From Bell's Theorem to Secure Quantum Key Distribution

Antonio Acín,[1] Nicolas Gisin,[2] and Lluis Masanes[3]

[1]*ICFO—Institut de Ciencies Fotoniques, Mediterranean Technology Park, 08860 Castelldefels (Barcelona), Spain*
[2]*GAP-Optique, University of Geneva, 20 Rue de l'École de Médecine, CH-1211 Geneva 4, Switzerland*
[3]*School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom*

(Received 21 October 2005; published 20 September 2006)

The first step in any quantum key distribution (QKD) protocol consists of sequences of measurements that produce correlated classical data. We show that these correlation data must violate some Bell inequality in order to contain distillable secrecy, if not they could be produced by quantum measurements performed on a separable state of larger dimension. We introduce a new QKD protocol and prove its security against any individual attack by an adversary only limited by the no-signaling condition.

DOI: 10.1103/PhysRevLett.97.120405 PACS numbers: 03.65.Ud, 03.67.Dd

Since its invention in 1984 by Bennett and Brassard [1], quantum key distribution (QKD) has led to a change of paradigm with respect to the existing classical cryptographic schemes. While the latter ones base their security on mathematical assumptions, the security of QKD relies on the fact that the devices of the legitimate parties, Alice and Bob, and of the eavesdropper, Eve, are governed by quantum physics. Then, security proofs of QKD schemes exploit well-established quantum features such as the no-cloning theorem, or the monogamy (i.e., nonshareability) of entanglement. The common structure in all these schemes is the following: Alice and Bob first establish some correlation, for instance by performing some measurements on a quantum state or, equivalently, by Alice sending some quantum states that are later measured by Bob [2]. Next, exploiting the quantum formalism, Alice and Bob bound Eve's information and distill a secret key by public communication. Formally, a correlation is a conditional probability distribution $P(a, b|x, y)$, where a and b are Alice and Bob's output data, respectively, and x and y are their choices of inputs. For instance, x and y could be their choice of measurement settings and a and b the obtained results.

There is, however, an extra assumption in most of the existing QKD protocols: Alice and Bob know how the correlation $P(a, b|x, y)$ has been established. For example, in a photon-polarization implementation of the Bennett-Brassard 1984 (BB84) protocol [1]. Alice knows she sends photons in two polarization bases, z and x, that are measured by Bob in the same bases. In other words, Alice and Bob *trust their devices* [3]. This often implicit assumption, as discussed below, is crucial for the security of standard QKD. Therefore, all security proofs of QKD are based on (i) the validity of the quantum formalism *plus* the assumption that (ii) the legitimate partners perfectly know how their correlation is established, e.g., they know the dimensions of the Hilbert space describing their quantum systems. Indeed, all the security proofs of existing QKD schemes heavily exploit the Hilbert-space artillery of quantum physics. Experimentally, additional Hilbert-space dimensions can correspond to "side channels," i.e., to degrees of freedom coded accidentally. For example, in photon-polarization coding, the wavelength could be accidentally correlated to the state of polarization.

It is desirable to have key distribution schemes where assumption (ii) is not needed for the security. In such schemes, Alice and Bob should just exploit a well-established physical principle to extract a key from some observed correlation, $P(a, b|x, y)$, without having to care about the practical details needed for the correlation distribution [4]. Standard QKD proofs do not fit into this desired picture. Indeed, an immediate *necessary condition* for some correlation to be secure in this device-independent scenario is that $P(a, b|x, y)$ should not allow a description in terms of local classical variables: $P(a, b|x, y) \neq \sum_\lambda p_\lambda P(a|x, \lambda)P(b|y, \lambda)$. If not, the adversary Eve may hold a copy of the λ's. This necessary condition implies that the correlation $P(a, b|x, y)$ must violate a Bell inequality [5]. Thus, any security proof of QKD in this more general scenario should make a direct use of quantum nonlocality [6,7]. Note that the correlation corresponding to the well-known BB84 protocol [1] does not satisfy this condition.

In this Letter, we present a new 4-state QKD protocol directly built from data violating the Clauser-Horne-Shimony-Holt (CHSH) Bell inequality [8]:

$$P(a_0 = b_0) + P(a_0 = b_1) + P(a_1 = b_0) + P(a_1 \neq b_1) \leq 3, \tag{1}$$

where $P(a_j = b_k) = P(a = b = 0|x = j, y = k) + P(a = b = 1|x = j, y = k)$. Hence we name our new protocol the CHSH protocol. Following the ideas introduced in [9], we demonstrate its security against any individual attack by any adversary only limited by the no-signaling condition. The no-signaling condition says that local probabilities are independent of distant partner's inputs, e.g.,

$$P(a|x, y) = \sum_b P(a, b|x, y) = P(a|x). \tag{2}$$

This principle severely limits the set of possible correlations. For finite alphabets of inputs and outputs, the set of all these correlations is convex with a finite number of extreme points; hence it is a polytope.

The choice of the no-signaling principle for proving the security of our protocol is twofold. First, it is perhaps the strongest physical principle. Actually, Eve could be *supra-quantum*, since there are nonsignaling correlations not achievable using quantum states [10]. Second, it allows one to study key distribution based on physical assumptions beyond quantum physics, as shown in [9]. Indeed, thanks to the seminal paper by Popescu and Rohrlich [10], it was realized that one can study *nonlocality without Hilbert space* using the no-signaling principle. Several recent papers explore this new avenue [11–14].

Before proceeding, let us illustrate on the BB84 protocol the necessity of a Bell inequality violation for security. Ideally, in the noise-free case, the BB84 correlation satisfies $P(a = b|x = y) = 1$ and $P(a = b|x \neq y) = 1/2$. If such correlation results from measurements in the x and z bases on qubit pairs, then the state of these two qubits is necessarily maximally entangled and security follows. However, the same correlation can also be reproduced by the four-qubit state:

$$\rho_{AB} = \tfrac{1}{4}(|00\rangle\langle 00|_z + |11\rangle\langle 11|_z) \otimes (|00\rangle\langle 00|_x + |11\rangle\langle 11|_x). \tag{3}$$

Here, Alice holds the first and third qubit. Whenever she measures in the z (x) basis, she is actually measuring the first (third) qubit in this basis. The same happens for Bob, with the second and fourth qubit. Clearly, their measurement results are completely correlated when the bases agree and uncorrelated otherwise, precisely as for the ideal BB84 case. However, their state is separable, so a secure key cannot be established. The BB84 protocol becomes insecure even in the ideal noise-free situation.

It is now time to present the proposed CHSH protocol. Alice and Bob have for each realization the choice between two measurements with binary outcomes. It is essential that the obtained correlation violates the CHSH inequality (1). For example, Alice and Bob could share a Werner state $\rho_W = W P_{\phi^+} + (1 - W)\mathbb{1}/4$, with visibility W, where P_{ϕ^+} denotes the projector onto $|\phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, and perform the measurements that maximize the violation of the CHSH inequality (1). However, any other way to obtain data violating (1) is equally good. Violation of (1) implies that in three out of the four measurement choices, Alice and Bob are correlated [the three first terms in (1)], though not necessarily maximally, while in the fourth case they are anticorrelated. Therefore, the analog of basis reconciliation goes as follows: Bob announces all his measurement settings y, and Alice keeps all her data, but for the case of anticorrelation, she flips her bit. Let us denote by $P(a, b|x, y)$ and $\tilde{P}(a, b|x, y)$ the correlation before and after this basis reconciliation. Compared to the BB84 protocol, the partners keep all data; however, even in the ideal noise-free case, the correlation is not maximal: $\tilde{P}(a = b|x, y) < 1$. In the sequence, we limit our analysis to isotropic raw correlations with visibility V, without loss of generality [15]:

$$P(a, b|x, y) = V\tfrac{1}{2}\delta(a + b = xy) + (1 - V)\tfrac{1}{4}, \tag{4}$$

where $\delta(r = s) = 1$ whenever the equality holds modulo 2, and 0 otherwise. For $V \leq 1/\sqrt{2}$, such correlations can be distributed by quantum physics (e.g., by a Werner state with $W = \sqrt{2}V$). For $V > 1/2$ they violate the Bell inequality (1), so there are no local variables that Eve could hold.

As usual, we conservatively assume that the distribution of the correlation is done by Eve. Any attack consists thus of a three-party distribution $P(a, b, e|x, y, z)$ whose marginal is (4)

$$\begin{aligned} P(a, b|x, y) &= \sum_e P(a, b, e|x, y, z) \\ &= \sum_e P(e|z) P(a, b|x, y, z, e), \end{aligned} \tag{5}$$

where for the second equality we used the no-signaling condition: Eve's output e is independent of Alice and Bob's inputs x and y. Note that the no-signaling condition implies that even if Eve and Bob collaborated, they could not get any information about Alice's input x [6]. We can restrict our considerations to attacks where Eve prepares extreme points of Alice and Bob's no-signaling polytope. Indeed, consider an attack where some of the terms appearing in Eq. (5) do not correspond to extreme points of this polytope. These terms can be expressed as a convex combination of extreme points:

$$P(a, b|x, y, z, e) = \sum_\lambda P_{\text{ext}}(a, b|x, y, z, e, \lambda) P(\lambda). \tag{6}$$

Giving the knowledge of λ to Eve, one has an attack consisting of extreme points that is, at least, as good as the previous one, since [cf. (5)],

$$P(a, b|x, y) = \sum_{e,\lambda} P(e, \lambda|z) P_{\text{ext}}(a, b|x, y, z, e, \lambda). \tag{7}$$

We need to recall now some facts about nonsignaling correlations with binary input and output. Barrett and coworkers proved that in the binary case, the number of extreme nonsignaling correlations is very limited [11]. If, moreover, one concentrates on the correlations that violate the CHSH inequality, one finds a unique extreme correlation that violates it; this is the isotropic correlation (4) with $V = 1$. This correlation appears in the literature as Popescu-Rohrlich box [11], or nonlocal machine [12] or unit of nonlocality [13]. Finally, there are 8 local extreme deterministic correlations, $P(a, b|x, y) = P_D(a|x) P_D(b|y)$, that saturate the inequality (1); see Fig. 1.

For the local correlations, Eve knows the measurement outcomes a_0, a_1, b_0, and b_1. However, if Alice and Bob

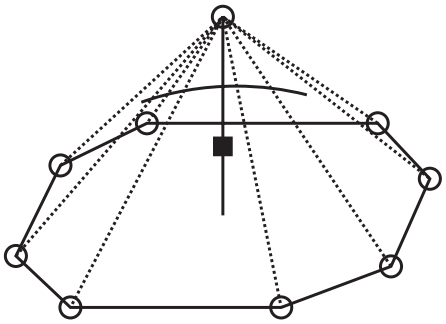

FIG. 1. Pictorial representation of nonsignaling correlations $P(a, b|x, y)$ for binary inputs and outputs. The closed thick line defines one of the facets of the polytope of local correlations, corresponding to the CHSH inequality. All the extreme points lying on this facet are also extreme points of the more general polytope of nonsignaling correlations. Only one extreme point is on top of the CHSH facet, given by the nonlocal machine. The curved line schematically represents the region of points achievable using quantum states. Isotropic correlations (4) lie along the vertical line starting from the nonlocal machine and entering the local polytope through the center of the CHSH facet. In the optimal eavesdropping attack, Eve simulates Alice and Bob's correlation, square point, by the mixture (i.e., convex combination) of extreme points of the nonsignaling polytope, circles in the figure.

share a nonlocal machine, they have the guarantee of perfect monogamy, because Eve cannot be correlated at all [11]. Eve's optimal attack then consists of the combination of extreme points that mimics Alice and Bob's correlation with the minimal weight for nonlocal points. Therefore, she prepares only those local points that are closer to Alice and Bob's correlation. This can be easily understood in Fig. 1: in order to reproduce the quantum correlation observed by Alice and Bob, represented by a square, Eve should send an equal mixture of the eight local points lying on the facet, plus the nonlocal machine on top of it. In what follows p_{NL} denotes the probability that Eve prepares a nonlocal machine. The Bell violation observed by Alice and Bob fixes the value of $p_{NL} = 2V - 1$. When the observed data are local, Eve can mimic them with deterministic local points. However, when the correlation is nonlocal, Eve is forced to sometimes send a nonlocal machine, in which case she holds no information about Alice's symbol a because of the no-signaling principle.

The resulting tripartite probability distribution, after basis reconciliation, is summarized in Table I, where $p_L = 1 - p_{NL}$. Eve's information on Alice and Bob's symbols is represented by two variables (e_a, e_b). The value at each position of the table gives the probability for the corresponding outcomes, e.g., $P(a = 0, b = 0, e = (?, 0)) = p_L/8$. Since only Bob announces his measurement, Eve sometimes has full information on Bob's but not on Alice's symbol after the basis reconciliation, even if her preparation was local. For instance, this is the case when Bob announces $y = 1$ and Eve has prepared a local point where $a_0 = a_1$. Moreover, one can see that, due to the properties

TABLE I. Eve's optimal individual attack. Alice and Bob's variables are binary, while Eve's information can be represented by two ternary variables, $e_a, e_b = 0, 1, ?$. The value "?" denotes those cases where she has no information. For example, $(?, 0)$ means that Eve knows $b = 0$ but not a.

b / a (e)	0	1
0	$(0, 0)$ $p_L/4$ $(?, 0)$ $p_L/8$ $(?, ?)$ $p_{NL}/2$	$(?, 0)$ $p_L/8$
1	$(?, 1)$ $p_L/8$	$(1, 1)$ $p_L/4$ $(?, 1)$ $p_L/8$ $(?, ?)$ $p_{NL}/2$

of the local points lying on the CHSH facet, Eve knows both symbols only when $a = b$.

Once the optimal individual attack has been determined, one can study the secrecy properties of the resulting probability distribution. The detailed calculation of these results will be given in a forthcoming paper [16]. Here we merely summarize our findings.

A positive secret key rate is achievable with one-way communication protocols. Since Bob announces his bases y, Eve's information on his symbol is larger than on Alice's. Consequently, the flow of information has to go from Alice to Bob. From Table I, one has that Bob's error probability is $\varepsilon_B = p_L/4$, while $I(A{:}E) = p_L/2$. Accordingly, the one-way key rate, $K^{\rightarrow}$, satisfies [17]

$$K^{\rightarrow} \geq I(A{:}B) - I(A{:}E) = 1 - h(p_L/4) - \frac{p_L}{2}, \qquad (8)$$

where h is the binary entropy. This quantity is positive for $p_{NL} \gtrsim 0.318$. The quantum region is given by $p_{NL} \leq \sqrt{2} - 1 \simeq 0.414$. This implies that some correlations that can be distributed by quantum states provide secret key secure against any no-signaling adversary (not necessarily restricted to quantum physics).

The secret key rate achievable with the two-way advantage distillation protocol of [18] is positive if $p_{NL} > 1/5$.

All correlations violating the CHSH inequality contain secrecy, not in the sense that a secret key can necessarily be generated from them, but in the weaker sense that they could not be broadcast without consuming secret bits. Indeed, the intrinsic information $I_{\downarrow} = I(A{:}B \downarrow E)$, which measures the amount of secret correlations in a probability distribution [19], is positive for the whole region of the Bell violation:

$$I_{\downarrow} = \left(1 - \frac{p_L}{2}\right)\left[1 - h\left(\frac{p_L}{4 - 2p_L}\right)\right]. \qquad (9)$$

Preprocessing by Alice and Bob, consisting in applying random flips to their data as studied in Ref. [20], helps. Using preprocessing, the one-way key rate as a function of the disturbance D in the quantum channel is shown in Fig. 2. The disturbance is defined in the standard way, namely, $D = 0$ corresponds to a perfect channel, and

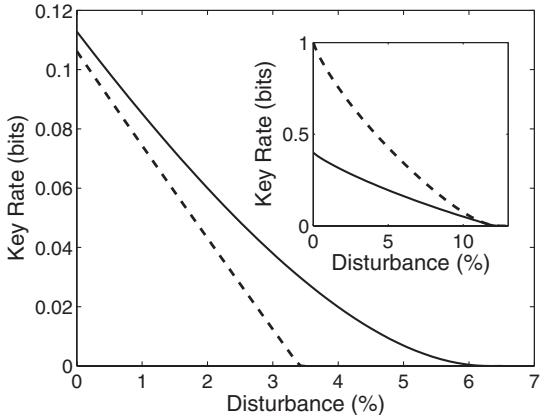

FIG. 2. Key rates using one-way communication against nonsignaling individual attacks. The dashed line corresponds to the situation where no preprocessing is employed, while the thick one to the optimal preprocessing. In the inset, the key rate against a standard quantum eavesdropper is given, compared to the BB84 protocol (dashed line).

$p_{NL} = \sqrt{2}(1 - 2D) - 1$. The critical disturbance is $D \lesssim 6.3\%$. In the case of two-way communication with preprocessing, a positive key rate is obtained when $D \lesssim 11.36\%$, or $p_{NL} \gtrsim 0.093$, still not sufficient to cover the whole region of the Bell violation.

All this analysis leaves as an open question whether secret key distillation is possible for all $p_{NL} > 0$. This is a difficult question because no lower bound on the secret key rate achievable with two-way communication is known. We investigated all known protocols and found that the optimal consists in the combination of preprocessing followed by advantage distillation, which gives $p_{NL} \gtrsim 0.093$. A consequence of these findings is the following interesting alternative: the probability distribution of Table I, for a small Bell violation, either contains bipartite bound information [21] or is distillable using a new technique.

To summarize, usual security proofs of QKD are based on (i) quantum physics and (ii) the perfect knowledge of the physical devices used for the correlation distribution. If one would like to remove this second assumption and construct device-independent key distribution protocols, the Alice-Bob correlation must violate a Bell inequality. If this is not the case, the data are insecure already against a classical eavesdropper. We presented a QKD protocol aimed at producing data that violate the CHSH inequality. We proved its security against the most general individual attack without signaling, independently of any assumption about Hilbert spaces. To our knowledge, our results represent the first step towards the characterization of optimal nonsignaling eavesdropping attacks. Compared to the results in [9], our analysis covers the noisy situation, but it restricts Eve to an individual attack. Moreover, our protocol is simpler and can easily be implemented using today's technology, contrary to the protocol in [9], which uses many arbitrarily close quantum states. We would like to conclude with a comment on the role played by Bell inequalities. It is often said that they are just examples of entanglement witnesses. However, as shown here, they are more than this, since they witness useful correlations independent of the Hilbert-space structure.

This work is supported by a Spanish MEC "Ramón y Cajal" grant, the Swiss NCCR "Quantum Photonics" and OFES within the EU project RESQ (No. IST-2001-37559), and the U.K. EPSRC (IRC QIP).

[1] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), p. 175.
[2] C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).
[3] D. Mayers and A. Yao, quant-ph/9809039.
[4] Of course, it is assumed that there is no information leakage from the devices.
[5] J. S. Bell, Physics (Long Island City, N.Y.) **1**, 195 (1964).
[6] Note that, similar to what happens for entanglement and standard QKD schemes [2], the spacelike configuration is not required for the realization of the protocol.
[7] This is somehow closer to Ekert's protocol [A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991)]. However, the Bell violation plays a more fundamental role in our scenario [2].
[8] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[9] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).
[10] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[11] J. Barrett *et al.*, Phys. Rev. A **71**, 022101 (2005).
[12] N. J. Cerf, N. Gisin, S. Massar, and S. Popescu, Phys. Rev. Lett. **94**, 220403 (2005).
[13] N. S. Jones and Ll. Masanes, quant-ph/0506182; J. Barrett and S. Pironio, Phys. Rev. Lett. **95**, 140401 (2005).
[14] Ll. Masanes, A. Acín, and N. Gisin, Phys. Rev. A **73**, 012112 (2006).
[15] Indeed, Alice and Bob can map any binary correlation into this isotropic correlation by local operations and classical communication keeping the CHSH violation [14].
[16] V. Scarani *et al.*, quant-ph/0606197 [Phys. Rev. A (to be published)].
[17] I. Csiszár and J. Körner, IEEE Trans. Inf. Theory **24**, 339 (1978).
[18] U. M. Maurer, IEEE Trans. Inf. Theory **39**, 733 (1993).
[19] U. Maurer and S. Wolf, IEEE Trans. Inf. Theory **45**, 499 (1999).
[20] B. Kraus, N. Gisin, and R. Renner, Phys. Rev. Lett. **95**, 080501 (2005).
[21] N. Gisin and S. Wolf, in *Proceeding of CRYPTO 2000*, Lecture Notes in Computer Science Vol. 1880 (Springer-Verlag, Berlin, 2000), p. 482.