Article

Device-independent security of quantum cryptography against collective attacks

ACIN, Antonio, *et al*.

Abstract

We present the optimal collective attack on a quantum key distribution protocol in the "device-independent" security scenario, where no assumptions are made about the way the quantum key distribution devices work or on what quantum system they operate. Our main result is a tight bound on the Holevo information between one of the authorized parties and the eavesdropper, as a function of the amount of violation of a Bell-type inequality.

Reference

UNIVERSITÉ DE GENÈVE

PHYSICAL REVIEW LETTERS

Device-Independent Security of Quantum Cryptography against Collective Attacks

Antonio Acín,[1,2] Nicolas Brunner,[3] Nicolas Gisin,[3] Serge Massar,[4] Stefano Pironio,[1] and Valerio Scarani[3]

[1]*ICFO-Institut de Ciencies Fotoniques, Mediterranean Technology Park, 08860 Castelldefels (Barcelona), Spain*
[2]*ICREA-Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain*
[3]*Group of Applied Physics, University of Geneva, CH-1211 Geneve 4, Switzerland*
[4]*Laboratoire d'information quantique, CP 225, Université Libre de Bruxelles, 1050 Brussels, Belgium*

We present the optimal collective attack on a quantum key distribution protocol in the ''device-independent'' security scenario, where no assumptions are made about the way the quantum key distribution devices work or on what quantum system they operate. Our main result is a tight bound on the Holevo information between one of the authorized parties and the eavesdropper, as a function of the amount of violation of a Bell-type inequality.

Quantum key distribution (QKD) allows two parties, Alice and Bob, to generate a secret key in the presence of an eavesdropper, Eve [1]. All QKD schemes rely for security on several assumptions. The basic one is that any eavesdropper, however powerful, must obey the laws of quantum physics. In addition to it, there are two other requirements, without which no shared secret key can be established. The first one is the freedom and secrecy of measurement settings: on each particle, both Alice and Bob should be allowed to choose freely among at least two measurement settings [e.g., the two bases of the Bennett-Brassard 1984 (BB84) protocol [2]] and this choice should not be known to Eve, at least as long as she can act on the incoming quantum states (in BB84, the bases are revealed, but only after the measurements are performed). The second requirement, even more obvious, is the secrecy of outcomes: at no stage should there be a leakage of information about the final key. These two requirements can be summarized by saying that no unwanted classical information must leak out of Alice's and Bob's laboratories. If an implementation has a default in this point (e.g., if a Trojan Horse attack is possible, or if Eve can access Bob's computer), no security can be guaranteed.

In addition to these essential requirements, existing security proofs [3–5] assume that Alice and Bob have (almost) perfect control of the state preparation and of the measurement devices. This assumption is often critical: for instance, the security of the BB84 protocol is entirely compromised if Alice and Bob, instead of sharing qubits as usually assumed, share four-dimensional systems [6,7].

At first sight, control of the apparatuses seems to be an inescapable requirement. Remarkably, this is not the case: we present here a device-independent security proof against collective attacks by a quantum Eve for the protocol described in Ref. [8]. Our proof holds under no other requirements than the essential ones listed above. It is therefore ''device independent'' in the sense that it needs no knowledge of the way the QKD devices work, provided quantum physics is correct and provided Alice and Bob do

not allow any unwanted signal to escape from their laboratories.

In a collective attack, Eve applies the same attack on each particle of Alice and Bob, but no other limitations are imposed to her. In particular, she can keep her systems in a quantum memory and perform a (coherent) measurement on them at any time. Collective attacks are very meaningful in QKD because a bound on the key rate for these attacks becomes automatically a bound for the most general attacks if a de Finetti theorem can be applied, as is the case in the usual security scenario [9].

The physical basis for our device-independent security proof is the fact that measurements on entangled particles can provide Alice and Bob with nonlocal correlations, i.e., correlations that cannot be reproduced by shared randomness (local variables), as detected by the violation of Bell-type inequalities. Considered in the perspective of QKD, the fact that Alice's and Bob's symbols are correlated in a nonlocal way, whatever be the underlying physical details of the apparatuses that produced those symbols, implies that Eve cannot have full information about them, otherwise her own symbol would be a local variable able to reproduce the correlations.

This intuition was at the origin of Ekert's 1991 proposal [10] and implicit in subsequent works [11,12]. Quantitative progress has been possible, however, only recently, thanks to the pioneering work of Barrett, Hardy, and Kent [13] and to further extensions [6,8,14]. For conceptual interest and mathematical simplicity, all these works studied security against a supra-quantum Eve, who could perform any operation compatible with the no-signaling principle. The proof of Ref. [13] applies only to the zero-error case; those in Refs. [6,8] allow for errors but restrict Eve to perform individual attacks; Masanes and Winter [14] proved non-universally composable security under the assumption that Eve's attack is arbitrary but is not correlated with the classical post-processing of the raw key. In this Letter, we focus on the more realistic situation in which Eve is constrained by quantum physics, and we prove universally composable security against collective attacks.

The protocol.—The protocol that we study is a modification of the Ekert 1992 protocol [10] proposed in Ref. [8]. Alice and Bob share a quantum channel consisting of a source that emits pairs of entangled particles. On each of her particles, Alice chooses between three possible measurements A_0, A_1, and A_2, and Bob between two possible measurements B_1 and B_2. All measurements have binary outcomes labeled by a_i, $b_j \in \{+1, -1\}$ (note, however, that the quantum systems may be of dimension larger than 2). The raw key is extracted from the pair $\{A_0, B_1\}$. In particular, the quantum bit error rate (QBER) is $Q = \text{prob}(a_0 \neq b_1)$. As mentioned in the introduction, Eve's information is bounded by evaluating Bell-type inequalities, since these are the only entanglement witnesses which are independent of the details of the system. In our case, Alice and Bob use the measurements A_1, A_2, B_1, and B_2 on a subset of their particles to compute the Clauser-Horne-Shimony-Holt (CHSH) polynomial [15]

$$\mathcal{S} = \langle a_1 b_1 \rangle + \langle a_1 b_2 \rangle + \langle a_2 b_1 \rangle - \langle a_2 b_2 \rangle, \quad (1)$$

which defines the CHSH inequality $\mathcal{S} \leq 2$. We note that there is no *a priori* relation between the value of $\mathcal{S}$ and the value of Q: these are the two parameters which are available to estimate Eve's information. Without loss of generality, we suppose that the marginals are random for each measurement, i.e., $\langle a_i \rangle = \langle b_j \rangle = 0$ for all i and j. Were this not the case, Alice and Bob could achieve it *a posteriori* through public one-way communication by agreeing on flipping a chosen half of their bits. This operation would not change the value of Q and $\mathcal{S}$ and would be known to Eve.

Eavesdropping.—In the device-independent scenario, Eve is assumed not only to control the source (as in usual entanglement-based QKD), but also to have fabricated Alice's and Bob's measuring devices. The only data available to Alice and Bob to bound Eve's knowledge are the observed relation between the measurement settings and outcomes, without any assumption on how the measurements are actually carried out or on what system they operate. In complete generality, we may describe this situation as follows. Alice, Bob, and Eve share a state $|\Psi\rangle_{ABE}$ in $\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B^{\otimes n} \otimes \mathcal{H}_E$, where n is the number of bits of the raw key. The dimension d of Alice and Bob Hilbert spaces $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^d$ is unknown to them and fixed by Eve. The measurement M_k yielding the kth outcome of Alice is defined on the kth subspace of Alice and chosen by Eve. This measurement depends on the kth setting A_{j_k} chosen by Alice, but possibly also on all previous settings and outcomes: $M_k = M(A_{j_k}, \bar{A}_{k-1}, \bar{a}_{k-1})$, where $\bar{A}_{k-1} = (A_{j_1}, \ldots, A_{j_{k-1}})$ and $\bar{a}_{k-1} = (a_{j_1}, \ldots, a_{j_{k-1}})$. The situation is similar for Bob.

Collective attacks.—In this Letter, we focus on collective attacks where Eve applies the same attack to each system of Alice and Bob. Specifically, we assume that the total state shared by the three parties has the product form $|\Psi_{ABE}\rangle = |\psi_{ABE}\rangle^{\otimes n}$ and that the measurements are a

function of the current setting only, e.g., for Alice $M_k = M(A_{j_k}) \equiv A_{j_k}$.

For collective attacks, the secret-key rate r under one-way classical post-processing from Bob to Alice is lower bounded by the Devetak-Winter rate [16],

$$r \geq r_{\text{DW}} = I(A_0 : B_1) - \chi(B_1 : E), \quad (2)$$

which is the difference between the mutual information between Alice and Bob, $I(A_0 : B_1) = 1 - h(Q)$ (h is the binary entropy), and the Holevo quantity between Eve and Bob, $\chi(B_1 : E) = S(\rho_E) - \frac{1}{2}\sum_{b_1 = \pm 1} S(\rho_{E|b_1})$. Note that the rate is given by (2) because $\chi(A_0 : E) \geq \chi(B_1 : E)$ holds for our protocol [8]; it is therefore advantageous for Alice and Bob to do the classical post-processing with public communication from Bob to Alice.

Upper bound on the Holevo quantity.—To find Eve's optimal collective attack, we must find the largest value of $\chi(B_1 : E)$ compatible with the observed parameters without assuming anything about the physical systems and the measurements that are performed. Our main result is the following.

Theorem.—Let $|\psi_{ABE}\rangle$ be a quantum state and $\{A_1, A_2, B_1, B_2\}$ a set of measurements yielding a violation $\mathcal{S}$ of the CHSH inequality. Then after Alice and Bob have symmetrized their marginals,

$$\chi(B_1 : E) \leq h\left(\frac{1 + \sqrt{(\mathcal{S}/2)^2 - 1}}{2}\right). \quad (3)$$

Before presenting the proof of this bound, we give an explicit attack which saturates it; this example clarifies why the bound (3) is independent of Q. Eve sends to Alice and Bob the two-qubit Bell-diagonal state

$$\rho_{AB}(\mathcal{S}) = \frac{1 + \mathcal{C}}{2} P_{\Phi^+} + \frac{1 - \mathcal{C}}{2} P_{\Phi^-}, \quad (4)$$

where $P_{\Phi^\pm}$ are the projectors on the Bell states $|\Phi^\pm\rangle = (|00\rangle \pm |11\rangle)\sqrt{2}$ and $\mathcal{C} = \sqrt{(\mathcal{S}/2)^2 - 1}$. She defines the measurements to be $B_1 = \sigma_z$, $B_2 = \sigma_x$, and $A_{1,2} = \frac{1}{\sqrt{1+\mathcal{C}^2}}\sigma_z \pm \frac{\mathcal{C}}{\sqrt{1+\mathcal{C}^2}}\sigma_x$. Any value of Q can be obtained by choosing A_0 to be σ_z with probability $1 - 2Q$ and to be a randomly chosen bit with probability $2Q$. This attack is impossible within the usual assumptions because here not only the state ρ_{AB}, but also the measurements taking place in Alice's apparatus depend explicitly on the observed values of $\mathcal{S}$ and Q. The state (4) has a nice interpretation: it is the two-qubit state which gives the highest violation $\mathcal{S}$ of the CHSH inequality for a given value of the entanglement, measured by the concurrence $\mathcal{C}$ [17].

We now present the proof of the Theorem stated above, in four steps; more details will be given in a forthcoming paper.

Proof, Step 1.—It is not restrictive to suppose that Eve sends to Alice and Bob a mixture $\rho_{AB} = \sum_c p_c \rho_{AB}^c$ of two-qubit states, together with a classical ancilla (known to her) that carries the value c and determines which measurements A_i^c and B_j^c are to be used on ρ_{AB}^c.

The proof of this first statement relies critically on the simplicity of the CHSH inequality (two binary settings on each side). We present the argument for Alice; the same holds for Bob. First, we may assume that the two measurements $A_{1,2}$ of Alice are von Neumann measurements, if necessary by including ancillas in the state ρ_{AB} shared by Alice and Bob. Thus A_1 and A_2 are Hermitian operators on $\mathbb{C}^d$ with eigenvalues ± 1. It follows from this that $A_1 A_2$ is a unitary, hence diagonalizable, operator. In the basis of $\mathbb{C}^d$ formed by the eigenvectors of $A_1 A_2$, one can show that A_1 and A_2 are block diagonal, with blocks of size 1×1 or 2×2 [18]. In other words, $A_j = \sum_c P_c A_j P_c$, where the P_cs are projectors of rank 1 or 2. From Alice's standpoint, the measurement of A_i thus amounts at projecting in one of the (at most) two-dimensional subspaces defined by the projectors P_c, followed by a measurement of the reduced observable $P_c A_i P_c = \vec{a}_i^c \cdot \vec{\sigma}$. Clearly, it cannot be worse for Eve to perform the projection herself before sending the state to Alice and learning the value of c. The same holds for Bob. We conclude that in each run of the experiment Alice and Bob receive a two-qubit state. The deviation from usual proofs lies in the fact that the measurements to be applied can depend explicitly on the state.

Proof, Step 2.—Each state ρ_{AB}^c can be taken to be a Bell-diagonal state and the measurements A_i^c and B_j^c to be measurements in the (x, z) plane.

To reduce the problem further in this way, we use some freedom in the labeling together with two applications of a usual argument. For fixed c (we now omit the index c), let us first choose the axis of the Bloch sphere on Alice's side in such a way that $\vec{a}_1$ and $\vec{a}_2$ define the (x, z) plane, and similarly on Bob's side. Eve is *a priori* distributing any two-qubit state ρ of which she holds a purification. Now, recall that we have supposed, without loss of generality, that all the marginals are uniformly random. Here comes an argument which is typical of QKD [4]: knowing that Alice and Bob are going to symmetrize their marginals, Eve does not lose anything in providing them a state with the suitable symmetry. The reason is as follows. First note that since the (classical) randomization protocol that ensures $\langle a_i \rangle = \langle b_j \rangle = 0$ is done by Alice and Bob through public communication, we can as well assume that it is Eve who does it; i.e., she flips the value of each outcome bit with probability one half. But because the measurements of Alice and Bob are in the (x, z) plane, we can equivalently, i.e., without changing Eve's information, view the classical flipping of the outcomes as the quantum operation $\rho \to \tilde{\rho} = (\sigma_y \otimes \sigma_y)\rho(\sigma_y \otimes \sigma_y)$ on the state ρ. We conclude that it is not restrictive to assume that Eve is in fact sending the mixture $\bar{\rho} = \frac{1}{2}(\rho + \tilde{\rho})$, i.e., that she is sending a state invariant under $\sigma_y \otimes \sigma_y$. Now, through an appropriate choice of basis that leaves invariant the (x, z) plane, and corresponding to the freedom to define the orientation of $\hat{y}$ and the direction of $\hat{x}$ for both Alice and Bob, every $\sigma_y \otimes \sigma_y$ invariant two-qubit state can be written in the Bell basis, ordered as $\{|\Phi^+\rangle, |\Psi^-\rangle, |\Phi^-\rangle, |\Psi^+\rangle\}$, in the canonical form

$$\bar{\rho} = \begin{pmatrix} \lambda_{\Phi^+} & ir_1 & & \\ -ir_1 & \lambda_{\Psi^-} & & \\ & & \lambda_{\Phi^-} & ir_2 \\ & & -ir_2 & \lambda_{\Psi^+} \end{pmatrix}, \quad (5)$$

with $\lambda_{\Phi^+} \geq \lambda_{\Psi^-}$, $\lambda_{\Phi^+} \geq \lambda_{\Phi^-} \geq \lambda_{\Psi^+}$ and r_1, r_2 real.

Finally, we repeat an argument similar to the one given above: since $\bar{\rho}$ and its conjugate $\bar{\rho}^*$ produce the same statistics for Alice and Bob's measurements and provide Eve with the same information, we can suppose without loss of generality that Alice and Bob rather receive the mixture $\frac{1}{2}(\bar{\rho} + \bar{\rho}^*)$, which is Bell diagonal.

Proof, Step 3.—For a Bell-diagonal state ρ_λ with eigenvalues λ ordered as above and for measurements in the (x, z) plane,

$$\chi_\lambda(B_1{:}E) \leq F(\mathcal{S}_\lambda) = h\left(\frac{1 + \sqrt{(\mathcal{S}_\lambda/2)^2 - 1}}{2}\right), \quad (6)$$

where $\mathcal{S}_\lambda = 2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Psi^-})^2 + (\lambda_{\Phi^-} - \lambda_{\Psi^+})^2}$ is the largest violation of the CHSH inequality by the state ρ_λ.

This step is mainly computational; we sketch it here and refer to a forthcoming paper for details. For Bell-diagonal states, for any choice of $B_1 = \cos\varphi \sigma_z + \sin\varphi \sigma_x$, one has $S(\rho_{E|b_1=0}) = S(\rho_{E|b_1=1}) \geq h(\lambda_{\Phi^+} + \lambda_{\Phi^-})$ with equality if and only if $B_1 = \sigma_z$. It follows that $\chi_\lambda(B_1{:}E) \leq H(\lambda) - h(\lambda_{\Phi^+} + \lambda_{\Phi^-})$. The right-hand side of this expression is in turn bounded by the function $F(\mathcal{S}_\lambda)$ appearing in (6). It now suffices to notice that $\mathcal{S}_\lambda = 2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Psi^-})^2 + (\lambda_{\Phi^-} - \lambda_{\Psi^+})^2}$ is the maximal violation of the CHSH inequality by the state ρ_λ [17,19]; it is achieved for $B_1 = \sigma_z$, $B_2 = \sigma_x$, and A_1 and A_2 depending explicitly on the λ's.

Proof, Step 4.—To conclude the proof, note that if Eve sends a mixture of Bell-diagonal states $\sum_\lambda p_\lambda \rho_\lambda$ and chooses the measurements to be in the (x, z) plane, then $\chi(B_1{:}E) = \sum_\lambda p_\lambda \chi_\lambda(B_1{:}E)$. Using (6), we then find $\chi(B_1{:}E) \leq \sum_\lambda p_\lambda F(\mathcal{S}_\lambda) \leq F(\sum_\lambda p_\lambda \mathcal{S}_\lambda)$, where the last inequality holds because F is concave. But since the observed violation $\mathcal{S}$ of CHSH is necessarily such that $\mathcal{S} \leq \sum_\lambda p_\lambda \mathcal{S}_\lambda$ and since F is a monotonically decreasing function, we find $\chi(B_1{:}E) \leq F(\mathcal{S})$.

Key rate.—Given the bound (3), the key rate (2) can be computed for any values of Q and $\mathcal{S}$. As an illustration, we study correlations satisfying $\mathcal{S} = 2\sqrt{2}(1 - 2Q)$, and which arise from the state $|\Phi^+\rangle$ after going through a depolarizing channel, or through a phase-covariant cloner, or more generally from any Bell-diagonal ρ_{AB} such that $\lambda_{\Phi^+} \geq \lambda_{\Psi^-}$ and $\lambda_{\Phi^-} = \lambda_{\Psi^+}$, when doing the measurements $A_0 = B_1 = \sigma_z$, $B_2 = \sigma_x$, $A_1 = (\sigma_z + \sigma_x)/\sqrt{2}$, and $A_2 = (\sigma_z - \sigma_x)/\sqrt{2}$. We consider these correlations because of their experimental significance, but it is important to stress that Alice and Bob do not need to assume that they perform the above qubit measurements. The corresponding key rate is plotted in Fig. 1 as a function of Q. For the sake of comparison, we have also plotted the key rate under the usual assumptions of QKD for the same set of correlations.
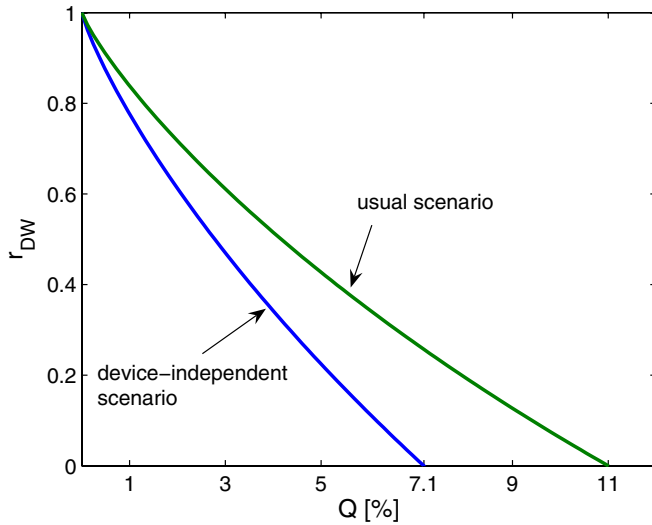
FIG. 1 (color online). Extractable secret-key rate against collective attacks in the usual scenario [$\chi(B_1{:}E)$ given by Eq. (7)] and in the device-independent scenario [$\chi(B_1{:}E)$ given by Eq. (3)], for correlations satisfying $\mathcal{S} = 2\sqrt{2}(1 - 2Q)$.

In this case, Alice and Bob have a perfect control of their apparatuses, which we have assumed to faithfully perform the qubit measurements given above. The protocol is then equivalent to Ekert's, which in turn is equivalent to the entanglement-based version of BB84, and one finds

$$\chi(B_1{:}E) \le h(Q + \mathcal{S}/2\sqrt{2}). \qquad (7)$$

If $\mathcal{S} = 2\sqrt{2}(1 - 2Q)$, this expression yields the well-known critical QBER of 11% [3], to be compared to 7.1% in the device-independent scenario (Fig. 1). [Note that the key rate given by Eq. (3) is much higher than the one against a no-signaling eavesdropper obtained by applying the security proof of [14].]

Final remarks.—Through its remarkable generality, our device-independent security proof allows us to ignore the detailed implementation of the QKD protocol and therefore applies in a simple way to situations where the quantum apparatuses are noisy or where uncontrolled side channels are present. It also applies to the situation where the apparatuses are entirely untrusted and provided by the eavesdropper herself. In this latter case, the proof cannot be applied to any existing device yet, because of the detection loophole which arises due to inefficient detectors and photon absorption. These processes imply that sometimes Alice's and Bob's detectors will not fire. A possible strategy to apply our proof to this new situation is for Alice and Bob to replace the absence of a click by a chosen outcome, in effect replacing detection inefficiency by noise. However, the amount of detection inefficiency that can be tolerated in this way is much lower than the one present in current quantum communication experiments. In Bell tests, this problem is often circumvented by invoking additional assumptions such as the fair sampling hypothesis—a very reasonable one if the aim is to constrain possible models of Nature, but hardly justified if the device is

provided by an untrusted Eve. In the light of the present work, the "detection loophole" thus becomes a meaningful issue in applied physics.

In conclusion, we have found the optimal collective attack on a QKD protocol in the device-independent scenario, in which no other assumptions are made than the validity of quantum physics and the absence of any leakage of classical information from Alice's and Bob's laboratories. If a suitable de Finetti-like theorem can be demonstrated in this scenario, the bound that we have presented here will in fact be the bound against the most general attacks.

[1] N. Gisin *et al.*, Rev. Mod. Phys. **74**, 145 (2002); M. Dušek, N. Lütkenhaus, and M. Hendrych, in *Progress in Optics*, edited by E. Wolf (Elsevier, New York, 2006), Vol. 49, p. 381.
[2] 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), pp. 175–179.
[3] P. W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000).
[4] B. Kraus, N. Gisin, and R. Renner, Phys. Rev. Lett. **95**, 080501 (2005); R. Renner, N. Gisin, and B. Kraus, Phys. Rev. A **72**, 012332 (2005).
[5] M. Koashi and J. Preskill, Phys. Rev. Lett. **90**, 057902 (2003); D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, Quantum Inform. Comput. **5**, 325 (2004).
[6] A. Acín, N. Gisin, and L. Masanes, Phys. Rev. Lett. **97**, 120405 (2006); V. Scarani *et al.*, Phys. Rev. A **74**, 042339 (2006).
[7] F. Magniez *et al.*, arXiv:quant-ph/0512111, Appendix A.
[8] A. Acín, S. Massar, and S. Pironio, New J. Phys. **8**, 126 (2006).
[9] R. Renner, arXiv:quant-ph/0512258.
[10] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[11] C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).
[12] D. Mayers and A. Yao, Quantum Inform. Comput. **4**, 273 (2004).
[13] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).
[14] L. Masanes and A. Winter, arXiv:quant-ph/0606049.
[15] J. F. Clauser *et al.*, Phys. Rev. Lett. **23**, 880 (1969).
[16] I. Devetak and A. Winter, Proc. R. Soc. A **461**, 207 (2005).
[17] F. Verstraete and M. M. Wolf, Phys. Rev. Lett. **89**, 170401 (2002).
[18] L. Masanes, Phys. Rev. Lett. **97**, 050503 (2006).
[19] R. Horodecki, P. Horodecki, and M. Horodecki, Phys. Lett. A **200**, 340 (1995).