger, Horne, and Zeilinger (1989) [see also Greenberger *et al.* (1990) and Mermin (1990b)] involves three players Alice, Bob, and Charlie. Each player receives a binary input, denoted by A_i, B_i, and C_i, with $i = 1, 2$. For each input, players should provide a binary output ± 1. With a slight abuse of notation, we denote by $A_i = \pm 1$ the answer to the question A_i and so on. Suppose that the players share a state of the form $|\mathrm{GHZ}\rangle = (1/\sqrt{2})(|000\rangle + |111\rangle)$, and upon receiving input “1” (“2”) they perform a local Pauli measurement σ_x (σ_y).

It is not difficult to see that their measurement outcomes will always satisfy the following relations:

$$\begin{aligned} &A_1 B_1 C_1 = +1, A_1 B_2 C_2 = -1, \\ &A_2 B_1 C_2 = -1, A_2 B_2 C_1 = -1. \end{aligned} \tag{39}$$

We contrast these quantum predictions with those of a local model, where the answer of each party depends only on the question he receives and on some shared random data λ. Since the correlations in Eq. (39) are perfect (i.e., exactly $+1$ or -1), each answer must clearly be a deterministic function of the local question and of λ. For fixed λ, a local model thus amounts to assigning a definite value ± 1 to all of the variables A_i, B_i, and C_i. But then this is in direct contradiction with Eqs. (39). To see this, consider the product of all four left-hand-side terms. Since $A_i^2 = B_i^2 = C_i^2 = 1$, this product is necessary equal to 1, but the product of the right-hand side is -1. This argument demonstrates in a simple way the incompatibility between the predictions of quantum theory and those of any local model.

Note that the above GHZ paradox can be recast as the violation of Mermin's inequality, given in Eq. (38), i.e., the GHZ correlations (39) violate the inequality (38) up to its algebraic maximum 4. In the language of nonlocal games, it provides an example of a game for which there exists a quantum strategy that wins it with probability $p_{\mathrm{win}} = 1$ (see Sec. II.B.4). GHZ paradoxes of the above types are also known as “pseudotelepathy” games (Brassard, Broadbent, and Tapp, 2005) or “Kochen-Specker” games (Mermin, 1993; Renner and Wolf, 2004). Other multipartite GHZ-type paradoxes, as well as a more detailed discussion of the nonlocal correlations of GHZ states, can be found in Sec. VI.D.2. Notable examples of nonlocality proofs without inequalities of the GHZ type but in the bipartite case have been presented by Cabello (2001), Aravind (2002), and Aolita, Gallego, Acín *et al.* (2012).

Another interesting demonstration of quantum nonlocality without inequalities was given by Hardy (1993). Consider a bipartite Bell test, in which each observer chooses between two dichotomic measurements. Hardy considered a situation in which the joint probability distribution satisfies the following relations:

$$\begin{aligned} &p(+1, +1|A_1, B_1) = 0, \\ &p(+1, -1|A_2, B_1) = 0, \\ &p(-1, +1|A_1, B_2) = 0. \end{aligned} \tag{40}$$

For any distribution that is local, it then follows that

$$p_{\mathrm{Hardy}} \equiv p(+1, +1|A_2, B_2) = 0. \tag{41}$$

Hardy realized that this logical implication does not hold in quantum mechanics.

Consider an entangled state of two qubits of the form

$$|\psi\rangle = \alpha(|01\rangle + |10\rangle) + \beta|00\rangle, \tag{42}$$

where $2|\alpha|^2 + |\beta|^2 = 1$. Both parties perform the same measurements. The first measurement is in the computational basis, with result $+1$ for state $|0\rangle$, and -1 for state $|1\rangle$. For the second measurement, the result $+1$ corresponds to a projection on the qubit state $|\phi\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle$, while the result -1 is associated with the orthogonal projector. Setting $\alpha = \beta\tan\theta$, one obtains Hardy's paradox: all three equations (40) are satisfied, but we obtain $p_{\mathrm{Hardy}} = 2\beta\sin^2\theta > 0$ if $0 < |\beta| < 1$. An interesting aspect of this construction is that the paradox occurs for any entangled state of two qubits, with the notable exception of the maximally entangled state ($\beta = 0$). This represents one of the first hints that entanglement and nonlocality are not monotonically related (see Sec. III.A.7).

The strongest demonstration of Hardy's paradox gives $p_{\mathrm{Hardy}} = (5\sqrt{5} - 11)/2 \approx 9\%$ (Hardy, 1993), which results in the maximal possible value in quantum theory (Rabelo, Zhi, and Scarani, 2012). That is, Hardy's paradox cannot be strengthened by using higher-dimensional quantum entangled states. For interesting extensions of Hardy's paradox, see Fritz (2011), and references therein.

F. Quantifying nonlocality

So far, we have mainly discussed the problem of detecting nonlocal correlations, i.e., determining whether given correlations $\mathbf{p}$ belong to $\mathcal{L}$ or not. Another relevant question is how to quantify nonlocality.

A common choice for quantifying nonlocality is through the amount of violation of a Bell inequality, i.e., $\mathbf{p}$ is more nonlocal than $\mathbf{q}$ if $\mathbf{s}\cdot\mathbf{p} > \mathbf{s}\cdot\mathbf{q}$ for some Bell expression $\mathbf{s}$. The problem with this approach is that there can be another Bell expression $\mathbf{s}'$ such that $\mathbf{s}'\cdot\mathbf{q} > \mathbf{s}'\cdot\mathbf{p}$. Another problem is that a given Bell inequality can be written in many equivalent ways using the normalization conditions $\mathbf{1}\cdot\mathbf{p} = \sum_{abxy} p(ab|xy) = m^2$ (recall that m denotes the number of possible inputs x and y). For instance, let $\mathbf{s}$ be the CHSH expression (4), for which $S_l = 2$ and $S_q = 2\sqrt{2}$. Consider the Bell expression $\mathbf{s}_\alpha = \alpha\mathbf{s} + [(1-\alpha)/2]\mathbf{1}$ obtained from the CHSH expression through irrelevant rescaling and addition of an offset. For any $\mathbf{p}$, we thus have $\mathbf{s}_\alpha\cdot p = \alpha\mathbf{s}\cdot\mathbf{p} - 2\alpha + 2$, which implies that the local bound $S_l^\alpha = 2$ of the new Bell expression is identical to the one of the original CHSH expression, but now its maximal quantum violation $S_q^\alpha = 2 + 2(\sqrt{2}-1)\alpha$ can (artificially) be made arbitrarily large by increasing α.

If the amount of violation of Bell inequalities is used to quantify nonlocality, this amount of violation must thus first be normalized in some proper way. If this normalization is well chosen, one can then often relate the amount of violation of Bell inequalities to an operational measure of nonlocality.

Possible operational measures of nonlocality are simply given by the tolerance of nonlocal correlations to the addition of noise, such as white noise (Kaszlikowski *et al.*, 2000; Acín *et al.*, 2002), local noise (Pérez-Garcia *et al.*, 2008; Junge *et al.*, 2010), or detection inefficiencies (Massar, 2002; Massar *et al.*, 2002). In particular, it was shown by Pérez-Garcia *et al.* (2008) (see also Sec. III.B.2) that the tolerance of $\mathbf{p}$ to any local noise, defined as the minimal value of r such that $r\mathbf{p} + (1-r)\mathbf{q}\notin\mathcal{L}$ for all $\mathbf{q}\in\mathcal{L}$, is given by $r = 2/(\nu+1)$, where ν is the maximal possible violation by $\mathbf{p}$ of a Bell inequality, defined in the following way:

$$\nu = \max_{\mathbf{s}} \frac{|\mathbf{s}\cdot\mathbf{p}|}{\max_{\mathbf{q}\in\mathcal{L}}|\mathbf{s}\cdot\mathbf{q}|}. \tag{43}$$

Note that taking the ratio and the absolute value is crucial for a meaningful definition of this amount of violation. If instead of the ratio one takes for instance the difference, a change of scale $\mathbf{s}\to\gamma\mathbf{s}$ would lead to arbitrary violations. If one removes instead the absolute value, the same happens via an offset, as in the example discussed above.

Another operational measure of the nonlocality of correlations $\mathbf{p}$ is given by the amount of classical communication between the two wings of the Bell experiment by which a local model has to be supplemented for reproducing these correlations. This approach was adopted by Maudlin (1992), Brassard, Cleve, and Tapp (1999), Steiner (2000), Bacon and Toner (2003), and Toner and Bacon (2003) (see also the discussion in Sec. III.C). Pironio (2003)showed that any Bell expression $\mathbf{s}$ can be rewritten in a normalized form $\mathbf{s}*$, through an appropriate change of scale and an offset, such that the minimal average amount of classical communication $C(\mathbf{p})$ necessary to reproduce $\mathbf{p}$ is given by $C(\mathbf{p}) = \max_{\mathbf{s}*}\mathbf{s}*\cdot\mathbf{p}$. Techniques for estimating the communication complexity of arbitrary no-signaling correlations and their relation to Bell violations were further developed by Degorre *et al.* (2011).

TABLE I. Kullback-Leibler (KL) divergence for the most common quantum Bell tests. ME stands for maximally entangled.

Bell inequality	Quantum state	KL divergence (bits)
CHSH	ME 2-qubit	0.046
CGLMP	ME 2-qutrit	0.058
CGLMP	Optimal 2-qutrit	0.077
Mermin GHZ	GHZ 3-qubit	0.208

Finally, a third proposed approach to measure nonlocality is through its "statistical strength" (van Dam, Grunwald, and Gill, 2005): that is, the amount of confidence that the measurement outcomes of n independent Bell experiments governed by a behavior $\mathbf{p}$ could not have been reproduced by a local behavior. Indeed, statistical fluctuations on a finite sample allow for the possibility of apparent Bell inequality violations even by a local model (this issue for the interpretation of experimental results of a Bell test is specifically discussed in Sec. VII.B.3). In an experiment, the goal is to test in a finite number of trials whether the system obeys a Bell local model (hypothesis LOC) or whether it is governed by some quantum model that is nonlocal (hypothesis QM). The statistical tool that quantifies the asymptotic average amount of support in favor of QM against LOC per independent trial is the Kullback-Leibler (KL) (or relative entropy) divergence (van Dam, Grunwald, and Gill, 2005). This quantity can be seen as a distance $D(\mathbf{p})$ between a given behavior $\mathbf{p}$ and the set of local behaviors.

The statistical strength of the most common nonlocality tests have been estimated by Acín, Gill, and Gisin (2005) and van Dam, Grunwald, and Gill (2005) and are summarized here in Table I. It is worth noting that the CHSH scenario is the strongest test among bipartite Bell tests involving qubits (van Dam, Grunwald, and Gill, 2005). Considering higher-dimensional systems, optimal tests (Acín, Gill, and Gisin, 2005) involve partially entangled states (rather than maximally entangled ones), illustrating the astonishing relation between entanglement and nonlocality (see Sec. III.A.7). Finally, the Mermin-GHZ test (see Sec. II.D), involving three qubits, appears to be much stronger than the considered bipartite Bell tests (van Dam, Grunwald, and Gill, 2005).

G. Multiple rounds and parallel repetition

So far, we have characterized the predictions $\mathbf{p} = \{p(a,b|x,y)\}$ of local, quantum, or no-signaling systems in *single-round* Bell experiments where a single choice of input pair (x, y) is made and the two devices produce a single output pair (a, b). More generally, we also consider *multiple-round* Bell experiments in which a sequence $(x_1, y_1), \ldots, (x_n, y_n)$ of input pairs is used in the two devices, resulting in a sequence $(a_1, b_1), \ldots, (a_n, b_n)$ of output pairs. A physical model for such an experiment will thus be characterized by the joint probabilities $\mathbf{p}_n = \{p(a_1 b_1 \cdots a_n b_n | x_1 y_1 \cdots x_n y_n)\}$. The motivation for considering such multiple-round Bell scenarios is clear: it corresponds to the situation of real experimental tests of Bell inequalities in which the two quantum devices are probed many times to gather sufficient measurement statistics.

Three general multiple-round scenarios can be distinguished (Barrett *et al.*, 2002). First, the n output pairs can be obtained by measuring n independent[6] systems. Effectively, this means that the measurement settings are applied sequentially, i.e., the next input pair is introduced in the two devices after outcomes have been produced for the previous round, and furthermore the devices have no memory of the previous round. In this scenario, we say that $\mathbf{p}_n$ is local, which we denote $\mathbf{p}_n \in \mathcal{L}_n^I$, if

$$p(a_1 b_1 \cdots a_n b_n | x_1 y_1 \cdots x_n y_n) = \sum_\lambda q_\lambda p_1(a_1 b_1 | x_1 y_1, \lambda) \times \cdots \times p_n(a_n b_n | x_n y_n, \lambda)$$

(similar definitions apply to the quantum and no-signaling cases).

In the second scenario, the measurement settings are also applied sequentially, but the devices' behavior in a given round can depend on the previous measurement settings and outputs, i.e., the devices have a memory of the previous rounds.[7] In this case, we say that $\mathbf{p}_n$ is local, which we denote $\mathbf{p}_n \in \mathcal{L}_n^M$, if we write

$$p(a_1 b_1 \cdots a_n b_n | x_1 y_1 \cdots x_n y_n) = \sum_\lambda q_\lambda p_1(a_1 b_1 | x_1 y_1, \lambda) p_2(a_2 b_2 | x_2 y_2, w_1, \lambda) \times \cdots \times p_n(a_n b_n | x_n y_n, w_{n-1}, \lambda),$$

where $w_i = (a_1 b_1 \cdots a_i b_i, x_1 y_1 \cdots x_n y_n)$denotes all inputs and outputs up to round i. This situation is the most general one that characterizes usual experimental tests of Bell inequality.

Finally, we also consider a third scenario in which Alice and Bob apply all their n inputs at the same time, and then, at a later time, the device produces all n outputs. We then say that $\mathbf{p}_n$ is local, which we denote $\mathbf{p}_n \in \mathcal{L}_n^S$, if

$$p(a_1 b_1 \cdots a_n b_n | x_1 y_1 \cdots x_n y_n) = \sum_\lambda q_\lambda p(a_1 b_1 \cdots a_n b_n | x_1 y_1 \cdots x_n y_n, \lambda).$$

In this case, the devices can exhibit a collective behavior where the outputs a_i of Alice's device at round i depend on the values of inputs and outputs of her device at any other round, and similarly for Bob's device. This multiple-round model is formally equivalent to a single-round model with "big" inputs $x = x_1 \cdots x_n$ and $y = y_1 \cdots y_n$ and outputs $a = a_1 \cdots a_n$ and $b = b_1 \cdots b_n$.

The memory model $\mathcal{L}_n^M$ and the simultaneous model $\mathcal{L}_n^S$ are strictly more powerful than the independent model $\mathcal{L}_n^I$. It is easy to see though that local strategies exploiting such memory or collective effects cannot reproduce nonlocal correlations (Barrett *et al.*, 2002), which necessarily require some genuine nonlocal resource, such as an entangled quantum state.

Another potential problem though is that in experimental tests the correlations $\mathbf{p}_n$, which characterize the probabilities with a different set of events, are not directly observable. Instead one observes a finite number of events, representing only one particular realization of the set of possibilities encoded in $\mathbf{p}_n$. If the local models $\mathcal{L}_n^I$, $\mathcal{L}_n^M$, or $\mathcal{L}_n^S$ cannot reproduce nonlocal correlations on average, it could nevertheless be possible that clever choices of such multiple-round strategies, in particular, exploiting memory or collective effects, could increase the chance of a statistical fluctuation resulting in an apparent violation of a Bell inequality. In the cases of the independent and memory models, which are the most relevant to experimental tests and applications of quantum nonlocality, such statistical fluctuations are harmless and can easily be controlled (Barrett *et al.*, 2002; Gill, 2003). (See Sec. VII.B.3 for a more detailed discussion.) This is due to the fact that at any given round i, independent and memory local models are constrained to satisfy the Bell inequalities, even when conditioning on events up to round $i - 1$. That is, if $\mathbf{p}_{i|w_{i-1}}$ denote the correlations at round i conditioned on the past variables w_{i-1}, we necessarily have $\mathbf{s} \cdot \mathbf{p}_{i|w_{i-1}} \leq S_l$ for every i, w_{i-1}, and Bell expression $\mathbf{s}$.

This last property can be nicely rephrased in the language of nonlocal games. It implies that to win n instances of a game, there is no better strategy than using each time the strategy that is optimal for a single round. This is not the case in the simultaneous model, where all inputs are given at the same time and all output produced at the same time. In this case, which is known as a *parallel repetition* of the game in computer science, there may exist collective strategies to win n instances of the games that are better than using each time the optimal single-round strategy. It is, in fact, known that, for example, the CHSH game can be played better locally over many rounds [see Barrett *et al.* (2002) for an explicit example in the case $n = 2$]; that is, when playing the CHSH game many times in parallel the gap between the local and quantum bounds *shrinks*.

The question of whether there exists a better strategy for parallel repetition of the game is particularly interesting from the perspective of computer science (Cleve *et al.*, 2007). However, it also tells us something about the

[6] Note that we allow some correlations between the different systems through some global shared randomness λ; see the definition of $\mathcal{L}_n^I$ later in the text. The n systems are only independent with respect to sharing the inputs and outputs.

[7] Formally, we consider here two-sided memory models, where each device has a memory of every previous input and output, including those of the other devices. One can also consider one-sided memory models, where each device has only a memory of the inputs and outputs relative to his side of the Bell experiment but not the other one (Barrett *et al.*, 2002).

strength of correlations between physical systems when Alice and Bob hold many particles to be measured simultaneously.

Note that if there exists a strategy that lets the players win with probability $p_{\rm win}$ in a single round, then they can win with probability $p_{\rm win}^n$ when playing the game n times. The question is then whether there exists a strategy that beats this value. We speak of (strong) parallel repetition if there exists a nontrivial q such that the winning probability when playing the game n times is always upper bounded by q^n. The term perfect parallel repetition refers to the case where $q = p_{\rm win}$. It is known that for classical strategies, i.e., local models, parallel repetition holds (Raz, 1998). More precisely, if $p_{\rm win} = 1 - \epsilon$, then for all games $p_{\rm win}^n = (1-\epsilon^c)^{\Omega(n/s)}$ for some $c \geq 2$, where s is the length of the answers (Raz, 1998; Holenstein, 2007). A strong parallel repetition theorem has $c = 1$. It is furthermore known that for unique games, $p_{\rm win}^n = (1-\epsilon^2)^{\Omega(n)}$ (Rao, 2008). However, for the so-called odd cyle game we require $c \geq 2$, and thus strong parallel repetition does not always hold (Raz, 2008).

For no-signaling strategies, it is known (Holenstein, 2007) that parallel repetition also holds. As quantum and classical theory obey the no-signaling principle this also gives a bound for quantum and classical correlations. Yet, since for many games (e.g., unique games) we have $p_{\rm win} = 1$ in the no-signaling case, this bound is not always insightful. For quantum correlations, it is known that for XOR games (two-outcome correlation Bell inequalities), perfect parallel repetition holds (Cleve *et al.*, 2007). Again, this also gives a bound for classical correlations, but already for the CHSH game it is not known how tight this bound actually is. Parallel repetition in the quantum setting also holds for unique games (Kempe, Regev, and Toner, 2010). A more general result is known for quantum correlations (Kempe and Vidick, 2011); however, it requires the game to be modified slightly to include "check" rounds. A similar construction can be made for local correlations (Feige and Kilian, 2000).

III. NONLOCALITY AND QUANTUM THEORY

In this section, we analyze the quantum resources (in terms of entanglement or Hilbert space dimension) that are necessary to produce nonlocal correlations by performing local measurements on quantum states.[8] Here we focus on the case of bipartite states, whereas the nonlocal correlations of multipartite quantum states will be discussed in Sec. VI.

A. Nonlocality versus entanglement

In order to obtain nonlocal correlations from measurements on a quantum state, it is necessary that the latter is entangled. That is, the state cannot be written in the separable form

$$\rho_{AB} = \sum_\lambda p_\lambda \rho_A^\lambda \otimes \rho_B^\lambda. \tag{44}$$

Indeed, if a state is of the above form, the correlations obtained by performing local measurements on it are given by

$$\begin{aligned} p(ab|xy) &= \mathrm{tr}\left[\sum_\lambda p_\lambda(\rho_A^\lambda \otimes \rho_B^\lambda) M_{a|x} \otimes M_{b|y}\right] \\ &= \sum_\lambda p_\lambda \mathrm{tr}(\rho_A^\lambda M_{a|x}) \mathrm{tr}(\rho_B^\lambda M_{b|y}) \\ &= \sum_\lambda p_\lambda p(a|x,\lambda) p(b|y,\lambda), \end{aligned} \tag{45}$$

which is of the local form (11). Hence the observation of nonlocal correlations implies the presence of entanglement.

It is interesting to investigate whether this link can be reversed. That is, do all entangled states lead to nonlocality? In the case of pure states, the answer is positive. Specifically, for any entangled pure state, it is possible to find local measurements such that the resulting correlations violate a Bell inequality,[9] in particular, the CHSH inequality. This was shown for the case of two-qubit states by Capasso, Fortunato, and Selleri (1973) and for bipartite states of arbitrary Hilbert space dimension by Gisin (1991) and Home and Selleri (1991).[10] Therefore, all pure entangled states are nonlocal. The only pure states that do not violate Bell inequalities are the product states $|\Psi\rangle = |\psi\rangle_A \otimes |\phi\rangle_B$.

For mixed states, it turns out that the relation between entanglement and nonlocality is much more subtle, and in fact not yet fully understood. First, Werner (1989) discovered a class of mixed entangled states which admits a local model [i.e., of the form (11)] for any possible local measurements. Hence the resulting correlations cannot violate any Bell inequality. While Werner considered only projective measurements, his results were later extended to the case of general measurements (POVMs) by Barrett (2002).

The situation is complicated by the fact that directly performing measurements on a mixed state ρ is not always the best way to reveal its nonlocality. For instance, it may be necessary to perform joint measurements on several copies of the state, that is, considering the state $\rho \otimes \rho \otimes \cdots \otimes \rho$ (Palazuelos, 2012a). Alternatively one may need to apply a judicious preprocessing to ρ, for instance, a filtering, before performing the measurements (Popescu, 1995). Therefore, there exist different possible scenarios for revealing the nonlocality of mixed entangled states, some examples of which are represented in Fig. 5 and are discussed in more detail below. Importantly a state may lead to nonlocal correlations in a given scenario but not in others. It is also worth mentioning that when many copies of a state can be jointly preprocessed before the measurements, the problem becomes closely related to entanglement distillation. Indeed, any state from which pure bipartite entanglement can be distilled will lead to nonlocality. For undistillable (or bound)

[8] Another resource that can be considered is the time required to achieve a certain Bell inequality violation, given the range of energy available during the measurements (Doherty and Wehner, 2011).

[9] Note that this result also holds for all multipartite pure entangled states (Popescu and Rohrlich, 1992), as discussed in more detail in Sec. VI.

[10] This result was also stated, without giving an explicit construction, by Werner (1989).

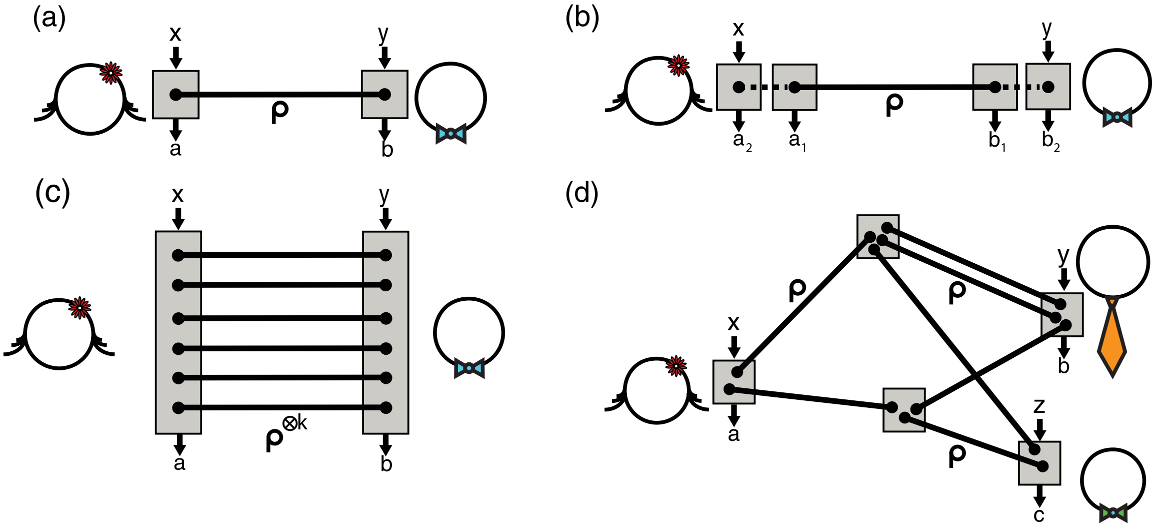

FIG. 5 (color online). The nonlocality of a quantum state ρ can be revealed in different scenarios. (a) The simplest scenario: Alice and Bob directly perform local measurements on a single copy of ρ. (b) The hidden nonlocality scenario: Alice and Bob first apply a filtering to the state; upon successful operation of the filter, they perform the local measurements for the Bell test. (c) Many-copy scenario: Alice and Bob measure collectively many copies of the state ρ. (d) Network scenario: several copies of ρ are distributed in a quantum network, where each observer performs local measurements.

bipartite entangled states, it is not yet known whether Bell inequality violations can be obtained, or whether these states admit a local model, as conjectured by Peres (1999). Nevertheless, recent results suggest that nonlocality might in fact be generic for all entangled states (Masanes, Liang, and Doherty, 2008).

1. Single-copy nonlocality

The simplest possibility to reveal nonlocality of an entangled state ρ is to find suitable local measurements such that the resulting correlations violate a Bell inequality. In the case of pure states this is always sufficient to reveal nonlocality. In particular, as mentioned above, all pure entangled states violate the CHSH inequality (Gisin, 1991; Home and Selleri, 1991). For mixed states, a necessary and sufficient condition for any two-qubit state to violate the CHSH inequality was given by Horodecki, Horodecki, and Horodecki (1995). This criterion works as follows. Associate with any two-qubit state ρ a correlation matrix T_ρ with entries $t_{ij} = \mathrm{tr}[\rho(\sigma_i \otimes \sigma_j)]$ for $i, j = 1, 2, 3$, where σ_i are the Pauli matrices. The maximum CHSH value S for ρ (considering the most general measurements) is then simply given by

$$S_\rho = 2\sqrt{m_{11}^2 + m_{22}^2}, \tag{46}$$

where m_{11}^2 and m_{22}^2 are the two largest eigenvalues of the matrix $T_\rho T_\rho^T$ (T_ρ^T denotes the transpose of T_ρ). Using the above criterion, it is possible to relate the entanglement of ρ, as measured by its concurrence, to its maximal violation of CHSH (Verstraete and Wolf, 2002).

From the above criterion it is straightforward to see that not every entangled two-qubit mixed state violates the CHSH inequality. However, contrary to the case of pure states, it is here not enough to focus on the CHSH inequality. In particular, there exist two-qubit states which do not violate CHSH, but violate a more sophisticated Bell inequality [I_{3322}, see Eq. (21)] involving three measurement settings per party (Collins and Gisin, 2004). Another example is the two-qubit Werner state, given by a mixture of a maximally entangled state $|\phi_+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ and the maximally mixed state, i.e.,

$$\rho_W = p|\phi_+\rangle\langle\phi_+| + (1-p)\frac{\mathbb{1}}{4}. \tag{47}$$

This state is separable for $p \leq 1/3$ (and thus does not violate any Bell inequality) and entangled otherwise. Using the criterion (46) one finds that $S = p2\sqrt{2}$, which leads to a violation of CHSH for $p > 1/\sqrt{2} \approx 0.707$. However, it was shown by Vértesi (2008) that the state (47) violates a Bell inequality involving 465 settings per party for $p \gtrsim 0.7056$.

If explicit Bell inequality violations yield upper bounds on the critical value of p necessary to reveal the nonlocality of the state (47), it is also possible to obtain lower bounds by constructing explicit local models. Werner (1989) showed that the correlations resulting from projective measurements on the state (47) admit a local model if $p \leq 1/2$, even though the state is entangled for $p > 1/3$. Entangled states admitting a local model are usually termed *local states*. Here we describe Werner's model, following the presentation of Popescu (1994). Note first that it is sufficient to construct a local model for $p = 1/2$, since then the model can be extended for any $p < 1/2$ by mixing with completely uncorrelated and random data. Let Alice and Bob measure the spin polarization of their particles in the $\hat{n}_A$ and $\hat{n}_B$ directions, respectively, where vectors describe the measurements on the Bloch sphere. The probability that both Alice and Bob get the outcome “0” is given by

$$p(00|\hat{n}_A, \hat{n}_B) = \tfrac{1}{4}(1 - \tfrac{1}{2}\cos\alpha), \tag{48}$$

where α is the angle between $\hat{n}_A$ and $\hat{n}_B$. Now we give a local hidden variable model that gives the same statistics. Here the hidden variable, shared by Alice and Bob, is a vector on the Bloch sphere $\hat{\lambda} = (\sin\theta\cos\phi\hat{x} + \sin\theta\sin\phi\hat{y} + \cos\theta\hat{z})$. In each run of the experiment a different $\hat{\lambda}$ is sent, chosen according to the uniform distribution $dq(\hat{\lambda}) = \sin\theta d\theta d\phi/4\pi$. After receiving $\hat{\lambda}$, Alice gives the outcome 0 with probability

$$p_A(0, \hat{n}_A, \hat{\lambda}) = \cos^2(\alpha_A/2), \tag{49}$$

where α_A is the angle between $\hat{n}_A$ and $\hat{\lambda}$. At the same time, Bob gives the outcome 0 with probability

$$p_B(0|\hat{n}_B, \hat{\lambda}) = \begin{cases} 1, & \text{if } 2\cos^2(\alpha_B/2) < 1, \\ 0, & \text{if } 2\cos^2(\alpha_B/2) > 1, \end{cases} \tag{50}$$

where α_B is the angle between $\hat{n}_B$ and $\hat{\lambda}$. Now one can check that the joint probability distribution obtained by Alice and Bob using this local model is given by

$$p_{LHV}(00|\hat{n}_A, \hat{n}_B) = \int dq(\hat{\lambda}) p_A(0, \hat{n}_A, \hat{\lambda}) p_B(0, \hat{n}_B, \hat{\lambda}), \tag{51}$$

which is indeed equal to Eq. (48). It is straightforward to check that the above model reproduces the desired correlations for all measurement outcomes.

Later on, it was proven that two-qubit Werner states are local for $p \lesssim 0.66$ by Acín, Gisin, and Toner (2006), using a connection to the Grothendieck constant (see Sec. III.B.2). Furthermore, Barrett (2002) extended the result of Werner to the most general (nonsequential) quantum measurements (so-called POVMs), where a local model is given for $p \leq 5/12$. For the interval $0.66 \lesssim p \lesssim 0.7056$ (or $5/12 < p \lesssim 0.7056$ if considering POVMs) it is not known whether the nonlocality of the state (47) can be revealed by performing measurements on a single copy of the state at a time.

Werner and Barrett also derived a local model for a family of states generalizing the two-qubit state (47) to arbitrary Hilbert space dimension d. These are called Werner states, given by

$$\rho_W = p\frac{2P_{\text{anti}}}{d(d-1)} + (1-p)\frac{\mathbb{1}}{d^2}, \tag{52}$$

where $\mathbb{1}$ is a $d \times d$ identity matrix, and P_{anti} denotes the projector on the antisymmetric subspace. These states have a particular symmetry, being invariant under unitary operations of the form $U \otimes U$. The values of α for which ρ_W is entangled and admits a local model (for projective or general measurements) are summarized in Table II.

The local models discussed above were further extended by Almeida *et al.* (2007) and Wiseman, Jones, and Doherty (2007) to another family of states generalizing the two-qubit state (47), namely, the isotropic states

$$\rho_{\text{iso}} = p|\Phi_+\rangle\langle\Phi_+| + (1-p)\frac{\mathbb{1}}{d^2}, \tag{53}$$

where $|\Phi_+\rangle = (1/\sqrt{d})\sum_{i=0}^{d-1}|ii\rangle$ is a maximally entangled state of local dimension d. Again, for such states there exist a range of the parameter p for which ρ_{iso} is entangled but admits a local model (see Table II). Note also that ρ_{iso} violates the CGLMP inequality [see Eq. (22)] when p is above a critical value p_{NL} that decreases with the local dimension d. In particular, $p_{NL} \to 0.67$ when $d \to \infty$ (Collins, Gisin, Linden *et al.*, 2002) (see Fig. 6).

TABLE II. Separability and locality bounds for Werner states (52) and for isotropic states (53). For Werner states, bounds for projective measurements were derived by Werner (1989) and by Barrett (2002) for POVMs. For isotropic states, bounds were derived by Almeida *et al.* (2007) and by Wiseman, Jones, and Doherty (2007) for projective measurements.

	Werner state (52)	Isotropic state (53)
Separable	$p \leq \frac{1}{d-1}$	$p \leq \frac{1}{d+1}$
Local for general measurements	$p \leq \frac{(d-1)^{(d-1)}(3d-1)}{(d+1)d^d}$	$p \leq \frac{(d-1)^{(d-1)}(3d-1)}{(d+1)d^d}$
Local for projective measurements	$p \leq \frac{d-1}{d}$	$p \leq \frac{(-1+\sum_{k=1}^{d} 1/k)}{d-1}$

More generally, the approach of Almeida *et al.* (2007) allows one to construct a local model for general states, of the form $\rho = p|\psi\rangle\langle\psi| + (1-p)\mathbb{1}/d^2$, where $|\psi\rangle$ is an arbitrary entangled pure state in $\mathbb{C}^d \otimes \mathbb{C}^d$. It is found that for $p \leq \Theta[\log(d)/d^2]$ the state ρ admits a local model for projective measurements. Interestingly there is a $\log(d)$ gap in the asymptotic limit between the above bound and the separability limit, which is given by $p \leq \Theta(1/d^2)$. An upper bound on p follows from the result of Acín *et al.* (2002), where it is shown that a state of the form

$$\rho = p|\phi_+\rangle\langle\phi_+| + (1-p)\frac{\mathbb{1}}{d^2}$$

(where $|\phi_+\rangle$ denotes a two-qubit maximally entangled state) violates the CHSH inequality for

$$p \geq \Theta\left(\frac{4}{(\sqrt{2}-1)d}\right),$$

which tends to zero when $d \to \infty$. This shows that there exist entangled states embedded in $\mathbb{C}^d \otimes \mathbb{C}^d$ which are highly robust against white noise.

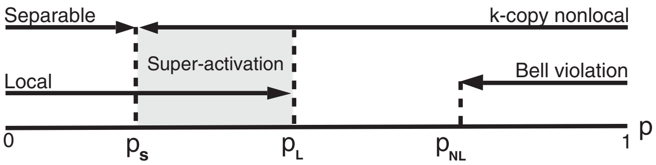

FIG. 6. Nonlocal properties of the isotropic state (53). The state is separable for $p \leq p_s = 1/(d+1)$, admits a local model for $p \leq p_L$ (Almeida *et al.*, 2007), and violates a Bell inequality for $p_L < p_{NL} < p$ (Collins, Gisin, Linden *et al.*, 2002). In the interval $p_L < p < p_{NL}$, it is not known whether the state admits a local model or, on the contrary, violates a Bell inequality. Finally, when several copies of the isotropic state can be measured jointly, nonlocality is obtained whenever a single copy of the state is entangled, that is, if $p > p_s$ (D. Cavalcanti *et al.*, 2013). In the gray region, superactivation of quantum nonlocality occurs (Palazuelos, 2012a).

Finally, it is worth pointing out the connection (Werner, 1989) between the fact that a quantum state admits a local model and the existence of a symmetric extension (Doherty, Parrilo, and Spedalieri, 2002) for this state. A bipartite state ρ_{AB} has a k-symmetric extension (with respect to part B) if there exists a quantum state of $k+1$ parties, $\rho'_{AB_1\cdots B_k}$, such that $\rho'_{AB_i}=\rho_{AB}$ for every $i=1,\ldots,k$, where ρ'_{AB_i} denotes the reduced state of subsystems A and B_i. Terhal, Doherty, and Schwab (2003) showed that if Alice and Bob share a state ρ_{AB} that has a k-symmetric extension, every experiment where Bob uses at most k measurement settings (independently of the number of outputs) can be simulated by a local model. Note that there is no restriction on the number of measurement settings for Alice. This result can be understood as follows: consider a Bell scenario where Alice chooses among m measurements, represented by operators $M_{a|x}$ with $x=1,\ldots,m$ and Bob among k measurements, given by $M_{b|y}$ with $y=1,\ldots,k$. Since ρ_{AB} has a k-symmetric extension, for each measurement x of Alice, the joint probability distribution $p(a,b_1\cdots b_k|x,y_1=1\cdots y_k=k)$ is well defined via the Born rule[11]

$$\begin{aligned}&p(a,b_1\cdots b_k|x,y_1=1\cdots y_k=k)\\&\quad=\mathrm{tr}[\rho'_{AB_1\cdots B_k}(M_{a|x}\otimes M_{b_1|1}\otimes\cdots\otimes M_{b_k|k})].\end{aligned}\tag{54}$$

From these distributions one can then define a joint probability distribution for all possible measurements as

$$\begin{aligned}&p(a_1\cdots a_m,b_1\cdots b_k|x=1\cdots x=m,y_1=1\cdots y_k=k)\\&=\frac{\prod_{i=1}^m p(a_ib_1\cdots b_k|x_iy_1=1\cdots y_k=k)}{p(b_1\cdots b_k|y_1=1\cdots y_k=k)^{m-1}}.\end{aligned}\tag{55}$$

This joint probability distribution provides the local model.[12] Note that if a state has a ∞-symmetric extension it is separable (Doherty, Parrilo, and Spedalieri, 2004).

2. Hidden nonlocality

Popescu (1995) proposed a more general way of obtaining nonlocal correlations from an entangled quantum state. Instead of performing a single measurement (in each run of the test), each observer now performs a sequence of measurements. For instance, the observers may first perform a local filtering to their systems before performing a standard Bell test, as in Fig. 5(b). That is, they each apply some physical operation (e.g., a measurement) to their system and proceed with the standard Bell test only if that physical operation yields a desired outcome. If one (or both) local operations do not yield the desired outcome, the parties discard this run of the test. Popescu demonstrated the power of this sequential scenario by showing explicitly that certain entangled states admitting a local model can display nonlocality if a judicious local filtering is performed. Hence, the filtering reveals the "hidden nonlocality" of the state. In particular, Popescu showed that this occurs for Werner states [see Eq. (52)] of local dimension $d\geq 5$.

One can intuitively understand hidden nonlocality in the following way. Alice and Bob share a mixed entangled state ρ. Importantly, even if ρ is local, it may be viewed as a statistical mixture involving one (or more) nonlocal states. In order to extract nonlocality from ρ, Alice and Bob first apply a local measurement for which a given outcome can occur only (or most likely) for a nonlocal state in the mixture. Hence, by postselecting those events in which this particular measurement outcome occurs, Alice and Bob can filter out the nonlocal state. Finally, by performing appropriate local measurements, they can violate a Bell inequality.

In order to exclude the existence of a local model reproducing this sequential measurement scenario, it is essential that Alice and Bob choose the measurement basis of the final measurement after a successful operation of the filter. If this is not the case, a local strategy could fake Bell inequality violation by adapting the outcome of the first measurement based on the knowledge of which basis has been chosen for the second measurement. A formal account of this argument can be found in Teufel *et al.* (1997) and Zukowski *et al.* (1998). A general framework for Bell tests with sequential measurements was discussed by Teufel *et al.* (1997) and Gallego *et al.* (2013).

A question left open in the work of Popescu (1995) is whether hidden nonlocality can also be demonstrated for an entangled state admitting a local model for the most general nonsequential measurements. Note that Popescu (1995) considered Werner states, which admit a local model for projective measurements, but are not known to be local when POVMs are considered. This question was answered recently by Hirsch *et al.* (2013), where it is shown that there exist entangled states featuring genuine hidden nonlocality, that is, states which admit a local model for nonsequential POVMs, but violate a Bell inequality using judicious filtering.

Other examples of hidden nonlocality were reported. Gisin (1996) showed that there exist two-qubit states which do not violate the CHSH inequality, but do violate CHSH after a judicious local filtering is applied. Peres (1996) demonstrated that five copies of a two-qubit Werner state (47) admitting a local model for projective measurements display hidden nonlocality. It is worth noting that these works on hidden nonlocality eventually lead to the concept of distillation of entanglement, a central notion in quantum information theory.

Finally, an important question in this area is whether all entangled states feature nonlocality when local filtering is considered. Although this question is yet to be answered, important progress was recently achieved. Masanes, Liang, and Doherty (2008) showed that for every entangled state ρ, there exists another state σ which does not violate the CHSH

[11] Note that the same argument holds if, instead of a k-symmetric extension, ρ_{AB} has a k-symmetric quasiextension, where, instead of a state of $k+1$ parties $\rho'_{AB_1\cdots B_k}$, one has an entanglement witness of $k+1$ parties $W_{AB_1\cdots B_k}$, with unit trace and such that the reduced states satisfy $W_{AB_i}=\rho_{AB}$ for all i.

[12] Indeed, it is easy to see that the existence of a joint distribution for all possible measurements that Alice and Bob can make is equivalent to the existence of a local model (Fine, 1982). Simply think of $\lambda=(a_1\cdots a_m,b_1\cdots,b_k)$ as the hidden variable instructing which outcome every party must output for any measurements that they perform and the joint probability $p(a_1\cdots a_m,b_1\cdots b_k|x=1\cdots x=m,y_1=1\cdots y_k=k)$ as the distribution $q(\lambda)$'s over hidden variables.

inequality, such that $\rho \otimes \sigma$ violate CHSH after local filtering (Liang, Masanes, and Rosset, 2012). In particular, if one chooses ρ such that it does not violate CHSH, a phenomenon of "activation" of CHSH nonlocality occurs.

3. Multicopy nonlocality

Another relevant scenario consists of allowing the parties to perform measurements on several copies of the state ρ in each run of the Bell test. However, here no initial filtering is allowed, contrary to the scenario of hidden nonlocality. In the multicopy scenario, represented in Fig. 5(c), Alice and Bob can perform measurements on k copies of the state ρ, that is, they measure effectively a state of the form $\rho^{\otimes k} = \rho \otimes \rho \otimes \cdots \otimes \rho$ (k times). The key point here is that the parties can now perform joint measurements on their k subsystems, that is, measurements featuring eigenstates which are entangled. Remarkably, the maximal violation of the CHSH inequality for certain states can be increased if several copies of the state are jointly measured (Liang and Doherty, 2006). In fact, there exist states ρ which do not violate the CHSH inequality, but $\rho^{\otimes 2}$ does (Navascués and Vértesi, 2011).

In more general terms, the possibility of performing measurements on several copies of a state leads to a phenomenon of *activation of nonlocality*. Notably, it was recently demonstrated that quantum nonlocality can be superactivated (Palazuelos, 2012a), that is, the combination of a number of local quantum states can become nonlocal. This demonstrates that nonlocality is not an additive quantity. Specifically, it was shown by Palazuelos (2012a) that by performing joint measurements on many copies of a local isotropic state ρ_{iso} [see Eq. (53)] of local dimension $d = 8$ it is possible to violate a Bell inequality, without involving any preprocessing. This is remarkable given that the initial state ρ_{iso} admits a local model for the most general measurements (i.e., including POVMs).

More recently, it was shown that for every state $\rho \in \mathbb{C}^d \otimes \mathbb{C}^d$, with singlet fidelity[13] larger than $1/d$, there exist a number of copies k of ρ such that $\rho^{\otimes k}$ is nonlocal (D. Cavalcanti *et al.*, 2013). This result implies that every entangled isotropic state (53) is a nonlocal resource and establishes a direct connection between the usefulness of a state in quantum teleportation and its nonlocality (see Sec. III.A.6). Whether superactivation of nonlocality is possible for any entangled state admitting a local model is an interesting open question.

4. More general scenarios

It is also relevant to investigate the case in which several copies of a bipartite entangled state ρ are distributed in a network of n observers, as shown in Fig. 5(d). It turns out that here a phenomenon of activation of nonlocality can also occur. That is, by judiciously placing several copies of a state ρ admitting a local model, nonlocal correlations among the n observers can be obtained. The state ρ is then termed a "nonlocal resource." Again, activation of nonlocality is possible here due to the fact that one (or more) observer can perform a joint measurement on several subsystems (see Sec. III.A.3).

Examples of activation of nonlocality in networks were reported. First, by concatenating many copies of a state which does not violate the CHSH inequality in an entanglement swapping scenario one obtains a state which violates CHSH (De *et al.*, 2005; Klobus *et al.*, 2012). Second, it was shown that many copies of a two-qubit Werner state (47) distributed in a star network violate a Bell inequality for $p \gtrsim 0.64$, hence for states which admit a local model for projective measurements (De *et al.*, 2005; Cavalcanti, Almeida *et al.*, 2011). The cases of isotropic states, as well as other examples of activation of nonlocality, were discussed by Cavalcanti, Rabelo, and Scarani (2012).

5. Entanglement distillation and nonlocality

As mentioned in Sec. III.A.2, the notion of hidden nonlocality is intimately related to entanglement distillation. For instance, in Peres (1996), the local filtering that is applied on several copies of a state can be used to distill entanglement. Hence the protocol of Peres (1996) can be decomposed as entanglement distillation followed by a standard (single-copy) Bell test. In this sense, every entangled state that is distillable can be used to obtain nonlocal correlations.

An interesting question then arises concerning bound entangled states, i.e., states from which no entanglement can be distilled (Horodecki, Horodecki, and Horodecki, 1998). In fact, a long-standing open conjecture—referred to as the Peres conjecture—is that every state with a positive partial transposition (PPT), hence undistillable, admits a local model (Peres, 1999). More generally, the goal is to understand the link between distillability and nonlocality. Notably, several works established a partial link between both concepts, showing that important classes of Bell inequalities cannot be violated by any PPT state (Werner and Wolf, 2000; Salles, Cavalcanti, and Acín, 2008). For instance, the violation of the CHSH (and more generally of all Mermin inequalities) certifies that the state can be distilled (Acín, 2001; Acín, Scarani, and Wolf, 2003; Masanes, 2006). More recently, a method for upper bounding the possible violation of a given Bell inequality for PPT states (in arbitrary Hilbert space dimension) was presented by Moroder *et al.* (2013), from which it can be shown that many bipartite Bell inequalities cannot be violated by PPT states. Finally, note that in the case of more parties, it is proven that nonlocality does not imply distillability of entanglement (Vértesi and Brunner, 2012), hence disproving the Peres conjecture in the multipartite case.

6. Nonlocality and teleportation

Quantum teleportation (Bennett *et al.*, 1993) is another "nonlocal phenomenon" based on quantum entanglement. As is the case with nonlocality, it turns out that not every entangled state is useful for teleportation, in the sense of outperforming classical strategies (Horodecki, Horodecki, and Horodecki, 1999). It is then natural to ask if the fact that a state is useful for teleportation is related to its nonlocality.

This question was first raised by Popescu (1994), who noticed that certain two-qubit entangled Werner states

[13] The singlet fidelity (or equivalently entanglement fraction) of a state ρ is defined as the maximal fidelity SF of ρ with a maximally entangled state (MES), i.e., $SF(\rho) = \max_{|\psi\rangle \in \text{MES}} \langle\psi|\rho|\psi\rangle$.

admitting a local model can nevertheless be useful for teleportation. This led to the conclusion that usefulness in teleportation and nonlocality are unrelated. Interestingly, this difference vanishes when considering more general scenarios for revealing nonlocality. In particular, it was recently shown that in the multicopy scenario, where several copies of the state can be jointly measured, any state that is useful for teleportation is a nonlocal resource (D. Cavalcanti *et al.*, 2013). Hence, this work establishes a direct link between teleportation and nonlocality.

Note also that a more qualitative relation between the amount of CHSH violation and usefulness for teleportation was derived by Horodecki, Horodecki, and Horodecki (1996). Specifically, the maximal violation S_ρ of the CHSH inequality of a two-qubit state ρ is shown to lower bound its average fidelity for teleportation as follows:

$$F_{\text{telep}} \geq \frac{1}{2}\left(1 + \frac{S_\rho^2}{12}\right). \tag{56}$$

Note that the optimal classical strategy achieves $F_{\text{telep}} = 2/3$ in the qubit case. For a device-independent version, see Ho, Bancal, and Scarani (2013).

7. More nonlocality with less entanglement

As discussed previously, the relation between entanglement and nonlocality is subtle. Another interesting question is to see whether a quantitative link can be established between both concepts. Astonishingly, it turns out that in certain cases, and depending on which measure of nonlocality is adopted, less entanglement can lead to more nonlocality.

An example is provided by certain Bell inequalities, whose maximal violation can be achieved only with partially entangled states (Acín *et al.*, 2002) (considering states of a given Hilbert space dimension). More importantly, there exist simple Bell inequalities, the maximal violation of which cannot be obtained from maximally entangled states of any dimension, but requires partially entangled states (Liang, Vértesi, and Brunner, 2011; Vidick and Wehner, 2011). Also, it is known that there exist Bell inequalities for which partially entangled states give violations which are arbitrarily larger compared to maximally entangled states (Junge and Palazuelos, 2011; Regev, 2012) (see Sec. III.B.2).

Interestingly it turns out that this phenomenon, sometimes referred to as an *anomaly of nonlocality* [see Méthot and Scarani (2007) for a short review], occurs for other measures of nonlocality as well. Notably, this effect was discovered by Eberhard (1993), who showed that weakly entangled two-qubit states are more resistant to the detection loophole compared to maximally entangled states (see Sec. VII.B.1.c). Moreover, the anomaly of nonlocality was also shown to occur when considering the statistical strength of Bell tests (Acín, Gill, and Gisin, 2005), and the simulation of quantum correlations with nonlocal resources (Brunner, Gisin, and Scarani, 2005).

B. Nonlocality versus Hilbert space dimension

In this section, we consider the link between nonlocality and another property of quantum systems: the dimension of the Hilbert space in which the quantum state and measurements are defined. Indeed, the Hilbert space dimension generally represents a resource, in the sense that higher-dimensional Hilbert spaces contain more complex quantum states.

Formally, we say that the correlations $p(ab|xy)$ have a d-dimensional representation if there exists a state ρ_{AB} in $\mathbb{C}^d \otimes \mathbb{C}^d$, and measurement operators $M_{a|x}$ and $M_{b|y}$ acting on $\mathbb{C}^d$, such that

$$p(ab|xy) = \text{tr}(\rho_{AB} M_{a|x} \otimes M_{b|y}). \tag{57}$$

In some cases, it is also admitted that $p(ab|xy)$ has a d-dimensional representation if $p(ab|xy)$ can be written as a convex combination of correlations of Eq. (57).

In the following, we discuss two natural questions. First, what is the minimal dimension d necessary to reproduce a given set of correlations $p(ab|xy)$? This question is closely related to the concept of "dimension witnesses." Second, how much nonlocality can correlations of Eq. (57) contain as a function of d?

1. Minimal Hilbert space dimension and dimension witnesses

Here the general question is to determine what quantum resources, in terms of Hilbert space dimension, are necessary to reproduce certain quantum correlations. For instance, if we consider a Bell scenario with a given number of inputs and outputs, what is the minimal dimension d such that all quantum correlations (i.e., all correlations attainable in quantum mechanics) can be reproduced? This is in general a very difficult problem. In the case of binary inputs and outputs, we know that qubits ($d = 2$) are sufficient, if convex combinations are taken into account (Cirel'son, 1980). However, beyond this simple case, very little is known. In fact, we do not even know if a finite d is sufficient for a scenario involving a finite number of measurements and outcomes. Actually, recent work suggests that this might not be the case (Pál and Vértesi, 2010), giving evidence that the maximal violation of the I_{3322} Bell inequality (see Sec. II.B.3) can be attained using only a quantum state of infinite dimension.

A related question is the following. Given some correlations originating from measurements on a quantum system, can we place a lower bound on the Hilbert space dimension of the state and measurements necessary to reproduce them? That is, can we show that certain correlations are impossible to obtain with arbitrary quantum states and measurements of a given dimension? The concept of a dimension witness allows one to address this question. Specifically, a dimension witness for quantum systems of dimension d is a linear function of the probabilities $p(ab|xy)$ described by a vector $\mathbf{w}$ of real coefficients w_{abxy}, such that

$$W \equiv \sum_{a,b,x,y} w_{abxy} p(ab|xy) \leq w_d \tag{58}$$

for all probabilities of Eq. (57) with ρ_{AB} in $\mathbb{C}^d \otimes \mathbb{C}^d$, and such that there exist quantum correlations for which $W > w_d$

(Brunner, Pironio *et al.*, 2008). When some correlations violate Eq. (58), they can thus be established only by measuring systems of local dimension strictly larger than d. The simplest examples of dimension witnesses involve Bell inequalities featuring measurements with ternary outcomes, the maximal violation of which cannot be reached with qubits, but requires qutrits (Brunner, Pironio *et al.*, 2008). Other examples are discussed in Sec. III.B.2.

It is also possible to devise entropic dimension witnesses (Wehner, Christandl, and Doherty, 2008), which were discussed in the context of information-theoretic tasks. Finally, the dimension of a single system can be witnessed in a prepare and measure scenario (Gallego *et al.*, 2010). Note however that, since this approach is not based on nonlocal correlations, it is not possible to separate quantum and classical behavior in general; indeed, any quantum behavior can be simulated classically by using systems of high enough dimension.

2. Grothendieck's constant and Bell inequalities with unbounded violation

As mentioned in Sec. II.F, there exist several possible measures of nonlocality. A natural option consists of quantifying the strength of a given nonlocal correlation $\mathbf{q}$ through the following quantity:

$$\nu(\mathbf{q}) \equiv \sup_{\mathbf{s}} \frac{|\langle \mathbf{s}, \mathbf{q}\rangle|}{\sup_{\mathbf{p}\in\mathcal{L}}|\langle \mathbf{s}, \mathbf{p}\rangle|}. \tag{59}$$

This represents the ratio between the maximal quantum value for a Bell expression $\mathbf{s}$ (i.e., $|\langle \mathbf{s}, \mathbf{q}\rangle|$) and its local bound (i.e., $\sup_{\mathbf{p}\in\mathcal{L}}|\langle \mathbf{s}, \mathbf{p}\rangle|$), maximized over all possible Bell expressions $\mathbf{s}$. Note that the absolute value is important here, otherwise the quantity could be ill defined. This quantity quantifies how much local noise (considering any possible local noise) must be added to $\mathbf{q}$ such that the global distribution becomes local (Pérez-Garcia *et al.*, 2008; Junge *et al.*, 2010). An interesting feature of this quantity is that it provides a unified measure of nonlocality, allowing one to compare the violations of different Bell inequalities.

Tsirelson (1987) pointed out a connection between Grothendieck's inequality, which arose in the study of tensor norms, and the quantum violation of certain Bell inequalities. Tsirelson showed that $\nu(\mathbf{q})$ is upper bounded by Grothendieck's constant K_G for any two-outcome correlation Bell inequality (i.e., XOR games). Although the exact value of the latter is not known, it is proven that

$$1.6769 \leq K_G \leq \frac{\pi}{2\log(1+\sqrt{2})} \approx 1.7822.$$

Importantly, this bound holds for quantum systems of arbitrary dimension.

Moreover, Tsirelson showed that, when restricting to qubits, one has that $\nu(Q) \leq K_3$, where K_3 is Grothendieck's constant of the order of 3. Since it is known that $K_3 < K_G$, it follows that there exist two-outcome correlation Bell inequalities which are dimension witnesses for qubits (Brunner, Pironio *et al.*, 2008). Explicit examples have been constructed by Vértesi and Pal (2008). Moreover, it was proven that dimension witnesses for any Hilbert space dimension d can be obtained from XOR games (Vértesi and Pál, 2009; Briët, Buhrman, and Toner, 2011).

Tsirelson also raised the question of whether it would be possible to have unbounded violations of Bell inequalities. That is, does there exist a family of Bell scenarios for which the quantity $\nu(Q)$ is unbounded?

The first result in this direction is due to Mermin (1990a), who considered a multipartite scenario. Specifically, he introduced a family of Bell inequalities for an arbitrary number of parties n (now referred to as the Mermin inequalities, see Sec. II.D), and showed that by performing measurements on an n-party GHZ state one obtains a violation of these inequalities that grows exponentially with n, while the local bound remains constant.

A natural question is then whether unbounded Bell violations can also occur in the case of a fixed number of parties. This is however a very hard problem, mainly because of the difficulty of finding Bell inequalities and to estimate their quantum violations. It was discovered recently that the abstract concepts of operator space theory and tensor norms provide a useful framework for the study of violations of Bell inequalities in quantum mechanics [see Junge *et al.* (2010) for an introduction]. This line of research was started by Pérez-Garcia *et al.* (2008), where the existence of tripartite correlation Bell inequalities with unbounded quantum violations was proven. Later they showed that similar results hold for (noncorrelation) bipartite Bell inequalities (Junge *et al.*, 2010). More formally, these studies focus on the quantity $\nu(Q)$ [see Eq. (59)], i.e., the maximal quantum violation of any Bell inequality $\mathbf{s}$, as a function of the number of measurement settings, outcomes, and Hilbert space dimension. Remarkably they showed that $\nu(Q)$ can be upper and lower bounded by ratios of different norms of the Bell expression $\mathbf{s}$ (here viewed as a functional), which have been studied in operator space theory.

While the works mentioned above give nonconstructive proofs of the existence of Bell inequalities with unbounded violations, explicit examples have also been found. The strongest result is due to Junge and Palazuelos (2011) who explicitly constructed (up to random choices of signs) a bipartite Bell inequality featuring a violation of the order of $\sqrt{k}/\log k$, where each party has k possible measurements with k outcomes, considering quantum systems of dimension $d = k$. Notably, this constructions appears to be close to optimal, as a separation between this violation and known upper bounds is only quadratic in k (Junge *et al.*, 2010; Junge and Palazuelos, 2011). Recently, an explicit and simplified presentation of these Bell inequalities was given by Regev (2012), based on standard quantum information techniques. Other explicit examples of Bell inequalities with unbounded violations have been presented (Buhrman *et al.*, 2011). Note that, while the construction of Buhrman *et al.* (2011) uses maximally entangled states, the works of Junge and Palazuelos (2011) and Regev (2012) considered entangled states with low entanglement. Finally, an upper bound on the maximum Bell violation (for any possible Bell inequality) of a given quantum state was derived by Palazuelos (2012b).

C. Simulation of quantum correlations

So far we examined which quantum resources are necessary to produce nonlocal correlations, in terms of entanglement or Hilbert space dimension of quantum states. Here we discuss the converse question. How can we use nonlocal resources to characterize and quantify the nonlocality of entangled quantum states? If a state violates a Bell inequality, we know that its measurement statistics cannot be reproduced by a local model. However, we can simulate its correlations if we have access to a nonlocal resource, such as classical communication or nonlocal resources such as the PR box. The minimal amount of nonlocal resources required can then be considered as a measure of the nonlocality of the state. Here we give a brief review of progress in this direction.

1. Simulating the singlet state

A classical simulation protocol of a given quantum state $|\psi\rangle$ aims at reproducing the correlations obtained from local measurements on $|\psi\rangle$, using only shared randomness and classical communication. For definiteness, we focus here on the singlet state of two qubits, i.e., $|\psi\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$, which is also the most studied case.

Alice and Bob first receive as input a unit vector on the Bloch sphere, i.e., $\hat{n}_A$, $\hat{n}_B \in \mathbb{R}^3$, representing projective measurements $\hat{n}_A \cdot \vec{\sigma}$ and $\hat{n}_B \cdot \vec{\sigma}$, where $\vec{\sigma}$ is the vector of Pauli matrices. Then they are allowed to exchange classical communication. Finally, they must produce binary outcomes a, $b = \pm 1$ reproducing the expected statistics, i.e., $p(ab|\hat{n}_A\hat{n}_B) = 1/4(1 - ab\hat{n}_A \cdot \hat{n}_B)$.

It is then interesting to look for the model using the least classical communication, since the smallest number of bits required to simulate $|\psi\rangle$ can be considered as a measure of the nonlocality of $|\psi\rangle$. This approach was proposed independently by Maudlin (1992), Brassard, Cleve, and Tapp (1999), and Steiner (2000). These first partial results were superseded by Toner and Bacon (2003), where it is shown that a single bit of communication is sufficient to exactly simulate the correlations of local projective measurements on a singlet state. Note that in this model Alice and Bob use infinite shared randomness, which is proven to be necessary for models with finite communication (Massar *et al.*, 2001).

It is also interesting to investigate simulation models using only nonsignaling resources, such as the PR box (see Sec. II.C.2). Remarkably, a single PR box is enough to simulate the singlet correlations (Cerf *et al.*, 2005). The latter model is even more economical than the model of Toner and Bacon (2003), since a PR box is a strictly weaker nonlocal resource; indeed, while it is possible to get a PR box from one bit of communication, the opposite is impossible since the PR box is nonsignaling.

Finally, it is also possible to devise a simulation model of the singlet state in which postselection is allowed (Gisin and Gisin, 1999); that is, the parties are not required to provide an output in all runs of the protocol. Indeed postselection should be considered as a nonlocal resource, giving rise to the detection loophole (see Sec. VII.B.1).

A unified presentation of all above models can be found in Degorre, Laplante, and Roland (2005).

2. Other quantum states

The simulation of quantum correlations of arbitrary bipartite entangled quantum states has also been investigated. Notably, Regev and Toner (2007) showed that the correlations obtained from local measurements with binary outputs on any $\rho_{AB} \in \mathbb{C}^d \otimes \mathbb{C}^d$ can be simulated with only two bits of communication, which is proven to be necessary in general (Vértesi and Bene, 2009). Note, however, that this model focuses on the correlations between the outcomes of Alice and Bob and does not, in general, reproduce the expected quantum marginals.

A case of particular interest is that of partially entangled qubit states, i.e., $|\psi_\theta\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle$. While it is shown that its correlations (including marginals) can be perfectly simulated with two bits of communication for any θ (Toner and Bacon, 2003), it is not known whether a single bit of communication would suffice. It is, however, proven that a simulation model using a single PR box does not exist (Brunner, Gisin, and Scarani, 2005) for weakly entangled states ($\theta \leq \pi/7.8$). Thus it appears that less entangled states require more nonlocal resources to be simulated compared to maximally entangled ones, illustrating the subtle relation between entanglement and nonlocality (see Sec. III.A.7). A simulation model for states $|\psi_\theta\rangle$ using only nonsignaling resources has also been presented (Brunner, Gisin *et al.*, 2008).

Moreover, Brassard, Cleve, and Tapp (1999) established the fact that the simulation of measurements with d outcomes on a maximally entangled state in $\mathbb{C}^d \otimes \mathbb{C}^d$ requires classical communication of order d bits. Therefore, there exist families of quantum nonlocal correlations requiring an arbitrarily large amount of classical communication for being simulated. Much less is known on the simulation of multipartite entangled states. Branciard and Gisin (2011) presented a simulation model for equatorial measurements on the three-qubit GHZ state which requires three bits of communication, or eight PR boxes. The simulation of the protocol of entanglement swapping, which combines entangled states and entangled measurements, was also discussed (Branciard, Brunner *et al.*, 2012).

More generally the problem of simulating quantum nonlocal correlations is intimately related to the field of communication complexity. Thus, many results on communication complexity are relevant in the context of nonlocality. For more details on communication complexity and on the procedure for converting communication complexity problems into nonlocal tasks, see Buhrman *et al.* (2010).

3. Elitzur-Popescu-Rohrlich decomposition

A different perspective on simulating quantum correlations was presented by Elitzur, Popescu, and Rohrlich (1992), often referred to as the EPR2 approach. They proposed decomposing a quantum probability distribution $p_q(ab|xy)$ into local and nonlocal parts. Formally, that means writing p_q as a convex combination of a local distribution (p_l) and a nonlocal one (p_{ns}):

$$p_q(ab|xy) = wp_l(ab|xy) + (1 - w)p_{\text{ns}}(ab|xy), \qquad (60)$$

with $0 \leq w \leq 1$. Note that, since p_q and p_l are no-signaling distributions, p_{ns} is also no signaling (hence the subscript "ns"). Clearly, any distribution can be written in this way (take, for instance, $w = 0$ and $p_q = p_{\rm ns}$). To find the EPR2 decomposition, one then finds the maximum of w among all possible decompositions of Eq. (60). This quantity, denoted $w_{\max}$, defines the local content of the distribution p_q. The EPR2 decomposition can be understood as a simulation of the distribution p_q where, with probability $w_{\max}$ a local distribution is used, and with probability $1 - w_{\max}$ a nonlocal (no-signaling) distribution is used. Note that $q_{\max}$ can also be considered as a measure of the nonlocality of the distribution p_q: if $w_{\max} = 1$, p_q is local; if $w_{\max} < 1$, p_q is nonlocal; if $w_{\max} = 0$, p_q is fully nonlocal.

One can bound $w_{\max}$, for a given distribution p_q, through the violation of a Bell inequality $\mathbf{s} \cdot \mathbf{p} \leq S_l$ (Barrett, Kent, and Pironio, 2006). Denote Q, the Bell value of distribution p_q. It is straightforward to see that

$$w_{\max} \leq \frac{S_{\rm ns} - Q}{S_{\rm ns} - S_l}, \tag{61}$$

where $S_{\rm ns}$ is the maximal value of the Bell expression s for any no-signaling distribution. Notice that if p_q reaches the maximal value allowed by no signaling (i.e., $Q = S_{\rm ns}$), then $w_{\max} = 0$. This means that the quantum distribution is maximally nonlocal according to the EPR2 decomposition; hence no local part can be extracted.

It is also possible to define the local content of a quantum state. To do this, consider all possible measurements that can be applied to a quantum state and then derive the local content for the distribution obtained from these measurements. Originally, Elitzur, Popescu, and Rohrlich (1992) showed that the maximally entangled state of two qubits has zero local content, i.e., it is fully nonlocal. This result was then generalized to any bipartite maximally entangled state (Barrett, Kent, and Pironio, 2006), via a generalization of the chained Bell inequality (see Sec. II.B.3), showing that such states can provide maximally nonlocal and monogamous correlations, which is relevant for instance in quantum cryptography.

The local content of other quantum states has also been discussed. In particular, for the case of two-bit entangled pure states $|\psi_\theta\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle$, with $\theta \in [0, \pi/4]$, it was proven that $q_{\max} = 1 - \cos(2\theta)$ (Portmann, Branciard, and Gisin, 2012). The EPR2 decomposition of pure entangled two-qutrit states was also sketched by Scarani (2008).

Finally, note that these ideas were generalized to the multipartite case by Almeida, Cavalcanti *et al.* (2010).

IV. APPLICATIONS OF QUANTUM NONLOCALITY

When considering nonlocality as a potential resource for information processing, two intuitive ideas immediately come to mind. First, since the existence of nonlocal correlations between the two wings of a Bell experiment seems to imply some connection between these two distant wings, one can hope to exploit this connection to communicate, and, in particular, to communicate faster than light. Second, since a local model for a Bell experiment is equivalent, as seen in Sec. II.B, to a deterministic model in which a definite outcome $a(x)$ and $b(y)$ is assigned in advance to every measurement x and to every measurement y, nonlocality then suggests, in contrast, that these measurement outcomes are fundamentally undetermined and thus that they could be used to establish cryptographic keys. Both ideas are partly true and partly misleading. In both cases, the no-signaling principle plays a fundamental role.

A. Communication complexity

In the first example discussed previously, no signaling acts as a limitation: we have already seen that the no-signaling conditions (7), which are satisfied by any set of correlations arising from measurement on quantum systems, imply that Bob's outcome does not reveal any information about Alice's input x and the other way around. Thus, no signaling prevents the use of nonlocal correlations as a substitute for direct communication between Alice and Bob. It may then come as a surprise that nonlocality can nevertheless be exploited to reduce the amount of communication in certain distributed computing tasks, in both information theory and the study of communication complexity. In the setting of communication complexity, Alice receives an n-bit string x and Bob receives an n-bit string y and the aim is for Bob to compute some function $f(x, y)$ with as little communication between Alice and Bob. This can always be achieved if Alice sends her n-bit string x to Bob, but for certain functions less communication is sufficient. The minimum number of bits that must be exchanged between Alice and Bob for Bob to determine $f(x, y)$ is known as the *communication complexity* of f. Cleve and Buhrman (1997) realized that if Alice and Bob share systems exhibiting nonlocal correlations, then they can compute certain functions with less communication than would be required without such nonlocal systems. This phenomenon does not violate the no-signaling principle because the knowledge that Bob obtains about Alice's input through $f(x, y)$ is no greater than what is already conveyed by Alice's communication. The field of communication complexity is an active field of research in computer science, in which strong connections with nonlocality have been discovered since Cleve and Buhrman (1997). For more details, see Buhrman *et al.* (2010).

B. Information theory

Nonlocal correlations can also enhance communication power in the context of information theory. Consider two parties, Alice and Bob, communicating via a noisy communication channel. If we care only about the rate at which information is transmitted from Alice to Bob such that the error rate goes to zero in the large block length limit, then this transmission rate cannot be increased using entanglement (Bennett *et al.*, 2002) or even no-signaling correlations (Cubitt *et al.*, 2011). However, the situation changes when we care about the rate at which information can be sent without any error at all (Cubitt *et al.*, 2011). The maximum such rate is known as the zero-error capacity of a noisy-communication channel. For example, it is known that if Alice and Bob share certain no-signaling correlations, zero-error transmission becomes possible through a noisy channel, even if that channel's zero-error capacity is zero without the ability to use such correlations (Cubitt *et al.*, 2010). Notably, even certain quantum correlations are useful in this context, in particular, those achieving a unit winning probability in a pseudotelepathy game (see Sec. II.B.4).

C. Quantum cryptography

In our second intuition discussed previously, in which the violation of Bell inequalities guarantees the presence of randomness, the no-signaling principle is no longer a limitation, but a prerequisite. Indeed, in the same way that every local model can be seen to be equivalent to a local deterministic model, every nonlocal model is equivalent to a nonlocal deterministic model where, for each run of the Bell test, definite outputs $a(x,y)$ and $b(x,y)$ are assigned to every pair of inputs (x,y). In full generality, the violation of Bell inequalities does not therefore guarantee by itself any indeterminacy in the outcomes a and b (as already stressed in the Introduction, nonlocality is—as its name indicates—about the violation of locality, not about the violation of determinism). However, every nonlocal deterministic model is necessarily signaling: if $a(x,y)$ depends nontrivially on both x and y, then Alice can recover some information about Bob's input y from the knowledge of the output a and her choice x. In a model that reproduces nonlocal correlations and which is intrinsically no signaling, the measurement outcomes cannot therefore be fully determined in each run of the Bell test and they must necessarily exhibit some randomness. This intuition is at the basis of *device-independent cryptography* in which the violation of a Bell inequality, which can be asserted without any detailed physical assumptions on the working of the devices, guarantees the production of cryptographic keys that are genuinely random and secure to any adversary limited by quantum theory or, more generally, by the no-signaling principle.

1. Initial developments

One of the earliest connections between nonlocality and cryptography is due to Herbert (1975), who interpreted the 0 and 1 outcomes produced by two distant quantum devices as correlated binary random messages. By considering the error rates in such messages, he presented an elementary derivation of Bell's theorem, but he did not go as far as deducing that quantum nonlocality could be exploited for a secure cryptographic scheme.

The practical application of Bell nonlocality to cryptography was first realized by Ekert in his celebrated paper (Ekert, 1991), which represents more generally one of the founding articles of quantum cryptography. The problem of establishing a secure, encrypted communication between two parties can be reduced to the problem of generating a secure, cryptographic key, i.e., sufficiently long strings of random bits that are shared between Alice and Bob, but unknown to any potential eavesdropper Eve. Ekert presented a protocol for this key distribution problem which is based on the CHSH inequality and uses a source of two-qubit maximally entangled states $|\phi_+\rangle$; here we present a slight variation of this protocol introduced by Acín, Massar, and Pironio (2006). Each party repeatedly receives one qubit from the source and performs a measurement on it. In each run, Alice chooses among three possible measurements $x = 0, 1, 2$ and obtains an outcome bit a; Bob chooses among two possible measurements $y = 0, 1$ and obtains an outcome bit b. Once all states have been measured, Alice and Bob publicly announce the settings they have chosen for each particular measurement and divide their results into two groups. The subset of the results corresponding to the measurements $x = 1, 2$ and $y = 0, 1$ is used to evaluate the CHSH inequality violation. Hence these measurements are chosen such as to maximize this violation (see Sec. I.A): for instance, Alice measures in the direction 0, $\pi/2$ in the x-z plane of the Bloch sphere for $x = 1, 2$ and Bob in the directions $-\pi/4, \pi/4$ for $y = 0, 1$ (also in the x-z plane). The subset of results corresponding to the choices $x = 0$ and $y = 0$ are used to generate the shared key. Hence the measurement $x = 0$ is chosen in the same direction $-\pi/4$ as Bob's measurement $y = 0$, in such a way that the key bits a and b are perfectly correlated.

The CHSH violation guarantees, as discussed earlier, that the key bits are undetermined and, in particular, that Eve could not have fixed them in advance. More generally, Eve could attempt to obtain information about the values of a and b by performing delayed measurements (after the public disclosure of Alice and Bob's settings) on a system of her own correlated with Alice and Bob's systems. As remarked by Ekert, the protocol is also secure against such attacks as a maximal CHSH violation guarantees that the state shared by Alice and Bob is (essentially equivalent) to a pure entangled $|\phi_+\rangle$ state, which cannot be correlated to any system in Eve's possession. In a realistic implementation, Alice and Bob's key bits will not be perfectly correlated and the CHSH violation will not be maximal, implying that Eve can obtain some finite information on these key bits. But provided that these imperfections are not too important, it should be possible to distill a shared secret key from the raw data of Alice and Bob by applying error-correction and privacy amplification protocols.

The intuition for security in the Ekert protocol is based on the violation of a Bell inequality which can be assessed independently of the protocol's implementation, but this aspect was not fully recognized at the time. When assuming that Alice and Bob's devices perform measurements on qubits in complementary bases, Ekert's protocol was found to be equivalent to an entanglement-based version of the Bennett-Brassard 1984 protocol (Bennett, Brassard, and Mermin, 1992). This was important in establishing entanglement as a central concept for quantum key distribution (QKD), but it also implied that the subsequent security proofs used "qubits" and "complementary bases" as implicit assumptions.[14] One crucial point was (understandably) missed in those early days: the fact that the implicit qubits and complementary base assumptions requires a very good control of, and ultimately some trust for, the physical implementation.[15]

[14] Quantitative relations between security bounds and the violation of Bell inequalities were pointed out (Scarani and Gisin, 2001); but this link turned out to be an artifact of the assumption called "individual attacks" and did not survive in stricter security proofs, which were rather derived from the notion of entanglement distillation.

[15] It is interesting to notice, though, the note added to the Bennett, Brassard, and Mermin (1992) paper, which comes close to explicitly recognizing the device-independent aspect of the Ekert scheme.

However, the violation of Bell inequalities can be established without such knowledge. Therefore, a cryptography protocol based on nonlocality requires fewer assumptions: the devices can in principle be tested and the security of the protocol certified without any detailed characterization of the devices; to some extent, the devices could even be malicious and have been prepared by the eavesdropper. This is called a *device-independent* (DI) assessment.

The idea of DI quantum cryptography was first made explicit by Mayers and Yao (1998, 2004), who called it self-testing. Although their analysis is not directly based on Bell inequalities, it obviously exploits correlations that are nonlocal. The breakthrough that pushed the recent development of device-independent quantum key distribution (DIQKD) came from Barrett, Hardy, and Kent who introduced a QKD protocol based on the chained Bell inequality (see Sec. II.B.3) and proved it to be secure against "superquantum" eavesdroppers that may violate the law of quantum physics but which are constrained by the no-signaling principle (Barrett, Hardy, and Kent, 2005). A practical protocol based on the CHSH inequality was then introduced by Acín, Gisin, and Masanes (2006) (although it was proved secure only against a restricted family of attacks), where it was also noticed that proving security assuming only the no-signaling principle implies, in particular, that one can do away with the "device-dependent" assumptions of standard QKD. The DI potential of such a QKD scheme based on Bell inequalities was then fully perceived by Acín *et al.* (2007), who introduced a DI security proof for collective attacks of the variation of Ekert's protocol presented above against an eavesdropper constrained by the entire quantum formalism and not only the no-signaling principle.

Finally, the ideas from DIQKD have been adapted to the simpler task of DI randomness generation (DIRNG) by Colbeck (2007) and Pironio *et al.* (2010) and to distrustful quantum cryptography by Silman *et al.* (2011), where a scheme for the device-independent implementation of (imperfect) bit commitment and coin tossing was introduced.

In the following we discuss in more detail the status of current security proofs for DIQKD and DIRNG, the assumptions on the devices that underlie them, and the prospects for experimental implementations. We first briefly discuss the quantitative aspects of the relation between randomness and nonlocality since it is at the basis of many security proofs and protocols. Note that the development of DIQKD and the recent attacks on standard QKD protocols such as Lydersen *et al.* (2010) have led to a series of feasible proposals for QKD that are intermediate between device-dependent and device-independent schemes; see, for instance, Lydersen *et al.* (2010), Pawlowski and Brunner (2011), Branciard, Cavalcanti *et al.* (2012), Braunstein and Pirandola (2012), Lo, Curty, and Qi (2012), Tomamichel *et al.* (2012), and Lim *et al.* (2013). We do not review this work here, as it does not directly rely on nonlocality as a resource.

2. Randomness versus nonlocality

a. Quantitative measures of randomness

Imagine Alice holds a measurement device that produces outcomes a when performing a measurement, where we let R_A denote the random variable of the outcome. When can we say that a is random? One way to think about randomness is by means of introducing an observer Eve, who tries to guess Alice's measurement outcome a—the better the guess the less random a is. In order to guess a, Eve may perform an arbitrary measurement on a system E, which is possibly correlated with the one of Alice. We use z to label her measurement setting and e to label her measurement outcome. For any given z, Eve's best guess for a corresponds to the most probable outcome, the one maximizing $p(a|ez)$. The *guessing probability* of Eve is then defined as her average probability to correctly guess a, maximized over all her possible measurements

$$p_{\text{guess}}(R_A|E) := \max_z \sum_e p(e|z) \max_a p(a|e,z). \tag{62}$$

This guessing probability can also be expressed as the minimum entropy $H_{\min}(R_A|E) = -\log_2 p_{\text{guess}}(R_A|E)$ (König, Renner, and Schaffner, 2009). It takes on values between 0 and $\log|R_A|$, corresponding to the cases where Eve can guess perfectly, and where Eve's probability of guessing is no better than for the uniform output distribution $1/|R_A|$, respectively.

The minimum entropy is a good measure of how random Alice's measurement outputs are because it tells us exactly how many uniform classical random bits ℓ can in principle be obtained from a classical string $a \in R_A$ by applying some function $f_r: R_A \to \{0,1\}^\ell$. It is easy to see that if we have a guarantee only about the min entropy of the so-called source R_A, then no randomness can be obtained using just one deterministic function f. However, if we are willing to invest some perfect randomness labeled $R = r$ from an initial seed, and choose a function f_r depending on it, then we can obtain randomness. This process is known as *randomness extraction* and enjoys a long history in computer science [see Vadhan (2012) for a survey]. Formally, a (strong) extractor produces an output $\rho_{F(R_A)E}$ that is close to uniform and uncorrelated from Eve $\|\rho_{F(R_A)ER} - \mathbb{1}_{2^\ell}/2^\ell \otimes \rho_{ER}\|_1 \leq \epsilon$ for some small ϵ, even if Eve later learns which function f_r we applied. In the context of cryptography, this is also called privacy amplification. If Eve holds only classical side information about Alice's system it is known that randomness extraction is possible, where the maximum output size obeys $\ell \approx H_{\min}(R_A|E)$ (Impagliazzo, Levin, and Luby, 1989). This is also true if Eve holds quantum side information (Renner, 2008; De *et al.*, 2009; Ta-Shma, 2009). More generally the full quantum minimum entropy (König, Renner, and Schaffner, 2009) has been shown to exactly characterize how much randomness can be obtained by making measurements on A by Berta, Fawzi, and Wehner (2012). However, no such general result is known if Eve holds arbitrary no-signaling (i.e., supraquantum) side information (Hänggi and Renner, 2010); see Sec. IV.C.4 for a more detailed discussion.

b. Randomness and Bell violations

In order to discuss quantitative links between randomness and the violations of Bell inequalities, it is useful, as in the previous discussion, to introduce an additional observer and thus consider nonlocal correlations shared between Alice, Bob, and Eve. In such a tripartite setting, the correlations are characterized by the probabilities $p(abe|xyz)$. If Eve measures z and obtains e, then Eve's characterization of Alice's device is now given by the conditional probability distributions $p(a|xez)$. If Eve learns x, then for any given z her best guess for a corresponds to the most probable outcome maximizing $p(a|xez)$. Maximizing over z thus means that Eve can guess a with probability $p_{\text{guess}}(R_A|EX=x)$. In the case where a can take on two values and Alice and Bob's devices are characterized by a CHSH expectation value S, it was shown by Pironio *et al.* (2010) [see also Masanes, Pironio, and Acín (2011) for an alternative derivation] that independently of the devices' behaviors and Eve's strategy

$$p_{\text{guess}}(R_A|EX=x) \le \tfrac{1}{2}(1+\sqrt{2-S^2/4}). \tag{63}$$

In particular, when $S=2\sqrt{2}$, we get as expected $P_{\text{guess}}(A|EX=x) \le 1/2$ corresponding to 1 bit of minimum entropy $H_{\min}(R_A|EX=x)$, implying that Alice's output is fully random. When the CHSH expectation achieves the local bound $S=2$, we get the trivial bound $p_{\text{guess}}(A|EX=x) \le 1$. Using the SDP hierarchy (Navascues, Pironio, and Acín, 2007, 2008) (see Sec. II.C.1.d) it is possible to derive analogous bounds for arbitrary Bell inequalities; see Pironio *et al.* (2010) for details. One can also compute upper bounds on the guessing probability not only for the local randomness (corresponding to the output a alone), but also of the global randomness (corresponding to the pair of outcomes a and b). At the point of maximal CHSH violation, for instance, one finds $p_{\text{guess}}(R_A R_B|EX=xY=y) \le 1/4 + \sqrt{2}/8 \simeq 0.427$ corresponding to 1.23 bits of minimum entropy (Pironio *et al.*, 2010; Acín, Massar, and Pironio, 2012), where the random variable R_B corresponds to Bob's outcome.

The above bounds on the guessing probability are obtained by assuming that the devices and the eavesdropper obey quantum theory. Similar bounds can be obtained assuming only the no-signaling principle. In this case, one obtains the following tight bound for the CHSH inequality (Barrett, Kent, and Pironio, 2006; Masanes *et al.*, 2009; Pironio *et al.*, 2010):

$$p_{\text{guess}}(R_A|EX=x) \le \frac{3}{2} - \frac{S}{4}. \tag{64}$$

At the point $S=2\sqrt{2}$ of maximal quantum violation, one finds $p_{\text{guess}}(A|EX=x) \le 0.79$ which is, as expected, less constraining than the quantum bound (63). Maximal randomness $p_{\text{guess}}(A|EX=x) \le 1/2$ is obtained now only at the maximal no-signaling violation $S=4$ of the CHSH inequality, corresponding to a PR box. The above bound has also been generalized for the Δ-output–m-input chained inequality (see Sec. II.B.3) by Barrett, Kent, and Pironio (2006)

$$p_{\text{guess}}(R_A|EX=x) = \frac{1}{d} + \frac{d}{4} S^{(\Delta,m)}_{\text{chained}}. \tag{65}$$

For $m \to \infty$ the maximal quantum violation of the chained inequality tends to the maximal no-signaling violation $S_{\text{chained}} = 0$; note that the Bell inequality is written as $S^{(\Delta,m)} \ge \Delta - 1$, and thus a "high" violation means a lower value for $S^{(\Delta,m)}_{\text{chain}}$. In this limit, one thus gets $p_{\text{guess}}(R_A|EX=x) \le 1/\Delta$, i.e., the outcome can be certified to be fully random even assuming no signaling alone. This property is central to the security of the QKD protocol introduced by Barrett, Hardy, and Kent (2005) and was further developed by Colbeck and Renner (2008, 2011) to show that some extensions of quantum theory cannot have improved predictive power.

Naively one would expect that less nonlocality in a Bell-type experiment implies less randomness. In the quantum setting, this intuition is not always correct. In the case of two-output Bell scenarios, the maximal amount of local randomness (characterizing the single outcome a) corresponds to 1 bit of minimum entropy and the maximal amount of global randomness (characterizing the joint outcome pair a, b) corresponds to 2 bits. Acín, Massar, and Pironio (2012) showed that, through a family of (nonfacet) two-input–two-output Bell inequalities, that such values can be attained with nonlocal correlations that are arbitrarily close to the local region or which arise from states with arbitrarily little entanglement. This work suggests that while nonlocality is necessary to certify the presence of randomness, its quantitative aspects are related to the extremality of nonlocal correlations. Extremality was already identified by Franz, Furrer, and Werner (2011) as a key property for characterizing the behaviors which are independent of any measurement results of an eavesdropper. This work also presents different tools to certify and to find extremal behaviors for particular Bell scenarios. Finally, Dhara, Prettico, and Acín (2013) showed that maximal global randomness can be obtained in a variety of scenarios (including multipartite ones) from the violation of certain Bell inequalities.

3. Device-independent randomness generation

The above relations between nonlocality and randomness immediately suggest using Bell-violating devices to certify the generation of random numbers in a DI manner (Colbeck, 2007). This idea was further developed by Pironio *et al.* (2010), where a practical protocol for randomness generation was introduced, the first quantitative bounds on the randomness produced where shown, and a proof-of-principle experimental demonstration was performed.

The bounds that we presented above relate only the randomness and expected Bell violation of a pair of quantum devices for a single use of the devices. In an actual protocol for DIRNG, however, the devices are used n times in succession. A typical protocol consists of three main steps (Colbeck, 2007; Pironio *et al.*, 2010; Colbeck and Kent, 2011): a measurement step, where the successive pairs of inputs $(x_1, y_1), \ldots, (x_n, y_n)$ are used in the devices, yielding a sequence of outputs $(a_1, b_1), \ldots, (a_n, b_n)$; an estimation step, where the raw data are used to estimate a Bell parameter (if this parameter is too low, the protocol may abort); and a randomness extraction step, where the raw output string is processed to obtain a smaller final string $r = r_1, \ldots, r_m$ which is uniformly random and private with respect to any potential adversary. In addition to the Bell-violating devices, the protocol may also consume some initial random seed for choosing the inputs in the measurement step and for processing the raw data in the randomness extraction step. If more randomness is generated than is initially consumed, one has then achieved DI randomness expansion.

Pironio *et al.* (2010) introduced a generic family of protocols based on arbitrary Bell inequalities and achieving quadratic expansion. These protocols are robust to noise and generate randomness for any amount of violation (up to statistical errors). The analysis of the randomness that is produced is based on an extension of the single-copy bounds of Eqs. (63) and (64) to the n-copy case. A proof-of-principle implementation using two entangled atoms separated by about 1 m was also reported (see Sec. VII). The security of these protocols has been proven against quantum or no-signaling adversaries with *classical-side* information. The technical tools for proving security were already introduced by Pironio *et al.* (2010), but this was rigorously established only by Fehr, Gelles, and Schaffner (2013) and Pironio and Massar (2013). In these later works, it was further shown how to achieve superpolynomial randomness expansion by repeatedly using the randomness of a pair of devices as input for another pair. A scheme based on the CHSH inequality secure against adversaries with *quantum-side* information and achieving superpolynomial expansion with a single pair of quantum devices was obtained by Vazirani and Vidick (2012a). This scheme, though, requires a high violation of the CHSH inequality and is not noise tolerant.

The security of the above protocols relies on a series of minimal assumptions. First, the devices and the eavesdropper are constrained by quantum theory or at least by the no-signaling principle. Second, the initial randomness seed is independent and uncorrelated from the devices' behavior. Third, the two quantum devices are noninteracting during each successive measurement.[16] Fourth, it is also implicit of course that the devices can be secured, in the sense that they do not directly leak unwanted information to the adversary. Apart from these basic requirements, the devices are mostly uncharacterized. In particular, no assumptions are made on the specific measurements that they implement, on the quantum state that is being measured, on the Hilbert space dimension, etc.

The level of confidence in the realization of the above assumptions in an actual implementation or the measures that must be taken to enforce them may vary depending on the adversary model that one is considering. For instance, it depends on whether the devices are considered to be outright malicious and programmed by a dishonest provider (i.e., the adversary itself) or whether the manufacturer of the device is assumed to be honest and the concept of DI is merely used to account for limited control of the apparatus or unintentional flaws in the devices (Pironio and Massar, 2013). In the latter case, in particular, a weak source of randomness, such as a pseudorandom generator, may be sufficient for all practical purposes to generate the initial seed (in which case the protocol, which produces strong cryptographically secure randomness, is best viewed as a randomness generation protocol than an expansion one). Note that, in the honest-provider scenario, the adversary may be considered to be disentangled from the quantum devices, implying that proving security against classical-side information as in Fehr, Gelles, and Schaffner (2013) and Pironio and Massar (2013) is already sufficient.

Recently, protocols and security analysis have also been introduced where some of the above assumptions are relaxed. In Silman, Pironio, and Massar (2013), the separation assumption is relaxed and a small amount of cross talk between the devices is allowed. This opens up the possibility of using existent experimental systems with high data rates, such as Josephson phase qubits on the same chip.

Colbeck and Renner (2012) introduced the problem of *randomness amplification*, which aims at extracting perfect (or arbitrarily close to perfect) randomness from an initial source that is partly correlated with the devices and the adversary. It was shown that if one is given access to certain so-called Santha-Vazirani (SV) sources, then randomness amplification against an adversary limited only by the no-signaling principle is possible for certain parameters of the source. Improving on this first result, Gallego, Masanes *et al.* (2013) showed that an arbitrarily SV source can be amplified using certain multipartite quantum correlations. Finally, less stringent models of a compromised random seed than SV have been considered (Hall, 2011) and the conditions for Bell-based randomness expansion against an adversary preparing independent and identically distributed correlations have been studied by Koh *et al.* (2012).

4. Device-independent quantum key distribution

The protocols, the underlying assumptions, and the security proofs for DIQKD are similar in spirit to DIRNG with the added complication that DIQKD involves two remote parties that must communicate over a public channel to establish the shared secret key. A typical DIQKD protocol consists of the following steps: a measurement step, where Alice and Bob measure a series of entangled quantum systems; an estimation step, in which Alice and Bob publicly announce a fraction of their measurement results to estimate the violation of a Bell inequality and the error rate in their raw data; an error-correction step, in which these errors are corrected using a classical protocol that involves public communication; and finally, a privacy-amplification step in which a shorter, secure key is distilled from the raw key based on a bound on the eavesdropper's information deduced from the Bell violation estimation.

The first DIQKD protocol proven secure against general attacks by a no-signaling eavesdropper was introduced by Barrett, Hardy, and Kent (2005). The protocol is based on the chained Bell inequality (23) and produces a single secure key bit. It represents mostly a proof-of-principle result as the protocol is inefficient and unable to tolerate reasonable levels

[16] Note that this does not necessarily imply that the measurements should be spacelike separated in the relativistic sense. This spacelike separation is required to close the locality loophole in fundamental tests of Bell inequalities, where the aim is to rule out alternative models of nature that can go beyond present-day physics. In the context of DIRNG, we assume however from the beginning the validity of quantum theory and use Bell inequalities as a tool to quantify in a DI way the randomness of quantum theory. Once we assume quantum theory, there are many ways to ensure that the two systems are not interacting other than placing them in spacelike intervals, e.g., by shielding the devices (Pironio *et al.*, 2010).

of noise. In Barrett, Hardy, and Kent (2005) security is proven assuming that each of the n entangled pairs measured in the protocol is isolated from the other pairs. The protocol thus requires that Alice and Bob have n separate pairs of devices, rather than a single pair of devices that they use repeatedly n times. The no-signaling conditions are required to hold between each of the $2n$ systems of Alice and Bob. This assumption was removed by Barrett, Colbeck, and Kent (2012), where security is proven in the situation where Alice and Bob have only one device each, which they repeatedly use. Instead of full no-signaling correlations among the $2n$ systems of Alice and Bob, the security is thus based on *time-ordered* no-signaling conditions, where no signaling is required only from future inputs to previous inputs, but where later outputs can depend arbitrarily on previous inputs.

Efficient and noise-tolerant protocols were introduced by Acín, Gisin, and Masanes (2006) and Scarani *et al.* (2006) [see also Acín, Massar, and Pironio (2006)], where however the security analysis was restricted to individual attacks against no-signaling eavesdroppers. General security against no-signaling eavesdroppers was later proven by Masanes (2009), Masanes *et al.* (2009), and Hänggi, Renner, and Wolf (2010) under the assumption, as in Barrett, Hardy, and Kent (2005), that Alice and Bob use n separated pairs of devices constrained by full no-signaling conditions. The question of whether it is possible to prove the security of an efficient and noise-tolerant protocol in the case where Alice and Bob repeatedly use a single pair of devices constrained by time-ordered no-signaling conditions is still open. One of the difficulties in obtaining such a result is related to the possibility of performing privacy amplification against a no-signaling eavesdropper. Hänggi, Renner, and Wolf (2009) showed that if no signaling is imposed between only Alice's device and Bob's, but signaling within each device is allowed (so that the output of a device can depend on the inputs of other devices used later in the protocol), then privacy amplification is not possible for protocols based on the CHSH inequality. This result was further extended by Arnon-Friedman, Hänggi, and Ta-Shma (2012) for a set of more general conditions, but still less restrictive than the desired time-ordered no-signaling conditions. Recently, Arnon-Friedman and Ta-Shma (2012) showed that superpolynomial privacy amplification for protocols based on the chained inequality is impossible under the assumption of time-ordered no-signaling conditions. This work still leaves open the question of exponential privacy amplification for protocols based on a different Bell inequality or whether linear privacy amplification is possible.

Another line of results, concerned with security against eavesdroppers that are constrained by the entire quantum formalism and not only the no-signaling principle, was initiated by Acín *et al.* (2007) . The advantage in this case is that better key rates and noise resistance can be expected [as illustrated by the difference between the randomness bounds (63) and (64)] and that privacy amplification is possible and well studied. The work of Acín *et al.* (2007) and Pironio *et al.* (2009) proved the security of the CHSH-based protocol introduced by Acín, Massar, and Pironio (2006) against collective attacks by a quantum eavesdropper. This proof was extended to a slightly more general setting by McKague (2010b). General security proofs of protocol based on arbitrary Bell inequalities under the assumption that the devices of Alice and Bob are memoryless (or equivalently that they use n noninteracting pairs of devices instead of a single one) were introduced by Hänggi and Renner (2010) and Masanes, Pironio, and Acín (2011). The memory assumption on the device was removed by Pironio *et al.* (2013), but security was proven only against quantum adversaries with classical-side information, a condition that is satisfied if the eavesdropper has access only to short-term quantum memories. The key rates in Hänggi and Renner (2010), Masanes, Pironio, and Acín (2011), and Pironio *et al.* (2013) are simple expressions expressed in term of single-copy bounds on the randomness of the form (63). The general security of a CHSH-based protocol with no memory assumptions on the devices or the eavesdropper was reported by Reichardt, Unger, and Vazirani (2012, 2013), albeit polynomially inefficient and does not tolerate noisy devices. The security is obtained as a corollary of a more general strong testing result that allows the shared quantum state and operators of the two untrusted devices to be completely characterized. Finally, a complete DI proof of security of QKD that tolerates a constant noise rate and guarantees the generation of a linear amount of key was given by Vazirani and Vidick (2012b) for a protocol that is a slight variant of Ekert's protocol. It is an open question whether this approach can lead to trade-offs between the noise rate and the key rate as good as the ones that have been shown to be achievable under additional memory assumptions on the devices or the eavesdropper.

The general assumptions that underlie the above proofs are similar to the ones for DIRNG: the validity of quantum theory or the no-signaling principle, access to a random seed independent of the devices and the eavesdropper, a separation assumption on the behavior of the devices, and the implicit assumption that the devices do not directly leak out unwanted information to the eavesdropper. Apart from that, the devices are mostly uncharacterized and no assumptions are made on the Hilbert space dimension, the specific measurements that are implemented, etc.

Note that in the dishonest-provider scenario, where the devices are outright malicious and assumed to have been prepared by the eavesdropper, repeated implementations of a protocol using the same devices can render an earlier generated key insecure due to device-memory-based attacks (Barrett, Colbeck, and Kent, 2013). In such attacks, untrusted devices may record their inputs and outputs and reveal information about them via publicly announced outputs during later implementations of the protocol. See Barrett, Colbeck, and Kent (2013) for a thorough discussion of the general scope of such attacks, including the possibilities of countering them by refined protocols. A countermeasure relying on an encryption scheme which allows Alice and Bob to exchange data without the devices leaking information about previously generated keys to Eve was presented by McKague and Sheridan (2012).

Finally, we say a few words about experimental perspectives for DIQKD. The implementation of a DIQKD protocol requires a genuine Bell violation over large distances. Genuine here means with the detection loophole closed (at least if one is considering complete DI with no further assumptions on the devices); see Sec. VII.B.1. Transmission losses in optical fibers, however, represent a fundamental limitation for the realization of a detection-loophole free Bell test on any distance relevant for QKD. Approaches to circumvent the problem of transmission losses have been proposed based on heralded qubit amplifiers (Gisin, Pironio, and Sangouard, 2010; Pitkanen *et al.*, 2011) and standard quantum relays based on entanglement swapping with linear optics (Curty and Moroder, 2011), but an experimental demonstration still represents a great challenge. Quantum repeaters may also provide a possible solution. More recently, another approach based on spin-photon interactions in cavities was also discussed (Brunner *et al.*, 2013; Mattar, Brask, and Acin, 2013). Improved data postprocessing has also been proposed to increase the tolerance to lost photons (Ma and Lütkenhaus, 2012).

D. Other device-independent protocols

In a quantum experiment, the violation of a Bell inequality reveals the presence of entanglement in a device-independent way. In fact, in some cases a much stronger statement can be made. Certain quantum correlations can be reproduced only by performing specific local measurements on a specific entangled state. Hence the observation of such correlations allows one to characterize an unknown source of quantum states, as well as the measurement devices, in a device-independent manner. For instance, the observation of the maximal violation of the CHSH inequality implies that the underlying quantum state is necessarily equivalent to a two-qubit singlet state (Cirel'son, 1980). Moreover, the measurement settings of both Alice and Bob must anticommute (Braunstein, Mann, and Revzen, 1992; Popescu and Rohrlich, 1992). Another method, developed by Mayers and Yao (2004), allows one to reach the same conclusion. Such procedures are termed *self-testing* of the singlet state.

More formally, these works show the following. Consider an experiment involving a state $|\psi\rangle$ and measurement operators M_A^i and M_B^j, with i, $j = 1, 2$. If a CHSH value of $S = 2\sqrt{2}$ is achieved, then the state is equivalent (up to local isometries) to a singlet state $|\psi_-\rangle$ and the measurement are to anticommuting Pauli operators σ_A^i for Alice with $\{\sigma_A^i, \sigma_A^k\} = 2\delta_{ik}\mathbb{1}$ (and similarly for Bob σ_B^j), in the sense that

$$\Phi(|\psi\rangle) = |\text{junk}\rangle \otimes |\psi_-\rangle, \tag{66}$$

$$\Phi(M_A^i M_B^j|\psi\rangle) = |\text{junk}\rangle \otimes \sigma_A^i \sigma_B^j |\psi_-\rangle, \tag{67}$$

where $\Phi = \Phi_A \otimes \Phi_B$ is a local isometry, and $|\text{junk}\rangle$ is a state shared by Alice and Bob.

For a self-testing protocol to be practical it should be robust to small deviations from the ideal case, due for instance to experimental imperfections. The first proof of the robustness of the Mayers-Yao scheme was derived by Magniez *et al.* (2006), and later considerably simplified by McKague and Mosca (2011). McKague, Yang, and Scarani (2012) presented a framework for studying the robust self-testing of the singlet state, which can be used to device independently certify the entanglement fraction of a source (Bardyn *et al.*, 2009). More generally, it was shown in the ground-breaking work of Reichardt, Unger, and Vazirani (2012, 2013) that self-testing can be achieved in the CHSH scenario even if the devices feature a quantum memory. Loosely speaking, this means that the only way to achieve a violation of the CHSH inequality close to $2\sqrt{2}$ is if the measured bipartite states are close to the tensor product of singlet states, and the measurements are the optimal CHSH measurements.

Self-testing of other quantum states was also discussed. In particular, the case of partially entangled bipartite states was addressed by Yang and Navascues (2013). In the multipartite setting, the case of graph states was discussed by McKague (2010a), while Miller and Shi (2012) considered self-testing in XOR games. Also, the device-independent certification of "entangled measurements" was investigated (Rabelo *et al.*, 2011; Vértesi and Navascues, 2011).

An interesting development of these ideas is the possibility of self-testing a quantum computation. This consists of self-testing a quantum state and a sequence of operations applied to this state. This approach was introduced by Magniez *et al.* (2006). A full analysis of such a protocol, with a reduced set of assumptions compared to Magniez *et al.* (2006), was recently given by Reichardt, Unger, and Vazirani (2013).

Moving away from self-testing, an interesting development is the device-independent assessment of multipartite quantum entanglement. Notably, techniques for devising device-independent witnesses of genuine multipartite entanglement (Bancal, Gisin *et al.*, 2011) were developed. Moreover, Brunner, Sharam, and Vértesi (2012) discussed how the structure of multipartite entangled states can be characterized using Bell inequalities; that is, how different classes of multipartite entangled states can be distinguished from each other from their nonlocal correlations.

V. INFORMATION-THEORETIC PERSPECTIVE ON NONLOCALITY

As seen in Sec. IV, nonlocality can be seen as a resource for information processing and communication tasks and the no-signaling principle plays a fundamental role in this respect. We have also seen in Sec. II that there exist no-signaling correlations that are more nonlocal than those of quantum theory, as pointed out by Popescu and Rohrlich (1994). If Alice and Bob had access to such PR boxes they could implement many of the protocols discussed earlier, from communication complexity to cryptography, often with much higher efficiency than what quantum correlations allow (van Dam, 2005). No-signaling nonlocal correlations can thus be viewed as information-theoretic resources and investigated as such (Barrett, Linden *et al.*, 2005). This new perspective raises two general questions: Can we develop a resource theory of nonlocality, similar to the resource theory of entanglement? What distinguishes quantum correlations from more general no-signaling correlations in this information-theoretic context? To answer them it is first useful to identify the physical properties which are generic to all no-signaling nonlocal theories.

A. Properties of no-signaling correlations

Remarkably, it turns out that many features of quantum mechanics, usually thought of as counterintuitive and genuinely quantum, are in fact general features of any no-signaling theory featuring nonlocality (Masanes, Acín, and Gisin, 2006; Barrett, 2007). These include a no-cloning theorem, the monogamy of correlations, a disturbance versus information gain trade-off in measurements, the inherent randomness of measurement outcomes, the complementarity of measurements, and uncertainty relations. These physical properties are clearly relevant from an information-theoretic point of view; consider for instance the role that the no-cloning theorem or the monogamy of entanglement plays in quantum information science. The fact that such properties are generic to all no-signaling nonlocal theories thus already suggests that such theories offer interesting possibilities for information processing.

We already gave the intuition in Sec. IV.C of why measurement outcomes must be random in any nonlocal no-signaling theory. We now illustrate some of the other above properties with simple examples based on Popescu-Rohrlich–type correlations. Consider that Alice and Bob share a PR box, i.e., correlations of the form

$$p(ab|xy) = \begin{cases} \frac{1}{2}, & a\oplus b = xy, \\ 0, & \text{otherwise}, \end{cases} \tag{68}$$

where $\oplus$ is addition mod 2, and $x, y \in \{0, 1\}$ denote the inputs and $a, b \in \{0, 1\}$ the outputs. The impossibility of having a perfect cloning machine is here easily derived here by contradiction. Assume such a machine exists. Then Bob could apply it to its subsystem, resulting in a tripartite probability distribution $p(ab_1b_2|xy_1y_2)$ satisfying

$$a\oplus b_1 = xy_1, \qquad a\oplus b_2 = xy_2, \tag{69}$$

with a, b_1, and b_2 locally uniformly distributed. Combining Eqs. (69) leads to

$$b_1\oplus b_2 = x(y_1\oplus y_2), \tag{70}$$

showing that Bob's marginal probability distribution directly depends on x, the input of Alice, when Bob uses inputs such that $y_1\oplus y_2 = 1$. Thus, Alice can signal to Bob, which contradicts our basic hypothesis that the theory is nonsignaling. Therefore, we concluded that a perfect cloning machine cannot exist in a theory featuring PR-box correlations. General and rigorous proofs can be found in Masanes, Acín, and Gisin (2006) and Barrett (2007). The impossibility of broadcasting no-signaling nonlocal correlations has been discussed by Barnum *et al.* (2007) and Joshi *et al.* (2013).

The above simple example also indicates that no-signaling correlations are constrained by monogamy relations (see Sec. VI.C). In particular, a PR box being an extremal point of the no-signaling set must be decoupled from any other system (Barrett, Linden *et al.*, 2005; Masanes, Acín, and Gisin, 2006).

For the last example, we illustrate the existence of a notion of complementarity of measurements in generalized nonsignaling theories (Masanes, Acín, and Gisin, 2006). Considering again PR-box correlations, the two possible measurements on Bob's side (corresponding to $y = 0$ and $y = 1$) cannot be compatible; that is, there cannot be a single joint measurement Y returning outcomes b_0 and b_1 corresponding, respectively, to $y = 0$ and $y = 1$. Indeed, this implies the existence of a distribution $P(ab_0b_1|xY)$ satisfying $b_0\oplus b_1 = x$ (since $a\oplus b_0 = 0$ and $a\oplus b_1 = x$), thus violating no signaling as in the above example.

B. Nonlocality measures, interconversion, and distillation

If nonlocal boxes can be viewed as an information-theoretic resource, can we define a theoretical framework, analogous, e.g., to the framework that has been developed for the study of entanglement, that would allow us to answer unambiguously questions such as can two given sets of nonlocal correlations be considered equivalent resources or what is a good measure of nonlocality?

A prerequisite for addressing these issues is to understand interconversion between nonlocal boxes, that is, the simulation of a given nonlocal box using a supply of other nonlocal boxes. In this context, separated parties are allowed to perform local operations on their boxes. They can relabel the inputs and outputs, and also “wire” several boxes, using for instance the output of one box as the input for another box. Importantly, classical communication is not allowed, as it represents a nonlocal resource, which allows trivially for the simulation of any nonlocal box.

The interconversion of bipartite boxes has been studied by Barrett, Linden *et al.* (2005), Jones and Masanes (2005), and Forster and Wolf (2011) and is by now relatively well understood. The main conclusion to be drawn from these works is that the PR box represents a good unit of bipartite nonlocality (much like the singlet in the case of entanglement) in the sense that any bipartite no-signaling box can be simulated to an arbitrary precision using a supply of PR boxes (Forster and Wolf, 2011). In the multipartite case, the situation is more complicated. On the one hand, several classes of extremal nonlocal boxes can be simulated exactly using PR boxes (Barrett, Linden *et al.*, 2005; Barrett and Pironio, 2005). On the other hand, there exist nonsignaling boxes which can be proven to not be approximated using an arbitrarily large supply of PR boxes (Barrett and Pironio, 2005; Pironio, Bancal, and Scarani, 2011). In particular, there exist quantum nonlocal correlations with this property (Barrett and Pironio, 2005). It is still an open question whether there exists a unit of multipartite nonlocality; in fact, even proposing a good candidate is challenging given the complexity of the set of multipartite nonsignaling correlations (see Sec. II.D).

Another relevant issue is whether nonlocality can be distilled. That is, from a supply of weakly nonlocal boxes is it possible to obtain via local operations (i.e., relabelings and wirings) one copy of a box featuring more nonlocality, in the sense that it violates more a given Bell inequality than the original boxes? Interestingly, nonlocality distillation is possible for certain classes of nonlocal boxes (Forster, Winkler, and Wolf, 2009). Moreover, maximally nonlocal PR-box correlations can be distilled out of certain boxes with arbitrarily weak nonlocality (Brunner and Skrzypczyk, 2009), i.e., violating a Bell inequality by an arbitrarily small amount. The existence of such distillation protocols has important consequences from an information-theoretic point of view. For instance, if a certain class of boxes can be distilled to a PR box, then all boxes in this class inherit the information-theoretic power of the PR box. Note also a series of negative results, concerning, in particular, the impossibility of distilling isotropic nonlocal correlations. Such correlations (mixtures of PR box and white noise) are of particular importance, since any nonlocal box can be "depolarized" to an isotropic without decreasing its nonlocality (Masanes, Acín, and Gisin, 2006). Partial no-go theorems have been derived (Dukaric and Wolf, 2008; Short, 2009; Forster, 2011), but a full proof is still missing.

These developments have opened novel possibilities for defining natural measures of nonlocality, such as the "distillable nonlocality" (Forster, Winkler, and Wolf, 2009; Brunner *et al.*, 2011) of a nonlocal box, the maximal amount of nonlocality that can be extracted from an arbitrarily large supply of such boxes. The first steps toward establishing a more general resource theory of nonlocality have recently been taken (Brunner *et al.*, 2011; Gallego, Würflinger *et al.*, 2012).

Finally, it is interesting to look for sets of correlations which are invariant under local operations. A set is said to be closed under wirings if, by combining correlations of this set via local operations, it is impossible to generate correlations outside the set. The study of such sets was initiated by Allcock, Brunner, Linden *et al.* (2009). Clearly the sets of local, quantum, and no-signaling correlations are all closed under wirings. Finding other closed sets appears to be a nontrivial problem. An interesting open problem is whether there exists, in the CHSH scenario, a strict subset of the no-signaling polytope that is closed under wirings and features more nonlocality than quantum mechanics (i.e., violating Tsirelson's bound).

C. Consequences of superstrong nonlocality

The existence of no-signaling correlations stronger than quantum mechanical ones raises fundamental questions. Why is nonlocality limited in quantum theory? Would there be unlikely consequences from a physical or information-theoretical point of view if supraquantum correlations were available? Can we identify reasonable principles that allow us to characterize the boundary that separates quantum from supraquantum correlations? The work discussed next addresses such questions. We first deal with information-theoretic consequences of supraquantum nonlocality and then use more physical concepts.

1. Information-theoretic consequences

a. Communication complexity and nonlocal computation

The first result showing a sharp difference between quantum and superquantum correlations in their capability of performing information-theoretic tasks was given by van Dam (2005) in the context of communication complexity. As discussed in Sec. IV.A, communication complexity deals with the problem of determining the number of bits that Alice and Bob need to exchange to compute the value $f(x,y)$ of a function whose inputs x and y are distributed among Alice and Bob. The amount of communication that is required depends on the particular function f and the resources that are available to Alice and Bob. Consider binary (or boolean) functions $f(x,y):\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ taking n-bit strings $x = x_1 \cdots x_n$ and $y = y_1 \cdots y_n$ as inputs. It was proven that some of these functions have high communication complexity, basically Alice must send her entire bit string x to Bob, even if Alice and Bob are allowed to share unlimited prior entanglement. An example of such a function is the inner product function $f(x,y) = x \cdot y = \sum_i x_i y_i$ (Cleve *et al.*, 1999). In contrast, if unlimited PR boxes were available to Alice and Bob, then a single bit of classical communication from Alice to Bob is sufficient for Bob to evaluate any binary function, that is, communication complexity collapses.

Consider again the inner product function. Suppose that Alice and Bob share n PR boxes and receive inputs x and y. They input x_i and y_i in box i, and then get outcomes a_i and b_i satisfying $a_i \oplus b_i = x_i y_i$. The inner product function can be expressed as

$$f(x,y) = \sum_i x_i y_i = \sum_i a_i \oplus b_i = \underbrace{\sum_i a_i}_{\text{Alice's side}} \oplus \underbrace{\sum_i b_i}_{\text{Bob's side}} .$$

Thus Alice can compute locally $c = \sum_i a_i$, and send the single bit c to Bob who then outputs $c \oplus b$, where $b = \sum_i b_i$, which is indeed the inner product. The inner product function is of particular importance, since any binary function f can be decomposed into inner products, from which the result of van Dam follows.

This idea was later generalized to the context of probabilistic communication complexity where Alice and Bob must compute f with a minimum probability of success (Brassard *et al.*, 2006). It was shown that certain noisy PR boxes, with CHSH value $S > 4\sqrt{2/3} \approx 3.266$, make communication complexity trivial in this scenario. Finally, using nonlocality distillation, it can be shown that (nonquantum) boxes with an arbitrarily small amount of nonlocality can nevertheless collapse communication complexity (Brunner and Skrzypczyk, 2009).

Linden *et al.* (2007) introduced a task closely related to communication complexity, termed *nonlocal computation*. The binary function f that Alice and Bob must compute has the special form $f(x,y) = g(x \oplus y) = g(z)$ where $g(z)$ is a boolean function taking as input an n-bit string z (with $z_i = x_i \oplus y_i$ and x_i uniform for $i = 1,..,n$). Thus each party has locally no information about the function's input z. Alice and Bob are asked to output one bit, respectively, a and b, such that $a \oplus b = f(x,y) = g(z)$. The figure of merit is then the average success probability of Alice and Bob. While strategies based on quantum correlations offer no advantage over classical ones for the nonlocal computation of an arbitrary function, it turns out that certain superquantum correlations provide an advantage. Remarkably, if one considers as a function the nonlocal AND of two bits $g(z_1, z_2) = z_1 z_2$, then the limit at which noisy PR boxes stop providing an advantage over classical and quantum correlations corresponds exactly to Tsirelson's bound. Note, however, that when the distribution of inputs is not perfectly uniform, i.e., when Alice and Bob have partial knowledge (even arbitrarily small) about the function's input z, quantum correlations provide an advantage over classical ones (Allcock, Buhrman, and Linden, 2009).

b. Information causality

Suppose Alice sends an m-bit message to Bob. How much information is potentially available to Bob? A natural guess is that the amount of information potentially available to Bob is equal to what he receives, that is, m bits. This is in essence the principle of information causality: the amount of information potentially available to Bob about Alice's data is not higher than the amount of information Alice sends to him (Pawlowski *et al.*, 2009). While information causality is satisfied in both classical and quantum physics, this is not the case in general, if supraquantum correlations are available. Hence information causality can be viewed as a strengthening of the no-signaling principle.

To see how superquantum correlations can violate information causality, suppose that Alice is given two classical bits x_0 and x_1, uniformly distributed. Bob is interested in learning one of these two bits, but Alice does not know which one. To make the task nontrivial, Alice is allowed to send only one bit to Bob. Can they devise a protocol such that Bob can always retrieve the desired bit? In a scenario where Alice and Bob share only classical or quantum correlations, the answer is no. However, if Alice and Bob share a PR box, the task becomes possible (Wolf and Wullschleger, 2005). Alice first inputs $x_0 \oplus x_1$ in her end of the PR box and gets outcome a. She then sends the one-bit message to Bob: $m = a \oplus x_0$. Bob, who is interested in bit x_k of Alice, inputs k on his end of the PR box and gets outcome b. Upon receiving Alice's message m, Bob makes his guess $G = b \oplus m = x_k$. Hence, Bob's guess is always correct.

The principle of information causality allows one to recover part of the boundary between quantum and superquantum correlations (Allcock, Brunner, Pawlowski, and Scarani, 2009; Pawlowski *et al.*, 2009). Notably, any theory that allows for the violation of Tsirelson's bound violates information causality.

Finally, note that an extension of information causality was recently formulated for quantum information (Pitalua-Garcia, 2013).

c. Limitations on multipartite correlations

The principles discussed previously focus on bipartite correlations. A nonlocal game termed *guess your neighbor's input* was introduced by Almeida, Bancal *et al.* (2010), which reveals an intriguing separation between quantum and superquantum correlations in a multipartite context. Consider n distant parties placed on a ring. Each party i is given an input bit x_i according to a joint prior probability distribution $p(x_1 \cdots x_n)$. As the name of the game suggests, each party is then asked to give a guess a_i of his right neighbor's input, i.e., such that $a_i = x_{i+1}$ for all $i = 1, \ldots, n$. Since a high probability of success at this game would lead to signaling, it is not surprising that quantum resources provide no advantage over classical ones, for any distribution of the inputs. However, it turns out that certain no-signaling superquantum correlations outperform classical and quantum strategies for certain distributions of the inputs. Remarkably, some of these games correspond to facet Bell inequalities. Hence "guess your neighbor's input" identifies a portion of the boundary of the quantum set which is of maximal dimension. Moreover, this quite innocuous game has several rather surprising applications, related to generalizations of Gleason's theorem (Acín *et al.*, 2010; Barnum *et al.*, 2010) and to unextendible product basis (Augusiak *et al.*, 2011).

The motivation for many of the results discussed previously is to identify general properties or a set of principles that potentially single out quantum correlations. Gallego *et al.* (2011) showed that any such principles must be genuinely multipartite. More specifically, there exist tripartite superquantum correlations which are local among every possible bipartition (even if many copies of them are available and wirings are performed) (Gallego *et al.*, 2011; Yang *et al.*, 2012). Thus, no bipartite principle can ever rule out these correlations. Such superquantum correlations can nevertheless be ruled out by a novel principle termed "local orthogonality" (Fritz *et al.*, 2013), inspired from the game of "guess your neighbor's input."

2. Physical consequences

a. Macroscopic locality

Loosely speaking, macroscopic locality is a principle requiring that nonlocal correlations admit a classical limit. More specifically, in a Bell test involving a large number of pairs of particles, the statistics of coarse-grained measurements (not resolving discrete particles) should admit an explanation in terms of a local model, i.e., should not violate any Bell inequality (Navascués and Wunderlich, 2010). This is the case in quantum mechanics (Bancal *et al.*, 2008; Navascués and Wunderlich, 2010), but not in general no-signaling theories. Notably, the set of correlations satisfying macroscopic locality can be completely characterized. It corresponds to the set Q_1, the first approximation to the set of quantum correlations in the hierarchy of semidefinite programs (Navascues, Pironio, and Acín, 2007) discussed in Sec. II.C.1.d. This set is, however, strictly larger than the quantum set. Thus, there are superquantum correlations that still satisfy macroscopic locality. Yang *et al.* (2011) showed that analytical quantum Bell inequalities can be derived from macroscopic locality. Finally, note that there exist correlations satisfying macroscopic locality which nevertheless violate information causality (Cavalcanti, Salles, and Scarani, 2010).

b. Uncertainty and information

Wehner, Christandl, and Doherty (2008) showed that one can reformulate any Bell inequality in the language of information, which for projection nonlocal games (see Sec. II.B.4.c) works as follows. For every question x and answer a of Alice, one can write down a string $s_{x,a} = (s_{x,a}^{(1)}, \ldots, s_{x,a}^{(m)})$, where $s_{x,a}^{(y)} = b$ is the answer that Bob must return for question y in order for them to win the game. Written in this way, one can think of the state of Bob's system conditioned on Alice measuring x and obtaining outcome a as an encoding of the string $s_{x,a}$ from which Bob must retrieve entry $s_{x,a}^{(y)}$ correctly. Oppenheim and Wehner

(2010) furthermore showed that for any physical theory uncertainty relations can be understood as imposing limits on how well we can retrieve information from an encoding. This information-theoretic perspective is the essential idea behind the relation between nonlocality and uncertainty found in Oppenheim and Wehner (2010), which holds for any physical theory. It should be noted that the aim of Oppenheim and Wehner (2010) was not to derive limits on nonlocality by appealing to intuitive notions on how we expect information to behave, but rather to link it to another concept already existing within quantum mechanics.

c. Local quantum mechanics

Acín *et al.* (2010) and Barnum *et al.* (2010) showed that the correlations of bipartite systems that can be described locally by quantum mechanics cannot be stronger than quantum correlations. More precisely, if the no-signaling principle holds, and Alice and Bob are locally quantum, then all possible correlations between them admit a quantum mechanical description. However, the situation is different in the multipartite case. There exist tripartite correlations which are locally quantum, which are nevertheless stronger than any quantum correlations (Acín *et al.*, 2010).

D. Nonlocality in generalized probabilistic theories

The idea of investigating the information-theoretic power of nonlocal correlations more general than quantum ones led, following Hardy (2001) and Barrett (2007), to a very active line of research in which information processing has been considered in the broader framework of "general probabilistic theories" (GPT) or "convex-operational" formalism. This framework allows one to define full-fledged theories (i.e., that include notions of states, evolution, measurements, and not only "correlations") in which classical and quantum theories are merely two special cases. Given such a formalism, one can compare and contrast quantum theory with other alternative theoretical models. The hope is to better understand quantum theory and identify in what ways it is special. To date, much work has focused on information processing in GPT, investigating for instance cloning, broadcasting, teleportation, or entanglement swapping. Even if these works connect and partly overlap with many of the issues mentioned above, we do not review this fruitful work here as it does not directly take nonlocality as a starting point. We refer the interested reader instead to Barnum and Wilce (2012) for a short review. In what follows, we only mention work that explicitly considers Bell nonlocality in the context of GPT.

Steeg and Wehner (2009) showed that superstrong random access encodings exist in certain theories that violate the CHSH inequality beyond Tsirelson's bound. A quantum random access code is an encoding of an n-bit string $x = x_1, \ldots, x_n \in \{0,1\}^n$ into a quantum state $\rho_x \in \mathcal{B}(\mathcal{H})$ such that each bit x_j can be retrieved from ρ_x with some probability p_j. Nayak (1999) showed that if the state has dimension at most $\dim(\mathcal{H}) = d$, then the success probabilities are bounded as $\sum_x[1 - h(p_j)] \leq \log d$, where $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$ is the binary entropy. Steeg and Wehner (2009) showed that this inequality can be violated for some theories that allow stronger than quantum correlations, i.e., superstrong random access encodings exist in such theories. In particular, there exists generalized "states" in a Hilbert space of dimension d which effectively contains more than d bits of information.

Janotta *et al.* (2011) showed that there is a connection between the strength of nonlocal correlations in a physical theory and the structure of the state spaces of individual systems. In particular, a class of GPTs is presented that allows one to study the transition between classical, quantum, and superquantum correlations by varying only the local state space. It was shown that the strength of nonlocal correlations depends strongly on the geometry. As the amount of uncertainty in a theory bounds the geometry of the state space, this provides insight into the work of Oppenheim and Wehner (2010). An intriguing consequence of these results is the existence of models that are locally almost indistinguishable from quantum mechanics, but can nevertheless generate maximally nonlocal correlations (Janotta *et al.*, 2011).

VI. MULTIPARTITE NONLOCALITY

In the multipartite case, nonlocality displays a much richer and more complex structure compared to the case of two parties. This makes the study and the characterization of multipartite nonlocal correlations an interesting, but challenging problem. It comes thus to no surprise that our understanding of nonlocality in the multipartite setting is much less advanced than in the bipartite case.

The study of multipartite nonlocality was initiated by the ground-breaking work of Svetlichny (1987). Svetlichny introduced the concept of genuine multipartite nonlocality, derived a Bell-type inequality for testing it, and showed that this strong form of nonlocality occurs in quantum mechanics. Later, in particular, with the advent of quantum information science, the concepts and tools introduced by Svetlichny were further developed.

In this section, we start by defining various notions of multipartite nonlocality (with a particular focus on genuine multipartite nonlocality) and discuss the detection of multipartite nonlocality. Next, we discuss the notion of monogamy of nonlocality, which limits nonlocality between different subsets of parties. Finally, we discuss the nonlocality of multipartite quantum systems.

A. Defining multipartite nonlocality

The notion of Bell nonlocality that we introduced in Secs. I and II in the case of two separated observers readily extends to three or more observers. For simplicity, we consider in this section the case of three separated observers Alice, Bob, and Charlie. Their measurement settings are denoted x, y, z and their outputs by a, b, c, respectively. The experiment is thus characterized by the joint probability distribution $p(abc|xyz)$. We say that these correlations are local if they can be written in the form

$$p(abc|xyz) = \int d\lambda q(\lambda) p_\lambda(a|x) p_\lambda(b|y) p_\lambda(c|z), \tag{71}$$

where λ is a shared local random variable and $\int d\lambda q(\lambda) = 1$, and that they are nonlocal otherwise. This represents the natural generalization of Bell's locality condition (3) to the multipartite case. The set of correlations that can be written in Eq. (71) is denoted $\mathcal{L}$.

However, in the multipartite case, there exist several possible refinements of this notion of nonlocality. For instance, consider a joint distribution of the form $p(abc|xyz) = p(ab|xy) \times p(c|z)$, i.e., Charles is uncorrelated to Alice and Bob. These correlations can clearly violate the locality condition (71) if $p(ab|xy)$ is nonlocal, although no nonlocality at all is exhibited between Alice, Bob, and Charles. In other words, such correlations exibibit only bipartite nonlocality. In contrast, one can consider a situation where all three parties are nonlocally correlated. This is referred to as genuine multipartite nonlocality, which represents the strongest form of multipartite nonlocality. The main purpose of this section is to discuss the problem of defining formally, in the spirit of Bell's definition, this concept of genuine multipartite nonlocality.

1. Genuine multipartite nonlocality *à la* Svetlichny

The first definition of genuine multipartite nonlocality was proposed by Svetlichny (1987). To describe it suppose that $p(abc|xyz)$ can be written in the form

$$\begin{aligned} p(abc|xyz) = &\int d\lambda q(\lambda) p_\lambda(ab|xy) p_\lambda(c|z) \\ &+ \int d\mu q(\mu) p_\mu(bc|yz) p_\mu(a|x) \\ &+ \int d\nu q(\nu) p_\nu(ac|xz) p_\nu(b|y), \end{aligned} \tag{72}$$

where $\int d\lambda q(\lambda) + \int d\mu q(\mu) + \int d\nu q(\nu) = 1$. This represents a convex combination of three terms, where in each term at most two of the parties are nonlocally correlated. For instance, the term $\int d\lambda q(\lambda) p_\lambda(ab|xy) p_\lambda(c|z)$ represents correlations where Charles is locally correlated (through the hidden variable λ) with the joint system of Alice and Bob. The correlations between Alice and Bob, however, are arbitrary, and, in particular, can be nonlocal. Operationally, we can think of such correlations as describing a situation where Alice and Bob are free to share arbitrary nonlocal resources between themselves or are able to communicate freely, while they are prevented to do so with Charles. The convex combination (72) thus represents a situation where only two parties share a nonlocal resource or communicate in any measurement run. We say that they are two-way nonlocal. On the other hand, if $p(abc|xyz)$ cannot be written in the above form, then necessarily the three parties Alice, Bob, and Charles must share some common nonlocal resource. We then say that they are three-way nonlocal or genuinely tripartite nonlocal. Detecting such a form of multipartite nonlocality is an important issue. As for detecting standard nonlocality, it is possible to write down Bell inequalities, the violation of which guarantee that the correlations are genuinely multipartite (see Sec. VI.B).

Operationally, we define local correlations as those that can be generated by separated classical observers that have access to share randomness but who cannot communicate, two-way correlations as those where arbitrary communication is allowed between two parties, and three-way as those where arbitrary communication is allowed between all parties. One can also consider more refined definitions based on more general communication patterns (particularly in the multipartite case with a large number of parties). For instance, we can consider the case where Alice is allowed to communicate to Bob and to Charles, while Bob and Charles cannot communicate to anyone. Such generalizations of Svetlichny's approach were considered by Jones, Linden, and Massar (2005) and Bancal *et al.* (2009).

While Svetlichny's notion of genuine multipartite nonlocality is often used in the literature, it has certain drawbacks discussed next.

2. Beyond Svetlichny's model

In Svetlichny's definition of genuine multipartite nonlocality, parties that are allowed to share nonlocal resources can display arbitrary correlations. In particular, this includes signaling probability distributions. For instance, considering again the above tripartite example, the bipartite probability distributions, e.g., $p_\lambda(ab|xy)$, entering decomposition (72) are unconstrained, apart from normalization. In particular, this means that we have not imposed the no-signaling constraints:

$$p_\lambda(a|xy) = p_\lambda(a|xy') \quad \forall\ a, x, y, y', \tag{73}$$

$$p_\lambda(b|xy) = p_\lambda(b|x'y) \quad \forall\ b, x, x', y, \tag{74}$$

where $p_\lambda(a|xy) = \sum_b p_\lambda(ab|xy)$ is Alice's marginal probability distribution, and similarly for Bob. These conditions guarantee that, even given the knowledge of λ, Alice cannot send a message to Bob by choosing her measurement setting, and vice versa. If at least one of the above constraints is not satisfied, then this allows for signaling. Signaling from Alice to Bob occurs when Eq. (74) is not satisfied. Similarly, signaling from Bob to Alice occurs when Eq. (73) is not satisfied.

Such signaling terms in Svetlichny's definition (72) are inconsistent from a physical perspective (they lead to grandfather-type paradoxes) as well as from an operational point of view (Gallego, Würflinger *et al.*, 2012; Barrett, Pironio *et al.*, 2013). To give a rough idea of why this is so [see Gallego, Würflinger *et al.* (2012) and Barrett, Pironio *et al.* (2013) for more details], consider, for instance, Svetlichny's definition from the perspective of classical simulations of quantum correlations in terms of shared random data and communication. The decomposition (72) corresponds to simulation models where all parties receive their measurement setting at the same time, then there are several rounds of communication between only two of the parties, say Alice and Bob, and finally, all parties produce a measurement outcome. During the communication step, Alice and Bob can establish arbitrary correlations in Svetlichny's model, in particular, they can violate the two above no-signaling conditions. But consider now a slightly different simulation model where measurements are given to the parties in a sequence that is arbitrary and not fixed in advance. Upon receiving a measurement setting, a party must produce an output immediately, as happens when measuring a real quantum state. But then if Alice received her measurement choice before Bob, she must determine her output without having received any communication from Bob and thus Eq. (73), imposing no signaling from Bob to Alice, cannot be violated. If, in another round, it is Bob that receives his measurement before Alice, then it is Eq. (74), imposing no signaling from Alice to Bob, that cannot be violated.

To address such shortcomings of Sveltichny's definition, there are two alternatives. The most immediate one is to require that all bipartite correlations, e.g., $p_\lambda(ab|xy)$, appearing in the decomposition (72), satisfy the no-signaling conditions (Almeida, Cavalcanti *et al.*, 2010; Barrett, Pironio *et al.*, 2013). The set of correlations that can be written that admits such a decomposition is denoted $\mathcal{S}_{2|1}^{\rm ns}$. Correlations that cannot be written in this form can then be considered to be genuinely tripartite nonlocal.

However, there is a more interesting definition of genuine multipartite nonlocality based on time ordering. Basically, one now requires that in the decomposition (72), all bipartite correlations are time ordered. Specifically, the set $\mathcal{S}_{2|1}^{\rm to}$ of two-way time-ordered correlations contains all distributions that can be written in the form

$$\begin{aligned} p(abc|xyz) = & \int d\lambda q(\lambda) p_\lambda^{T_{AB}}(ab|xy) p_\lambda(c|z) \\ & + \int d\mu q(\mu) p_\mu^{T_{AC}}(ac|xz) p_\mu(b|y) \\ & + \int d\nu q(\nu) p_\nu^{T_{BC}}(bc|yz) p_\nu(a|x), \end{aligned} \tag{75}$$

where $p_\lambda^{T_{AB}}(ab|xy)$ denotes a probability distribution that is time-order dependent: when Alice measures before Bob, we have that $p_\lambda^{T_{AB}}(ab|xy) = p_\lambda^{A<B}(ab|xy)$; when Bob measures before Alice, we have that $p_\lambda^{T_{AB}}(ab|xy) = p_\lambda^{B<A}(ab|xy)$. It is then required that $p_\lambda^{A<B}(ab|xy)$ and $p_\lambda^{B<A}(ab|xy)$ are both (at most) one-way signaling; $p_\lambda^{A<B}(ab|xy)$ is such that only Alice can signal to Bob, while $p_\lambda^{B<A}(ab|xy)$ is such that only Bob can signal to Alice. These requirements avoid the problems discussed above. According to this definition, a probability distribution $p(abc|xyz)$ that cannot be written in the form (75) is then said to be genuine multipartite nonlocal.

All three definitions of genuine multipartite nonlocality introduced in this section are nonequivalent (Gallego, Würflinger *et al.*, 2012; Barrett, Pironio *et al.*, 2013) and we have the strict relations

$$\mathcal{L} \subset \mathcal{S}_{2|1}^{\rm ns} \subset \mathcal{S}_{2|1}^{\rm to} \subset \mathcal{S}_{2|1}^{\rm Svet}. \tag{76}$$

Thus while violation of Svetlichny's decomposition (72) always guarantees that the correlations $p(abc|xyz)$ are genuinely tripartite nonlocal, there exist some correlations whose tripartite character only manifests itself when considering the weaker definitions $\mathcal{S}_{2|1}^{\rm ns}$ and $\mathcal{S}_{2|1}^{\rm to}$.

B. Detecting genuine multipartite nonlocality

After having defined the concept of genuine multipartite nonlocality, we now briefly discuss how one can detect it through the violation of appropriate Bell inequalities.

1. Svetlichny's inequality

The first inequality for detecting genuine multipartite nonlocality was introduced by Svetlichny (1987). Focusing on a tripartite system, Svetlichny derived a Bell-type inequality which holds for any distribution of Eq. (72). Thus a violation of such inequality implies the presence of genuine tripartite nonlocality. It should be noted that this in turn implies the presence of genuine tripartite entanglement.

We focus now on the case where each party j performs one out of two possible measurements denoted x_j and x_j'. All measurements are dichotomic, hence their results are denoted by $a_j = \pm 1$ and $a_j' = \pm 1$. Svetlichny then proved that the inequality

$$\begin{aligned} S_3 = & a_1 a_2 a_3' + a_1 a_2' a_3 + a_1' a_2 a_3 - a_1' a_2' a_3' + a_1' a_2' a_3 + a_1' a_2 a_3' \\ & + a_1 a_2' a_3' - a_1 a_2 a_3 \le 4 \end{aligned} \tag{77}$$

holds for any probability distribution of Eq. (72). Note that the above polynomial should be understood as a sum of expectation values; for instance, $a_1 a_2 a_3'$ stands for the expectation value of the product of the measurement outcomes when the measurements are x_1, x_2, and x_3'.

To get more intuition about Svetlichny's inequality, and to prove that its violation implies the presence of genuine multipartite nonlocality, we follow the simple approach of Bancal, Brunner *et al.* (2011). We first rewrite the inequality as follows:

$$S_3 = S a_3' + S' a_3 \le 4, \tag{78}$$

where $S = a_1 a_2 + a_1 a_2' + a_1' a_2 - a_1' a_2'$ is the CHSH expression, and $S' = a_1' a_2' + a_1' a_2 + a_1 a_2' - a_1 a_2$ is one of its equivalent forms obtained by permuting primed and nonprimed measurements. Now observe that it is the input setting of Charlie that defines which version of the CHSH game Alice and Bob are playing. When Charlie gets the input x_3', then Alice and Bob play the standard CHSH game; when Charlie gets the input x_3, Alice and Bob play its symmetry. Hence it follows that $S_3 \le 4$ holds for any bipartition model of Eq. (72). Consider the bipartition $A|BC$. Bob knows which version of the CHSH game he is supposed to play with Alice, since he is together with Charlie. However, CHSH being a nonlocal game, Alice and Bob cannot achieve better than the local bound (i.e., $S = 2$ or $S' = 2$) as they are separated. Thus it follows that $S_3 \le 4$ for the bipartition $A|BC$. Note that the same reasoning holds for the bipartition $B|AC$. Finally, since the polynomial is symmetric under permutation of the parties, it follows that $S_3 \le 4$ for all bipartitions. The inequality (77) detects the genuine multipartite nonlocality of important classes of quantum states, such as GHZ and W (see Sec. VI.D).

2. Generalizations to more parties, measurements, and dimensions

Svetlichny's inequality has been generalized to a scenario featuring an arbitrary number of parties n (Collins, Gisin, Popescu *et al.*, 2002; Seevinck and Svetlichny, 2002). By repeating the procedure which allowed us to build Svetlichny's inequality from CHSH [see Eq. (78)] we get

$$S_n = S_{n-1}a'_n + S'_{n-1}a_n \le 2^{n-1}, \tag{79}$$

where S'_{n-1} is obtained from S_{n-1} by applying the mapping $a_1 \to a'_1$ and $a'_1 \to a_1$ (Bancal, Brunner *et al.*, 2011). Note also that generalizations to the most general scenario, featuring an arbitrary number of parties, measurements, and systems of arbitrary dimensions, were derived by Bancal, Brunner *et al.* (2011); see also Aolita *et al.* (2012a).

Finally, note that Bell inequalities detecting notions of genuine multipartite nonlocality more refined than that of Sveltichny (see Sec. VI.A.2) were presented by Barrett, Pironio *et al.* (2013).

C. Monogamy

The monogamy of nonlocal correlations is nicely illustrated by considering the CHSH inequality in a tripartite scenario. Let Alice, Bob, and Charlie have two possible dichotomic measurements, represented by observables A_x, B_y, and C_z with $x, y, z \in \{0, 1\}$. We can now evaluate the CHSH expression for Alice to Bob and Alice to Charlie. Denote by $\mathcal{B}_{AB}$ and $\mathcal{B}_{AC}$ the corresponding Bell operators for the CHSH inequality as defined in Sec. II.C.1.a. It is important to note that Alice's measurements are the same for both inequalities. Scarani and Gisin (2001) showed that, for any three-qubit state shared by the parties, if $\langle\mathcal{B}_{AB}\rangle > 2$ then $\langle\mathcal{B}_{AC}\rangle \le 2$. That is, if the statistics of Alice and Bob violate the CHSH inequality, then the statistics of Alice and Charlie will not. More generally, Toner and Verstraete (2006) showed that for an arbitrary quantum state shared by the three parties, we have

$$\langle\mathcal{B}_{AB}\rangle^2 + \langle\mathcal{B}_{AC}\rangle^2 \le 8. \tag{80}$$

Note again that if Alice and Bob violate their CHSH inequality, then Alice and Charlie do not. Moreover, if Alice and Bob observe maximal CHSH violation (i.e., a CHSH value of $2\sqrt{2}$), then $\langle\mathcal{B}_{AB}\rangle^2 = 8$ and hence by Eq. (80) the data of A and C are uncorrelated. Monogamy of correlations, however, is not specific to the CHSH inequality but applies to essentially all bipartite Bell inequalities. In the language of games (Sec. II.B.4), this has been used by Kempe *et al.* (2008) and Ito and Vidick (2012) to “immunize” a nonlocal game against the use of entanglement.

It is interesting to note that even no-signaling correlations are monogamous (Barrett, Linden *et al.*, 2005; Masanes, Acín, and Gisin, 2006; Pawlowski and Brukner, 2009) (see Sec. IV.C). In particular, Toner (2009) showed that $|\langle\mathcal{B}_{AB}\rangle| + |\langle\mathcal{B}_{AC}\rangle| \le 4$, which is tight if we consider no-signaling correlations.

The fact that QKD protocols based on nonlocality can be proven secure (see Sec. IV.C) can also be understood as a consequence of the monogamy of quantum correlations among Alice, Bob, and Eve, and was indeed one of the factors motivating its study.

Underlying the monogamy of correlations in the quantum setting is an inherent monogamy of entanglement (Terhal, 2004). Understanding the exact relation between both forms of monogamy is an interesting open problem.

D. Nonlocality of multipartite quantum states

1. Multipartite nonlocality versus multipartite entanglement

In this section we discuss the relation between quantum nonlocality and entanglement in the multipartite setting. Similarly to the bipartite case, the two concepts are intimately related, although precisely understanding the link is a challenging problem.

Note that all pure entangled n-partite states are nonlocal (Popescu and Rohrlich, 1992). That is, their measurement statistics cannot be decomposed in Eq. (71). This follows from the fact that it is always possible for $n-2$ parties to project (via a local projection) the remaining two parties in a pure entangled state. Since the latter is nonlocal, the result follows. It should be stressed that this result does not rely on any form of postselection.

In entanglement theory, a concept of particular importance is that of genuine multipartite entanglement. A quantum state features genuine multipartite entanglement when it cannot be decomposed as a convex combination of biseparable states (states which are separable on at least one bipartition of the parties). Indeed, this notion is somehow analogous to that of genuine multipartite nonlocality, and it is not surprising that both are related. In particular, genuine multipartite quantum nonlocality can be obtained only if measurements on a genuine multipartite entangled state are made. Thus, the presence of genuine multipartite nonlocality witnesses the presence of genuine multipartite entanglement. Importantly this is achieved in a device-independent way; that is, genuine multipartite entanglement is here certified without placing any assumptions about the devices used in the experiment, contrary to usual methods such as entanglement witnesses and quantum tomography. Note that it is possible to design even better device-independent techniques for witnessing genuine multipartite, the violation of which does not imply the presence of genuine multipartite nonlocality (Bancal, Gisin *et al.*, 2011) [see also Nagata, Koashi, and Imoto (2002) and Uffink (2002)].

It is, however, not known whether all pure genuine multipartite entangled states are genuine multipartite nonlocal. It has been shown (Almeida, Cavalcanti *et al.*, 2010) that all connected graph states are fully genuine nonlocal, in the no-signaling approach discussed in Sec. VI.A.2. Moreover, it was also shown that the tangle, a specific measure of multipartite entanglement, is closely related to the violation of Svetlichny's inequality (Ghose *et al.*, 2009; Ajoy and Rungta, 2010). In particular, from this connection it can be shown that there exist pure entangled states in the GHZ class which do not violate Svetlichny's inequality.

Finally, it is worth noting that the connection between genuine multipartite entanglement and nonlocality may depend on which definition of genuine multipartite nonlocality is used. Using the definition based on time ordering [see Eq. (75)], numerical evidence suggests that all pure genuine tripartite entangled qubit states are genuine tripartite nonlocal (Barrett, Pironio *et al.*, 2013). More recently, Yu and Oh (2013) proved that all pure genuinely tripartite entangled states are tripartite nonlocal with respect to the definition based on no signaling (see Sec. VI.A.2). Tripartite nonlocality of Gaussian states was discussed by Adesso and Piano (2014).

2. Greenberger-Horne-Zeilinger states

GHZ states are today arguably the most studied, and possibly the best understood, multipartite quantum states from the point of view of entanglement and nonlocality. GHZ states display one of the most striking forms of nonlocality in the context of the Mermin-GHZ paradox (see Sec. II.D). By performing local measurements on a tripartite GHZ state

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle), \tag{81}$$

one obtains correlations which are maximally nonlocal, since the predictions of quantum mechanics are here in full contradiction with those of local models. Interestingly, it turns out however that these particular GHZ correlations do not feature genuine multipartite nonlocality (Cereceda, 2002; Mitchell, Popescu, and Roberts, 2004), as they can be reproduced by a biseparable model of Eq. (72).

It is nevertheless possible to generate genuine multipartite nonlocal correlations from local measurements on a tripartite GHZ state (Svetlichny, 1987). In particular, one can get violation of Svetlichny's inequality (77) of $S_3 = 4\sqrt{2} > 4$, which turns out to be the largest possible violation in quantum mechanics (Mitchell, Popescu, and Roberts, 2004). This violation can be intuitively understood by considering again Eq. (78) of Svetlichny's inequality. Since it is Charlie's measurement setting that dictates which version of the CHSH game Alice and Bob are playing, the best strategy for Charlie consists of remotely preparing (by performing a measurement on her qubit) a state for Alice and Bob that is optimal for the violation of the corresponding CHSH game (Bancal, Brunner *et al.*, 2011).

The nonlocal correlations of generalized GHZ states, of the form

$$|\mathrm{GHZ}_n^d\rangle = \frac{1}{\sqrt{d}}\sum_{j=0}^{d-1}|j\rangle^{\otimes n} \tag{82}$$

featuring n parties and systems of local dimension d, have also been investigated. First, analogs of the Mermin-GHZ paradox were reported (Zukowski and Kaszlikowski, 1999; Cerf, Massar, and Pironio, 2002) for certain combinations of n and d. More recently, a general construction for arbitrary n and d was given by Ryu *et al.* (2013). A Mermin-GHZ–type paradox was also presented for the case of continuous variable systems (Massar and Pironio, 2001; van Loock and Braunstein, 2001).

The genuine multipartite nonlocality of generalized GHZ states has also been investigated. It was first shown that all qubit GHZ states (i.e., $|\mathrm{GHZ}_n^2\rangle$) violate the generalization (79) of Svetlichny's inequality for an arbitrary number of parties and hence display genuine multipartite nonlocal correlations (Collins, Gisin, Popescu *et al.*, 2002; Seevinck and Svetlichny, 2002). Recently, it was shown that the correlations of any state of Eq. (82) are fully genuinely multipartite nonlocal, as well as monogamous and locally random (Aolita *et al.*, 2012b). The robustness of GHZ nonlocality against local noise was investigated by Laskowski *et al.* (2010) and Chaves *et al.* (2012).

3. Graph states

Graph states (Hein, Eisert, and Briegel, 2004) form an important family of multipartite quantum states (including GHZ and cluster states) useful for applications in quantum information science. In particular, all code word states used in the standard quantum error correcting codes correspond to graph states, and one-way quantum computation uses graph states as a resource. Here we discuss the nonlocality of graph states (for GHZ states see Sec. VI.D.2).

Graph states are defined as follows. Let G be a graph featuring n vertices and a certain number of edges connecting them. For each vertex i, we define neigh(i) as the neighborhood of i, which represents the set of vertices which are connected to i by an edge. Next, one associates with each vertex i a stabilizing operator

$$g_i = X_i \bigotimes_{j\in\mathrm{neigh}(i)} Z_j, \tag{83}$$

where X_i, Y_i, and Z_i denote the Pauli matrices applied to qubit i. The graph state $|G\rangle$ associated with graph G is then the unique common eigenvector to all stabilizing operators g_i, i.e., $g_i|G\rangle = |G\rangle$ for all $i \in \{1,\ldots,n\}$. From a physical point of view, the graph G describes all the perfect correlations of the state, since $\langle G|g_i|G\rangle = 1$ for all $i \in \{1,\ldots,n\}$. By considering the set of operators that can be obtained from products of stabilizer operators (83), one obtains a commutative group featuring 2^n elements. This is the stabilizer group, defined as

$$S(G) = \{s_j\}_{j=1,\ldots,2^n}, \quad \text{where} \quad s_j = \prod_{i\in I_j(G)} s_i, \tag{84}$$

where $I_j(G)$ denotes any of the 2^n subsets of the vertices of the graph G.

Interestingly, this fundamental structure of graph states underpins a strong form of nonlocality (Gühne *et al.*, 2005; Scarani *et al.*, 2005). It turns out that all graph states feature nonlocal correlations (Gühne *et al.*, 2005). In order to prove this, the main idea consists of constructing Bell inequalities by adding all elements of the stabilizer group $S(G)$. Thus we consider the operator

$$\mathcal{B}(G) = \sum_{i=1}^{2^n} s_i = \sum_{i=1}^{2^n} \bigotimes_{j=1}^{n} O_j^i, \tag{85}$$

where operators $O_j^i \in \{\mathbb{1}, X_j, Y_j, Z_j\}$ are from the Pauli basis.

It is then possible to define a Bell inequality based on the above Bell operator and to compute its local bound

$$L(G) = \max_{\text{LHV}} |\langle \mathcal{B} \rangle|. \tag{86}$$

While the graph state $|G\rangle$ reaches the value of 2^n for such Bell inequality (indeed $s_i|G\rangle = |G\rangle$ for all $i \in \{1, \ldots, 2^n\}$), it turns out that $L(G) < 2^n$ for any graph G. Thus, for all graph states it is possible to construct a Bell inequality, which the state then maximally violates. Indeed this demonstrates that nonlocality is a generic feature of all graph states. Moreover, for certain families of graph states, basically states based on tree graphs (featuring no closed loops), the violation of the Bell inequality grows exponentially with the number of vertices (Gühne *et al.*, 2005; Toth, Gühne, and Briegel, 2006).

While the generality of the above approach is remarkable, it is possible for certain important classes of graph states, in particular, for cluster states, to derive stronger proofs of nonlocality (Scarani *et al.*, 2005). Cluster states form a subclass of graph states based on square lattice graphs. For simplicity and clarity we discuss here the case of a four qubit cluster state on a one-dimensional lattice,[17] which is locally equivalent to

$$|\text{Cl}_4\rangle = \tfrac{1}{2}(|0000\rangle + |0011\rangle + |1100\rangle - |1111\rangle). \tag{87}$$

The state $|\text{Cl}_4\rangle$ is defined by the stabilizer relations

$$\begin{aligned} X Z \mathbb{1}\, \mathbb{1} &= 1 \quad (E1), \\ Z X Z \mathbb{1} &= 1 \quad (E2), \\ \mathbb{1}\, Z X Z &= 1 \quad (E3), \\ \mathbb{1}\, \mathbb{1}\, Z X &= 1 \quad (E4). \end{aligned} \tag{88}$$

By multiplying certain of these four relations, we get

$$\begin{aligned} (E1) \times (E3)\colon\ X \mathbb{1} X Z &= 1, \\ (E2) \times (E3)\colon\ Z Y Y Z &= 1, \\ (E1) \times (E3) \times (E4)\colon\ X \mathbb{1} Y Y &= 1, \\ (E2) \times (E3) \times (E4)\colon\ Z Y X Y &= -1. \end{aligned} \tag{89}$$

Note that here we used the Pauli algebra, which explains the emergence of a minus sign in the last relation above. It can be readily checked that for any deterministic local model, i.e., attributing ±1 values to each measurement (X, Y, Z), it is impossible to satisfy simultaneously all four relations above; at least one of them will not hold. Check, for instance, that by simply multiplying (using standard multiplication) the first three relations in Eq. (89), one obtains the fourth relation. Therefore, we obtain a perfect contradiction between quantum and classical predictions, in the spirit of the GHZ paradox.

Similar to the Mermin-GHZ case (see Sec. II.D), this logical contradiction can be rephrased as a Bell inequality. By considering the four relations in Eq. (89), we get

$$|a_1 a_3' a_4 + a_1 a_3 a_4' + a_1' a_2 a_3 a_4 - a_1' a_2 a_3' a_4'| \le 2. \tag{90}$$

Notice that by grouping the first two parties one obtains the Mermin inequality (38). Performing measurements on the state $|\text{Cl}_4\rangle$, the algebraic maximum of 4 can be obtained for the left-hand side of Eq. (90). Finally note that an interesting feature of the Bell inequality (90) is that it cannot be violated by the four-qubit GHZ state. Thus the inequality is a strong entanglement witness[18] for the cluster $|\text{Cl}_4\rangle$. The above construction can be generalized to cluster states of an arbitrary number of qubits and of arbitrary local dimension, as well as to certain classes of graph states (Scarani *et al.*, 2005).

The nonlocality of graph states can also be revealed by using sets of local measurements that are not stabilizers (Gühne and Cabello, 2008). An interesting issue is to understand whether there exists a link between the nonlocality of cluster states and the computational power they offer. Although such a connection has not been clearly established yet, progress has been made (Hoban *et al.*, 2011). Another important class of graph states is code word states, the nonlocality of which has been discussed by DiVincenzo and Peres (1997).

Finally, it is important to note that all nonlocality proofs discussed in this section concern Bell nonlocality. Unfortunately much less is known concerning the genuine multipartite nonlocality of graph states. It is, however, known that all states based on connected graphs (graphs in which any two vertices are connected, although not necessarily in a direct manner) display fully genuine multipartite nonlocality (Almeida, Cavalcanti *et al.*, 2010).

4. Nonlocality of other multipartite quantum states

In the multipartite case, entanglement displays a rich structure, with many inequivalent classes of states. Although we know that all multipartite entangled pure states are nonlocal, very little is known beyond the case of graph states.

An important class of multipartite entangled states are Dicke states, that is, states with a fixed number of excitations and symmetric under permutation of the parties, which are central in the context of the interaction of light and matter (Dicke, 1954). The symmetric state of n particles with a single excitation, known as the W state, reads

$$|W_n\rangle = \frac{1}{\sqrt{n}}(|0\cdots01\rangle + \cdots + |10\cdots0\rangle). \tag{91}$$

Such states are relevant to the description of various physical systems, such as quantum memories. One possibility for detecting the nonlocality of W states consists of having $n-2$ parties performing a measurement in the logical basis $\{|0\rangle, |1\rangle\}$. When all project onto the $|0\rangle$ eigenstate, which happens with a fairly large probability (increasing with n), they prepare for the remaining two parties a (two-qubit) Bell

[17] Note that for two and three qubits, the 1D cluster state is equivalent to a Bell state and to a GHZ state, respectively.

[18] Notice however that, as written, the inequality (90) can also be maximally violated by a three-partite entangled state, since party 2 has only one setting. This deficiency can be overcome by symmetrizing the inequality over the parties.

state, on which the CHSH inequality can then be tested and violated (Sen(De) *et al.*, 2003). Another manifestation of the nonlocality of the Dicke states is based on their robustness to losses. Indeed when $k \ll n$ particles are lost, the state remains basically unchanged. For instance, for W states one has that $\text{tr}_k(|W_n\rangle\langle W_n|) \approx |W_{n-k}\rangle\langle W_{n-k}|$, where tr_k denotes the partial trace on the k particles which have been lost. The W state thus has a high "persistency" of nonlocality (Brunner and Vértesi, 2012), in the sense that a large number of particles must be lost in order to destroy all nonlocal correlations. This appears to be a generic feature of Dicke states.

Another relevant problem is whether one can distinguish different classes of multipartite entangled states via their nonlocal correlations. This can be done using judiciously designed Bell inequalities (Schmidt *et al.*, 2008; Brunner, Sharam, and Vértesi, 2012). For instance, the resistance to losses of W states can be exploited to distinguish their nonlocal correlations from those of GHZ states.

The nonlocal properties of more general classes of states have been discussed. The nonlocality of symmetric qubit states was first investigated by Wang and Markham (2012). Exploiting the Majorana representation, they derived Hardy-type nonlocality proofs (see Sec. II.E) for arbitrary symmetric pure entangled states. Also, the resistance to noise has been evaluated numerically for a large class of multipartite quantum states (Gruca *et al.*, 2010).

The relation between entanglement distillability and nonlocality was also investigated in the multipartite case. Dür (2001) and Augusiak and Horodecki (2006) showed that a multiqubit bound entangled state can violate the Mermin inequalities. However, the states considered in these works become distillable when several parties can group. In fact, Acín (2001) showed that the violation of the Mermin inequalities implies that distillability between groups of parties. More recently, Vértesi and Brunner (2012) presented an example of a fully bound entangled state (for which no entanglement can be distilled even when parties are allowed to group) which violates a Bell inequality. This shows that nonlocality does not imply the presence of distillable entanglement and refutes the Peres conjecture in the multipartite case (see Sec. III.A.5).

VII. EXPERIMENTAL ASPECTS

Violations of Bell inequalities have been observed experimentally in a variety of physical systems, giving strong evidence that nature is nonlocal. Nevertheless, all experiments suffer from various loopholes, opened by technical imperfections, which makes it in principle possible for a local model to reproduce the experimental data, even if in a highly contrived way. In recent years, an intense research effort has been devoted to the design and realization of a loophole-free Bell experiment, which should be within experimental reach in the near future. Besides its fundamental interest, closing some of these loopholes (in particular, the detection loophole) is important from the perspective of practical applications of nonlocality such as device-independent quantum information processing. Indeed, while the idea that nature is exploiting such loopholes to fake nonlocal correlations may sound conspiratorial, the perspective is entirely different when we consider the possibility that they are exploited by an adversary to break a cryptography protocol.

In this section we review the main achievements and challenges in this area. For a more exhaustive discussion on Bell experiments, we refer the interested reader to recent reviews (Genovese, 2005; Pan *et al.*, 2012).

A. Bell experiments

1. Photons

Tremendous experimental progress in quantum optics during the 1960s opened the door to possible tests of quantum nonlocality in the laboratory. First, using atomic cascades, it became possible to create pairs of photons entangled in polarization. Second, the polarization of single photons could be measured using polarizers and photomultipliers. Only 3 years after the proposal of CHSH (Clauser *et al.*, 1969), Freedman and Clauser (1972) reported the first conclusive test of quantum nonlocality, demonstrating a violation of the CHSH Bell inequality by 6 standard deviations.

During the following years, other experiments (Fry and Thompson, 1976; Aspect, Grangier, and Roger, 1981, 1982b) were performed, giving further confirmation of the predictions of quantum mechanics. However, the main drawback of all these experiments was that they were performed with static setups in which the polarization analyzers were held fixed, so that all four correlation terms had to be estimated one after the other. In principle, the detector on one side could have been aware of the measurement setting chosen on the other side, thus opening a loophole[19] (see Sec. VII.B.2).

Crucial progress came in 1982, when Aspect, Dalibard, and Roger (1982) performed the first Bell experiment with time-varying polarization analyzers. The settings were changed during the flight of the particle and the change of orientation on one side and the detection event on the other side were separated by a spacelike interval, thus closing the locality loophole (see Sec. VII.B.2). It should be noted though that the choice of measurement settings was based on acousto-optical switches, and thus governed by a quasiperiodic process rather than a truly random one. Nevertheless the two switches on the two sides were driven by different generators at different frequencies and it is very natural to assume that their functioning was uncorrelated. The experimental data turned out to be in excellent agreement with quantum predictions and led to a violation of the CHSH inequality by 5 standard deviations.

The advent of quantum information in the 1990s triggered renewed interest in experimental tests of quantum nonlocality. In 1998, violation of Bell inequalities with photons separated by more than 10 km was reported (Tittel *et al.*, 1998). That same year, another experiment demonstrated violation of Bell inequalities with the locality loophole closed and using a quantum random number generator to generate the measurement settings (Weihs *et al.*, 1998). In turn, both of these

[19] Moreover, by performing Bell tests with all correlation terms measured successively with the settings held fixed, it is not unusual to observe experimentally, because of slow drifts in the setup, apparent violations of Bell inequalities above Tsirelson's bound or even violation of the no-signaling conditions (7) (Afzelius, 2011).

experiments were adapted to implement quantum key distribution based on nonlocal quantum correlations (Jennewein *et al.*, 2000; Tittel *et al.*, 2000), following Ekert's idea (see Sec. IV.C).

Demonstrations of quantum nonlocality in photonic systems have been reported using various types of encoding apart from polarization. Bell inequality violations based on phase and momentum of photons have been achieved (Rarity and Tapster, 1990). Franson (1989) proposed a test of quantum nonlocality based on the energy-time uncertainty principle. This encoding, used for instance in the experiment of Tittel *et al.* (1998), led to the concept of time-bin encoding (Tittel *et al.*, 1999) which turned out to be particularly well suited for the distribution of entanglement on long distances. Bell inequality violation has also been demonstrated using photons entangled in orbital angular momentum (Mair *et al.*, 2001). An important advantage of both time-bin and orbital angular momentum encodings is that they allow for the realization of higher-dimensional quantum systems, whereas polarization is limited to qubits. Nonlocal correlations of qutrits have been reported with time bins (Thew *et al.*, 2004), while Bell violation with orbital angular momentum has recently been reported using systems of dimensions up to 11 (Dada *et al.*, 2011). Another possibility for creating higher-dimensional entanglement consists of generating pairs of photons entangled in several degrees of freedom, so called hyper-entangled photons (Kwiat, 1997). Bell experiments have been performed with such systems (Barreiro *et al.*, 2005; Ceccarelli *et al.*, 2009), combining polarization, spatial mode, and energy-time degrees of freedom. Finally, continuous variable systems have also been investigated. In particular, Babichev, Appel, and Lvovsky (2004) demonstrated the nonlocality of a single photon using homodyne measurements.

Other interesting aspects of quantum nonlocality have been investigated experimentally. Notably, the phenomenon of hidden nonlocality (see Sec. III.A.3) was observed by Kwiat *et al.* (2001), and Hardy's paradox (see Sec. II.E) was realized by White *et al.* (1999). It is also worth mentioning the experiment of Fedrizzi *et al.* (2009) which demonstrated violation of the CHSH inequality over a free-space link of 144 km.

Multipartite quantum nonlocality has also been demonstrated experimentally. Bell inequality violations were achieved with three photons, generating both GHZ (Pan *et al.*, 2000) and W (Eibl *et al.*, 2004) states, and with four-photon GHZ states (Eibl *et al.*, 2003; Zhao *et al.*, 2003) and cluster states (Walther *et al.*, 2005). Genuine multipartite nonlocality of three-photon GHZ states was demonstrated by Lavoie, Kaltenbaek, and Resch (2009).

Note also that nowadays Bell experiments can even be envisaged for pedagogical purposes (Dehlinger and Mitchell, 2002). In particular, ready-to-use setups are available commercially (Qutools, 2005), which are fully operational, even from the perspective of research (Pomarico, Bancal *et al.*, 2011).

Finally, it is important to keep in mind that all the Bell experiments discussed above are plagued by the detection loophole (see Sec. VII.B.1). This is because the photon detection efficiency in these experiments is low (typically 10%–20%) which makes it possible, in principle, for a local model to reproduce the raw data. It is only under the assumption that the probability of detecting or nondetecting a photon is independent of the choice of measurement (the so-called "fair-sampling" assumption, allowing one to discard inconclusive events) that the experimental data lead to Bell inequality violations.

Recently though experimental violation of Bell inequalities with the detection loophole closed were reported by Christensen *et al.* (2013) and Giustina *et al.* (2013). It should be noted, however, that the data analysis of Giustina *et al.* (2013) is affected by the time-coincidence loophole (see Sec. VII.B.1), and is thus not fully satisfactory. This point was subsequently addressed by Larsson *et al.* (2013). Since both of these experiments are table top, using relatively slow detectors, they are still plagued by the locality loophole.

2. Atoms

Besides photons, Bell experiments have also been conducted with atomic systems. Such systems offer an important advantage from the point of view of the detection, with efficiencies typically close to unity. Therefore, atomic systems are well adapted for performing Bell experiments free of the detection loophole. Such an experiment was first realized by Rowe *et al.* (2001), using two Be$^+$ ions in a magnetic trap. In this experiment, the two ions were placed in the same trap, separated only by 3 μm. The locality loophole was thus left wide open, since each ion can feel the light field aimed at measuring the state of the other ion.

More recently, quantum nonlocality was demonstrated between two Yb$^+$ ions sitting in separated traps, 1 m apart (Matsukevich *et al.*, 2008). This was further improved to a distance of 20 m using rubidium atoms (Hofmann *et al.*, 2012). Although this distance is still insufficient to close the locality loophole (a distance of 300 m is required using the fastest procedure to measure the atomic state of the atoms) the cross talk between the two atoms is now completely suppressed. Here the entanglement between the distant atoms is achieved using an "event-ready" scheme (Simon and Irvine, 2003), shown in Fig. 7, which is based on entanglement swapping (Zukowski *et al.*, 1993). Each atom is first transferred to an excited state. The ion is deexcited by emitting a photon. The structure of the atomic levels is chosen such that the polarization of the emitted photon is maximally entangled with the state of the atom. The emitted photons are then collected in single mode optical fibers. Finally a partial Bell state measurement is performed on the two photons, using a simple 50:50 beam splitter followed by single photon detectors. A coincidence detection of two photons at the detectors indicates that the photons were in a given Bell state. In this case entanglement swapping is achieved, that is, the initial atom-photon entanglement has been converted to atom-atom entanglement. Upon successful detection of the photons, local measurements are performed on the atoms. The procedure is repeated until enough data have been taken in order to obtain good statistics. Important advantages of such an event-ready experiment is its robustness to photon losses and to the coincidence-time loophole (Larsson and Gill, 2004). Recently, such an experiment was used to conduct a

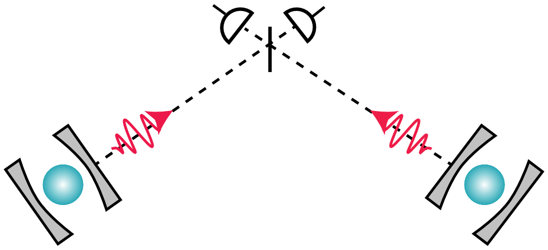

FIG. 7 (color online). A Bell test based on distant entangled atoms. Each atom is entangled with an emitted photon. Upon successful projection of the two photons onto a Bell state, the two atoms become entangled. The scheme is therefore "event ready," which makes it robust to photon losses in the channel. Moreover, since atomic measurements have an efficiency close to 1, this scheme is free of the detection loophole. This setup has been implemented experimentally with a distance of 20 m between the atoms (Hofmann *et al.*, 2012), and used for device-independent randomness expansion. From Pironio *et al.*, 2010.

proof-of-principle demonstration of device-independent randomness expansion (Pironio *et al.*, 2010) (see Sec. IV.C.3).

3. Hybrid schemes and other systems

Finally, we mention that Bell inequality violations have also been reported using atom-photon entanglement (Moehring *et al.*, 2004) and entanglement between a photon and a collective atomic excitation (Matsukevich *et al.*, 2005). Nonlocality was also demonstrated in Josephson phase superconducting qubits. In particular, violation of the CHSH inequality was achieved by Ansmann *et al.* (2009), whereas the GHZ paradox was demonstrated by DiCarlo *et al.* (2010) and Neeley *et al.* (2010).

B. Loopholes

1. Detection loophole

In a large class of Bell experiments, in particular, those carried out with photons, measurements do not always yield conclusive outcomes. This is due either to losses between the source of particles and the detectors or to the fact that the detectors themselves have nonunit efficiency. A measurement apparatus, used, e.g., to test the CHSH inequality, has then three outcomes instead of two: it can as usual give the outcomes -1 or $+1$, or it can give a "no-click" outcome, denoted $\perp$. The simplest way to deal with such "inconclusive" data is simply to discard them and evaluate the Bell expression on the subset of "valid" ± 1 measurement outcomes. As pointed out by Pearle (1970) and Clauser and Horne (1974), this way of analyzing the results is consistent only under the assumption that the set of detected events is a *fair sample*, i.e., that the accepted data are representative of the data that would have been recorded if the detectors had unit efficiency. More generally, one can consider local models where this fair-sampling assumption fails and in which the probability to obtain a no-click outcome $\perp$ depends on the choice of measurement (Pearle, 1970; Clauser and Horne, 1974; Santos, 1992). If the detection efficiency is too low (below a certain threshold), such local models can completely reproduce the observed data, opening the so-called detection loophole. The threshold efficiency required to close this detection loophole is typically high for practical Bell tests. As a consequence, most optical realizations of Bell tests performed so far are plagued by the detection loophole.

Another closely related loophole is the time-coincidence loophole (Larsson and Gill, 2004). This loophole exploits timing issues in Bell tests, which in turn can affect detection efficiency. Christensen *et al.* (2013) showed how this loophole can affect real experiments.

a. Faking Bell inequality violations with postselection

Throwing away no-click outcomes and keeping only the valid outcomes ± 1 is an example of postselection. In general, allowing for postselection in a given theory allows one to achieve tasks which would be impossible without it. In particular, postselection makes it possible to fake the violation of a Bell inequality, even in a purely local theory.

To illustrate this idea, we see how it is possible for a local model to fake maximal violation of the CHSH inequality. In particular, we show how to generate Popescu-Rohrlich correlations $a \oplus b = xy$, where $x, y, a, b \in \{0, 1\}$ (see Sec. II.C.2), starting from shared randomness and allowing the detectors on Alice's side to produce a no-click outcome $\perp$. The model is the following. The shared randomness corresponds to two uniform random bits x_{guess} and a. Given measurement setting y, Bob's detector outputs $b = a \oplus x_{\mathrm{guess}} y$. Alice's detectors output a whenever her measurement setting is $x = x_{\mathrm{guess}}$ and output $\perp$ when $x \neq x_{\mathrm{guess}}$. Focusing on the conclusive outcomes (e.g., ± 1), Alice and Bob have achieved maximally nonlocal PR correlations, i.e., achieving a CHSH value of $S = 4$. The probability for Alice to obtain a conclusive outcome is $1/2$, which is the probability that $x = x_{\mathrm{guess}}$, while Bob always obtains a conclusive outcome. With additional shared randomness, it is possible to symmetrize the above model, such that Alice and Bob's detection probability is $2/3$ (Massar and Pironio, 2003). Therefore, if the detection efficiency in a CHSH Bell experiment is below $2/3$, no genuine Bell inequality violation can be obtained, since the above local strategy could have been used by the measurement apparatuses. More generally, the minimum detection efficiency required for successfully violating a given Bell inequality depends on the number of parties and measurements involved (see Sec. VII.B.1.b).

Interestingly, recent experiments demonstrated fake violations of Bell inequalities using classical optics (Gerhardt *et al.*, 2011), positive Wigner function states and quadrature measurements (Tasca *et al.*, 2009), a classical amplification scheme (Pomarico, Sanguinetti *et al.*, 2011), and high-dimensional analyzers (Romero *et al.*, 2013). These experiments are performed under the same conditions as standard Bell experiments, but exploit side channels. This illustrates the importance of closing the detection loophole in Bell tests, in particular, for the perspective of implementing device-independent protocols.

b. Taking into account no-click events

The previous discussion shows that in order to close the detection loophole no-click outcomes cannot be discarded without making further assumptions. The most general way to take no-click events into account is simply to treat them as an additional outcome and instead of a Δ-outcome Bell inequality (if the number of "conclusive" outcomes is Δ) use a $(\Delta+1)$-outcome Bell inequality. A possible way to obtain an effective $(\Delta+1)$-outcome Bell inequality from a Δ-outcome one is simply to merge the no-click outcome with one of the valid outcomes,[20] i.e., systematically assign one of the valid outcomes to the no-click events. In particular, the Clauser-Horne inequality (Clauser and Horne, 1974), which is often used in Bell tests with inefficient detectors, is nothing but the CHSH inequality where the -1 outcome and the no-click outcome $\perp$ have been merged into one effective -1 outcome.

Assigning one of the valid outcomes to the no-click outcome $\perp$ is often the optimal way to treat no-click events, although there is no general proof of this and a counterexample exists for no-signaling correlations (Wilms *et al.*, 2008). In the case where Δ detectors are used to register the Δ outcomes of a measurement, assigning one of the conclusive outcomes to the no-click events has also the technical advantage that the detector associated with that particular outcome is no longer needed, i.e., only $\Delta-1$ detectors are sufficient since no distinction is being made between obtaining the Δth outcome and not detecting anything.

c. Threshold efficiencies

When treating the no-click outcome as described previously, one generally finds that a Bell violation is obtained only if the detector efficiencies are above a certain threshold. The minimal threshold efficiency η^*, required to close the detection loophole, depends generally on the number of parties, measurements, and outcomes involved in the Bell test. Moreover, η^* may also vary depending on the exact set of correlations that is considered. Thus, in quantum Bell tests, η^* may also depend on which entangled state and which measurement settings are considered. Next we review the efficiency thresholds for the most important Bell inequalities and for the most common quantum entangled states.

We start by deriving η^* for the CHSH Bell inequality using a two-qubit maximally entangled state. Performing judicious local measurement on this state, one obtains a CHSH value of $S=2\sqrt{2}$ (the maximum value possible in quantum mechanics). Now, we assume that Alice and Bob have imperfect detectors with efficiency η and that when a no-click result $\perp$ is obtained, they assign to it the $+1$ outcome. When both detectors click, which happens with probability η^2, Alice and Bob achieve $S=2\sqrt{2}$. When only one detector clicks, the outcomes are completely uncorrelated leading to $S=0$. Finally, when no detectors click, which happens with probability $(1-\eta)^2$, Alice and Bob both always output $+1$, thus achieving the local bound $S=2$. In order to close the detection loophole, we must ensure that the entire data of the experiment violate the CHSH inequality, i.e., that

$$\eta^2 2\sqrt{2}+(1-\eta)^2 2>2. \tag{92}$$

This leads to the condition that

$$\eta>\eta^*=\frac{2}{1+\sqrt{2}}\approx 82.8\%. \tag{93}$$

Therefore, in order to get a detection loophole free CHSH violation with a two-qubit maximally entangled state, it is sufficient to have a detection efficiency larger than ~82.8% (Mermin, 1986). This efficiency is also necessary: an explicit local model can be built which reproduces the experimental data when $\eta<2/(1+\sqrt{2})$.

Remarkably, it turns out that this threshold efficiency can be lowered by considering partially entangled states, of the form $|\psi_\theta\rangle=\cos\theta|00\rangle+\sin\theta|11\rangle$, as discovered by Eberhard (1993). In particular, in the limit of a product state (i.e., $\theta\to 0$) one obtains the fact that $\eta^*\to 2/3$. This astonishing result was the first demonstration that sometimes less entanglement leads to more nonlocality (see Sec. III.A.7).

From an experimental perspective, it is relevant to see how the previous results are affected by the presence of background noise. In general this amounts to a considerable increase of the threshold efficiencies. Even for very low levels of background noise, the threshold efficiency usually increases by a few percent. A detailed analysis can be found in Eberhard (1993). Another point concerns events in which no detection happened on either side. In certain cases, joint losses are not detrimental for the demonstration of nonlocality (Massar *et al.*, 2002).

Beyond the CHSH case, discussed in detail by Branciard (2011), it is known that lower threshold efficiencies can be tolerated for Bell inequalities featuring more measurement settings. A lower bound for the threshold efficiency is given by

$$\eta^*\geq\frac{m_A+m_B-2}{m_A m_B-1}, \tag{94}$$

where m_A (m_B) denotes the number of settings of Alice (Bob) (Massar and Pironio, 2003). While it is not known whether this bound can be obtained in general with quantum correlations, improvements over the threshold efficiencies of the CHSH inequalities have been obtained by considering Bell scenarios with more measurement settings. For qubit states only minor improvements were found (Massar *et al.*, 2002; Brunner and Gisin, 2008; Pal and Vértesi, 2009). Massar (2002) showed that, when considering systems of higher Hilbert space dimension d, the threshold efficiency can become arbitrarily close to zero. Unfortunately, this result is of limited practical interest since improvements over the CHSH case are obtained only for systems with $d\gtrsim 1600$. Also, the number of measurements becomes exponentially large, namely, 2^d. More recently a Bell test involving entangled quqats ($d=4$) and four (binary) measurement settings was shown to tolerate detection efficiencies as low as ~61.8% (Vértesi, Pironio, and Brunner, 2010).

[20] Inequalities obtained in this way are *liftings* of the original inequality (Pironio, 2005). If the original inequality is facet defining for the Δ-outcome Bell polytope, then the lifted inequality is facet defining for the $(\Delta+1)$-outcome polytope. However, the $(\Delta+1)$-outcome Bell polytope has also in general facets that cannot be viewed as liftings of Δ-outcome inequalities.

Such a scheme could be implemented optically using hyperentanglement.

Threshold efficiencies have also been derived for certain multipartite Bell tests (with n parties), using qubit GHZ states. Based on a combinatorial study, Buhrman *et al.* (2003) showed that an arbitrarily small efficiency can be tolerated as n becomes large. Threshold efficiencies approaching 50% in the limit of a large n were demonstrated for the Mermin inequalities (Cabello, Rodriguez, and Villanueva, 2008) and for a multipartite generalization of the CH inequality (Larsson and Semitecolos, 2001). More recently, multipartite Bell tests tolerating efficiencies significantly below 50% were reported by Pal, Vértesi, and Brunner (2012).

Finally, detection efficiencies have also been considered in asymmetric Bell experiments. Consider first the case in which Alice and Bob feature different detection efficiencies (η_A and η_B, respectively). In particular, results have been obtained for the case where $\eta_A < 1$ and $\eta_B = 1$, which is relevant for Bell tests based on atom-photon entanglement (Alice holds the photon and Bob the atom). It has been shown that for the CHSH inequality the efficiency of Alice's detector can be lowered to 50% (Brunner *et al.*, 2007; Cabello and Larsson, 2007). Moreover, this efficiency can be further lowered to ~43% by considering a three-setting Bell inequality (Brunner *et al.*, 2007). Considering Bell tests with d measurement settings and d-dimensional systems, an efficiency as low as $1/d$ can be tolerated, which is optimal (Vértesi, Pironio, and Brunner, 2010). Another asymmetric scenario is the case in which the local measurements have different efficiencies. Let η_{A_0} and η_{A_1} be the efficiencies of Alice's measurements and η_{B_0} and η_{B_1} the efficiencies of Bob's measurements. If one of the measurements of each party has unit efficiency (say $\eta_{A_0} = \eta_{B_0} = 1$), then the CHSH inequality can be violated for an arbitrarily low efficiency for the other measurement, i.e., $\eta_{A_1} = \eta_{B_1} \rightarrow 0$ (Garbarino, 2010). Such an approach fits Bell tests using hybrid measurements, such as homodyne (high efficiency) and photodetection (low efficiency); see Sec. VII.C for more detail.

2. Locality loophole

The locality condition (3) is motivated by the absence of communication between the two measurement sites of a Bell experiment. This seems well justified if the two sites are sufficiently separated so that the measurement duration is shorter than the time taken by a signal traveling at the speed of light, to travel from one site to the other. If this condition is not satisfied, one could in principle conceive a purely "local" mechanism (i.e., involving slower-than-light speed signals) underlying the observed correlations (Aspect, 1975, 1976; Bell, 1977a).

In addition to the requirement that the two measurement sites are spacelike separated, it must also be the case that the measurement setting on one side is not determined by an earlier event that could be correlated with the measurement setting on the other side; in particular, it should not be correlated with the hidden variables λ characterizing the source of particles. That is, the measurement settings must correspond to "random" or "free" choices, which are independent from the other side and from the hidden state of the particle pairs (Shimony, Horne, and Clauser, 1976; Bell, 1977b).

Mathematically, the above requirements correspond to the following conditions:

$$p(a|x,y,b,\lambda) = p(a|x,\lambda), \quad p(b|y,x,a,\lambda) = p(b|y,\lambda), \tag{95}$$

and

$$q(\lambda|xy) = q(\lambda) \tag{96}$$

from which the locality decomposition (3) follows. Failure to satisfy them is known as the *locality loophole*. The failure to specifically address the independence condition (96) between measurement choices and hidden variables is sometimes called the "freedom of choice" (Scheidl *et al.*, 2010) or "measurement-independence" loophole (Hall, 2010; Barrett and Gisin, 2011).

The experiment of Aspect, Dalibard, and Roger (1982) using entangled photons was the first to convincingly address the locality loophole. It involved on each side a switching device for the incoming photons followed by two polarizers aligned along different orientations. A change of direction occurred approximately every 10 ns. The distance $L = 13$ m between the two switches was large enough so that the time of travel of a signal between the switches at light speed $L/c =$ 43 ns was larger than the delay of 10 ns between two switchings. The switching of the polarizers was done through a home-built device, based on the acousto-optical interaction of the incoming light with an ultrasonic standing wave in water. Using this mechanism it should be noted that not all photons were submitted to forced switching. In addition, the switches were not truly random, since the acousto-optical were driven by periodic generators. The two generators on the two sides, however, were functioning in a completely uncorrelated way, since they were operated by different rf generators at different frequencies with uncorrelated frequency drifts.

The experiment of Aspect, Dalibard, and Roger (1982) was the only one involving fast changes of the measurement settings until the one of Weihs *et al.* (1998), which used high-speed electro-optic modulators to switch between two polarization measurement settings on each side. The two modulators where controlled on a nanosecond time scale by two independent quantum random number generators, excluding any light-speed influence between the two sides, which were separated by a distance of a few hundred meters. Leaving aside the possibility that the outputs of the quantum random number generators are predetermined at some hidden level, this setup is often regarded as having conclusively closed the locality loophole. In Scheidl *et al.* (2010), spacelike separation was not only enforced between the outputs of the two random generators, but also between them and the emission of photons from the laser source generating the entangled particles, implying that these three processes are independent from each other, provided that they are not themselves determined by some earlier events.

At this point, it should be stressed that, contrary to the detection loophole, the locality loophole can never be "completely" closed. Strictly speaking, it requires spacelike separation between the event determining the choice of measurement setting on one side and the event corresponding to the output of the measurement device on the other side. The first problem is that this requires one to know precisely the time at which these two events happen. But if we use some random process that outputs a random value at time, say $t = 0$, how do we know that this value was precisely determined at this time and not at some earlier time $t = -\delta$? Similarly, how do we know precisely when a measurement is completed without making some assumptions on the collapse of the wave function (Kent, 2005)? This last issue was addressed by Salart *et al.* (2008), where the violation of Bell inequalities for events that are spacelike separated according to a simple model of gravitational collapse has been reported.

The second problem is that we can never be sure that the choices of measurements are really random or free. For instance, in the experiments (Weihs *et al.*, 1998; Scheidl *et al.*, 2010) the measurement choices are decided by processes that are genuinely random according to standard quantum theory. But this need not be the case according to some deeper theory. Some have argued that a better experiment for closing the locality loophole would be to arrange the choice of measurement setting to be determined directly by humans or by photons arriving from distant galaxies from opposite directions, in which case any local explanation would involve a conspiracy on the intergalactic scale (Vaidman, 2001).

The point of this discussion is that an experiment "closing" the locality loophole should be designed in such a way that any theory salvaging locality by exploiting weaknesses of the above type should be sufficiently conspiratorial and contrived that it is reasonably not worth considering it.

Finally, it is worth noting that in device-independent applications of quantum nonlocality the experimental requirements for satisfying conditions (95) and (96) are sensibly different than in fundamental tests of nonlocality, since one usually assumes the validity of quantum theory, that the quantum systems are confined in well-defined measurement devices that can be shielded from the outside world, that the inputs are under the control of the users, etc. This stance was used for instance by Pironio *et al.* (2010), where the atoms were separated from each other, although by no means fulfilling any of the spacelike separation prescriptions required for a fundamental, locality loophole-free Bell test.

3. Finite statistics

Since it is expressed in terms of the *probabilities* for the possible measurement outcomes in an experiment, a Bell inequality is formally a constraint on the expected or average behavior of a local model. In an actual experimental test, however, the Bell expression is estimated only from a finite set of data and one must take into account the possibility of statistical deviations from the average behavior. The conclusion that Bell locality is violated is thus necessarily a statistical one. In most experimental papers reporting Bell violations, the statistical relevance of the observed violation is expressed in terms of the number of standard deviations separating the estimated violation from its local bound. There are several problems with this analysis, however. First, it lacks a clear operational significance. Second, it implicitly assumes some underlying Gaussian distribution for the measured systems, which is justified only if the number of trials approaches infinity. It also relies on the assumption that the random process associated with the kth trial is independent of the chosen settings and observed outcomes of the previous $k-1$ trials. In other words, the devices are assumed to have no memory, which is a questionable assumption (Accardi and Regoli, 2000).

A better measure of the strength of the evidence against local models is given by the probability with which the observed data could have been reproduced by a local model. For instance, consider the CHSH test and let $\langle a_x b_y \rangle_{\rm obs}$ be the means of the observed products of a and b when measurements x and y are chosen computed over N trials. Gill (2012) showed that the probability that two devices behaving according to a local model gives rise to a value

$$S_{\rm obs} = \langle a_0 b_0 \rangle_{\rm obs} + \langle a_0 b_1 \rangle_{\rm obs} + \langle a_1 b_0 \rangle_{\rm obs} - \langle a_1 b_1 \rangle_{\rm obs} > 2 + \varepsilon$$

is

$$p(S_{\rm obs} > 2 + \varepsilon) \leq 8 e^{-4N(\varepsilon/16)^2}. \tag{97}$$

This statement assumes that the behavior of the devices at the kth trial does not depend on the inputs and outputs in previous runs. But this memoryless assumption can be avoided and similar statements taking into account arbitrary memory effects can be obtained (Gill, 2003). As discussed in Sec. II.G, it is easy to convince oneself that, in the limit of infinitely many runs, devices with memory cannot fake the violation of a Bell inequality (Barrett *et al.*, 2002; Gill, 2003). Indeed, for any given local variable strategy, there is always at least one set of settings for which that strategy fails in the corresponding Bell game. But in any run, the settings are chosen at random, independently of the source and of the past: therefore, even taking the past into account, the local variables cannot avoid the possibility that the wrong settings are chosen. This reasoning can be extended to the finite statistics case through the use of martingales (Gill, 2003). A better general method to deal with memory effects and finite statistics which is asymptotically optimal in the limit of large trials was proposed by Zhang, Glancy, and Knill (2011) and Zhang, Knill, and Glancy (2013).

C. Toward a loophole-free Bell test

1. Photons

The main drawback of photonic experiments for performing a loophole-free Bell test is the limited detection efficiency of single photon detectors. Nevertheless, considerable progress has been achieved in the past years, in particular, with the development of detectors based on superconducting materials, which can achieve detection efficiencies above 95%. Such detectors were recently used in Bell-type experiments. Smith *et al.* (2012) reported on a detection loophole-free demonstration of quantum steering (see Sec. VIII.C). More recently, an experiment demonstrated violation of a Clauser-Horne Bell inequality with the detection loophole closed (Christensen *et al.*, 2013; Giustina *et al.*, 2013). Here total efficiencies of 75% were achieved for each party. Note, however, that these experiments are tabletop and did not close the locality loophole. Since the detection process in superconducting detectors is typically slow (of the order of μs), achieving a loophole-free Bell violation requires a separation of the order of 300 m.

Another possibility was recently suggested by Cabello and Sciarrino (2012), which consists of precertifying the presence of a photon before the choice of local measurement is performed. This proposal appears, however, challenging from an experimental point of view.

2. Continuous variable systems

An interesting alternative for achieving high detection efficiencies using photon consists of using homodyne measurements, which measure a continuous degree of freedom (often called quadrature) of the optical mode. Such measurements are realized by mixing the optical mode with an intense reference oscillator on a beam splitter and can reach efficiencies close to unity nowadays. The first proposals for using homodyne measurements in Bell tests were presented by Grangier, Potasek, and Yurke (1988), Tan, Walls, and Collett (1991), and Gilchrist, Deuar, and Reid (1998). Since then, many alternative proposals were discussed. However, their experimental realization has remained elusive so far.

It is important to mention that a homodyne measurement has a continuous number of possible outcomes. Since Bell inequalities have typically a discrete number of outcomes (for instance, binary outcomes in the case of CHSH), one has to dichotomize the outcome of the homodyne measurement, a procedure referred to as *binning*. For instance, a natural dichotomization strategy is given by the sign binning, where one assigns the values $+1$ if the measurement returns a positive outcome, and -1 otherwise.

Homodyne measurements were shown to be able to detect the nonlocality of certain quantum states. However, all tests proposed so far present major difficulties for experimental realizations. First, several schemes consider quantum states which cannot be produced using current technology (Munro, 1999; Wenger *et al.*, 2003; Cavalcanti *et al.*, 2007; Acín *et al.*, 2009). Second, the proposals of García-Patron *et al.* (2004) and Nha and Carmichael (2004) [see also Garcia-Patron, Fiurasek, and Cerf (2005)] use states which could be realized experimentally but lead only to very small Bell inequality violations, hence requiring an extremely low level of noise, currently out of reach of an experimental point of view. Note that an experiment using homodyne measurements demonstrated a violation of the CHSH inequality (Babichev, Appel, and Lvovsky, 2004), but postselection was involved which thus opened the detection loophole.

An interesting alternative consists of devising hybrid schemes, which make use of homodyne measurements as well as photodetection. Cavalcanti, Brunner *et al.* (2011) showed that relatively high CHSH violations can be achieved using a state that can be experimentally produced.[21] Promising developments of hybrid schemes were recently discussed by Araújo *et al.* (2012), Brask *et al.* (2012), Brask and Chaves (2012), Quintino *et al.* (2012), and Teo *et al.* (2013).

Finally, several works also considered more complex measurements, such as parity measurements. In particular, Banaszek and Wódkiewicz (1998) [see also Banaszek and Wódkiewicz (1999)] demonstrated that such measurements can reveal the nonlocality of the famous EPR state, discussed by Einstein, Podolsky, and Rosen (1935). The use of measurements based on nonlinear local operations was proposed by Stobińska, Jeong, and Ralph (2007). However, realization of such measurements is still out of reach from current technologies.

3. Atom-atom and atom-photon entanglement

A promising avenue toward a loophole-free Bell test is based on the possibility to generate long-distance entanglement between two trapped atoms (Simon and Irvine, 2003). The procedure for entangling the two remote atoms consists of the joint detection of two photons, each coming from one of the atoms, in an entangled basis. In this way, the initial atom-photon entanglement is transformed into atom-atom entanglement via entanglement swapping.

This scheme was demonstrated experimentally using two trapped atoms separated by 1 m (Matsukevich *et al.*, 2008; Pironio *et al.*, 2010) and more recently up to 20 m (Hofmann *et al.*, 2012). The detection loophole was closed in these experiments, thanks to the near unit efficiency of atomic measurements. In order to close the locality loophole, a distance of the order of 300 m would be required using the fastest atomic measurement techniques available today (Rosenfeld *et al.*, 2009). This is still currently challenging but the perspectives for a loophole-free test are promising.

Bell tests based on the direct observation of atom-photon entanglement have also been proposed (Brunner *et al.*, 2007; Cabello and Larsson, 2007). However, the difficulty of efficiently collecting the photons emitted from the atom, and the relatively high detection efficiencies required for the photon detection in order to close the detection loophole, make this approach more delicate to implement. To overcome some of these problems, the use of continuous variable degrees of freedom of the light field combined with efficient homodyne measurements was recently explored (Sangouard *et al.*, 2011; Araújo *et al.*, 2012; Teo *et al.*, 2013).

More recently, it was proposed to achieve a loophole-free Bell test using spin photon interactions in cavities (Brunner *et al.*, 2013; Sangouard *et al.*, 2013). Here the entanglement of a pair of photons is mapped to two distant spins (e.g., carried by a single atom or a quantum dot). Importantly, this mapping can be achieved in a heralded way. By choosing the measurement settings only after successful heralding, the scheme is immune from the detection loophole, since spin systems can usually be measured with high efficiencies.

D. Bell tests without alignment

Bell inequality violations in quantum mechanics require the parties to perform judiciously chosen measurement settings on an entangled state. Experimentally, this requires a careful

[21] Note that the idea of considering hybrid measurements was first discussed by Ji *et al.* (2010), although the proposed scheme is not a proper Bell test (Cavalcanti and Scarani, 2011).

calibration of the measurement devices and alignment of a shared reference frame between the distant parties. Although such assumptions are typically made implicitly in theoretical works, establishing a common reference frame and aligning and calibrating measurement devices in experimental situations are never trivial issues. For instance, in quantum communications via optical fibers, unavoidable small temperature changes induce strong rotations of the polarization of photons in the fiber, which makes it challenging to maintain good alignment. In turn this may considerably degrade the implementation of quantum protocols, such as Bell tests.

This led several to investigate whether Bell tests, and more generally quantum communication protocols (Bartlett, Rudolph, and Spekkens, 2007), could be realized without the need of a common reference frame. The first approach proposed was based on decoherence-free subspaces (Cabello, 2003). The experimental realization of such ideas is challenging as it requires high-dimensional entanglement, although progress was recently reported (D'Ambrosio *et al.*, 2012).

A more recent approach investigated Bell tests performed with randomly chosen measurements (Liang *et al.*, 2010). This first theoretical work considered the CHSH Bell scenario, with qubit measurements chosen randomly and uniformly (according to the Haar measure) on the Bloch sphere, on a maximally entangled state of two qubits. When all four measurements are chosen at random, the probability that the obtained statistics will violate the CHSH inequality is ~28%. When unbiased measurements are used, this probability increases to ~42%. More recently it was shown however that if both parties perform three unbiased measurements (i.e., forming an orthogonal triad on the Bloch sphere), the probability of violating a Bell inequality becomes one (Shadbolt *et al.*, 2012; Wallman and Bartlett, 2012)). This scheme was realized experimentally demonstrating the robustness of the effect to experimental imperfections such as losses and finite statistics (Shadbolt *et al.*, 2012) [see also Palsson *et al.* (2012)]. From a more conceptual point of view, these works are interesting as they show that quantum nonlocality is much more generic than previously thought.

VIII. RELATED CONCEPTS

This section deals with variations around Bell's theorem, in which different notions of nonlocality (stronger or weaker than Bell's) are considered. Note that there also exist mathematical relations between local models and noncontextual models. We do not review this connection here; see Mermin (1993) for a short review of both concepts and their relation.

A. Bilocality and more general correlation scenarios

In a tripartite Bell scenario, the standard definition of locality is given by

$$p(abc|xyz) = \int d\lambda q(\lambda) p_\lambda(a|x) p_\lambda(b|y) p_\lambda(c|z), \qquad (98)$$

where λ is a shared local random variable and $\int d\lambda q(\lambda) = 1$. This represents the most general model in which the outcome of each observer is determined by their input and λ. Since λ is shared between all three parties, arbitrary prior correlations can be established between the parties.

In certain quantum experiments involving three separated observers, the distribution $p(abc|xyz)$ is obtained by performing measurements on independent quantum states, originating from different sources. A typical example is the protocol of entanglement swapping (Zukowski *et al.*, 1993)—also known as teleportation of entanglement—in which two systems that never interacted become entangled. Here one party (say Bob) shares initially an entangled state with both Alice (denoted system AB_1) and Charlie (system B_2C). That is, Alice and Bob share an entangled pair distributed by a source located between them, while a second source distributes entanglement between Bob and Charlie. Importantly, these two sources are completely independent from each other, hence systems AB_1 and B_2C share no prior correlations. It is then natural to ask whether the observed correlations $p(abc|xyz)$ can be reproduced using a local model with the same feature, that is, in which systems AB_1 and B_2C are initially uncorrelated. Formally such models can be written in the following form:

$$p(abc|xyz) = \int d\lambda d\mu q(\lambda) q(\mu) p_\lambda(a|x) p_{\lambda,\mu}(b|y) p_\mu(c|z), \qquad (99)$$

where Alice and Bob share the local random variable λ, while Bob and Charlie share μ. Since the variables λ and μ are assumed to be independent, their distribution factorizes, i.e., $q(\lambda,\mu) = q(\lambda)q(\mu)$. The above condition is termed *bilocality*, and correlations satisfying it are called bilocal. It turns out that not all local correlations [i.e., of Eq. (98)] can be written in the bilocal form. Hence the bilocality condition is a strictly stronger constraint than locality. It is then interesting to ask how to characterize the set of bilocal correlations, as this will lead to more stringent tests for revealing nonlocality in an entanglement swapping experiment.

The first exploratory work investigated this question in the context of the detection loophole (Gisin and Gisin, 2002; Zukowski *et al.*, 2008). More recently, a systematic approach was taken by Branciard, Gisin, and Pironio (2010) and Branciard, Rosset *et al.* (2012). In particular, these works present nonlinear inequalities for testing the bilocality condition, which are much more stringent compared to standard Bell inequalities. Note that the set of bilocal correlations is not convex in general, and hence its characterization requires nonlinear inequalities.

More generally, it is possible to consider an arbitrary correlation scenario, involving an arbitrary number of sources and observers, where certain systems can be initially correlated while others are independent. Similar to the previous discussion, when two systems are assumed to be independent, they are described by a product distribution (Branciard, Rosset *et al.*, 2012; Fritz, 2012b). In fact, a typical Bell experiment can be viewed in this picture, featuring three independent sources. These are the source that produces the entangled state, the source generating the measurement settings of Alice, and the source generating the setting of Bob. Indeed, if these sources are not independent, the Bell violation is plagued by the measurement-independence loophole. A related approach for considering locality in general correlation scenarios using the formalism of causal networks was put forward by Wood and Spekkens (2012).

B. No-go theorems for nonlocal models

The study of no-go theorems for *nonlocal models* is reduced to a few examples. On the one hand, it is obvious that some of these models will reproduce all observed correlations, so there is no hope of finding a result *à la* Bell which would falsify all of them. On the other hand, one needs to have good motivation in order to propose a specific example of a nonlocal model. In fact, basically two classes of models have been considered so far: we review them briefly here, but refer to the original articles for a detailed justification of the interest of each model.

1. Models *à la* Leggett

Pure entangled states are characterized by the fact that the properties of the pair are well defined, but those of the individual subsystems are not. Consider for instance a maximally entangled state of two qubits. Although the global state has zero entropy, the state being pure, the reduced state of each qubit is fully mixed thus having maximum entropy. An interesting question is whether one could devise alternative no-signaling models, reproducing quantum correlations, in which the individual properties are well defined, or at least where the model gives a higher degree of predictability compared to quantum predictions. The first work in this direction was presented by Leggett (2003), who discussed a specific nonlocal model and proved its inability to reproduce quantum correlations. Leggett's model was first tested experimentally by Gröblacher *et al.* (2007). A clear discussion of the scope and limitations of this type of models was given by Branciard *et al.* (2008).

In a nutshell, the question is whether the probability distribution predicted by quantum theory p_Q can be seen as a convex combination of more fundamental distributions $p_Q = \int d\lambda p_\lambda$. Because of Bell's theorem, for some λ at least, the distribution p_λ must be nonlocal; but we leave the correlations and their mechanism aside and concentrate on the marginals: we request that the p_λ specify well-defined individual properties. Focusing on the case of a maximally entangled qubit pair, Leggett proposed a model in which the marginals take the forms

$$p_\lambda(a|\vec{x}) = \tfrac{1}{2}(1 + a\vec{u} \cdot \vec{x})$$

and

$$p_\lambda(b|\vec{y}) = \tfrac{1}{2}(1 + b\vec{v} \cdot \vec{y}).$$

Here the hidden variables consist of two vectors $\lambda = (\vec{u}, \vec{v})$. The intuition behind this model is the following. Locally, say on Alice's side, the system behaves as if it had well-defined polarization given by $\vec{u}$. For a measurement direction $\vec{x}$, the marginal distribution is then given by Malus's law. Hence this model makes more definite predictions for individual properties compared to quantum theory. It turns out, however, that Leggett's model is unable to reproduce quantum correlations. In particular, from the no-signaling condition and the above marginals, one can derive constraints, in the form of inequalities, on the correlations. Quantum predictions violate these inequalities. Note that there is no direct relation between Leggett inequalities and Bell inequalities. In particular, the tests of Leggett inequalities known to date rely on the characterization of the measurement parameters and are therefore not device independent, contrary to Bell inequalities.

Finally, note that more general models were also discussed and demonstrated to be incompatible with quantum predictions. This shows that quantum correlations cannot be reproduced using no-signaling theories which make more accurate predictions of individual properties compared to quantum theory (Colbeck and Renner, 2008, 2012).

2. Superluminal signaling models

A second class of models addresses the possibility of explaining quantum correlation using some explicit *signaling* mechanism. Of course, this is problematic, because the signal should propagate faster than light: these models must thus postulate the existence of a preferred frame in which this signal propagates. Two cases have been considered.

In the first one, the preferred frame is universal. From Bell's theorem, it follows that the speed v of the signal must be infinite. Clearly, one can find a model with v being infinite which reproduces all quantum correlations. On the other hand, the predictions of any signaling model where v is finite will differ from those of quantum mechanics. For instance, in a bipartite Bell test, nonlocal correlations should vanish when the distance between the two observers exceeds a certain bound, since there is then simply not enough time for the signal to propagate. Experimental investigations could place lower bounds on v (Salart *et al.*, 2008). For a wide class of preferred frames, this bound exceeds the speed of light by several orders of magnitude.

Furthermore, it is in fact possible to rule out theoretically any communication model in which v is finite using certain assumptions. Specifically, consider a model that (1) reproduces the quantum predictions when there is enough time for a signal to propagate at speed v between the parties; (2) the model is no signaling on average, that is, at the level of the observed statistics the no-signaling conditions (7) are satisfied (i.e., any explicit signaling happens only at the hidden level). Then by considering a multipartite arrangement, it was shown by Bancal, Pironio *et al.* (2012), building on earlier work by Scarani and Gisin (2005) and Coretti, Hänggi, and Wolf (2011), that these two conditions are mutually incompatible. In other words, under the assumption that the observed statistics satisfy the no-signaling principle, quantum correlations cannot be reproduced by a model with finite speed signaling.

In the second type of models, the rest frame of each measurement device is its own preferred frame. In this case, if the measurement devices move away from one another, a *before-before* configuration can be created, in which each particle perceives that it is the first one to undergo the measurement: then, nonlocal correlations should disappear (Suarez and Scarani, 1997). This prediction has been falsified experimentally (Stefanov *et al.*, 2002, 2003; Zbinden *et al.*, 2001).

C. Steering

One of the most common ways of describing the effect of entanglement, noticed already in the seminal paper of Schrödinger (1936), uses the notion of *steering*: by making a measurement on her system, Alice can prepare at a distance Bob's state. More precisely, Alice cannot choose which state she prepares in each instance, because this would amount to signaling; however, if she sends to Bob the results of her measurements, Bob can check that indeed his conditional states have been steered.

Although often invoked in the field, this notion had not been the object of detailed studies until the work of Wiseman, Jones, and Doherty (2007). In this and subsequent work, they formalized steering as information-theoretic tasks and pointed out how it differs from nonlocality. Steering can be viewed as the distribution of entanglement from an untrusted party. Alice wants to convince Bob that she can prepare entanglement. Bob trusts his measurement device; hence he knows what observables he is measuring on his system. However, Bob does not trust Alice, whose device is then described by a black box. In this sense the task is intermediate between nonlocality (in which both Alice and Bob work with black boxes) and standard entanglement witnessing (in which both parties have perfect control of the observables which are measured). The protocol works as follows. Alice sends a quantum system to Bob, whose state, she claims, is entangled to her system. Upon receiving his system, Bob chooses a measurement setting (from a predetermined set of measurements) to perform on it. He then informs Alice about his choice of measurement and asks her to provide a guess for his measurement outcome. After repeating this procedure a sufficiently large number of times, Bob can estimate how strongly his system is correlated to that of Alice. If the correlations are strong enough, Bob concludes that the systems are indeed entangled and that Alice did indeed steer his state.

Interestingly, it turns out that entanglement is necessary but not sufficient for steering, while steering is necessary but not sufficient for nonlocality. Hence, steering represents a novel form of inseparability in quantum mechanics, intermediate between entanglement and nonlocality (Wiseman, Jones, and Doherty, 2007; Saunders *et al.*, 2010). The quantitative relation between steering, entanglement, and Bell nonlocality is yet to be fully understood.

Experimentally, steering can be tested using steering inequalities, similar to Bell inequalities. The first steering criterion were derived for continuous variable systems, mostly based on variances of observables (Reid, 1989) and entropic uncertainty relations (Walborn *et al.*, 2011); see Reid *et al.* (2009) for a review. More recently, steering inequalities were presented for discrete variables (Cavalcanti *et al.*, 2009). All these tests were investigated experimentally. Similar to Bell tests, experimental steering tests suffer from loopholes. Nevertheless, closing these loopholes appears to be less demanding compared to Bell tests, in particular, in terms of detection efficiency. A loophole-free steering experiment was recently reported (Wittmann *et al.*, 2012).

D. Semiquantum games

In the usual Bell test scenario, distant parties share a quantum state distributed from a source. The local measurements and their outcomes are represented by classical data. Recently Buscemi (2012) proposed a variant of Bell tests, termed semiquantum games, in which the inputs of each party are given by quantum states. That is, each party holds a black box in which the observer inputs a quantum state. Inside the box, this quantum input is then jointly measured with the quantum system coming from the source, and a classical output is produced. In case the input quantum states are orthogonal (hence perfectly distinguishable), the setup is simply equivalent to a standard Bell test with classical inputs. However, when the states are not mutually orthogonal, the setup becomes more general. Buscemi showed that, in this case, for any entangled state ρ there exists a semiquantum game, the statistics of which cannot be reproduced by a local model. Hence, semiquantum games highlight a nonlocal aspect of quantum states, which is however different from Bell nonlocality. More recently it was shown that semiquantum games and entanglement witnesses are intimately related. In particular, for detecting an entangled state ρ, a semiquantum game can be constructed directly from an entanglement witness detecting ρ (Branciard *et al.*, 2013). Applications of these ideas for the detection of entanglement (Branciard *et al.*, 2013), steering (Cavalcanti, Hall, and Wiseman, 2013), and the classical simulation of quantum correlations (Rosset *et al.*, 2013) were recently discussed.

IX. CONCLUSION

Fifty years after the publication of Bell's theorem (Bell, 1964), Bell nonlocality is still (perhaps more than ever) at the center of an active and intense research activity that spans the foundations of quantum theory, applications in quantum information science, and experimental implementations.

We covered in this review most of the recent developments, some of them happening while this review was being written. To give only three examples of recent progress on long-standing or important questions: Peres's conjecture that no bound entangled state can give rise to nonlocal correlations is now disproved in the tripartite case (Vértesi and Brunner, 2012) (but is still open in the bipartite case); it has been shown that nonlocal correlations can be exploited to perform arbitrary computations in a device-independent way (Reichardt, Unger, and Vazirani, 2013); on the experimental side, the detection loophole has finally been closed in full-optical implementations (Christensen *et al.*, 2013; Giustina *et al.*, 2013). We hope that this review will motivate further developments on this fascinating topic of Bell nonlocality.

ACKNOWLEDGMENTS

We thank all of our nonlocal colleagues for enlightening discussions over the recent years. Also, we thank Antonio Acín, Alain Aspect, Richard Gill, Nicolas Gisin, and Yeong-Cherng Liang for comments on this manuscript. This work is supported by the National Research Foundation and the Ministry of Education of Singapore (partly through the

Academic Research Fund Tier 3 MOE2012-T3-1-009), the UK EPSRC, the Swiss National Science Foundations (Grant No. PP00P2_138917), the EU CHISTERA Projects DIQIP and QCS, the FRS-FNRS Project DIQIP, and the Brussels-Capital Region through a BB2B Grant.

APPENDIX: GUIDE TO BELL INEQUALITIES

The goal of this Appendix is to orientate the reader looking for a particular type of Bell inequality. Here we classify Bell inequalities according to the number of parties n, the number of measurements for each party m, and the number of outcomes Δ.[22] Below, for any Bell test scenario given by the triple (n, m, Δ), we provide references to articles presenting relevant Bell inequalities. Note that we do not give the inequalities explicitly; some of these can nevertheless be found in parts of this review, in particular, in Secs. II and V.

1. Bipartite Bell inequalities

a. Binary outputs: $(2,m,2)$

For the case $m = 2$, CHSH is the only tight Bell inequality. Note that if one of the parties has more measurement, i.e., $m_A = 2$ and $m_B = m$, CHSH is still known to be the only tight inequality.

For $m = 3$, one additional tight inequality arises I_{3322} (see Sec. 2), discovered independently by Froissard (1981) and Collins and Gisin (2004).

For $m = 4, 5$, the complete list of tight Bell inequalities is unfortunately not known. Incomplete lists can be found in Brunner and Gisin (2008), Pal and Vértesi (2009), and Bancal, Gisin, and Pironio (2010). Note that for $m_A = 4$ and $m_B = 3$, the complete list of tight Bell inequalities was given by Collins and Gisin (2004).

For $m \geq 6$, much less is known. Incomplete lists of facets could be derived using sophisticated techniques from convex geometry (Avis *et al.*, 2004).

For arbitrary values of m, the family of inequalities I_{mm22} introduced by Collins and Gisin (2004) turns out to be tight in general (Avis and Ito, 2007). It is also worth mentioning the family of chained Bell inequalities (Pearle, 1970; Braunstein and Caves, 1990), although these inequalities are not tight for $m \geq 3$.

b. Arbitrary number of outputs: $(2,m,\Delta)$

In the cases $m = 2$ and $\Delta = 3$, the inequality of CGLMP (Collins, Gisin, Linden *et al.*, 2002) is the only tight inequality additional to CHSH. For $\Delta \geq 4$, the CGLMP inequality is known to be tight, but there exist additional facets in this case (Bancal, Gisin, and Pironio, 2010).

Note that the chained Bell inequalities have been generalized to this scenario (Barrett, Kent, and Pironio, 2006). Whether they are tight or not for $\Delta \geq 3$ is not known.

2. Multipartite Bell inequalities

a. Binary outputs: $(n,m,2)$

All tight correlation Bell inequalities are known for the case $m = 2$ (Werner and Wolf, 2001b; Zukowski and Brukner, 2002). Indeed, this set contains the inequalities of Mermin-Ardehali-Belinskiõ-Klyshko (Mermin, 1990a; Ardehali, 1992; Belinskii and Klyshko, 1993). For noncorrelation Bell inequality, the complete set of tight Bell inequalities in the cases $n = 3$ and $m = 2$ was given by Sliwa (2003). For arbitrary n and m, Laskowski *et al.* (2004) provided tight Bell inequalities.

b. Arbitrary number of outputs: (n,m,Δ)

A general family of Bell inequalities based on variance inequalities was derived by Cavalcanti *et al.* (2007), and further developed by He *et al.* (2009). More generally these inequalities are particular cases of inequalities discussed by Salles *et al.* (2010). Note that these inequalities are not tight as they are not linear. However, they can be conveniently used for continuous variables (i.e., $\Delta \to \infty$). A generalization of the Mermin inequalities for the scenario $(3,2,\Delta)$ was presented by Grandjean *et al.* (2012); these inequalities are known to be tight for $\Delta \leq 8$. For $n = 3$, $m = 2$, and $\Delta = 3$, 4, and 5, tight inequalities were given by Chen *et al.* (2008).

Note also that there exist functional Bell inequalities which can be defined for an infinite number of settings (Sen(De), Sen, and Zukowski, 2002).

c. Bell inequalities detecting genuine multipartite nonlocality

Svetlichny (1987) derived the first inequality for testing genuine multipartite nonlocality in the case (3,2,2). This inequality was later generalized to an arbitrary number of parties, e.g., to the scenario $(n,2,2)$ (Collins, Gisin, Popescu *et al.*, 2002; Seevinck and Svetlichny, 2002). Bancal, Gisin, and Pironio (2010) also introduced more inequalities for the simplest case (3,2,2).

Svetlichny's inequality was also generalized to a more general scenario $(n,2,\Delta)$ by Bancal, Brunner *et al.* (2011) and (n,m,Δ) by Bancal, Branciard *et al.* (2012).

Barrett, Pironio *et al.* (2013) discussed various definitions of the concept of genuine multipartite nonlocality and introduced inequalities for testing each of these models in the scenario (3,2,2).

REFERENCES

Accardi, L., and M. Regoli, 2000, Preprint 399, Volterra Institute, University of Rome II.

Acín, A., 2001, Phys. Rev. Lett. **88**, 027901.

Acín, A., R. Augusiak, D. Cavalcanti, C. Hadley, J. Korbicz, M. Lewenstein, and M. Piani, 2010, Phys. Rev. Lett. **104**, 140404.

Acín, A., N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, 2007, Phys. Rev. Lett. **98**, 230501.

Acín, A., N. J. Cerf, A. Ferraro, and J. Niset, 2009, Phys. Rev. A **79**, 012112.

Acín, A., T. Durt, N. Gisin, and J. Latorre, 2002, Phys. Rev. A **65**, 052325.

[22] Note that for Bell inequalities featuring different numbers of measurements or outcomes for different parties, m and Δ represent the maximum values.

Acín, A., R. Gill, and N. Gisin, 2005, Phys. Rev. Lett. **95**, 210402.

Acín, A., N. Gisin, and L. Masanes, 2006, Phys. Rev. Lett. **97**, 120405.

Acín, A., N. Gisin, and B. Toner, 2006, Phys. Rev. A **73**, 062105.

Acín, A., S. Massar, and S. Pironio, 2006, New J. Phys. **8**, 126.

Acín, A., S. Massar, and S. Pironio, 2012, Phys. Rev. Lett. **108**, 100402.

Acín, A., V. Scarani, and M. M. Wolf, 2003, J. Phys. A **36**, L21.

Adesso, G., and S. Piano, 2014, Phys. Rev. Lett. **112**, 010401.

Afzelius, M., 2011 (private communication).

Aharon, N., S. Machnes, B. Reznik, J. Silman, and L. Vaidman, 2013, Nat. Comput. **12**, 5.

Ajoy, A., and P. Rungta, 2010, Phys. Rev. A **81**, 052334.

Allcock, J., N. Brunner, N. Linden, S. Popescu, P. Skrzypczyk, and T. Vértesi, 2009, Phys. Rev. A **80**, 062107.

Allcock, J., N. Brunner, M. Pawlowski, and V. Scarani, 2009, Phys. Rev. A **80**, 040103.

Allcock, J., H. Buhrman, and N. Linden, 2009, Phys. Rev. A **80**, 032105.

Almeida, M., J.-D. Bancal, N. Brunner, A. Acín, N. Gisin, and S. Pironio, 2010, Phys. Rev. Lett. **104**, 230404.

Almeida, M., D. Cavalcanti, V. Scarani, and A. Acín, 2010, Phys. Rev. A **81**, 052111.

Almeida, M. L., S. Pironio, J. Barrett, G. Tóth, and A. Acín, 2007, Phys. Rev. Lett. **99**, 040403.

Ansmann, M., *et al.*, 2009, Nature (London) **461**, 504.

Aolita, L., R. Gallego, A. Acín, A. Chiuri, G. Vallone, P. Mataloni, and A. Cabello, 2012, Phys. Rev. A **85**, 032107.

Aolita, L., R. Gallego, A. Cabello, and A. Acín, 2012a, Phys. Rev. Lett. **108**, 100401.

Aolita, L., R. Gallego, A. Cabello, and A. Acín, 2012b, Phys. Rev. Lett. **108**, 100401.

Araújo, M., M. T. Quintino, D. Cavalcanti, M. F. Santos, A. Cabello, and M. T. Cunha, 2012, Phys. Rev. A **86**, 030101(R).

Aravind, P. K., 2002, Found. Phys. Lett. **15**, 397.

Ardehali, M., 1992, Phys. Rev. A **46**, 5375.

Arnon-Friedman, R., E. Hänggi, and A. Ta-Shma, 2012, arXiv:1205.3736.

Arnon-Friedman, R., and A. Ta-Shma, 2012, Phys. Rev. A **86**, 062333.

Aspect, A., 1975, Phys. Lett. **54A**, 117.

Aspect, A., 1976, Phys. Rev. D **14**, 1944.

Aspect, A., J. Dalibard, and G. Roger, 1982, Phys. Rev. Lett. **49**, 1804.

Aspect, A., P. Grangier, and G. Roger, 1981, Phys. Rev. Lett. **47**, 460.

Aspect, A., P. Grangier, and G. Roger, 1982, Phys. Rev. Lett. **49**, 91.

Augusiak, R., J. Stasinska, C. Hadley, J. Korbicz, M. Lewenstein, and A. Acín, 2011, Phys. Rev. Lett. **107**, 070401.

Augusiak, R., and P. Horodecki, 2006, Phys. Rev. A **74**, 010305.

Avis, D., H. Imai, T. Ito, and Y. Sasaki, 2004, quant-ph/0404014.

Avis, D., H. Imai, T. Ito, and Y. Sasaki, 2005, J. Phys. A **38**, 10 971.

Avis, D., and T. Ito, 2007, Discrete Appl. Math. **155**, 1689.

Avis, D., S. Moriyama, and M. Owari, 2009, IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **E92-A**, 1254.

Babai, L., L. Fortnow, and C. Lund, 1991, Comput. Complex. **1**, 3.

Babichev, S. A., J. Appel, and A. I. Lvovsky, 2004, Phys. Rev. Lett. **92**, 193601.

Bacon, D., and B. Toner, 2003, Phys. Rev. Lett. **90**, 157904.

Banaszek, K., and K. Wódkiewicz, 1998, Phys. Rev. A **58**, 4345.

Banaszek, K., and K. Wódkiewicz, 1999, Phys. Rev. Lett. **82**, 2009.

Bancal, J.-D., C. Branciard, N. Brunner, N. Gisin, and Y.-C. Liang, 2012, J. Phys. A **45**, 125301.

Bancal, J.-D., C. Branciard, N. Brunner, N. Gisin, S. Popescu, and C. Simon, 2008, Phys. Rev. A **78**, 062110.

Bancal, J.-D., C. Branciard, N. Gisin, and S. Pironio, 2009, Phys. Rev. Lett. **103**, 090503.

Bancal, J.-D., N. Brunner, N. Gisin, and Y.-C. Liang, 2011, Phys. Rev. Lett. **106**, 020405.

Bancal, J.-D., N. Gisin, Y.-C. Liang, and S. Pironio, 2011, Phys. Rev. Lett. **106**, 250404.

Bancal, J.-D., N. Gisin, and S. Pironio, 2010, J. Phys. A **43**, 385303.

Bancal, J.-D., S. Pironio, A. Acín, Y.-C. Liang, V. Scarani, and N. Gisin, 2012, Nat. Phys. **8**, 867.

Bardyn, C. E., T. C. H. Liew, S. Massar, M. McKague, and V. Scarani, 2009, Phys. Rev. A **80**, 062337.

Barnum, H., J. Barrett, M. Leifer, and A. Wilce, 2007, Phys. Rev. Lett. **99**, 240501.

Barnum, H., S. Beigi, S. Boixo, M. Elliot, and S. Wehner, 2010, Phys. Rev. Lett. **104**, 140401.

Barnum, H., and A. Wilce, 2012, arXiv:1205.3833.

Barreiro, J., N. Langford, N. Peters, and P. Kwiat, 2005, Phys. Rev. Lett. **95**, 260501.

Barrett, J., 2002, Phys. Rev. A **65**, 042302.

Barrett, J., 2007, Phys. Rev. A **75**, 032304.

Barrett, J., R. Colbeck, and A. Kent, 2012, Phys. Rev. A **86**, 062326.

Barrett, J., R. Colbeck, and A. Kent, 2013, Phys. Rev. Lett. **110**, 010503.

Barrett, J., D. Collins, L. Hardy, A. Kent, and S. Popescu, 2002, Phys. Rev. A **66**, 042111.

Barrett, J., and N. Gisin, 2011, Phys. Rev. Lett. **106**, 4.

Barrett, J., L. Hardy, and A. Kent, 2005, Phys. Rev. Lett. **95**, 010503.

Barrett, J., A. Kent, and S. Pironio, 2006, Phys. Rev. Lett. **97**, 170409.

Barrett, J., N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, 2005, Phys. Rev. A **71**, 022101.

Barrett, J., and S. Pironio, 2005, Phys. Rev. Lett. **95**, 140401.

Barrett, J., S. Pironio, J.-D. Bancal, and N. Gisin, 2013, Phys. Rev. A **88**, 014102.

Bartlett, S., T. Rudolph, and R. Spekkens, 2007, Rev. Mod. Phys. **79**, 555.

Belinskii, A., and D. Klyshko, 1993, Phys. Usp. **36**, 653.

Bell, J., 1964, Physics **1**, 195.

Bell, J. S., 1975, "The theory of local beables," reprinted in Bell, 2004.

Bell, J. S., 1977a, "Atomic-Cascade Photons and Quantum-Mechanical Nonlocality," reprinted in Bell, 2004.

Bell, J. S., 1977b, "Free Variables and Local Causality," reprinted in Bell, 2004.

Bell, J. S., 1990, "La nouvelle cuisine," reprinted in Bell, 2004.

Bell, J. S., 2004, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, England).

Bennett, C. H., G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, 1993, Phys. Rev. Lett. **70**, 1895.

Bennett, C. H., G. Brassard, and N. D. Mermin, 1992, Phys. Rev. Lett. **68**, 557.

Bennett, C. H., P. W. Shor, J. A. Smolin, and A. V. Thapliyal, 2002, IEEE Trans. Inf. Theory **48**, 2637.

Berta, M., O. Fawzi, and S. Wehner, 2012, Advances in Cryptology **7417**, 776.

Boyd, S., and L. Vandenberghe, 2004, *Convex Optimization* (Cambridge University Press, Cambridge, England).

Branciard, C., 2011, Phys. Rev. A **83**, 032123.

Branciard, C., N. Brunner, H. Buhrman, R. Cleve, N. Gisin, S. Portmann, D. Rosset, and M. Szegedy, 2012, Phys. Rev. Lett. **109**, 100401.

Branciard, C., N. Brunner, N. Gisin, C. Kurtsiefer, A. Linares, A. Ling, and V. Scarani, 2008, Nat. Phys. **4**, 681.

Branciard, C., E. Cavalcanti, S. Walborn, V. Scarani, and H. Wiseman, 2012, Phys. Rev. A **85**, 010301.

Branciard, C., and N. Gisin, 2011, Phys. Rev. Lett. **107**, 020401.

Branciard, C., N. Gisin, and S. Pironio, 2010, Phys. Rev. Lett. **104**, 170401.

Branciard, C., D. Rosset, N. Gisin, and S. Pironio, 2012, Phys. Rev. A **85**, 032119.

Branciard, C., D. Rosset, Y.-C. Liang, and N. Gisin, 2013, Phys. Rev. Lett. **110**, 060405.

Brask, J. B., N. Brunner, D. Cavalcanti, and A. Leverrier, 2012, Phys. Rev. A **85**, 042116.

Brask, J. B., and R. Chaves, 2012, Phys. Rev. A **86**, 010103.

Brassard, G., A. Broadbent, and A. Tapp, 2005, Found. Phys. **35**, 1877.

Brassard, G., H. Buhrman, N. Linden, A. Méthot, A. Tapp, and F. Unger, 2006, Phys. Rev. Lett. **96**, 250401.

Brassard, G., R. Cleve, and A. Tapp, 1999, Phys. Rev. Lett. **83**, 1874.

Braunstein, S., and C. Caves, 1988, Phys. Rev. Lett. **61**, 662.

Braunstein, S., and S. Pirandola, 2012, Phys. Rev. Lett. **108**, 130502.

Braunstein, S. L., and C. M. Caves, 1990, Ann. Phys. (N.Y.) **202**, 22.

Braunstein, S. L., A. Mann, and M. Revzen, 1992, Phys. Rev. Lett. **68**, 3259.

Briët, J., H. Buhrman, and B. Toner, 2011, Commun. Math. Phys. **305**, 827.

Briet, J., and T. Vidick, 2013, Commun. Math. Phys. **321**, 181.

Brunner, N., D. Cavalcanti, A. Salles, and P. Skrzypczyk, 2011, Phys. Rev. Lett. **106**, 020402.

Brunner, N., and N. Gisin, 2008, Phys. Lett. A **372**, 3162.

Brunner, N., N. Gisin, S. Popescu, and V. Scarani, 2008, Phys. Rev. A **78**, 052111.

Brunner, N., N. Gisin, and V. Scarani, 2005, New J. Phys. **7**, 88.

Brunner, N., N. Gisin, V. Scarani, and C. Simon, 2007, Phys. Rev. Lett. **98**, 220403.

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

Brunner, N., J. Sharam, and T. Vértesi, 2012, Phys. Rev. Lett. **108**, 110501.

Brunner, N., and P. Skrzypczyk, 2009, Phys. Rev. Lett. **102**, 160403.

Brunner, N., and T. Vértesi, 2012, Phys. Rev. A **86**, 042113.

Brunner, N., A. Young, C. Hu, and J. Rarity, 2013, New J. Phys. **15**, 105006.

Buhrman, H., R. Cleve, S. Massar, and R. de Wolf, 2010, Rev. Mod. Phys. **82**, 665.

Buhrman, H., P. Hoyer, S. Massar, and H. Röhrig, 2003, Phys. Rev. Lett. **91**, 047903.

Buhrman, H., O. Regev, G. Scarpa, and R. de Wolf, 2011, *Proceedings of the 26th IEEE Annual Conference on Computational Complexity (CCC)* (IEEE, New York), p. 157.

Buscemi, F., 2012, Phys. Rev. Lett. **108**, 200401.

Cabello, A., 2001, Phys. Rev. Lett. **86**, 1911.

Cabello, A., 2003, Phys. Rev. Lett. **91**, 230403.

Cabello, A., and J.-A. Larsson, 2007, Phys. Rev. Lett. **98**, 220402.

Cabello, A., D. Rodriguez, and I. Villanueva, 2008, Phys. Rev. Lett. **101**, 120402.

Cabello, A., and F. Sciarrino, 2012, Phys. Rev. X **2**, 021010.

Capasso, V., D. Fortunato, and F. Selleri, 1973, Int. J. Mod. Phys. Conf. Ser. **7**, 319.

Cavalcanti, D., A. Acín, N. Brunner, and T. Vértesi, 2013, Phys. Rev. A **87**, 042104.

Cavalcanti, D., M. L. Almeida, V. Scarani, and A. Acín, 2011, Nat. Commun. **2**, 184.

Cavalcanti, D., N. Brunner, P. Skrzypczyk, A. Salles, and V. Scarani, 2011, Phys. Rev. A **84**, 022105.

Cavalcanti, D., R. Rabelo, and V. Scarani, 2012, Phys. Rev. Lett. **108**, 040402.

Cavalcanti, D., A. Salles, and V. Scarani, 2010, Nat. Commun. **1**, 136.

Cavalcanti, D., and V. Scarani, 2011, Phys. Rev. Lett. **106**, 208901.

Cavalcanti, E., C. Foster, M. Reid, and P. Drummond, 2007, Phys. Rev. Lett. **99**, 210405.

Cavalcanti, E., M. Hall, and H. Wiseman, 2013, Phys. Rev. A **87**, 032306.

Cavalcanti, E., S. Jones, H. Wiseman, and M. Reid, 2009, Phys. Rev. A **80**, 032112.

Ceccarelli, R., G. Vallone, F. D. Martini, P. Mataloni, and A. Cabello, 2009, Phys. Rev. Lett. **103**, 160401.

Cereceda, J., 2002, Phys. Rev. A **66**, 024102.

Cerf, N., and C. Adami, 1997, Phys. Rev. A **55**, 3371.

Cerf, N., N. Gisin, S. Massar, and S. Popescu, 2005, Phys. Rev. Lett. **94**, 220403.

Cerf, N., S. Massar, and S. Pironio, 2002, Phys. Rev. Lett. **89**, 080402.

Chaves, R., D. Cavalcanti, L. Aolita, and A. Acín, 2012, Phys. Rev. A **86**, 012108.

Chaves, R., and T. Fritz, 2012, Phys. Rev. A **85**, 032113.

Chen, J.-L., C. Wu, L. Kwek, and C. Oh, 2008, Phys. Rev. A **78**, 032107.

Christensen, B., *et al.*, 2013, Phys. Rev. Lett. **111**, 130406.

Christof, T., and A. Lobel, 1997, porta, http://typo.zib.de/opt-long_projects/Software/Porta/.

Cirel'son, B. S., 1980, Lett. Math. Phys. **4**, 93.

Clauser, J. F., and M. A. Horne, 1974, Phys. Rev. D **10**, 526.

Clauser, J. F., M. A. Horne, A. Shimony, and R. A. Holt, 1969, Phys. Rev. Lett. **23**, 880.

Clauser, J. F., and A. Shimony, 1978, Rep. Prog. Phys. **41**, 1881.

Cleve, R., and H. Buhrman, 1997, Phys. Rev. A **56**, 1201.

Cleve, R., P. Hoyer, B. Toner, and J. Watrous, 2004, in *19th IEEE Conference on Computational Complexity* (IEEE, New York), p. 236.

Cleve, R., W. Slofstra, F. Unger, and S. Upadhyay, 2007, in *Proceedings of the 22nd Annual Conference on Computational Complexity* (IEEE, New York), pp. 109–114.

Cleve, R., W. van Dam, M. Nielsen, and A. Tapp, 1999, Lect. Notes Comput. Sci. **1509**, 61.

Colbeck, R., 2007, Quantum and relativistic protocols for secure multi-party computation, Ph.D. thesis, University of Cambridge.

Colbeck, R., and A. Kent, 2011, J. Phys. A **44**, 095305.

Colbeck, R., and R. Renner, 2008, Phys. Rev. Lett. **101**, 050403.

Colbeck, R., and R. Renner, 2011, Nat. Commun. **2**, 411.

Colbeck, R., and R. Renner, 2012, Nat. Phys. **8**, 450.

Collins, D., and N. Gisin, 2004, J. Phys. A **37**, 1775.

Collins, D., N. Gisin, N. Linden, S. Massar, and S. Popescu, 2002, Phys. Rev. Lett. **88**, 040404.

Collins, D., N. Gisin, S. Popescu, D. Roberts, and V. Scarani, 2002, Phys. Rev. Lett. **88**, 170405.

Condon, A., 1989, in *Proceedings of the 30th Annual Symposium on Foundations of Computer Science* (IEEE, New York), pp. 462–467.

Coretti, S., E. Hänggi, and S. Wolf, 2011, Phys. Rev. Lett. **107**, 100402.

Cubitt, T. S., D. Leung, W. Matthews, and A. Winter, 2010, Phys. Rev. Lett. **104**, 230503.

Cubitt, T. S., D. Leung, W. Matthews, and A. Winter, 2011, IEEE Trans. Inf. Theory **57**, 5509.

Curty, M., and T. Moroder, 2011, Phys. Rev. A **84**, 010304.

Dada, A., J. Leach, G. Buller, M. Padgett, and E. Andersson, 2011, Nat. Phys. **7**, 677.

D'Ambrosio, V., E. Nagali, S. Walborn, L. Aolita, S. Slussarenko, L. Marrucci, and F. Sciarrino, 2012, Nat. Commun. **3**, 961.

De, A., C. Portmann, T. Vidick, and R. Renner, 2009, arXiv:0912.5514.

De, A. S., U. Sen, C. Brukner, V. Buzek, and M. Zukowski, 2005, Phys. Rev. A **72**, 042310.

Degorre, J., M. Kaplan, S. Laplante, and J. Roland, 2011, Quantum Inf. Comput. **11**, 649.

Degorre, J., S. Laplante, and J. Roland, 2005, Phys. Rev. A **72**, 062314.

Dehlinger, D., and M. Mitchell, 2002, Am. J. Phys. **70**, 903.

de Oliveira Oliveira, M., 2010, arXiv:1009.0771.

Dhara, C., G. Prettico, and A. Acín, 2013, Phys. Rev. A **88**, 052116.

DiCarlo, L., M. Reed, L. Sun, B. Johnson, J. Chow, J. Gambetta, L. Frunzio, S. Girvin, M. Devoret, and R. Schoelkopf, 2010, Nature (London) **467**, 574.

Dicke, R., 1954, Phys. Rev. **93**, 99.

Dinur, I., and O. Reingold, 2006, SIAM J. Comput. **36**, 975.

DiVincenzo, D., and A. Peres, 1997, Phys. Rev. A **55**, 4089.

Doherty, A., and S. Wehner, 2011, New J. Phys. **13**, 073033.

Doherty, A. C., Y. C. Liang, B. Toner, and S. Wehner, 2008, in *IEEE Conference on Computational Complexity* (IEEE, New York), p. 199.

Doherty, A. C., P. A. Parrilo, and F. M. Spedalieri, 2002, Phys. Rev. Lett. **88**, 187904.

Doherty, A. C., P. A. Parrilo, and F. M. Spedalieri, 2004, Phys. Rev. A **69**, 022308.

Dukaric, D., and S. Wolf, 2008, arXiv:0808.3317.

Dür, W., 2001, Phys. Rev. Lett. **87**, 230402.

Eberhard, P., 1993, Phys. Rev. A **47**, R747.

Eibl, M., S. Gaertner, M. Bourennane, C. Kurtsiefer, M. Zukowski, and H. Weinfurter, 2003, Phys. Rev. Lett. **90**, 200403.

Eibl, M., N. Kiesel, M. Bourennane, C. Kurtsiefer, and H. Weinfurter, 2004, Phys. Rev. Lett. **92**, 077901.

Einstein, A., B. Podolsky, and N. Rosen, 1935, Phys. Rev. **47**, 777.

Ekert, A., 1991, Phys. Rev. Lett. **67**, 661.

Elitzur, A., S. Popescu, and D. Rohrlich, 1992, Phys. Lett. A **162**, 25.

Fedrizzi, A., R. Ursin, T. Herbst, M. Nespoli, R. Prevedel, T. Scheidl, F. Tiefenbacher, T. Jennewein, and A. Zeilinger, 2009, Nat. Phys. **5**, 389.

Fehr, S., R. Gelles, and C. Schaffner, 2013, Phys. Rev. A **87**, 012335.

Feige, U., and J. Kilian, 2000, SIAM J. Comput. **30**, 324.

Fine, A., 1982, Phys. Rev. Lett. **48**, 291.

Forster, M., 2011, Phys. Rev. A **83**, 062114.

Forster, M., S. Winkler, and S. Wolf, 2009, Phys. Rev. Lett. **102**, 120401.

Forster, M., and S. Wolf, 2011, Phys. Rev. A **84**, 042112.

Franson, J., 1989, Phys. Rev. Lett. **62**, 2205.

Franz, T., F. Furrer, and R. Werner, 2011, Phys. Rev. Lett. **106**, 250502.

Freedman, S., and J. Clauser, 1972, Phys. Rev. Lett. **28**, 938.

Fritz, T., 2011, Found. Phys. **41**, 1493.

Fritz, T., 2012a, Rev. Math. Phys. **24**, 1250012.

Fritz, T., 2012b, J. Math. Phys. (N.Y.) **53**, 072202.

Fritz, T., A. B. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín, 2013, Nat. Commun. **4**, 2263.

Froissard, M., 1981, Nuovo Cimento B **64**, 241.

Fry, E., and R. Thompson, 1976, Phys. Rev. Lett. **37**, 465.

Fukuda, K., 2003, cdd, http://www.inf.ethz.ch/personal/fukudak/cdd_home/.

Gallego, R., N. Brunner, C. Hadley, and A. Acín, 2010, Phys. Rev. Lett. **105**, 230501.

Gallego, R., L. Masanes, G. de la Torre, C. Dhara, L. Aolita, and A. Acín, 2013, Nat. Commun. **4**, 2654.

Gallego, R., L. Würflinger, A. Acín, and M. Navascues, 2011, Phys. Rev. Lett. **107**, 210403.

Gallego, R., L. Würflinger, A. Acín, and M. Navascues, 2012, Phys. Rev. Lett. **109**, 070401.

Gallego, R., L. Wurflinger, R. Chaves, A. Acin, and M. Navascues, 2014, New J. Phys. **16**, 033037.

Garbarino, G., 2010, Phys. Rev. A **81**, 032106.

Garcia-Patron, R., J. Fiurasek, and N. Cerf, 2005, Phys. Rev. A **71**, 022105.

García-Patron, R., J. Fiurasek, N. Cerf, J. Wenger, R. Tualle-Brouri, and P. Grangier, 2004, Phys. Rev. Lett. **93**, 130409.

Garg, A., and N. D. Mermin, 1984, Found. Phys. **14**, 1.

Genovese, M., 2005, Phys. Rep. **413**, 319.

Gerhardt, I., Q. Liu, A. Lamas-Linares, J. Skaar, V. Scarani, V. Makarov, and C. Kurtsiefer, 2011, Phys. Rev. Lett. **107**, 170404.

Ghose, S., N. Sinclair, S. Debnath, P. Rungta, and R. Stock, 2009, Phys. Rev. Lett. **102**, 250404.

Gilchrist, A., P. Deuar, and M. D. Reid, 1998, Phys. Rev. Lett. **80**, 3169.

Gill, R., 2003, *Mathematical Statistics and Applications: Festschrift for Constance van Eeden*, edited by M. Moore, S. Froda and C. Léger, IMS Lecture Notes—Monograph Series (IMS), Vol. 42, p. 133.

Gill, R., 2012, arXiv:1207.5103.

Gisin, N., 1991, Phys. Lett. A **154**, 201.

Gisin, N., 1996, Phys. Lett. A **210**, 151.

Gisin, N., and B. Gisin, 1999, Phys. Lett. A **260**, 323.

Gisin, N., and B. Gisin, 2002, Phys. Lett. A **297**, 279.

Gisin, N., S. Pironio, and N. Sangouard, 2010, Phys. Rev. Lett. **105**, 070501.

Giustina, M., *et al.*, 2013, Nature (London) **497**, 227.

Goldstein, S., T. Norsen, D. V. Tausk, and N. Zanghi, 2011, Scholarpedia **6**, 8378.

Grandjean, B., Y.-C. Liang, J.-D. Bancal, N. Brunner, and N. Gisin, 2012, Phys. Rev. A **85**, 052113.

Grangier, P., M. Potasek, and B. Yurke, 1988, Phys. Rev. A **38**, 3132.

Greenberger, D., M. Horne, A. Shimony, and A. Zeilinger, 1990, Am. J. Phys. **58**, 1131.

Greenberger, D. M., M. Horne, and A. Zeilinger, 1989, *Bell's Theorem, Quantum Theory, and Conceptions of the Universe* (Kluwer Academic, Dordrecht), p. 73.

Gröblacher, S., T. Paterek, R. Kaltenbaek, C. Brukner, M. Zukowski, M. Aspelmeyer, and A. Zeilinger, 2007, Nature (London) **446**, 871.

Gruca, J., W. Laskowski, M. Zukowski, N. Kiesel, W. Wieczorek, C. Schmid, and H. Weinfurter, 2010, Phys. Rev. A **82**, 012118.

Gühne, O., and A. Cabello, 2008, Phys. Rev. A **77**, 032108.

Gühne, O., G. Toth, P. Hyllus, and H. Briegel, 2005, Phys. Rev. Lett. **95**, 120405.

Hall, M., 2010, Phys. Rev. Lett. **105**, 250404.

Hall, M., 2011, Phys. Rev. A **84**, 022102.

Hänggi, E., and R. Renner, 2010, arXiv:1009.1833.

Hänggi, E., R. Renner, and S. Wolf, 2009, arXiv:0906.4760.

Hänggi, E., R. Renner, and S. Wolf, 2010, *Proceedings of Advances in Cryptology - EUROCRYPT 2010* (Springer, Heidelberg), p. 216.

Hardy, L., 1993, Phys. Rev. Lett. **71**, 1665.

Hardy, L., 2001, arXiv:quant-ph/0101012.

He, Q., E. Cavalcanti, M. Reid, and P. Drummond, 2009, Phys. Rev. Lett. **103**, 180402.

Hein, M., J. Eisert, and H. Briegel, 2004, Phys. Rev. A **69**, 062311.

Herbert, N., 1975, Am. J. Phys. **43**, 315.

Hirsch, F., M. Quintino, J. Bowles, and N. Brunner, 2013, Phys. Rev. Lett. **111**, 160402.

Ho, M., J.-D. Bancal, and V. Scarani, 2013, Phys. Rev. A **88**, 052318.

Hoban, M., E. Campbell, K. Loukopoulos, and D. Browne, 2011, New J. Phys. **13**, 023014.

Hofmann, J., M. Krug, N. Ortegel, L. Grard, M. Weber, W. Rosenfeld, and H. Weinfurter, 2012, Science **337**, 72.

Holenstein, T., 2007, in *Proceedings of the 39th Symposium on the Theory of Computing* (ACM, New York).

Home, D., and F. Selleri, 1991, Riv. Nuovo Cimento **14**, 1.

Horodecki, M., P. Horodecki, and R. Horodecki, 1998, Phys. Rev. Lett. **80**, 5239.

Horodecki, M., P. Horodecki, and R. Horodecki, 1999, Phys. Rev. A **60**, 1888.

Horodecki, R., M. Horodecki, and P. Horodecki, 1996, Phys. Lett. A **222**, 21.

Horodecki, R., P. Horodecki, and M. Horodecki, 1995, Phys. Lett. A **200**, 340.

Impagliazzo, R., L. Levin, and M. Luby, 1989, in *Proceedings of the 21st Symposium on the Theory of Computing* (ACM, New York), p. 12.

Ito, T., and T. Vidick, 2012, in *Proceedings of the 53rd Annual Symposium on the Foundations of Computer Science* (IEEE, New York), pp. 243–252.

Jain, R., Z. Ji, S. Upadhyay, and J. Watrous, 2010, in *Proceedings of the 42nd symposium on Theory of computing* (ACM, New York).

Janotta, P., C. Gogolin, J. Barrett, and N. Brunner, 2011, New J. Phys. **13**, 063024.

Jennewein, T., C. Simon, G. Weihs, H. Weinfurter, and A. Zeilinger, 2000, Phys. Rev. Lett. **84**, 4729.

Ji, S.-W., J. Kim, H.-W. Lee, and M. Z. H. Nha, 2010, Phys. Rev. Lett. **105**, 170404.

Jones, N., and L. Masanes, 2005, Phys. Rev. A **72**, 052312.

Jones, N. S., N. Linden, and S. Massar, 2005, Phys. Rev. A **71**, 042329.

Joshi, P., A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, and R. Horodecki, 2013, Quantum Inf. Comput. **7/8**, 567.

Junge, M., M. Navascues, C. Palazuelos, D. Perez-Garcia, V. Scholz, and R. Werner, 2011, J. Math. Phys. (N.Y.) **52**, 012102.

Junge, M., and C. Palazuelos, 2011, Commun. Math. Phys. **306**, 695.

Junge, M., C. Palazuelos, D. Pérez-Garcia, I. Villanueva, and M. Wolf, 2010, Phys. Rev. Lett. **104**, 170405.

Kaszlikowski, D., P. Gnacínski, M. Zukowski, W. Miklaszewski, and A. Zeilinger, 2000, Phys. Rev. Lett. **85**, 4418.

Kaszlikowski, D., and M. Zukowski, 2000, Phys. Rev. A **61**, 022114.

Kempe, J., H. Kobayashi, K. Matsumoto, B. Toner, and T. Vidick, 2008, in *Proceedings of the 49th Annual Symposium on the Foundations of Computer Science* (IEEE, New York), pp. 447–456.

Kempe, J., H. Kobayashi, K. Matsumoto, B. Toner, and T. Vidick, 2011, SIAM J. Comput. **40**, 848.

Kempe, J., O. Regev, and B. Toner, 2010, SIAM J. Comput. **39**, 3207.

Kempe, J., and T. Vidick, 2011, in *Proceedings of the 43rd Annual Symposium on the Theory of Computing* (ACM, New York), pp. 353–362.

Kent, A., 2005, Phys. Rev. A **72**, 012107.

Khalfin, L., and B. Tsirelson, 1992, Found. Phys. **22**, 879.

Khalfin, L. A., and B. S. Tsirelson, 1985, in *Symposium on the Foundations of Modern Physics*, edited by P. Lahti and P. Mittelstaedt (World Scientific Publishing, Singapore).

Klobus, W., W. Laskowski, M. Markiewicz, and A. Grudka, 2012, Phys. Rev. A **86**, 020302(R).

Koh, D. E., M. W. H. Hall, Setiawan, J. E. Pope, C. Marletto, A. Kay, V. Scarani, and A. Ekert, 2012, Phys. Rev. Lett. **109**, 160404.

König, R., R. Renner, and C. Schaffner, 2009, IEEE Trans. Inf. Theory **55**, 4337.

Kwiat, P., 1997, J. Mod. Opt. **44**, 2173.

Kwiat, P. G., S. Barraza-Lopez, A. Stefanov, and N. Gisin, 2001, Nature (London) **409**, 1014.

Landau, L., 1988, Found. Phys. **18**, 449.

Landau, L. J., 1987, Phys. Lett. A **123**, 115.

Larsson, J.-A., and R. Gill, 2004, Europhys. Lett. **67**, 707.

Larsson, J.-A., M. Giustina, J. Kofler, B. Wittmann, R. Ursin, and S. Ramelow, 2013, arXiv:1309.0712.

Larsson, J.-A., and J. Semitecolos, 2001, Phys. Rev. A **63**, 022117.

Laskowski, W., T. Paterek, C. Brukner, and M. Zukowski, 2010, Phys. Rev. A **81**, 042101.

Laskowski, W., T. Paterek, M. Zukowski, and C. Brukner, 2004, Phys. Rev. Lett. **93**, 200401.

Lasserre, J. B., 2001, SIAM J. Optim. **11**, 796.

Lavoie, J., R. Kaltenbaek, and K. Resch, 2009, New J. Phys. **11**, 073051.

Leggett, A., 2003, Found. Phys. **33**, 1469.

Liang, Y.-C., and A. Doherty, 2006, Phys. Rev. A **73**, 052116.

Liang, Y.-C., and A. C. Doherty, 2007, Phys. Rev. A **75**, 042103.

Liang, Y.-C., N. Harrigan, S. Bartlett, and T. G. Rudolph, 2010, Phys. Rev. Lett. **104**, 050401.

Liang, Y.-C., L. Masanes, and D. Rosset, 2012, Phys. Rev. A **86**, 052115.

Liang, Y.-C., T. Vértesi, and N. Brunner, 2011, Phys. Rev. A **83**, 052325.

Lim, C. C. W., C. Portmann, M. Tomamichel, M. Tomamichel, R. Renner, and N. Gisin, 2013, Phys. Rev. X **3**, 031006.

Linden, N., S. Popescu, A. J. Short, and A. Winter, 2007, Phys. Rev. Lett. **99**, 180502.

Lo, H.-K., M. Curty, and B. Qi, 2012, Phys. Rev. Lett. **108**, 130503.

Lydersen, L., C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, 2010, Nat. Photonics **4**, 686.

Ma, X., and N. Lütkenhaus, 2012, Quantum Inf. Comput. **12**, 0203.

Magniez, F., D. Mayers, M. Mosca, and H. Ollivier, 2006, in *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming*, Lecture Notes in Computer Science Vol. 4051 (Springer, Verlag), p. 72.

Mair, A., A. Vaziri, G. Weihs, and A. Zeilinger, 2001, Nature (London) **412**, 313.

Masanes, L., 2003, Quantum Inf. Comput. **3**, 345.

Masanes, L., 2006, Phys. Rev. Lett. **97**, 050503.

Masanes, L., 2009, Phys. Rev. Lett. **102**, 140501.

Masanes, L., A. Acín, and N. Gisin, 2006, Phys. Rev. A **73**, 012112.

Masanes, L., Y.-C. Liang, and A. Doherty, 2008, Phys. Rev. Lett. **100**, 090403.

Masanes, L., S. Pironio, and A. Acín, 2011, Nat. Commun. **2**, 238.

Masanes, L., R. Renner, M. Christandl, A. Winter, and J. Barrett, 2009, arXiv:quant-ph/0606049.

Massar, S., 2002, Phys. Rev. A **65**, 032121.

Massar, S., D. Bacon, N. Cerf, and R. Cleve, 2001, Phys. Rev. A **63**, 052305.

Massar, S., and S. Pironio, 2001, Phys. Rev. A **64**, 062108.

Massar, S., and S. Pironio, 2003, Phys. Rev. A **68**, 062109.

Massar, S., S. Pironio, J. Roland, and B. Gisin, 2002, Phys. Rev. A **66**, 052112.

Matsukevich, D., P. Maunz, D. Moehring, S. Olmschenk, and C. Monroe, 2008, Phys. Rev. Lett. **100**, 150404.

Matsukevich, D. N., T. Chanelière, M. Bhattacharya, S.-Y. Lan, S. Jenkins, T. Kennedy, and A. Kuzmich, 2005, Phys. Rev. Lett. **95**, 040405.

Mattar, A., J. Brask, and A. Acin, 2013, Phys. Rev. A **88**, 062319.

Maudlin, T., 1992, *Proceedings of the 1992 Meeting of the Philosophy of Science Association*, Vol. **1**, p. 404.

Maudlin, T., 2002, *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics* (Blackwell, Oxford).

Mayers, D., and A. Yao, 1998, *39th Annual Symposium on the Foundations of Computer Science* (IEEE, New York), pp. 503–509.

Mayers, D., and A. Yao, 2004, Quantum Inf. Comput. **4**, 273.

McKague, M., 2010a, arXiv:1010.1989.

McKague, M., 2010b, Ph.D. thesis, Universtiy of Waterloo, arXiv:1006.2352.

McKague, M., and M. Mosca, 2011, in *Proceedings of the 5th Conference on Theory of Quantum Computation, Communication & Cryptography (TQC2010)*, Lecture Notes in Computer Science (Springer, Heidelberg), p. 113.

McKague, M., and L. Sheridan, 2012, .

McKague, M., T. Yang, and V. Scarani, 2012, J. Phys. A **45**, 455304.

Mermin, N., 1986, in *Techniques and and Ideas in Quanturn Measurement Theory*, edited by D. M. Greenberger (New York Academy of Science, New York), p. 422.

Mermin, N. D., 1990a, Phys. Rev. Lett. **65**, 1838.

Mermin, N. D., 1990b, Phys. Today June, **43**, 9.

Mermin, N. D., 1993, Rev. Mod. Phys. **65**, 803.

Méthot, A., and V. Scarani, 2007, Quantum Inf. Comput. **7**, 157.

Miller, C., and Y. Shi, 2012, arXiv:1207.1819.

Mitchell, P., S. Popescu, and D. Roberts, 2004, Phys. Rev. A **70**, 060101.

Moehring, D., M. Madsen, B. Blinov, and C. Monroe, 2004, Phys. Rev. Lett. **93**, 090410.

Moroder, T., J.-D. Bancal, Y.-C. Liang, M. Hofmann, and O. Gühne, 2013, Phys. Rev. Lett. **111**, 030501.

Munro, W. J., 1999, Phys. Rev. A **59**, 4197.

Nagata, K., M. Koashi, and N. Imoto, 2002, Phys. Rev. Lett. **89**, 260401.

Navascues, M., T. Cooney, D. Perez-Garcia, and I. Villanueva, 2011, arXiv:1105.3373.

Navascues, M., S. Pironio, and A. Acín, 2007, Phys. Rev. Lett. **98**, 010401.

Navascues, M., S. Pironio, and A. Acín, 2008, New J. Phys. **10**, 073013.

Navascués, M., and T. Vértesi, 2011, Phys. Rev. Lett. **106**, 060403.

Navascués, M., and H. Wunderlich, 2010, Proc. R. Soc. A **466**, 881.

Nayak, A., 1999, in *Proceedings of the 40th Annual Symposium on Foundations of Computer Science* (IEEE, New York), pp. 369–376.

Neeley, M., *et al.*, 2010, Nature (London) **467**, 570.

Nha, H., and H. Carmichael, 2004, Phys. Rev. Lett. **93**, 020401.

Norsen, T., 2007, arXiv:0707.0401.

Norsen, T., 2009, Found. Phys. **39**, 273.

Oppenheim, J., and S. Wehner, 2010, Science **330**, 1072.

Pal, K., T. Vértesi, and N. Brunner, 2012, Phys. Rev. A **86**, 062111.

Pal, K. F., and T. Vértesi, 2009, Phys. Rev. A **79**, 022120.

Pál, K. F., and T. Vértesi, 2010, Phys. Rev. A **82**, 022116.

Palazuelos, C., 2012a, Phys. Rev. Lett. **109**, 190401.

Palazuelos, C., 2012b, arXiv:1206.3695.

Palsson, M., J. Wallman, A. Bennet, and G. J. Pryde, 2012, Phys. Rev. A **86**, 032322.

Pan, J.-W., D. Bouwmeester, M. Daniell, H. Weinfurter, and A. Zeilinger, 2000, Nature (London) **403**, 515.

Pan, J.-W., Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, and M. Zukowski, 2012, Rev. Mod. Phys. **84**, 777.

Parrilo, P., 2003, Mathematical Programming Ser. B **96**, 293.

Pawlowski, M., and C. Brukner, 2009, Phys. Rev. Lett. **102**, 030403.

Pawlowski, M., and N. Brunner, 2011, Phys. Rev. A **84**, 010302(R).

Pawlowski, M., T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, 2009, Nature (London) **461**, 1101.

Pearle, P., 1970, Phys. Rev. D **2**, 1418.

Peres, A., 1996, Phys. Rev. A **54**, 2685.

Peres, A., 1999, Found. Phys. **29**, 589.

Pérez-Garcia, D., M. Wolf, C. Palazuelos, I. Villanueva, and M. Junge, 2008, Commun. Math. Phys. **279**, 455.

Pironio, S., 2003, Phys. Rev. A **68**, 062102.

Pironio, S., 2004, Ph.D. thesis, Université Libre de Bruxelles.

Pironio, S., 2005, J. Math. Phys. (N.Y.) **46**, 062112.

Pironio, S., A. Acín, N. Brunner, N. Gisin, S. Massar, and V. Scarani, 2009, New J. Phys. **11**, 045021.

Pironio, S., J.-D. Bancal, and V. Scarani, 2011, J. Phys. A **44**, 065303.

Pironio, S., L. Masanes, A. Leverrier, and A. Acín, 2013, Phys. Rev. X **3**, 031007.

Pironio, S., and S. Massar, 2013, Phys. Rev. A **87**, 012336.

Pironio, S., M. Navascues, and A. Acín, 2010, SIAM J. Optim. **20**, 2157.

Pironio, S., *et al.*, 2010, Nature (London) **464**, 1021.

Pitalua-Garcia, D., 2013, Phys. Rev. Lett. **110**, 210402.

Pitkanen, D., X. Ma, R. Wickert, P. van Loock, and N. Lütkenhaus, 2011, Phys. Rev. A **84**, 022325.

Pitowsky, I., 1986, J. Math. Phys. (N.Y.) **27**, 1556.

Pitowsky, I., 1989, *Quantum Probability, Quantum Logic*, Lecture Notes in Physics (Springer, Heidelberg), Vol. 321.

Pomarico, E., J.-D. Bancal, B. Sanguinetti, A. Rochdi, and N. Gisin, 2011, Phys. Rev. A **83**, 052104.

Pomarico, E., B. Sanguinetti, P. Sekatski, H. Zbinden, and N. Gisin, 2011, New J. Phys. **13**, 063031.

Popescu, S., 1994, Phys. Rev. Lett. **72**, 797.

Popescu, S., 1995, Phys. Rev. Lett. **74**, 2619.

Popescu, S., and D. Rohrlich, 1992, Phys. Lett. A **166**, 293.

Popescu, S., and D. Rohrlich, 1994, Found. Phys. **24**, 379.

Portmann, S., C. Branciard, and N. Gisin, 2012, Phys. Rev. A **86**, 012104.

Quintino, M. T., M. Araújo, D. Cavalcanti, M. F. Santos, and M. T. Cunha, 2012, J. Phys. A **45**, 215308.

Qutools, 2005 [http://www.qutools.com].

Rabelo, R., M. Ho, D. Cavalcanti, N. Brunner, and V. Scarani, 2011, Phys. Rev. Lett. **107**, 050502.

Rabelo, R., L. Zhi, and V. Scarani, 2012, Phys. Rev. Lett. **109**, 180401.

Rao, A., 2008, in *Proceedings of STOC* (ACM, New York), pp. 1–10.

Rarity, J., and P. Tapster, 1990, Phys. Rev. Lett. **64**, 2495.

Rastall, P., 1985, Found. Phys. **15**, 963.

Raz, R., 1998, SIAM J. Comput. **27**, 763.

Raz, R., 2008, Proceedings of the 49th Foundations of Computer Science.

Regev, O., 2012, Quantum Inf. Comput. **12**, 9.

Regev, O., and B. Toner, 2007, IEEE Foundations of Computer Science.

Reichardt, B. W., F. Unger, and U. Vazirani, 2012, arXiv:1209.0448.

Reichardt, B. W., F. Unger, and U. Vazirani, 2013, Nature (London) **496**, 456.

Reid, M., 1989, Phys. Rev. A **40**, 913.

Reid, M. D., P. D. Drummond, W. P. Bowen, E. G. Cavalcanti, P. K. Lam, H. A. Bachor, U. L. Andersen, and G. Leuchs, 2009, Rev. Mod. Phys. **81**, 1727.

Renner, R., 2008, Int. J. Quantum. Inform. **06**, 1.

Renner, R., and S. Wolf, 2004, in *Proceedings of the International Symposium on Information Theory (ISIT)* (IEEE, New York).

Romero, J., Giovannini, D. Tasca, S. Barnett, and Padgett, 2013, New J. Phys. **15**, 083047.

Rosenfeld, W., M. Weber, J. Volz, F. Henkel, M. Krug, A. Cabello, M. Zukowski, and H. Weinfurter, 2009, Adv. Sci. Lett. **2**, 469.

Rosset, D., C. Branciard, Y.-C. Liang, and N. Gisin, 2013, New J. Phys. **15**, 053025.

Rowe, M., D. Kielpinski, V. Meyer, C. Sackett, W. Itano, C. Monroe, and D. Wineland, 2001, Nature (London) **409**, 791.

Ryu, J., C. Lee, M. Zukowski, and J. Lee, 2013, Phys. Rev. A **88**, 042101.

Salart, D., A. Baas, C. Branciard, N. Gisin, and H. Zbinden, 2008, Nature (London) **454**, 861.

Salles, A., D. Cavalcanti, and A. Acín, 2008, Phys. Rev. Lett. **101**, 040404.

Salles, A., D. Cavalcanti, A. Acín, D. Perez-García, and M. Wolf, 2010, Quantum Inf. Comput. **10**, 0703.

Sangouard, N., J.-D. Bancal, N. Gisin, W. Rosenfeld, P. Sekatski, M. Weber, and H. Weinfurter, 2011, Phys. Rev. A **84**, 052122.

Sangouard, N., J.-D. Bancal, P. Müller, J. Ghosh, and J. Eschner, 2013, New J. Phys. **15**, 085004.

Santos, E., 1992, Phys. Rev. A **46**, 3646.

Saunders, D. J., S. J. Jones, H. M. Wiseman, and G. J. Pryde, 2010, Nat. Phys. **6**, 845.

Scarani, V., 2008, Phys. Rev. A **77**, 042112.

Scarani, V., A. Acín, E. Schenk, and M. Aspelmeyer, 2005, Phys. Rev. A **71**, 042325.

Scarani, V., and N. Gisin, 2001, Phys. Rev. Lett. **87**, 117901.

Scarani, V., and N. Gisin, 2005, Braz. J. Phys. **35**, 2A.

Scarani, V., N. Gisin, N. Brunner, L. Masanes, S. Pino, and A. Acín, 2006, Phys. Rev. A **74**, 042339.

Scheidl, T., *et al.*, 2010, Proc. Natl. Acad. Sci. U.S.A. **107**, 19 708.

Schmidt, C., N. Kiesel, W. Laskowski, W. Wieczorek, M. Zukowski, and H. Weinfurter, 2008, Phys. Rev. Lett. **100**, 200407.

Scholz, V. B., and R. F. Werner, 2008, arXiv:0812.4305.

Schrijver, A., 1989, *Theory of linear and integer programming*, Wiley-Interscience Series in Discrete Mathematics (John Wiley & Sons, New York).

Schrödinger, E., 1936, Proc. Cambridge Philos. Soc. **32**, 446.

Seevinck, M., and G. Svetlichny, 2002, Phys. Rev. Lett. **89**, 060401.

Sen(De), A., U. Sen, M. Wiesniak, D. Kaszlikowski, and M. Zukowski, 2003, Phys. Rev. A **68**, 062306.

Sen(De), A., U. Sen, and M. Zukowski, 2002, Phys. Rev. A **66**, 062318.

Shadbolt, P., T. Vértesi, Y.-C. Liang, C. Branciard, N. Brunner, and J. O'Brien, 2012, Sci. Rep. **2**, 470.

Shimony, A., M. Horne, and J. Clauser, 1976, Epistemological Letters **13**, 17.1.

Short, A., 2009, Phys. Rev. Lett. **102**, 180502.

Silman, J., A. Chailloux, N. Aharon, I. Kerenidis, S. Pironio, and S. Massar, 2011, Phys. Rev. Lett. **106**, 220501.

Silman, J., S. Pironio, and S. Massar, 2013, Phys. Rev. Lett. **110**, 100504.

Simon, C., and W. Irvine, 2003, Phys. Rev. Lett. **91**, 110405.

Sliwa, C., 2003, Phys. Lett. A **317**, 165.

Smith, D., *et al.*, 2012, Nat. Commun. **3**, 625.

Steeg, G. V., and S. Wehner, 2009, Quantum Inf. Comput. **9**, 801.

Stefanov, A., H. Zbinden, N. Gisin, and A. Suarez, 2002, Phys. Rev. Lett. **88**, 120404.

Stefanov, A., H. Zbinden, N. Gisin, and A. Suarez, 2003, Phys. Rev. A **67**, 042115.

Steiner, M., 2000, Phys. Lett. A **270**, 239.

Stobińska, M., H. Jeong, and T. Ralph, 2007, Phys. Rev. A **75**, 052105.

Suarez, A., and V. Scarani, 1997, Phys. Lett. A **232**, 9.

Svetlichny, G., 1987, Phys. Rev. D **35**, 3066.

Tan, S., D. Walls, and M. Collett, 1991, Phys. Rev. Lett. **66**, 252.

Tasca, D., S. Walborn, F. Toscano, and P. S. Ribeiro, 2009, Phys. Rev. A **80**, 030101(R).

Ta-Shma, A., 2009, in *Proceedings of the 41st ACM STOC* (ACM, New York), p. 401.

Teo, C., M. Araújo, M. T. Quintino, J. Minár, D. Cavalcanti, V. Scarani, M. T. Cunha, and M. F. Santos, 2013, Nat. Commun. **4**, 2104.

Terhal, B., 2004, IBM J. Res. Dev. **48**, 71.

Terhal, B., A. Doherty, and D. Schwab, 2003, Phys. Rev. Lett. **90**, 157903.

Teufel, S., K. Berndl, D. Dürr, S. Goldstein, and N. Zanghı`, 1997, Phys. Rev. A **56**, 1217.

Thew, R., A. Acín, H. Zbinden, and N. Gisin, 2004, Phys. Rev. Lett. **93**, 010503.

Tittel, W., J. Brendel, H. Zbinden, and N. Gisin, 1998, Phys. Rev. Lett. **81**, 3563.

Tittel, W., J. Brendel, H. Zbinden, and N. Gisin, 1999, Phys. Rev. Lett. **82**, 2594.

Tittel, W., J. Brendel, H. Zbinden, and N. Gisin, 2000, Phys. Rev. Lett. **84**, 4737.

Tomamichel, M., S. Fehr, J. Kaniewski, and S. Wehner, 2012, arXiv:1210.4359.

Toner, B., 2009, Proc. R. Soc. A **465**, 59.

Toner, B., and D. Bacon, 2003, Phys. Rev. Lett. **91**, 187904.

Toner, B., and F. Verstraete, 2006, arXiv:quant-ph/0611001.

Toth, G., O. Gühne, and H. Briegel, 2006, Phys. Rev. A **73**, 022303.

Tsirelson, B., 1987, J. Sov. Math. **36**, 557.

Tsirelson, B. S., 1993, Hadronic J. Suppl. **8**, 329.

Uffink, J., 2002, Phys. Rev. Lett. **88**, 230406.

Vadhan, S., 2012, http://people.seas.harvard.edu/~salil/pseudorandomness/.

Vaidman, L., 2001, Phys. Lett. A **286**, 241.

van Dam, W., 2005, quant-ph/0501159v1.

van Dam, W., P. Grunwald, and R. Gill, 2005, IEEE Trans. Inf. Theory **51**, 2812.

van Dam, W., and P. Hayden, 2003, Phys. Rev. A **67**, 060302(R).

van Loock, P., and S. L. Braunstein, 2001, Phys. Rev. A **63**, 022106.

Vazirani, U., and T. Vidick, 2012a, *STOC '12 Proceedings of the 44th symposium on Theory of Computing* (ACM, New York), p. 61.

Vazirani, U., and T. Vidick, 2012b, Phil. Trans. R. Soc. A **370**, 3432.

Verstraete, F., and M. Wolf, 2002, Phys. Rev. Lett. **89**, 170401.

Vértesi, T., 2008, Phys. Rev. A **78**, 032112.

Vértesi, T., and E. Bene, 2009, Phys. Rev. A **80**, 062316.

Vértesi, T., and N. Brunner, 2012, Phys. Rev. Lett. **108**, 030403.

Vértesi, T., and M. Navascues, 2011, Phys. Rev. A **83**, 062112.

Vértesi, T., and K. Pál, 2009, Phys. Rev. A **79**, 042106.

Vértesi, T., and K. F. Pal, 2008, Phys. Rev. A **77**, 042106.

Vértesi, T., S. Pironio, and N. Brunner, 2010, Phys. Rev. Lett. **104**, 060401.

Vidick, T., and S. Wehner, 2011, Phys. Rev. A **83**, 052310.

Vongehr, S., 2012, arXiv:1207.5294.

Walborn, S. P., A. Salles, R. M. Gomes, F. Toscano, and P. H. S. Ribeiro, 2011, Phys. Rev. Lett. **106**, 130402.

Wallman, J., and S. Bartlett, 2012, Phys. Rev. A **85**, 024101.

Walther, P., M. Aspelmeyer, K. J. Resch, and A. Zeilinger, 2005, Phys. Rev. Lett. **95**, 020403.

Wang, Z., and D. Markham, 2012, Phys. Rev. Lett. **108**, 210407.

Wehner, S., 2006a, in *Proceedings of 23rd STACS*, LNCS Vol. **3884** (Springer, Heidelberg), pp. 162–171.

Wehner, S., 2006b, Phys. Rev. A **73**, 022110.

Wehner, S., M. Christandl, and A. Doherty, 2008, Phys. Rev. A **78**, 062112.

Weihs, G., T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, 1998, Phys. Rev. Lett. **81**, 5039.

Wenger, J., M. Hafezi, F. Grosshans, R. Tualle-Brouri, and P. Grangier, 2003, Phys. Rev. A **67**, 012105.

Werner, R. F., 1989, Phys. Rev. A **40**, 4277.

Werner, R. F., and M. M. Wolf, 2000, Phys. Rev. A **61**, 062102.

Werner, R. F., and M. M. Wolf, 2001a, Quantum Inf. Comput. **1**, 1.

Werner, R. F., and M. M. Wolf, 2001b, Phys. Rev. A **64**, 032112.

White, A., D. James, P. Eberhard, and P. Kwiat, 1999, Phys. Rev. Lett. **83**, 3103.

Wilms, J., Y. Disser, G. Alber, and I. C. Percival, 2008, Phys. Rev. A **78**, 032116.

Wiseman, H., S. J. Jones, and A. Doherty, 2007, Phys. Rev. Lett. **98**, 140402.

Wittmann, B., S. Ramelow, F. Steinlechner, N. Langford, N. Brunner, H. Wiseman, R. Ursin, and A. Zeilinger, 2012, New J. Phys. **14**, 053030.

Wolf, S., and J. Wullschleger, 2005, arXiv:quant-ph/0502030v1.

Wood, C., and R. Spekkens, 2012, arXiv:1208.4119.

Yang, T., D. Cavalcanti, M. Almeida, C. Teo, and V. Scarani, 2012, New J. Phys. **14**, 013061.

Yang, T., and M. Navascues, 2013, Phys. Rev. A **87**, 050102.

Yang, T. H., M. Navascués, L. Sheridan, and V. Scarani, 2011, Phys. Rev. A **83**, 022105.

Yu, S., and C. Oh, 2013, arXiv:1306.5330.

Zbinden, H., J. Brendel, N. Gisin, and W. Tittel, 2001, Phys. Rev. A **63**, 022111.

Zeilinger, A., 1999, Rev. Mod. Phys. **71**, S288.

Zhang, Y., S. Glancy, and E. Knill, 2011, Phys. Rev. A **84**, 062118.

Zhang, Y., E. Knill, and S. Glancy, 2013, Phys. Rev. A **88**, 052119.

Zhao, Z., T. Yang, Y.-A. Chen, A.-N. Zhang, M. Zukowski, and J.-W. Pan, 2003, Phys. Rev. Lett. **91**, 180401.

Ziegler, G. M., 1995, *Lectures on Polytopes*, Graduate Texts in Mathematics (Springer-Verlag, New York), Vol. 152.

Zukowski, M., and C. Brukner, 2002, Phys. Rev. Lett. **88**, 210401.

Zukowski, M., D. Greenberger, M. Horne, and A. Zeilinger, 2008, Phys. Rev. A **78**, 022111.

Zukowski, M., R. Horodecki, M. Horodecki, and P. Horodecki, 1998, Phys. Rev. A **58**, 1694.

Zukowski, M., and D. Kaszlikowski, 1999, Phys. Rev. A **59**, 3200.

Zukowski, M., D. Kaszlikowski, A. Baturo, and J.-A. Larsson, 1999, quant-ph/9910058 .

Zukowski, M., A. Zeilinger, M. Horne, and A. Ekert, 1993, Phys. Rev. Lett. **71**, 4287.