Fully Distrustful Quantum Cryptography

J. Silman,[1] A. Chailloux,[2] N. Aharon,[3] I. Kerenidis,[4,5] S. Pironio,[1] and S. Massar[1]

[1]*Laboratoire d'Information Quantique, Université Libre de Bruxelles, 1050 Bruxelles, Belgium*
[2]*LIAFA, Univ. Paris 7, F-75205 Paris, France; and Univ. Paris-Sud, 91405 Orsay, France*
[3]*School of Physics and Astronomy, Tel-Aviv University, Tel-Aviv 69978, Israel*
[4]*LIAFA, Univ. Paris 7 - CNRS; F-75205 Paris, France*
[5]*Centre for Quantum Technologies, National University of Singapore, Singapore 117543*

In the distrustful quantum cryptography model the different parties have conflicting interests and do not trust one another. Nevertheless, they trust the quantum devices in their labs. The aim of the device-independent approach to cryptography is to do away with the necessity of making this assumption, and, consequently, significantly increase security. In this paper we enquire whether the scope of the device-independent approach can be extended to the distrustful cryptography model, thereby rendering it 'fully' distrustful. We answer this question in the affirmative by presenting a device-independent (imperfect) bit-commitment protocol, which we then use to construct a device-independent coin flipping protocol.

Introduction – A quantum protocol is said to be device-independent if the reliability of its implementation can be guaranteed without making any assumptions regarding the internal workings of the underlying apparatus. The key idea is that the certification of a sufficient amount of nonlocality ensures that the underlying systems are quantum and entangled. By dispensing with the (mathematically convenient but physically untestable) notion of a Hilbert space of a fixed dimension, the device-independent approach does away with many cheating mechanisms and modes of failure, such as, for example, those exploited in [1, 2]. In fact, a device-independent protocol, in principle, remains secure even if the devices were fabricated by an adversary. So far, device-independent protocols have been proposed for quantum key-distribution [3–6], random number generation [7, 8], state estimation [9], and the self-testing of quantum computers [10].

In many everyday scenarios (e.g. the use of credit cards on the internet, secure identification, digital signatures), we need to ensure security not only against an eavesdropper, but crucially against malicious parties partaking in the protocol, i.e. when Alice and Bob do not trust each other. Many important results in quantum cryptography are related to the fundamental primitives in this setting: While, on the one hand, quantum weak coin flipping with arbitrarily small bias is possible [11], arbitrarily concealing and binding quantum bit-commitment is impossible [12–14]. However, less secure but non-trivial bit-commitment has been shown to be possible with trusted devices [15].

It is not a priori clear, whether the scope of the device-independent approach can be extended to cover cryptographic problems with distrustful parties. In particular, this setting presents us with a novel challenge: Whereas in device-independent quantum key-distribution Alice and Bob will cooperate to estimate the amount of nonlocality present, for protocols in the distrustful cryptography model, honest parties can rely only on themselves. In this paper we show that protocols in this model are indeed amenable to a device-independent formulation. As our aim is to provide a proof of concept, we concentrate on one of the simplest, yet most fundamental, primitives in this model, bit-commitment. We present a device-independent bit-commitment protocol, wherein after the commit phase Alice cannot control the value of the bit she wishes to reveal with probability greater than $\cos^2\left(\frac{\pi}{8}\right) \simeq 0.854$ and Bob cannot learn its value prior to the reveal phase with probability greater than $\frac{3}{4}$. We then use this protocol to construct a device-independent coin flipping protocol with bias $\lesssim 0.336$.

Bit-commitment – A bit-commitment protocol consists of two phases. In the commit phase, Alice interacts with Bob in order to commit to a bit. In the reveal phase, Alice reveals the value of the bit, possibly followed by some test that each party carries out to ensure that the other party has not cheated. In the time between the two phases, which may be of any duration, no actions are taken. The security of a protocol is always analyzed under the assumption that one of the parties is honest. We designate by P_{cont}, the maximum of the average of the probabilities with which Alice can reveal either value of the bit without being caught cheating, and by P_{gain} the maximum probability that dishonest Bob learns the value of bit before the reveal phase without being discovered, where these quantities are maximized over the set of possible cheating strategies available to Alice and Bob. The quantities $\epsilon_{\text{cont}} = P_{\text{cont}} - \frac{1}{2}$ and $\epsilon_{\text{gain}} = P_{\text{gain}} - \frac{1}{2}$ are termed 'Alice's control' and 'Bob's information gain'. A protocol with arbitrarily small ϵ_{cont} is called arbitrarily binding, while a protocol with arbitrarily small ϵ_{gain} is called arbitrarily concealing. As already mentioned, quantum mechanics does not allow for a protocol to be both arbitrarily binding and concealing at the same time. In fact, for a 'fair' protocol, in the sense that $\epsilon_{\text{cont}} = \epsilon_{\text{gain}}$, ϵ_{cont} is bounded from below by 0.207 [16]. The best known protocol gives $\epsilon_{\text{cont}} = \frac{1}{4}$ [15]. In contrast, in any classical protocol either Alice or Bob can cheat perfectly

($\epsilon_{\rm cont} = \frac{1}{2}$).

Device-independence – In our device-independent formulation, we assume that each honest party has one or several devices which are viewed as 'black boxes'. Each box allows for a classical input $s_i \in \{0,1\}$, and produces a classical output $r_i \in \{0,1\}$ (the index i designates the box). We make the assumption that the probabilities of the outputs given the inputs for an honest party can be expressed as $P(\mathbf{r}|\mathbf{s}) = \mathrm{Tr}\big(\rho \bigotimes_i \Pi_{r_i|s_i}\big)$, where ρ is some joint quantum state and $\Pi_{r_i|s_i}$ is a POVM element corresponding to inputting s_i in box i and obtaining the outcome r_i. Apart from this constraint we impose no restrictions on the boxes' behavior. In particular, we allow a dishonest party to choose the state ρ (which she can entangle with her system) and the POVM elements $\Pi_{r_i|s_i}$ for the other party's boxes.

The above assumption amounts to the most general modeling of boxes that (i) satisfy the laws of quantum theory, and (ii) are such that the physical process yielding the output r_i in box i depends solely on the input s_i, i.e. the boxes cannot communicate with one another. It is also implicit in our analysis that no unwanted information can enter or exit an honest party's laboratory. In a 'fully' distrustful setting, where the devices too cannot be trusted, these conditions can be satisfied by shielding the boxes. In particular, it is not necessary to carry out measurements in space-like separated locations to guarantee (ii), as in fundamental tests of nonlocality (see [8, 17]). This observation is important because relativistic causality is by itself sufficient for perfect bit-commitment and coin flipping [18, 19]. Hence, the fact that we do not rely on space-like measurements makes the conceptual implications of our work clearer and the quantum origin of the security evident.

The protocol – Our protocol is based on the Greenberger-Horne-Zeilinger (GHZ) paradox [20, 21]. We consider three boxes A, B, and C with binary inputs, s_A, s_B and s_C, and outputs r_A, r_B and r_C, respectively. The GHZ paradox consists of the fact that if the inputs satisfy $s_A \oplus s_B \oplus s_C = 1$, we can always have the outputs satisfy $r_A \oplus r_B \oplus r_C = s_A s_B s_C \oplus 1$. This relation can be guaranteed if the three boxes implement measurements on a three-qubit GHZ state $\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$, where $s_i = 0$ (1) corresponds to measuring σ_y (σ_x). In contrast, for local boxes this relation can only be satisfied with $\frac{3}{4}$ probability at most. The nonlocal and pseudo-telepathic nature of the GHZ paradox – the non-occurrence of certain input-output pairs that would necessarily occur in any local theory – are key, both to ensure that when both parties are honest the protocol does not abort, and to ensure that a dishonest party always has a non-zero probability of being caught cheating.

The protocol runs as follows. Alice has a box, A, and Bob has a pair of boxes, B and C. The three boxes are supposed to satisfy the GHZ paradox. *Commit phase*: Alice inputs into her box the value of the bit she wishes to commit to. Denote the input and output of her box by s_A and r_A. She then selects a classical bit a uniformly at random. If $a = 0$ ($a = 1$), she sends Bob a classical bit $c = r_A$ ($c = r_A \oplus s_A$) as her commitment. *Reveal phase*: Alice sends Bob s_A and r_A. Bob first checks whether $c = r_A$ or $c = r_A \oplus s_A$. He then randomly chooses a pair of inputs s_B and s_C, satisfying $s_B \oplus s_C = 1 \oplus s_A$, inputs them into his two boxes and checks that the GHZ paradox is satisfied. If any of these tests fails then he aborts. Note that if the parties are honest (and the boxes satisfy the GHZ paradox), then the protocol never aborts.

Alice's control – We consider the worst-case scenario, wherein (dishonest) Alice prepares (honest) Bob's boxes in any state she wants, possibly entangled with her own ancillary systems. Since the commit phase consists of Alice sending a classical bit c as a token of her commitment, without receiving any information from Bob, with no loss of generality we may assume that Alice decides on the value of c beforehand, and accordingly prepares Bob's boxes to maximize her control. Furthermore, since Alice's winning probability is invariant under the relabeling, $c \to c \oplus 1$, $r_A \to r_A \oplus 1$, $r_B \to r_B \oplus 1$, no value of c is preferable, and we assume that she sends $c = 0$.

Suppose now that Alice wishes to reveal 0 (i.e. she sends $s_A = 0$). She will then carry out some operation on her systems in order to decide the value of r_A to be sent. Bob will first check whether $r_A = 0$ or $r_A \oplus s_A = 0$, and since $s_A = 0$ it follows that Alice must send $r_A = 0$. Subsequently, Bob finds that the GHZ paradox is satisfied whenever $r_B \neq r_C$ for a choice of inputs such that $s_B \neq s_C$. Switching to a more compact notation in which $y_i = (-1)^{r_i}$ ($x_i = (-1)^{r_i}$) designates the output corresponding to $s_i = 0$ ($s_i = 1$), Alice's cheating probability in this case equals $\frac{1}{2}\left[P\left(y_B x_C = -1\right) + P\left(x_B y_C = -1\right)\right]$. On the other hand, suppose that Alice wishes to reveal 1. Then, r_A may take on any value (since Bob knows that in this case $r_A = 0$ or $r_A \oplus 1 = 0$), and hence, the only relevant test is the satisfaction of the GHZ paradox, i.e. whether $r_B \oplus r_C = s_B s_C \oplus 1 \oplus r_A$ for a choice of inputs such that $s_B = s_C$. Alice's cheating probability then equals $\frac{1}{2}\left[P\left(y_A y_B y_C = -1\right) + P\left(x_A x_B x_C = 1\right)\right]$. Hence, Alice's optimal cheating probability is obtained by maximizing over

$$\frac{1}{4}\big[P\left(y_B x_C = -1\right) + P\left(x_B y_C = -1\right) + P\left(x_A y_B y_C = -1\right) + P\left(x_A x_B x_C = 1\right)\big] \tag{1}$$

since we consider the average probability that Alice can reveal 0 and 1. As this expression involves only a single measurement setting for Alice's box, it admits a local description, implying that the maximum is obtained when Alice's box is deterministic. We see that in both cases (i.e. $x_A = \pm 1$), the problem reduces to maximizing the Clauser-Horne-Shimony-Holt (CHSH) inequality [22], so that $P_{\rm cont} = \cos^2\left(\frac{\pi}{8}\right) \simeq 0.854$.

Bob's information gain – Bob's most general strategy consists of sending Alice a box entangled with some ancillary system in his possession. Depending on the value of c he receives from Alice (which is uniformly random since Alice is honest), Bob carries out one of a pair of two-outcome measurements on his system. We denote Bob's binary input and output by m_B and g_B, where $m_B = 0$ ($m_B = 1$) corresponds to the measurement he carries out when Alice sends $c = 0$ ($c = 1$), and $g_B = 0$ ($g_B = 1$) corresponds to his guessing that Alice has committed to 0 (1). Bob's information gain is

$$\begin{aligned} P_{\text{gain}} &= \sum_{s_A, r_A, a} P(s_A, r_A, a) P\left(g_B = s_A \mid m_B = r_A \oplus (s_A \cdot a)\right) \\ &= \frac{1}{4} \sum_{s_A, r_A = 0, 1} P\left(r_A \mid s_A\right) \left[P\left(g_B = s_A \mid m_B = r_A\right) \right. \\ &\quad \left. + P\left(g_B = s_A \mid m_B = r_A \oplus s_A\right) \right] \\ &= \frac{1}{4} \sum_{s_A, r_A = 0, 1} \left[P\left(r_A, g_B = s_A \mid s_A, m_B = r_A\right) \right. \\ &\quad \left. + P\left(r_A, g_B = s_A \mid s_A, m_B = r_A \oplus s_A\right) \right]. \end{aligned} \tag{2}$$

Using the fact that $P(k, 0|0, k) + P(0, 1|1, k) + P(1, 1|1, k) \leq 1$ and $P(0, 0 \mid 0, 0) + P(1, 0 \mid 0, 1) \leq 1$, which follow from no-signaling (i.e. $\sum_{l=0,1} P(i_A, i_B \mid j_A, j_B) = P(i_A \mid j_A)$ and the same relation with $A \leftrightarrow B$) and normalization, we obtain that $P_{\text{gain}} = \frac{3}{4}$.

Optimal cheating strategies – Both Alice and Bob have a number of simple optimal cheating strategies available to them. Interestingly, both can optimally cheat using a three-qubit GHZ state and having the measurements of the honest party correspond to the measurement of σ_y and σ_x axes (corresponding to inputting 0 and 1), as in the GHZ paradox described above. This implies that the device-dependent version of our protocol, in which (honest) Alice and Bob share a GHZ state and measure σ_y and σ_x (recall that in the device-dependent setting an honest party can trust its measurement devices), does not afford more security. Our protocol has thus the curious property that its device-dependent version is essentially device-independent, in the sense that its security is not compromised in the event that an honest party cannot trust its measurement devices.

Using the GHZ state, dishonest Alice's strategy consists of measuring the polarization of her qubit along the axis $\hat{n} = \frac{1}{\sqrt{2}}(\hat{x} + \hat{y})$. If she obtains 0 then she knows she has 'prepared' Bob's boxes in the state $\frac{1}{\sqrt{2}}\left(e^{-i\pi/8}|00\rangle + e^{i\pi/8}|11\rangle\right)$, and she sends $c = 0$. If she wishes to reveal 0, she tells Bob she had input 0 and obtained 0. If she wishes to reveal 1, she tells Bob she had input 1 and obtained 0. Similarly, if she obtains 1, she sends $c = 1$, etc. It is straightforward to verify that this strategy gives rise to $P_{\text{cont}} = \cos^2\left(\frac{\pi}{8}\right) \simeq 0.854$.

Using the GHZ state, dishonest Bob's strategy consists of having Alice measure σ_y and σ_x according to the value of her commitment. Bob then measures the polarization of one of his qubits along the y axis and that of the other along the x axis. Whenever his outcomes are correlated, in the event that Alice sends $c = 0$ ($c = 1$) he guesses that she has input 1 (0), while whenever his outcomes are anti-correlated he guesses the reverse. It is straightforward to verify that this strategy gives rise to an information gain of $\frac{3}{4}$.

Device-independent coin flipping – (Strong) coin flipping is defined as the problem of two remote distrustful parties having to agree on a bit. If both parties are honest, then the outcome of the coin is uniformly random. The degree of security afforded by a protocol is quantified by the biases $\epsilon_i^A = P_i^A - \frac{1}{2}$ and $\epsilon_i^B = P_i^B - \frac{1}{2}$, where P_i^A (P_i^B) is Alice's (Bob's) maximal probability of biasing the outcome to i. The quantity $\epsilon = \max\left\{\epsilon_i^A, \epsilon_j^B\right\}_{i,j}$ is usually referred to as the bias of the protocol. A protocol is said to be fair whenever Alice and Bob enjoy the same bias. Like bit-commitment, and indeed most non-trivial protocols in distrustful cryptography, in the classical world its security is completely breached if no limits are placed on a dishonest party's computational power. In the quantum world the story is different [24], the optimal bias is $\epsilon = 0.207$ [25, 26] (a weaker version of coin flipping, on the other hand, allows for arbitrarily small bias [11]).

We remind the reader of a standard method to implement coin flipping using bit-commitment: Alice commits to a random bit a, Bob sends a random bit b to Alice, and then Alice reveals a. The outcome of the coin flip is just $a \oplus b$. In particular, $\epsilon_i^A = \epsilon_{\text{cont}}$ and $\epsilon_i^B = \epsilon_{\text{gain}}$. Using this construction with our device-independent bit-commitment protocol, we obtain a device-independent coin flipping protocol with biases $\epsilon_i^A = \cos^2\left(\frac{\pi}{8}\right) - \frac{1}{2} \simeq 0.354$ and $\epsilon_j^B = \frac{1}{4}$.

Since $\epsilon_i^A > \epsilon_j^B$, this construction advantages Alice. It is possible to lower the bias by equalizing the individual biases. Consider a new coin flipping protocol which consists of two repetitions of the above coin flipping protocol as follows. The result of the first (in which Alice commits) is used to determine who commits in the second. Say if the outcome is 0 (1), then Alice (Bob) commits in the second. It is no longer a priori clear what strategy Alice should adopt in the first repetition, since, in principle, it may be beneficial for her to adopt one in which she sometimes loses the first coin flip, but increases her chances of making it to the second repetition (by not getting caught cheating in the first repetition in which case Bob aborts). Nevertheless, it is evident that Alice's maximal cheating probability is bounded from above by $\cos^4\left(\frac{\pi}{8}\right) + \left(1 - \cos^2\left(\frac{\pi}{8}\right)\right) \cdot \frac{3}{4} \simeq 0.838$. On the other hand, Bob never gets caught cheating in the first repetition (though he may of course lose), therefore Bob's maximal cheating probability is just $\frac{3}{4}\cos^2\left(\frac{\pi}{8}\right)+\frac{1}{4}\cdot\frac{3}{4}\simeq 0.827$. By allowing for more repetitions (the $n-1$ th repetition determining who commits in the n th, etc.) we obtain that the biases ϵ_i^A and ϵ_j^B of the resulting protocol are bounded from above by $\simeq 0.336$.

Discussion – By introducing explicit device-independent bit-commitment and coin flipping protocols, we have shown that protocols in the distrustful cryptography model – wherein Alice and Bob do not cooperate to estimate the amount of nonlocality present – are amenable to a device-independent formulation. The fascinating connection between quantum nonlocality and cryptography, first noted by Ekert twenty years ago [27], is thus seen to apply also in the very rich field of cryptography with mutually distrustful parties (and devices), affording us with a novel perspective on the connection between cryptography and the foundations of quantum mechanics.

To conclude, we would like to point out some notable features of our protocols. (i) The protocols are single-shot and do not rely on any statistical estimation of the amount of nonlocality such as in the testing the degree of violation of a Bell inequality (even though their security is of course based on nonlocality). (ii) The device-dependent version of our protocol does not offer more security than the device-independent version. (iii) Since our security analysis is device-independent, it also covers the case where Alice's and Bob's outputs are affected by noise. Note that the analysis of noisy classical coin flipping in [28, 29] allows us to compute the quantum advantage in this case. (iv) The security afforded by our device-independent protocols is reasonably close to (though of course greater than) that of the best known device-dependent protocols. For the bit-commitment protocol we have $P_{\text{cont}}\simeq 0.854$ and $P_{\text{gain}}=\frac{3}{4}$, as compared to $P_{\text{cont}}=P_{\text{gain}}=\frac{3}{4}$ for the best known device-dependent protocol. The coin flipping protocol has a bias of $\lesssim 0.336$, as compared to 0.207 in the device-dependent case. (v) Our work allows the study of bit-commitment and coin flipping in the context of theories other than quantum mechanics. Indeed, it relies only on the GHZ paradox (to define the protocol in the honest case), on Tsirelson's bound on the CHSH inequality violation (which limits Alice's control) and on the no-signaling principle (which limits Bob's information gain). Curiously, the security of the protocol would increase if Tsirelson's bound were to decrease, reaching $P_{\text{cont}}=P_{\text{gain}}=\frac{3}{4}$ if it were equal to the local causal bound. In a theory constrained only by no-signaling, our protocol is no longer secure as PR boxes [30] allow to maximally violate the CHSH inequality, implying $P_{\text{cont}}=1$. Note that perfect bit-commitment was shown to be possible provided that honest parties have access to PR boxes and under the strong hypothesis (which we do not make) that a dishonest party cannot in any way tamper with the boxes [31]. It is an open question whether there exists a quantum bit-commitment protocol that is secure against dishonest parties limited only by the no-signaling principle, as is the case in quantum key-distribution [4, 32].

Acknowledgments – We acknowledge support from the BSF (grant no. 32/08) (N.A.), the Inter-University Attraction Poles Programme (Belgian Science Policy) under Project IAP-P6/10 (Photonics@be) (S.M., S.P., J.S), a BB2B grant of the Brussels-Capital region (S.P.), the FNRS (J.S.), the projects ANR-09-JCJC-0067-01, ANR-08-EMER-012 (A.C., I.K.), and the project QCS (grant 255961) of the E.U. (S.M., S.P., J.S., A.C., I.K.).

[1] F. Xu *et al.*, arXiv:1005.2376.
[2] L. Lydersen *et al.*, Nat. Photonics **4**, 686 (2010).
[3] D. Mayers and A. Yao, Quantum Inform. Comput. **4**, 273 (2004).
[4] J. Barrett *et al.*, Phys. Rev. Lett. **95**, 010503 (2005).
[5] A. Acín *et al.*, Phys. Rev. Lett. **97**, 120405 (2006).
[6] A. Acín *et al.*, Phys. Rev. Lett. **98**, 230501 (2007).
[7] R. Colbeck, PhD dissertation, Univ. Cambridge (2007), arXiv:0911.3814; R. Colbeck, A. Kent, arXiv:1011.4474.
[8] S. Pironio *et al.*, Nature **464**, 1021 (2010).
[9] C.-E. Bardyn *et al.*, Phys. Rev. A **80**, 062327 (2009).
[10] F. Magniez *et al.*, in *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming*, (Springer, 2006), p. 72.
[11] C. Mochon, arXiv:0711.4114.
[12] H.-K. Lo and H.F. Chau, Phys. Rev. Lett. **78**, 3410 (1997).
[13] D. Mayers, Phys. Rev. Lett. **78**, 3414 (1997).
[14] G.M. D'Ariano *et al.*, Phys. Rev. A **76**, 032328 (2007).
[15] R.W. Spekkens and T. Rudolph, Phys. Rev. A **65**, 012310 (2001).
[16] This bound follows from Kitaev's bound on the bias of strong coin flipping protocols [25].
[17] S. Pironio *et al.*, New J. Phys. **11**, 045021 (2009).
[18] A. Kent, Phys. Rev. Lett. **83**, 1447 (1999).
[19] A. Kent, Phys. Rev. Lett. **83**, 5382 (1999).
[20] D.M. Greenberger *et al.*, in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, edited by M. Kafatos (Kluwer, 1989), p. 74.
[21] N.D. Mermin, Phys. Today **43**, 9 (1990).
[22] J.F. Clauser *et al.*, Phys. Rev. Lett. **23**, 880 (1969).
[23] B.S. Cirel'son, Lett. Math. Phys. **4**, 93 (1980).
[24] D. Aharonov *et al.*, in *Proceedings of the 32nd Annual ACM Symposium on the Theory of Computing* (ACM Press, 2000), p. 705.
[25] A. Kitaev, unpublished. Proof reproduced in A. Ambainis *et al.*, in *Proceedings of the 19th Annual IEEE Conference on Computational Complexity*, (CS Press, 2004), p. 250.
[26] A. Chailloux and I. Kerenidis, in *Proceedings of the 50th Annual IEEE Symposium on the Foundations of Computer Science*, (CS Press, 2009), p. 527.
[27] A.K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[28] A.T. Nguyen *et al.*, New J. Phys. **10**, 083037 (2008).
[29] E.Hänggi and J. Wullschleger, arXiv:1009.4741.
[30] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[31] H. Buhrman *et al.*, Proc. R. Soc. A **462**, 1919 (2006).
[32] Ll. Masanes, Phys. Rev. Lett. **102**, 140501 (2009).