

# Experimental bit commitment based on quantum communication and special relativity

T. Lunghi,<sup>1</sup> J. Kaniewski,<sup>2</sup> F. Bussières,<sup>1</sup> R. Houlmann,<sup>1</sup>  
M. Tomamichel,<sup>2</sup> A. Kent,<sup>3,4</sup> N. Gisin,<sup>1</sup> S. Wehner,<sup>2</sup> and H. Zbinden<sup>1</sup>

<sup>1</sup>*Group of Applied Physics, University of Geneva, CH-1211 Genève 4, Switzerland*

<sup>2</sup>*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543*

<sup>3</sup>*Centre for Quantum Information and Foundations, DAMTP,*

*Centre for Mathematical Sciences, University of Cambridge, United Kingdom*

<sup>4</sup>*Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario, Canada N2L 2Y5*

(Dated: July 22, 2013)

Bit commitment is a fundamental cryptographic primitive in which Bob wishes to commit a secret bit to Alice. Perfectly secure bit commitment is impossible through asynchronous exchange of quantum information. Perfect security is however possible when Alice and Bob split into several agents exchanging classical and quantum communication at times and locations suitably chosen to satisfy specific relativistic constraints. Here we report an implementation of a bit commitment protocol using quantum communication and special relativity. Our protocol is based on [A. Kent, Phys. Rev. Lett. **109**, 130501 (2012)] and has the advantage that it is practically feasible with arbitrary large separations between the agents in order to maximize the commitment time. By positioning agents in Geneva and Singapore, we obtained a commitment time of 15 ms. A security analysis considering experimental imperfections and finite statistics is presented.

Bit commitment is a fundamental primitive with several applications such as coin tossing [1] and secure voting [2]. In a bit commitment protocol, Bob commits a secret bit to Alice at a given instant which he can choose to reveal some time later. Security here means that if Bob is honest, then his bit is perfectly concealed from Alice until he decides to open the commitment and reveal his bit. Furthermore, if Alice is honest, then it should be impossible for Bob to change his mind once the commitment is made. That is, the only bit he can reveal is the one he originally committed himself to. Information-theoretically secure bit commitment in a setting where the two parties exchange classical messages in an asynchronous fashion is impossible. An extensive amount of work was devoted to study asynchronous quantum bit commitment, for which perfect security was ultimately shown to be impossible [3, 4] as well (see also [5, 6]). This does not preclude the existence of protocols with partial (but less-than-complete) bias [7–10], which have been the subject of related experimental work [11–13]. Bit commitment was also demonstrated experimentally using the assumption of noisy quantum storage [14]. Alternatively, perfectly secure relativistic protocols based on the exchange of quantum and classical bits have been proposed [15–18]. We focus here on the protocol [18], which is proven secure in the ideal case (i.e. with perfect devices) [19, 20] and in the presence of losses [19]. In this Letter, we present the first experimental implementation of a secure bit commitment protocol that is based on quantum communication and relativistic constraints, along with a security proof taking into account its experimental imperfections.

Let us briefly describe the original protocol [18]. Fig. 1a shows the evolution of the protocol in a space-time diagram. Bob wants to commit a bit  $b$  to Alice. The protocol starts when Alice sends a group of  $N$  single photons to Bob (all photons arrive to Bob at the same time).

The quantum state of each photon is chosen at random among the four BB84 states. Bob then immediately and simultaneously measures all photons in the  $\{|0\rangle, |1\rangle\}$  basis to commit to  $b = 0$ , or in the  $\{|+\rangle, |-\rangle\}$  basis to commit to  $b = 1$ , where  $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ . This constitutes the commitment phase. Let us now denote  $r_i^{(b)}$  the measurement result of the  $i^{\text{th}}$  photon. Bob then immediately sends at light-speed the array  $r^{(b)} = \{b, r_1^{(b)}, \dots, r_N^{(b)}\}$  to trusted agents  $B_1$  and  $B_2$  that are symmetrically located on each side of Bob and that are separated by a straight-line distance  $d$ . Communication from Bob to each agent is encrypted using pre-shared key material. Once agents  $B_1$  and  $B_2$  have received  $r^{(b)}$ , they can (if they choose) later simultaneously communicate  $r^{(b)}$  to the trusted agents of Alice,  $A_1$  and  $A_2$ , who are sitting right next to  $B_1$  and  $B_2$ , respectively. This opens the commitment. The commitment is accepted when  $A_1$  and  $A_2$  later verify they have indeed received the same strings at the same time, and that the measurement results are consistent with the states she sent (hence security increases with  $N$ ). The duration of the commitment is  $d/2c$ , where  $c$  is the speed of light. Note that Bob's agents do not have to open the commitment as soon as they receive the string  $r^{(b)}$ , they can wait for as long as they wish. However, Alice has to consider that a cheating Bob could have stored the qubits in a quantum memory and retrieved them at the very last instant for the measurement. Hence, Alice's agents can only be sure that Bob was committed between  $t_o - d/2c$  and  $t_o$ , where  $t_o$  is the instant when they received  $r^{(b)}$ .

This original protocol presently appears impractical for terrestrial implementations with existing devices for direct bit commitments. The main reason is that imperfect preparation of true single photon states, combined with inefficient and slow single-photon detectors, seems unlikely to allow Alice to send, and Bob to detect, sufficiently many high-quality qubits in a time that is negli-

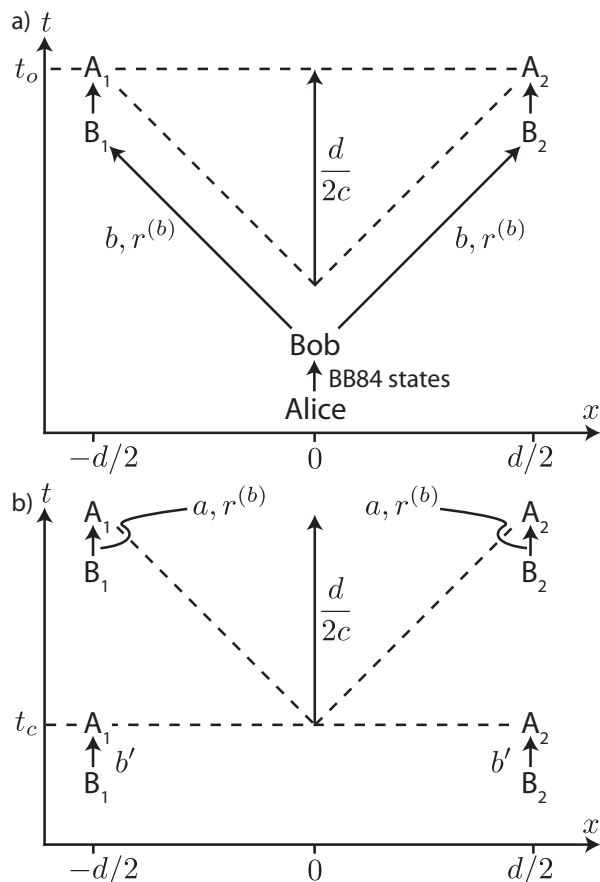


FIG. 1. a) Space-time ( $x$ - $t$ ) diagram of the protocol of [18]. b) The relativistic portion of a modified version of (a). The quantum exchange between Alice and Bob happens much before and is not shown.

gibly small compared to  $d/2c$  (i.e. at most 21.25 ms for  $d$  bounded by the Earth's diameter). Another problem is that it seems difficult for an honest Bob to communicate  $r^{(b)}$  to his agents at near light speed unless either there is a free-space line-of-sight communication channel between them, which requires  $d \lesssim 200$  km or so on Earth, or else neutrino pulses are sent through Earth to detectors, which requires expensive technology not widely available.

The protocol can be made more practical using a delayed-choice commitment; see Fig. 1-b. Specifically, the protocol starts when Alice and Bob exchange quantum bits, but the measurement basis of Bob, denoted  $b$ , is chosen at random and is not correlated with the bit he wishes to commit to. In this way, the quantum exchange can happen at any time and location before the commitment phase (defined below). In our implementation, after measuring all the quantum bits, Bob privately communicates the results to agents  $B_1$  and  $B_2$  separated by  $d$ , and also tells Alice which of the qubits she sent yielded a click in one of his detectors. This is important: Alice needs to know before the commitment phase which qubits were detected by Bob, otherwise there is a simple attack he can perform to break the protocol. Bob's

agents are now ready to start the commitment. For this, in our implementation, they simultaneously send at instant  $t_c$  a bit  $b' = b \oplus a$  to the nearby agents  $A_1$  and  $A_2$  of Alice, where  $a$  is the bit that Bob actually wants to commit himself to and  $b$  is the randomly committed bit. Note that as  $b$  is random, it effectively forms a one-time pad with which Bob encrypts  $a$ , and hence Alice cannot learn  $a$  from  $b'$ . At time  $t_c + d/2c$ ,  $B_1$  and  $B_2$  simultaneously reveal  $a$  and  $r^{(b)}$ , and  $A_1$  and  $A_2$  can check if the measurement results are consistent with  $b$ . The time  $t_c$  is chosen by Bob's agents.

Security against a malicious Alice in the ideal case is obvious, since Bob does not reveal any information about his quantum measurement until the opening of the commitment.

Security against a malicious Bob follows from special relativity [18–21], relying only on communication constraints that special relativity imposes. In relativistic cryptography, defining and analysing secure bit commitment requires specifying the possible spacetime coordinates of all information inputs and outputs [21], considering all possible spacetime locations of agents of each party. It is thus important to note that our protocol does not guarantee to Alice that either  $B_1$  or  $B_2$ , or any other agent of Bob's, knew the committed bit at time  $t_c$ . A technologically unlimited Bob could, for example locate an agent,  $B_0$ , halfway between  $B_1$  and  $B_2$ , who retains Alice's quantum states until time  $t_c$ , generates the bit  $b$  at that time, and sends the measurement outcomes at light speed to  $B_1$  and  $B_2$ .  $B_1$  and  $B_2$  could “blindly” initiate the commitment at time  $t_c$  by each sending a shared random bit  $b'$  to  $A_1$  and  $A_2$ , and later unveil by returning the measurement data corresponding to  $b$  together with the bit  $a = b' \oplus b$  effectively committed by  $B_0$ . However, our implementation *does* guarantee to Alice that Bob's agents collectively commit Bob to  $a$  by time  $t_c$  in the given frame, ensuring Bob is committed from at least  $t_c$  to  $t_c + d/2c$ .

It is important to realize that in the security proof it is enough to use the communication constraints imposed by special relativity. Specifically, we assume that no communication between  $B_1$  and  $B_2$  is possible *after the commitment phase is completed*. Quantum communication gives the advantage that Alice does need to share any secret data with her agents,  $A_1$  and  $A_2$ . On the other hand, if she tells them what quantum states were used in the commitment phase, then the dishonest Bob will be instantaneously caught cheating at the unveiling. For more information about the quantum advantage in these scenarios please refer to Appendix E.

Note that our protocol can straightforwardly be modified to use three (or more) separated agents of Bob's which gives certain advantages (for details refer to Appendix F). In the two-site protocol implemented, our security proof relies on the fact that not only is communication between  $B_1$  and  $B_2$  impossible between commitment and unveiling, but no other agent of Bob's can send classical or quantum information generated at any location

after  $t_c$  that reaches them both by  $t_c + (d/(2c))$ . Hence in the open phase  $B_1$  and  $B_2$  must generate their answers using disjoint quantum systems, which cannot have been jointly acted upon by any agent of Bob's after  $t_c$ .

We now discuss how experimental imperfections are taken into account in the security proof. Full details, including a step-wise description of all protocols, are given in the supplemental material. In practice, qubits sent by Alice can come from attenuated laser pulses yielding a Poisson distribution of the number of photons of mean value  $\mu$ . Pulses containing two or more photons prepared in the same quantum state can be split and measured by a malicious Bob in both bases. Using these results only, Bob's agents are no longer bound to a single quantum state for each declared detection, and they can choose to reveal the result (and the basis) of their choice, provided it is the same for  $B_1$  and  $B_2$ , without introducing any errors in  $r^{(b)}$ . Losses from Alice to Bob thus become important. Of all the pulses sent by Alice, only a fraction will yield a click in Bob's detectors. This fraction must therefore come from a majority of single-photon pulses, otherwise a cheating Bob can declare a detection only when Alice's pulse contains two or more photons at the output of her lab. Finally, Bob can try to cheat by measuring each photon in a basis that will minimize his probability of creating an error in  $r^{(b)}$  when he opens a bit which he was not committed to. While this probability can be minimized, it will always be strictly greater than zero.

Let us assume that Alice sends  $N$  pulses of light each containing on average  $\mu$  photons. Bob later declares that  $n$  of these pulses generated clicks in one of his detectors (and their labels, so that Alice knows the state she sent for each detection). Let  $p_{\text{det}} = n/N$  be the estimated detection probability declared by Bob. After Bob reveals  $r^{(b)}$ , Alice calculates the number of bits  $n'$  for which the preparation and measurement bases were the same, as well as the number of errors  $n_{\text{err}}$  on these bits (with respect to the states she sent). The qubit error rate (QBER) is defined as  $n_{\text{err}}/n'$ . In the limit  $N \rightarrow \infty$ , the protocol can be shown to be secure if

$$p_{\text{det}} > \frac{1 - e^{-\mu}(1 + \mu)}{1 - \frac{\text{QBER}}{\lambda}}, \quad (1)$$

where  $\lambda = \frac{1}{2} \left(1 - \frac{1}{\sqrt{2}}\right) \approx 0.146$ . To take into account the finite statistics, Alice and Bob both agree before the protocol on a maximal QBER value (denoted  $\delta$ ) and a minimal detection probability (denoted  $\gamma$ ): only those rounds of  $N$  pulses where  $p_{\text{det}} > \gamma$  will be used to do a commitment. Moreover, at the end of the protocol, Alice accepts Bob's commitment only if the  $\text{QBER} < \delta$  and if the timings of the two opening phases guarantee that no agent of Bob's can have communicated with both  $B_1$  and  $B_2$  between the commitment and unveiling.

Note that, on the one hand, suitable thresholds should be chosen to yield a robust protocol, i.e. the probability for it to fail when both parties are honest should be

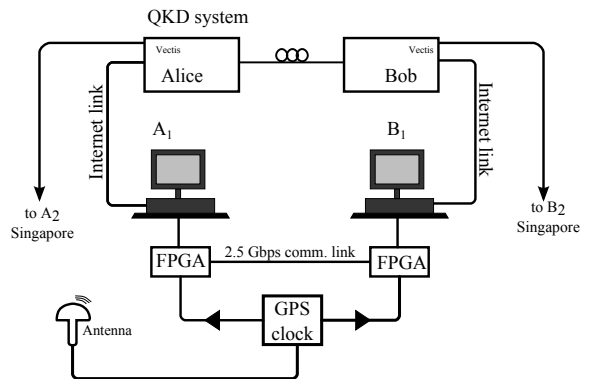


FIG. 2. Experimental setup located in Geneva. The setup located in Singapore is the same except for  $A_1$  and  $B_1$  that are replaced by  $A_2$  and  $B_2$ , and that there is no QKD system.

negligibly small. On the other hand, this increases the probability  $\epsilon$  for Bob to succeed in opening any value that he may desire after the commitment phase. The probability  $\epsilon$  can be calculated as shown in the supplemental material.

We implemented the modified protocol described above (Fig. 1-b). Thanks to its simplification, the quantum part can be implemented with a quantum key distribution (QKD) system using weak coherent pulses. In particular, we used a commercial quantum key distribution system “Vectis 5100” by ID Quantique. This system is based on the two-way “Plug-and-Play” configuration [22]: light pulses travel back-and-forth from Bob to Alice through a short optical fibre with negligible loss. When a pulse arrives on Alice side, she prepares the qubit and attenuates the light power down to the desired  $\mu$ . In order to prevent trojan-horse attacks, the incoming power is continuously monitored [23]. The system was installed in a office at the University of Geneva (see Fig. 2). The optical setup is divided in two quantum boxes that are respectively controlled by Alice and Bob, and that are connected by an optical fibre only. The quantum box of Alice (or Bob) records the relevant information for the protocol and communicates with agents  $A_1$  and  $A_2$  (or  $B_1$  and  $B_2$ ) through the internet.

One important modification to how the system operates in QKD needs to be done. Indeed, in QKD, Alice and Bob cooperate against an external eavesdropper, while in bit commitment they are potential adversaries. Hence, the software was modified to eliminate all the classical communication used for key sifting except for one thing, namely the reporting of which of the pulses sent by Alice generated a click in one of Bob's detectors. One side-channel attack then needs to be considered. Indeed, the probability  $p_{\text{det}}$  of Bob's honest apparatus might depend on the basis in which he measures. Alice could then exploit this difference to gain information about Bob's measurement basis, which is an issue for two-party protocols [24]. To eliminate this attack, Bob can test his

system for an imbalance, and correct it. In practice, we monitored during 40 hours the detection probability for each measurement basis, and found a ratio  $R = (95 \pm 1)\%$ . To eliminate this bias, we programmed Bob's quantum box to declare a detection in the most probable basis with a probability  $R$ . The measurement results of Bob are communicated to Bob's agents through the internet over a one-time-pad encrypted channel. Similarly, the list of states sent by Alice and for which Bob declared a detection is one-time-pad encrypted and sent to Alice's agents.

Classical agents  $A_1$  and  $B_1$  were also located in an office at the University of Geneva, while  $A_2$  and  $B_2$  were in an office on the campus of the National University of Singapore (see Fig. 2). The straight-line distance (through the Earth) between the two locations is about 9354 km, corresponding to a commitment time of  $d/2c = 15.6$  ms. This is close to the theoretical maximal value of  $\approx 21.2$  ms achievable with antipodal points on the surface of the Earth. A representation of the experimental setup is depicted in Fig. 2. Each of the classical agents is a standalone computer equipped with a field-programmable gate array (FPGA) programmed to execute the necessary steps of the protocol. Each FPGA is synchronized to universal time using a global positioning system (GPS) clock from Oscilloquartz. Communication between  $B_1$  and  $A_1$  (and similarly for  $B_2$  and  $A_2$ ) is done over a 2.5 Gbps optical link. The time required to communicate 7000 bits (which is a typical length for the string  $r^{(b)}$ ) is about  $3 \mu\text{s}$ , which is effectively negligible compared to 15 ms. The opening phase is thus considered to happen instantaneously. The time at which each step of the protocol occurred was measured with a time-to-digital converter and was found to be precise to within a fraction of a microsecond.

We realized fifty bit commitments by measuring in basis  $b = 0$ , and fifty more in basis  $b = 1$ , all of these with a  $\mu = (5.0 \pm 0.5) \times 10^{-2}$ , where the uncertainty corresponds to slightly overestimated daily fluctuations. Figures 3-a and 3-b display, respectively, the histogram of the observed  $p_{\text{det}}$  and of QBER for these commitments. The optical transmission from Alice to Bob, including detector efficiency was of the order of 6%, yielding a mean detection probability,  $p_{\text{det}}$ , of 0.32% (this includes the contribution of dark counts and multiphoton pulses). The QBER varied between 2.8 and 4.3% and averaged at 3.4%. The difference between the observed detection probabilities in bases "0" and "1" are attributed to statistical fluctuations.

Given the characterization described above, we choose to set the maximal tolerable value of the QBER to  $\delta = 5\%$  which, when combined with the average  $\mu$ , yield a value of 0.16% for the right-hand side of the asymptotic security condition (Eq. 1). For the minimum tolerable detection probability  $\gamma$ , we choose a value of 0.2% for with  $N = 220 \times 10^4$  pulses sent by Alice. The number of detections declared by Bob was about 7000 per commitment. The security parameter with these numbers is

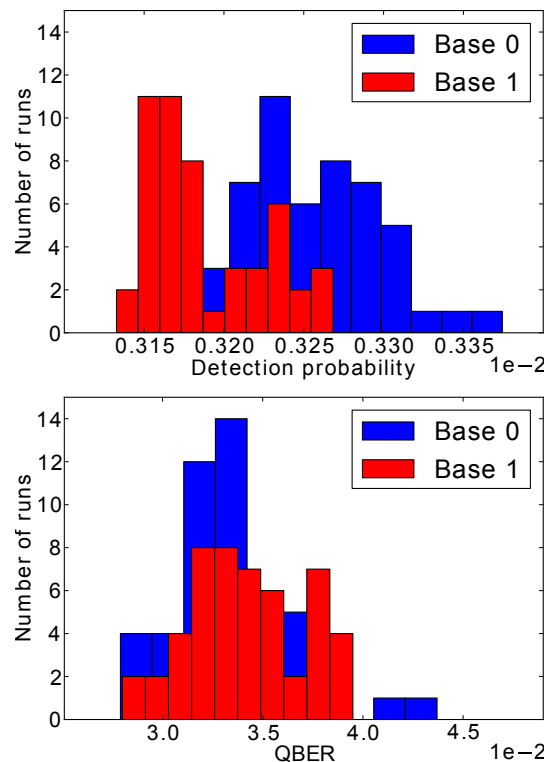


FIG. 3. a) Histogram of  $p_{\text{det}}$  for one hundred commitments. The blue (red) histogram shows the commitments made in basis  $b = 0$  ( $b = 1$ ). b) Histogram of the observed QBERs in each basis.

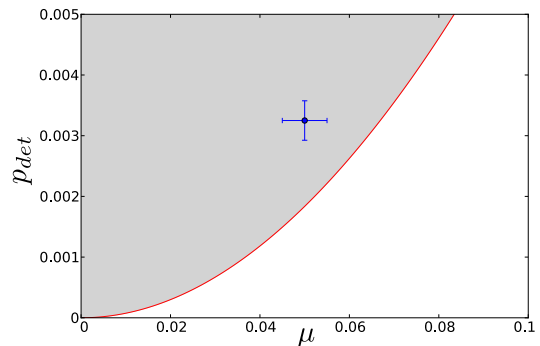


FIG. 4. Secure (grey) and insecure (white) regions corresponding to our experimental setup. The point corresponds to one of the commitments. The horizontal error bar correspond to the uncertainty on  $\mu$ , and the vertical one to the statistical uncertainty in estimating  $p_{\text{det}}$  for this particular commitment. Note that the point is well within the secure region, which was also the case for all the other commitments performed (not shown).

$\epsilon \leq 5.5 \times 10^{-8}$ . Figure 4 shows the asymptotic values of  $p_{\text{det}}$  as a function of  $\mu$  delimiting the secure region (in grey) from the insecure one (in white). The experimental point shown corresponds to one of the hundred commitments we realized. All the other ones are also within the secure region, which highlights the robustness of the

implementation.

We have demonstrated for the first time the possibility of implementing practical and secure bit commitment using quantum communication and special relativity, using a protocol with significant quantum advantages. Our implementation also demonstrates the possibility of implementing such commitments in real time for data acquired at a single location. This kind of system could potentially be useful in the high-speed trading stock market where short term commitments are sufficient. One of its main advantage is that the quantum part of the protocol can happen at any time before the committing phase. Hence, quantum data can be accumulated and communicated to agents located far away. The commitment time can in principle be multiplied [15] if Bob commits himself to all the bits of  $r^{(b)}$  instead of opening the commitment, and

this strategy can be refined [16] to sustain commitments indefinitely with constant communication rates (although security of indefinite commitments against general quantum attacks has not been proven). It would be an interesting theoretical and experimental challenge to extend the unconditionally secure commitment time significantly beyond the 15 ms obtained here.

*Acknowledgments.* This work was supported in part by the NCCR-QSIT, the European IP-SIQS, the Ministry of Education (MOE) and National Research Foundation Singapore through a Tier 3 Grant MOE2012-T3-1-009, and by the John Templeton Foundation. We thank IDQuantique and Oscilloquartz for useful discussions and technical help. We thank N. Walenta, C. Ci Wen Lim, P. Trinkler, M. Legré and D. Caselunghe for useful discussions.

- 
- [1] M. Blum, in *Advances in Cryptology: A Report on CRYPTO'81, Santa Barbara, California, USA* (1981) pp. 11–15.
- [2] A. Broadbent and A. Tapp, in *Proceedings of the IAVoSS Workshop On Trustworthy Elections (WOTE 2008)* (<http://eprint.iacr.org/2008/266>, 2008).
- [3] D. Mayers, *Phys. Rev. Lett.* **78**, 3414 (1997).
- [4] H.-K. Lo and H. F. Chau, *Phys. Rev. Lett.* **78**, 3410 (1997).
- [5] G. M. D'Ariano, D. Kretschmann, D. Schlingemann, and R. F. Werner, *Phys. Rev. A* **76**, 032328 (2007).
- [6] S. Winkler, M. Tomamichel, S. Heng, and R. Renner, *Phys. Rev. Lett.* **107** (2011), 10.1103/PhysRevLett.107.090502.
- [7] D. Aharonov, A. Ta-Shma, U. Vazirani, and A. C.-C. Yao, in *Proc. of the 32nd Annual ACM Symp. of Theory of Computing, Portland, Oregon, USA* (2000) pp. 705–714.
- [8] R. W. Spekkens and T. Rudolph, *Phys. Rev. A* **65**, 012310 (2001).
- [9] A. Kitaev, “Quantum coin-flipping,” Lecture delivered at the 2003 Annual Quantum Information Processing (QIP) Workshop, Mathematical Sciences Research Institute, Berkeley, CA.
- [10] A. Chailloux and I. Kerenidis, in *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on* (2011) pp. 354–362.
- [11] A. T. Nguyen, J. Frison, K. P. Huy, and S. Massar, *New Journal of Physics* **10**, 083037 (2008).
- [12] G. Berlin, G. Brassard, F. Bussi eres, N. Godbout, J. A. Slater, and W. Tittel, *Nat Commun* **2**, 561 (2011).
- [13] A. Pappa, P. Jouguet, T. Lawson, A. Chailloux, M. Legr e, P. Trinkler, I. Kerenidis, and E. Diamanti, [arXiv:1306.3368](https://arxiv.org/abs/1306.3368) (2013).
- [14] N. H. Y. Ng, S. K. Joshi, C. Chen Ming, C. Kurtsiefer, and S. Wehner, *Nat Commun* **3**, 1326 (2012).
- [15] A. Kent, *Phys. Rev. Lett.* **83**, 1447 (1999).
- [16] A. Kent, *Journal of Cryptology* **18**, 313 (2005).
- [17] A. Kent, *New Journal of Physics* **13**, 113015 (2011).
- [18] A. Kent, *Phys. Rev. Lett.* **109**, 130501 (2012).
- [19] S. Croke and A. Kent, *Phys. Rev. A* **86**, 052309 (2012).
- [20] J. Kaniewski, M. Tomamichel, E. Hanga, and S. Wehner, *Information Theory, IEEE Transactions on* **59**, 4687 (2013).
- [21] A. Kent, *Classical and Quantum Gravity* **29**, 224013 (2012).
- [22] G. Ribordy, J. D. Gautier, N. Gisin, O. Guinnard, and H. Zbinden, *Electronics Letters* **34**, 2116 (1998).
- [23] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, *Physical Review A* **73**, 022320 (2006).
- [24] N. H. Y. Ng, S. K. Joshi, C. C. Ming, C. Kurtsiefer, and S. Wehner, *Nat. Commun.* **3**, 1326 (2012).
- [25] H. Chernoff, *Ann. Math. Stat.* **23**, 493 (1952).
- [26] M. Ben-Or, S. Goldwasser, J. Kilian, and A. Wigderson, in *Proc. ACM STOC* (ACM Press, New York, New York, USA, 1988) pp. 113–131.
- [27] J.-R. Simard, *Classical and Quantum Strategies for Bit Commitment Schemes in the Two-Prover Model*, Master's thesis, McGill University (2007).

## Appendix A: Security Analysis

### 1. Outline

We begin our security analysis by describing how we model the honest and dishonest scenarios, i.e. what assumptions are made on the devices used to execute the protocol. In Section A 4 we describe a generalized protocol, in which the states emitted by Alice in the commit phase do not need to be *exactly* the four BB84 states and Bob's answer in the open phase does not need to match *exactly* the string encoded by Alice. For this fault-tolerant RBC protocol we present a security proof and compute an explicit security bound as a function of the number of qubits exchanged in the commit phase. This allows us to deal with errors on the communication channel in the experiment, but not yet with losses. Note that we do not simply extend the analysis of [20], but in fact our analysis is based on entirely new techniques that can - as a byproduct - also be used to obtain better parameters for the idealised setting. In Section A 5 we finally make our protocol robust against errors *and* losses. In particular, we show how the fault-tolerant RBC can be extended to deal with the presence of losses. We investigate how good the devices owned by the honest parties need to be to result in a protocol that is both robust and secure in the asymptotic limit. Finally, we show how to compute explicit security bounds for a finite-length protocol.

### 2. Modelling the experiment

Our model hinges on the following assumptions:

1. Alice uses a weak-coherent source with phase randomization.
2. Every photon is subject to independent noise and loss processes.
3. The detectors used by Bob for different bases have the same efficiency (independent of the state).

In the following we elaborate on what these assumptions mean for the honest and dishonest scenarios.

#### a. The honest scenario

Assuming that Alice uses a weak-coherent source with phase randomisation the ensemble emitted by the source is described by

$$\rho = \sum_{r=0}^{\infty} p_r |r\rangle\langle r|,$$

where  $p_r = e^{-\mu} \cdot \frac{\mu^r}{r!},$

$\mu$  is the average number of photons per pulse and  $|r\rangle$  is the Fock state of  $r$  photons. To model noise and losses between Alice and Bob for every photon we introduce an independent, binary random variable. Let  $\eta$  be the detection efficiency: every photon emitted by the source gets detected on Bob's side with probability  $\eta$ , otherwise it is lost. Applying such a process to the original state simply affects the probability distribution: the probability of Bob detecting  $r$  photons equals

$$p_r(\mu, \eta) = e^{-\mu\eta} \cdot \frac{(\mu\eta)^r}{r!}. \quad (\text{A1})$$

Now, suppose that  $n$  rounds are played (these are independent due to Assumption 2) and let  $N_r$  be the random variable counting the number of rounds in which exactly  $t$  photons were received. Then

$$\Pr[N_r = k] = \binom{n}{k} p_r^k (1 - p_r)^{n-k} \quad \text{and} \quad \mathbb{E}[N_r] = p_r n. \quad (\text{A2})$$

We model bit-flip errors in a similar way. For every photon detected on Bob's side and measured in the correct basis we introduce a bit flip error with probability  $Q$ .

*b. Dishonest scenarios*

In the case of one party being dishonest we need to assume that his/her devices are perfect — we want to achieve security against all-powerful adversaries. On the other hand, our assumptions still apply to the honest party.

If Alice is honest then we model her source as described before but now no losses or errors are guaranteed. Hence, what Bob sees are light pulses containing a Poisson-distributed (with parameter  $\mu$ ) number of photons. Moreover, we assume that Bob can do a perfect photon-counting experiment, which might help him to cheat. Bob can locate any number of agents anywhere in space-time, send classical or quantum information, including the quantum states sent by Alice, securely between his agents at light speed, and carry out arbitrary quantum operations on Alice's states and any ancillae.

If Bob is honest then we make no assumptions on the state emitted by Alice but we assume that the detectors Bob is using for the two different bases have the same efficiency (and that it does not depend on the state prepared by Alice).

### 3. Notation

We also fix some notation. Let  $[n] = \{1, 2, \dots, n\}$  and  $\mathcal{C}_n = \{0, 1\}^n$ . For  $\theta \in \mathcal{C}_n$  define the following sets

$$\begin{aligned} S &= \{k \in [n] : \theta_k = 0\}, \\ T &= \{k \in [n] : \theta_k = 1\}. \end{aligned}$$

Note that  $S \cup T = [n]$  and  $S \cap T = \emptyset$ . The (normalised) Hamming distance between two strings  $x, y \in \mathcal{C}_n$  is

$$d_H(x, y) := \frac{1}{n} |\{k \in [n] : x_k \neq y_k\}|.$$

The Hamming weight is defined as the distance from the string of all zeroes:

$$w_H(x) := d_H(x, 00 \dots 0).$$

### 4. Fault-tolerant RBC

The protocol presented in this section is essentially the original protocol in [18] with the following two extensions:

- In the original protocol Alice is required to create random BB84 states. In other words, she is required to generate two uniformly random bits,  $x, \theta \in \{0, 1\}$ , and create the state  $H^\theta|x\rangle$  of a qubit. We relax this constraint and simply require the two states for the same value of  $\theta$  to be orthogonal: if  $|\psi_x^\theta\rangle$  is the state corresponding to the random variables being  $x$  and  $\theta$  then we require

$$\langle \psi_0^0 | \psi_1^0 \rangle = \langle \psi_0^1 | \psi_1^1 \rangle = 0,$$

which for qubit states implies

$$|\psi_0^0\rangle\langle\psi_0^0| + |\psi_1^0\rangle\langle\psi_1^0| = |\psi_0^1\rangle\langle\psi_0^1| + |\psi_1^1\rangle\langle\psi_1^1| = \mathbb{1}. \quad (\text{A3})$$

The assumption that each pair forms a basis means that any prepare-and-measure scheme can be turned into an entanglement-based one. It will become evident from the security proof that the maximally incompatible states (e.g. the BB84 states) are the optimal choice from the security point of view.

- The strings that Bob's agents, Bill ( $B_1$ ) and Brian ( $B_2$ ), unveil in the open phase do not have to match Alice's string exactly. The opening will be accepted as long as Bill ( $B_1$ ) and Brian ( $B_2$ ) unveil the same string and it is close (in terms of Hamming distance) to the one measured by Alice.

*a. The protocol*

Let  $\delta$  be the error threshold and  $n$  be the number of rounds that Alice and Bob agree to play.

1. In every round Alice's agent Anne ( $A_1$ ) generates two uniformly random bits,  $x_k, \theta_k \in \{0, 1\}$ , and sends a pulse of photons in the state  $|\psi_x^\theta\rangle$  to Bill ( $B_1$ ).
2. (commit) Bill chooses the bit he wants to commit to, denoted by  $b$ . He measures all the incoming photons in the computational basis if  $b = 0$  or in the Hadamard basis if  $b = 1$ . Let  $y$  denote the string of measurement outcomes. Bill sends  $b$  and  $y$  to Brian ( $B_2$ ).
3. (open) To open the commitment Bill sends  $b$  and  $y$  to Anne and Brian does the same to Amy ( $A_2$ ). Let us denote the respective variables by  $b'$  and  $y'$ .
4. Amy forwards the string she received from Brian to Anne.
5. (verify) Let  $x_{S(T)}$  be the substring of  $x$  on the set  $S(T)$  as defined in Section A 3. Alice accepts the unveiling if the following three conditions are satisfied:
  - Bill and Brian unveil the same value,  $b = b'$ .
  - Bill and Brian submit the same outcome string,  $y = y'$ .
  - The string  $y$  is consistent with Alice's original encoding. More specifically, if  $b = 0$  we require  $d_H(x_S, y_S) \leq \delta$ , whereas if  $b = 1$  we require  $d_H(x_T, y_T) \leq \delta$ .

*b. Security for honest Alice*

A dishonest Bob can locate agents anywhere in space-time and transmit classical and quantum information between them securely at light speed. However, the relativistic signalling constraints implied by the timings in our protocol mean that the quantum states controlled respectively by Bill and Brian at the unveiling stage belong to disjoint Hilbert spaces that cannot have interacted since the commitment time  $t_c$ . We can thus restrict attention to the Hilbert spaces defining Bill and Brian's possible states at unveiling and to Bill and Brian's actions on these spaces.

We use the security model discussed in [20], in particular, we employ the global command scenario. Suppose that at the beginning of the open phase Bill and Brian receive a message asking them to unveil the bit  $b$  and let  $p_b$  be the probability that they succeed. We will show that for any strategy

$$p_0 + p_1 \leq 1 + \varepsilon, \tag{A4}$$

where  $\varepsilon$  decays exponentially in  $n$ , the number of rounds played.

The most general strategy by cheating Bob is to apply a quantum channel that ‘‘splits up’’ the quantum state into two parts: one of which is sent to Bill at or before unveiling, while the other is sent to Brian at or before unveiling. In the open phase, each of the agents will know which bit they are trying to unveil and will perform an arbitrary measurement on their respective systems to find out what answers they should send to Alice's agents.

Hence, the cheating strategy consists of a splitting channel and four POVMs (each agent has two different measurements depending on what they are trying to unveil). Let the measurement operators corresponding to Bill (Brian) trying to unveil  $b$  be  $\{P_y^b\}_{y \in \mathcal{C}_n}$  ( $\{Q_y^b\}_{y \in \mathcal{C}_n}$ ). Note that we can without loss of generality assume that these measurements are projective (by Neumark's dilation theorem since we make no assumptions on the dimension of the quantum state).

Moreover, let us purify the protocol. Generating one of the four states satisfying (A3) uniformly at random is equivalent to creating an EPR pair, and measuring one half of it in one of the two bases with equal probability. It is clear that the two protocols are equivalent from the security point of view but the purified version turns out to be easier to analyse.

Let  $\theta \in \mathcal{C}_n$  be the basis string that Alice picks uniformly at random. Then her measurement operators take the form  $|x^\theta\rangle = \bigotimes_{k=1}^n |\psi_{x_k}^{\theta_k}\rangle$  for  $x \in \mathcal{C}_n$ . Now, the checks that Alice performs can be written as projectors. Let  $\Pi_b^\theta$  be the projector corresponding to Alice accepting the unveiling of  $b$  given that her basis string is  $\theta$ . Then

$$\begin{aligned} \Pi_0^\theta &= \sum_{x \in \mathcal{C}_n} |x^\theta\rangle\langle x^\theta| \otimes \sum_{\substack{y \in \mathcal{C}_n \\ d_H(x_S, y_S) \leq \delta}} P_y^0 \otimes Q_y^0, \\ \Pi_1^\theta &= \sum_{x \in \mathcal{C}_n} |x^\theta\rangle\langle x^\theta| \otimes \sum_{\substack{y \in \mathcal{C}_n \\ d_H(x_T, y_T) \leq \delta}} P_y^1 \otimes Q_y^1. \end{aligned}$$



Let  $\rho$  be the tripartite state shared between Alice, Bill and Brian. Then the probability of successfully unveiling  $b$  can be written as

$$p_b = 2^{-n} \sum_{\theta \in \mathcal{C}_n} \text{tr}(\rho \Pi_b^\theta).$$

We want to bound  $p_0 + p_1$ , which corresponds to

$$p_0 + p_1 = 2^{-n} \sum_{\theta \in \mathcal{C}_n} \text{tr}(\rho[\Pi_0^\theta + \Pi_1^\theta]). \quad (\text{A5})$$

Adding the two projectors up yields

$$\Pi_0^\theta + \Pi_1^\theta = \sum_{x \in \mathcal{C}_n} |x^\theta\rangle\langle x^\theta| \otimes \left[ \sum_{\substack{y \in \mathcal{C}_n \\ d_H(x_S, y_S) \leq \delta}} P_y^0 \otimes Q_y^0 + \sum_{\substack{y \in \mathcal{C}_n \\ d_H(x_T, y_T) \leq \delta}} P_y^1 \otimes Q_y^1 \right].$$

The terms in the square bracket can be upper bounded by replacing one of the measurement operators by the identity matrix. Therefore,

$$\begin{aligned} \sum_{\substack{y \in \mathcal{C}_n \\ d_H(x_S, y_S) \leq \delta}} P_y^0 \otimes Q_y^0 + \sum_{\substack{y \in \mathcal{C}_n \\ d_H(x_T, y_T) \leq \delta}} P_y^1 \otimes Q_y^1 &\leq \left( \sum_{\substack{y \in \mathcal{C}_n \\ d_H(x_S, y_S) \leq \delta}} P_y^0 \right) \otimes \mathbb{1} + \mathbb{1} \otimes \left( \sum_{\substack{y \in \mathcal{C}_n \\ d_H(x_T, y_T) \leq \delta}} Q_y^1 \right) \\ &\leq \mathbb{1} \otimes \mathbb{1} + \sum_{\substack{y \in \mathcal{C}_n \\ d_H(x_S, y_S) \leq \delta}} P_y^0 \otimes \sum_{\substack{z \in \mathcal{C}_n \\ d_H(x_T, z_T) \leq \delta}} Q_z^1, \end{aligned}$$

where we applied the following operator inequality in the last step:

$$A \otimes \mathbb{1} + \mathbb{1} \otimes B = \mathbb{1} \otimes \mathbb{1} + A \otimes B - (\mathbb{1} - A) \otimes (\mathbb{1} - B) \leq \mathbb{1} \otimes \mathbb{1} + A \otimes B,$$

for any  $0 \leq A, B \leq \mathbb{1}$ . Therefore,

$$\Pi_0^\theta + \Pi_1^\theta \leq \mathbb{1} + \Pi_c^\theta,$$

where  $\Pi_c^\theta$  is the projector for the ‘‘cross-game’’, in which Bill has to unveil 0 and Brian has to unveil 1,

$$\Pi_c^\theta = \sum_{x \in \mathcal{C}_n} |x^\theta\rangle\langle x^\theta| \otimes \sum_{\substack{y \in \mathcal{C}_n \\ d_H(x_S, y_S) \leq \delta}} P_y^0 \otimes \sum_{\substack{z \in \mathcal{C}_n \\ d_H(x_T, z_T) \leq \delta}} Q_z^1.$$

Clearly,  $\text{tr}(\rho[\Pi_0^\theta + \Pi_1^\theta]) \leq 1 + \text{tr}(\rho \Pi_c^\theta)$ , which combined with (A5) gives

$$p_0 + p_1 \leq 1 + 2^{-n} \sum_{\theta \in \mathcal{C}_n} \text{tr}(\rho \Pi_c^\theta) = 1 + \text{tr}(\rho \langle \Pi_c^\theta \rangle) \leq 1 + \|\langle \Pi_c^\theta \rangle\|,$$

where  $\langle \cdot \rangle$  denotes averaging over  $\theta$ , i.e.  $\langle \Pi_c^\theta \rangle = 2^{-n} \sum_{\theta \in \mathcal{C}_n} \Pi_c^\theta$ , and  $\|\cdot\|$  denotes the Schatten  $\infty$ -norm. Changing the order of summation in  $\Pi_c^\theta$  gives

$$\Pi_c^\theta = \sum_{y, z \in \mathcal{C}_n} \sum_{\substack{x \in \mathcal{C}_n \\ d_H(x_S, y_S) \leq \delta \\ d_H(x_T, z_T) \leq \delta}} |x^\theta\rangle\langle x^\theta| \otimes P_y^0 \otimes Q_z^1.$$

Now, it is clear that only the  $x$ -dependent part needs to be averaged:

$$\langle \Pi_c^\theta \rangle = 2^{-n} \sum_{\theta \in \mathcal{C}_n} \Pi_c^\theta = \sum_{y, z \in \mathcal{C}_n} B_{yz} \otimes P_y^0 \otimes Q_z^1,$$

where

$$B_{yz} = 2^{-n} \sum_{\theta \in \mathcal{C}_n} \sum_{\substack{x \in \mathcal{C}_n \\ d_H(x_S, y_S) \leq \delta \\ d_H(x_T, z_T) \leq \delta}} |x^\theta\rangle\langle x^\theta|.$$

The product  $P_y^0 \otimes Q_z^1$  yields orthogonal projectors so the norm of  $\langle \Pi_c^\theta \rangle$  can be written as

$$\|\langle \Pi_c^\theta \rangle\| = \max_{y,z \in \mathcal{C}_n} \|B_{yz}\|. \quad (\text{A6})$$

To see which values of  $y$  and  $z$  maximize the norm we need to look more closely at the matrices  $B_{yz}$ . For every  $\theta$  define  $u(\theta) \in \mathcal{C}_n$  to be the string that satisfies  $[u(\theta)]_S = y_S$  and  $[u(\theta)]_T = z_T$ . Relabelling  $x \rightarrow x \oplus u(\theta)$  yields

$$B_{yz} = 2^{-n} \sum_{\theta \in \mathcal{C}_n} \sum_{\substack{x \in \mathcal{C}_n \\ w_H(x_S) \leq \delta \\ w_H(x_T) \leq \delta}} |(x \oplus u(\theta))^\theta \langle (x \oplus u(\theta))^\theta \rangle|.$$

The constraints on the second sum can be relaxed by noting that  $w_H(x_S) \leq \delta$  and  $w_H(x_T) \leq \delta$  implies  $w_H(x) \leq \delta$ . Hence,

$$B_{yz} \leq B'_{yz} = 2^{-n} \sum_{\theta \in \mathcal{C}_n} \sum_{\substack{x \in \mathcal{C}_n \\ w_H(x) \leq \delta}} |(x \oplus u(\theta))^\theta \langle (x \oplus u(\theta))^\theta \rangle|.$$

Now the second sum is independent of  $\theta$ , hence, the summation over  $\theta$  can be performed first. The product structure simplifies the summation

$$|x^\theta \langle x^\theta \rangle| = \bigotimes_{k=1}^n |\psi_{x_k}^{\theta_k} \langle \psi_{x_k}^{\theta_k} \rangle| \quad \text{and} \quad \sum_{\theta \in \mathcal{C}_n} \dots \iff \bigotimes_{k=1}^n \sum_{\theta_k \in \{0,1\}} \dots,$$

which implies that

$$2^{-n} \sum_{\theta \in \mathcal{C}_n} |(x \oplus u(\theta))^\theta \langle (x \oplus u(\theta))^\theta \rangle| = \bigotimes_{k=1}^n \rho_{x_k \oplus y_k, x_k \oplus z_k}.$$

where

$$\rho_{b,c} = \frac{1}{2} (|\psi_b^0 \rangle \langle \psi_b^0| + |\psi_c^1 \rangle \langle \psi_c^1|).$$

for  $b, c \in \{0,1\}$ . Note that  $\rho_{b,c} + \rho_{1-b,1-c} = \mathbb{1}$  so they are diagonal in the same eigenbasis. Therefore, without loss of generality we can write

$$\rho_{b,c} = \sum_{t \in \{0,1\}} \lambda_t^{b \oplus c} |e_t^{b \oplus c} \rangle \langle e_t^{b \oplus c}|,$$

for  $b, c \in \{0,1\}$ , where  $\lambda_0^{b \oplus c} + \lambda_1^{b \oplus c} = 1$ . In particular, we have

$$\left( \bigotimes_{k=1}^n \rho_{x_k \oplus y_k, x_k \oplus z_k} \right) \left( \bigotimes_{k=1}^n |e_{v_k}^{y_k \oplus z_k} \rangle \right) = \bigotimes_{k=1}^n \lambda_{x_k \oplus y_k \oplus v_k}^{y_k \oplus z_k} |e_{v_k}^{y_k \oplus z_k} \rangle.$$

Therefore, we also know the eigenbasis of

$$B'_{yz} = \sum_{\substack{x \in \mathcal{C}_n \\ w_H(x) \leq \delta}} \bigotimes_{k=1}^n \rho_{x_k \oplus y_k, x_k \oplus z_k},$$

and the highest eigenvalue equals simply

$$\|B'_{yz}\| = \max_{v \in \mathcal{C}_n} \sum_{\substack{x \in \mathcal{C}_n \\ w_H(x) \leq \delta}} \prod_{k=1}^n \lambda_{x_k \oplus y_k \oplus v_k}^{y_k \oplus z_k}.$$

Recall from (A6) that the expression we want to bound is

$$\max_{y,z \in \mathcal{C}_n} \|B'_{yz}\| = \max_{v,y,z \in \mathcal{C}_n} \sum_{\substack{x \in \mathcal{C}_n \\ w_H(x) \leq \delta}} \prod_{k=1}^n \lambda_{x_k \oplus y_k \oplus v_k}^{y_k \oplus z_k} = \max_{a,b \in \mathcal{C}_n} \sum_{\substack{x \in \mathcal{C}_n \\ w_H(x) \leq \delta}} \prod_{k=1}^n \lambda_{x_k \oplus b_k}^{a_k}.$$

Solving this maximization is easy (see Appendix B for details) and yields

$$\max_{y,z \in \mathcal{C}_n} \|B'_{yz}\| = \sum_{\substack{x \in \mathcal{C}_n \\ w_H(x) \leq \delta}} \prod_{k=1}^n \lambda_{x_k} = \sum_{k=0}^{\lfloor \delta n \rfloor} \binom{n}{k} \lambda_0^{n-k} \lambda_1^k,$$

where  $\lambda_0 = \max_{b,c} \lambda_b^c$  and  $\lambda_1 = 1 - \lambda_0$ . Finally, as we know that

$$\|\langle \Pi_c^\theta \rangle\| = \max_{y,z \in \mathcal{C}_n} \|B_{yz}\| \leq \max_{y,z \in \mathcal{C}_n} \|B'_{yz}\|$$

we obtain the security guarantee of the form presented in (A4) with

$$\varepsilon = \sum_{k=0}^{\lfloor \delta n \rfloor} \binom{n}{k} \lambda_0^{n-k} \lambda_1^k.$$

In the noiseless case ( $\delta = 0$ ) we obtain  $\varepsilon = \lambda_0^n$ , while for  $\delta > 0$  we use the Chernoff bound (D1) for the binomial distribution to get

$$\varepsilon \leq \exp\left(-\frac{1}{2}\left(\sqrt{\lambda_1} - \frac{\delta}{\sqrt{\lambda_1}}\right)^2 n\right), \quad (\text{A7})$$

for any  $\delta < \lambda_1$ . It is also clear that the cross-game becomes easy for the dishonest parties if  $\delta > \lambda_1$ . Hence,  $\lambda_1$  is the error threshold below which the cross-game becomes secure (asymptotically).

It is not clear whether the bit commitment game can be secure for any  $\delta > \lambda_1$  but if that is the case it cannot be proven using the cross-game reduction.

Finally, we want to connect  $\lambda_1$  to some measure of complementarity of the states emitted by Alice. We show in Appendix C that

$$\lambda_1 = \frac{1-t}{2},$$

where  $t = \max_{b,c \in \{0,1\}} |\langle \psi_b^0 | \psi_c^1 \rangle|$ . For example for the perfect BB84 states we get  $t = 1/\sqrt{2}$  and  $\lambda = \lambda_1 = 1/2 - 1/(2\sqrt{2}) \approx 0.146$ .

### c. Security for honest Bob

Since Alice does not receive any information from Bob before the open phase, she remains fully ignorant about the value of his commitment.

## 5. Fault-tolerant RBC with backreporting

If Alice were to implement the protocol presented in Section A 4 using a Poisson-type source she would need to set the mean number of photons very high so that the honest Bob detects clicks in the majority of rounds. However, then the probability of multi-photon emissions becomes significant which allows dishonest Bob to cheat. Here, we present a modified protocol, in which Bob is required to backreport the rounds he missed, which closes this security loophole.

### a. The protocol

The new protocol admits three parameters: the error allowance,  $\delta$ , the detection threshold,  $\gamma$ , and the mean number of photons per pulse emitted by Alice,  $\mu$ . We will see later that these parameters determine the trade-off between robustness and security.

1. In every round Alice's agent Anne ( $A_1$ ) generates two uniformly random bits,  $x_k, \theta_k \in \{0, 1\}$ , and sends a pulse of photons in the state  $|\psi_x^\theta\rangle$  to Bill ( $B_1$ ).

2. (commit) Bill chooses the bit he wants to commit to, denoted by  $b$ . He measures all the incoming photons in the computational basis if  $b = 0$  or in the Hadamard basis if  $b = 1$ . Let  $y$  denote the string of measurement outcomes.
3. (backreporting) Let  $\mathcal{M} \subseteq [n]$  be the *valid set*, namely the rounds in which Bill detected a click. Bill sends  $\mathcal{M}$  to Anne. Anne checks whether  $m = |\mathcal{M}| \geq \gamma n$ , otherwise she aborts the protocol. Denote the classical measurement outcomes on the valid set by  $y \in \{0, 1\}^m$ . Bill sends  $b$  and  $y$  to Brian ( $B_2$ ).
4. (open) To open the commitment Bill sends  $b$  and  $y$  to Anne and Brian does the same to Amy ( $A_2$ ). Let us denote the respective variables by  $b'$  and  $y'$ .
5. Amy forwards the string she received from Brian to Anne.
6. (verify) Let  $S$  and  $T$  be the following sets

$$S = \{k \in \mathcal{M} : \theta_k = 0\},$$

$$T = \{k \in \mathcal{M} : \theta_k = 1\}.$$

Note that these are essentially the sets defined in Section A 3 but truncated to the valid rounds only. Let  $x_{S(T)}$  be the substring of  $x$  on the set  $S(T)$ . Alice accepts the unveiling if the following three conditions are satisfied:

- Bill and Brian unveil the same value,  $b = b'$ .
- Bill and Brian submit the same outcome string,  $y = y'$ .
- The string  $y$  is consistent with Alice's original encoding. More specifically, if  $b = 0$  we require  $d_H(x_S, y_S) \leq \delta$ , whereas if  $b = 1$  we require  $d_H(x_T, y_T) \leq \delta$ .

The security for honest Bob follows straightforwardly from Assumption 3. If the probability of detecting a photon does not depend on the basis chosen by Bob then there is no correlation between the valid set and Bob's measurement choice. Therefore, despite having received the valid set Alice remains fully ignorant about Bob's measurement basis.

The remainder of the paper investigates the trade-off between robustness and security (for honest Alice). First, we consider the asymptotic limit (the number of rounds goes to infinity) and we establish simple conditions that the hardware and protocol parameters must satisfy for the protocol to be both robust and secure. These conditions define a region in the space of parameters, which we call the *feasible region*, and essentially tell us how good the equipment of the honest parties must be in order to implement the protocol securely. Finally, we show how to compute explicit security bounds for finite  $n$ .

### b. Asymptotic analysis

In the asymptotic limit all the random variables are concentrated around their average values. Therefore, to see for what values of the parameters that the protocol is in principle feasible it is enough to calculate the relevant averages.

*Robustness.* In the honest scenario if Bob has received at least one photon he will consider the round valid. For the protocol to go through he must have received at least  $\gamma n$  non-vacuum rounds. Hence, what we care about in the asymptotic limit is that the expected number of non-vacuum rounds (as observed by Bob) exceeds the detection threshold,  $\sum_{t=1}^{\infty} \mathbb{E}[N_r] > \gamma n$ . Expressing the expectation values through probabilities (A2) gives

$$\sum_{t=1}^{\infty} p_r > \gamma.$$

Replacing the sum by  $1 - p_0$  and evaluating it using (A1) we get

$$e^{-\mu\eta} + \gamma < 1.$$

If the first condition is satisfied then we only need to ensure that Bob's error rate is below the threshold. As the probability of making an error (obtaining the wrong outcome despite measuring in the right basis) in each round equals  $Q$ , the expected Hamming distance between what Alice encoded and what Bob measured also equals  $Q$ . Therefore, the second condition for robustness is  $Q < \delta$ .

*Security for honest Alice* In the case of dishonest Bob we assume that his fibre and detectors are perfect ( $\eta = 1$ ,  $Q = 0$ ) and, also, let us make the following modification that gives (slightly) more power to Bob but also makes the analysis simpler: we will replace all the vacuum rounds by single-photon emissions. Then,  $N_m := \sum_{k=2}^{\infty} N_k$ , the number of multi-photon emissions, becomes the only relevant parameter. If Bob receives two (or more) photons he will measure the first photon in the  $Z$  basis and the second one in the  $X$  basis and he will obtain valid answers to both questions (namely opening 0 and opening 1). This means that whichever bit he decides to unveil his answer will always be accepted by Alice. Therefore, the multi-photon rounds Bob can “win” for free and such rounds do not contribute to security. Hence, dishonest Bob’s optimal strategy will be to discard as many single-photon rounds as possible. It is clear that if Bob can discard all of them, he is left with multi-photon emissions only and no security is possible. Therefore, the first condition for security is: the number of multi-photon rounds is lower than Bob’s detection threshold

$$N_m < \lceil \gamma n \rceil.$$

After using up the entire backreporting allowance the number of valid rounds is  $m = \lceil \gamma n \rceil$  but there are only  $\lceil \gamma n \rceil - N_m$  (which is now guaranteed to be a positive number) single-photon rounds among them. Therefore, honest Alice thinks that they are playing a standard bit commitment game of  $\lceil \gamma n \rceil$  rounds but there is a certain number of multi-photon ones, which Bob can win “for free”. Hence, he can use his “error allowance” for the single-photon rounds and the security we achieve is that of playing a game of  $\lceil \gamma n \rceil - N_m$  rounds with the non-fractional error allowance of  $\delta \gamma n$ , which gives the effective (fractional) error allowance of

$$\delta' = \frac{\delta \gamma n}{\lceil \gamma n \rceil - N_m}.$$

In the proof (Section A 4 b) we find that we can only prove security if the effective (fractional) error allowance,  $\delta'$ , satisfies

$$\delta' < \lambda_1,$$

where  $\lambda$  measures how complementary the bases used by Alice are. Hence, in our case we require

$$\delta \gamma n < (\lceil \gamma n \rceil - N_m) \lambda_1. \quad (\text{A8})$$

In the asymptotic limit we look at the expectation value:

$$\mathbb{E}[N_m] = [1 - e^{-\mu}(1 + \mu)]n,$$

which applied to (A8) gives

$$e^{-\mu}(1 + \mu) + (1 - \frac{\delta}{\lambda_1})\gamma > 1.$$

*Asymptotically feasible region.* The asymptotically feasible region is defined as the region in the space of parameters that satisfies the robustness and security constraints, namely:

$$\begin{aligned} e^{-\mu\eta} + \gamma &< 1, \\ Q &< \delta, \\ e^{-\mu}(1 + \mu) + (1 - \frac{\delta}{\lambda_1})\gamma &> 1. \end{aligned}$$

It is clear that the first two inequalities (the robustness conditions) can be assumed to be equalities, which fixes  $\delta$  and  $\gamma$  and leaves us with only one, but fairly complicated, condition:

$$e^{-\mu}(1 + \mu) + (1 - \frac{Q}{\lambda_1})(1 - e^{-\mu\eta}) > 1.$$

This expression allows to verify whether for devices of given quality (quantified by  $\lambda_1$ ,  $Q$  and  $\eta$ ) there exists  $\mu$  that makes the protocol both robust and secure.

c. *Finite-size effects*

The asymptotic analysis is relevant as  $n \rightarrow \infty$  but in any practical scenario the number of rounds will be finite. Alice and Bob, having performed the protocol, will want explicit security guarantees (as a function of  $n$ ). Note that robustness is verified experimentally, hence, there is no need to calculate it and security for honest Bob is perfect by assumption. Therefore, we are only left with security for honest Alice.

Let  $E$  denote the event that Bob wins the cross-game. As shown in Section A 4 b,  $\Pr[E]$  is an upper bound on the security parameter of the bit commitment protocol. Conditioning on the number of multi-photon emissions yields

$$\Pr[E] = \sum_{k=0}^n \Pr[E|N_m = k] \Pr[N_m = k]. \quad (\text{A9})$$

Equation (A7) allows us to bound  $\Pr[E|N_m = k]$  as long as the number of multi-photon emissions is below the threshold,  $k_t = \gamma n(1 - \delta/\lambda_1)$ . For  $k < k_t$  we have

$$\Pr[E|N_m = k] \leq \exp\left(-\frac{1}{2}\left(\sqrt{(\gamma n - k)\lambda_1} - \frac{\delta\gamma n}{\sqrt{(\gamma n - k)\lambda_1}}\right)^2\right), \quad (\text{A10})$$

whereas for  $k \geq k_t$  there is no security and the trivial bound,  $\Pr[E|N_m = k] \leq 1$ , is the best we can hope for. Noting that the random variable  $N_m$  is Poisson-distributed with the parameter  $p_m = 1 - e^{-\mu}(1 + \mu)$  allows us to evaluate the bound numerically.

### Appendix B: A simple maximization problem

Suppose one is given two coins and is required to make  $n$  independent coin flips. The goal is to maximize the probability of guessing correctly at least a certain fraction of the outcome string. Which coin should be used in every round and what is the optimal guess? Intuitively, it is clear that one should pick the more biased coin and always guess its more probable outcome. For completeness we provide a rigorous proof of this statement.

For  $b, c \in \{0, 1\}$  let  $\lambda_b^c \in [0, 1]$  be such that  $\sum_b \lambda_b^c = 1$ . In other words,  $b$  corresponds to the outcome, while  $c$  tells us which coin has been used. Let  $y \in \mathcal{C}_n$  be the string denoting which coin was used in each round and let  $X$  be the random variable corresponding to the string of outcomes. Then for  $x \in \mathcal{C}_n$

$$\Pr[X = x] = \prod_{k=1}^n \lambda_{x_k}^{y_k}.$$

For an arbitrary string  $z \in \mathcal{C}_n$  the probability of being correct on at least  $n(1 - \delta)$  position equals

$$\Pr[\text{d}_H(X, z) \leq \delta] = \sum_{\substack{x \in \mathcal{C}_n \\ \text{d}_H(x, z) \leq \delta}} \prod_{k=1}^n \lambda_{x_k}^{y_k} = \sum_{\substack{x \in \mathcal{C}_n \\ \text{w}_H(x) \leq \delta}} \prod_{k=1}^n \lambda_{x_k \oplus z_k}^{y_k}.$$

Now, we want to see what values of  $y, z \in \mathcal{C}_n$  maximize this expression. Let us write down the summation corresponding to the  $j$ -th bit explicitly

$$\sum_{\substack{x \in \mathcal{C}_n \\ \text{w}_H(x) \leq \delta}} \prod_{k=1}^n \lambda_{x_k \oplus z_k}^{y_k} = \lambda_{z_j}^{y_j} \left[ \sum_{\substack{x \in \mathcal{C}_{n-1} \\ \text{w}_H(x) \leq n\delta/(n-1)}} \prod_{k \neq j} \lambda_{x_k \oplus z_k}^{y_k} \right] + (1 - \lambda_{z_j}^{y_j}) \left[ \sum_{\substack{x \in \mathcal{C}_{n-1} \\ \text{w}_H(x) \leq (n\delta-1)/(n-1)}} \prod_{k \neq j} \lambda_{x_k \oplus z_k}^{y_k} \right].$$

As the first square bracket is never smaller than the second one (because all the terms counted in the second bracket are also counted in the first one) we should always maximize  $\lambda_{z_j}^{y_j}$ . This observation is always true, regardless of the choices made for other rounds, and it applies to every round in the same way. Therefore, if  $\lambda_0 = \max_{b,c} \lambda_b^c$  and  $\lambda_1 = 1 - \lambda_0$  we find

$$\max_{z \in \mathcal{C}_n} \Pr[\text{d}_H(X, z) \leq \delta] = \sum_{\substack{v \in \mathcal{C}_n \\ \text{w}_H(v) \leq \delta}} \prod_{k=1}^n \lambda_{v_k} = \sum_{k=0}^{\lfloor \delta n \rfloor} \binom{n}{k} \lambda_0^{n-k} \lambda_1^k,$$

which recovers the intuitively correct solution.

### Appendix C: Explicit diagonalisation of a mixture of two pure states

Let  $|\psi_0\rangle$  and  $|\psi_1\rangle$  be two pure states such that  $\langle\psi_0|\psi_1\rangle = ce^{i\phi}$ . Then the eigenvalue decomposition of  $\rho = \frac{1}{2}(|\psi_0\rangle\langle\psi_0| + |\psi_1\rangle\langle\psi_1|)$  is

$$\rho = \lambda_+|e_+\rangle\langle e_+| + \lambda_-|e_-\rangle\langle e_-|,$$

where

$$\lambda_{\pm} = \frac{1 \pm c}{2},$$

$$|e_{\pm}\rangle = \frac{1}{\sqrt{2(1+c)}} \left[ |\psi_0\rangle \pm e^{-i\phi} |\psi_1\rangle \right].$$

### Appendix D: Chernoff bounds for the binomial distribution [25]

Let  $X_1, X_2, \dots, X_n$  be independent random variables taking on values 0 or 1. Then let  $X = \sum_{i=1}^n X_i$  and  $\mu$  be the expectation value of  $X$ . Then for any  $\delta > 0$  the following inequalities hold.

$$\Pr[X < (1 - \delta)\mu] < \left( \frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^{\mu} \leq \exp\left( -\frac{\mu\delta^2}{2} \right),$$

$$\Pr[X > (1 + \delta)\mu] < \left( \frac{e^{\delta}}{(1 + \delta)^{1+\delta}} \right)^{\mu} \leq \exp\left( -\frac{\mu\delta^2}{3} \right).$$

The bounds can be expressed more conveniently as

$$\Pr[X < s] < \exp\left( -\frac{1}{2} \left( \sqrt{\mu} - \frac{s}{\sqrt{\mu}} \right)^2 \right), \quad (D1)$$

$$\Pr[X > s] < \left( \frac{\mu}{s} \right)^s e^{s-\mu}. \quad (D2)$$

### Appendix E: Remarks on Quantum Advantage

#### 1. Remarks 1

Note that our protocol requires no communication from  $A_1$  or  $A_2$  at commitment or unveiling. These agents need trusted spacetime coordinates and reliable classical data storage, but do not need secure sites to keep data secret before or during the protocol: Alice may keep the secret data about her quantum states at some other secure site for later cross-checking. No unconditionally secure classical bit commitment protocol (relativistic or otherwise) allows this freedom, which is potentially valuable in the technologically asymmetric case where Alice is less able than Bob to maintain a network of secure sites. Conversely, if Alice shares the quantum state information securely with  $A_1$  and  $A_2$  and they keep it securely, Bob's cheating possibilities are quite constrained. If Bob cheats at commitment, i.e.  $B_1$  and  $B_2$  send different bit values for  $b'$ , he will be exposed as soon as  $A_1$  and  $A_2$  crosscheck, whether or not he unveils. If they send the same  $b'$  and Bob unveils at any time between  $t_c$  and  $t_c + (d/2c)$ , but tries to cheat at this stage, at least one of the  $A_i$  will learn instantly – without needing to cross-check – since he must send inconsistent information about  $b$  to at least one of them. Again, no unconditionally secure classical bit commitment protocol has this last feature, which allows (in a relativistically appropriate sense) instant action to be taken in the event of cheating at the unveiling stage.

#### 2. Remarks 2

In the main part of our manuscript, we already briefly touched upon different security models [20, 21]. In particular, one can see [20] our two site protocol as achieving security in a setting where communication is not allowed during certain phases of the protocol. For a bit commitment protocol, one can thereby distinguish several natural phases,

namely the commit phase, the wait phase, and finally the open and verification phase. One can then ask whether security is possible if communication is not possible in any of these phases.

This model is interesting because, while imposing minimum communication constraints necessary to evade the impossibility result [20, Sec. III.A], it is also sufficiently strong to allow for secure quantum bit commitment. One can show that no classical protocol implemented in this model can be secure [20, Sec. IV.B]. From the above description of the setup with multiple agents occupying different locations (see Fig. 1-b), we can readily deduce that the communication restraints are satisfied. However, the setup enforces even stronger communication constraints. For an in-depth discussion and illustrations concerning the different phases of bit commitment and communication models we refer to [20].

In [20] it was shown that there is a strict quantum advantage when we take the minimum number of phases in which communication is not allowed, namely the wait and open phase. Communication during the open phase, however, was possible. As mentioned before, the significant distance between Singapore and Geneva naturally guarantees such a lack of communication and thus enables secure bit commitment.

However, since of course the locations of our experimental labs are fixed to be in Geneva and Singapore, the distance also guarantees a lack of communication already during the commit phase itself. This scenario, when communication is never allowed, has long been studied in computer science [26], where in particular it is known that there exists a classical protocol for bit commitment that is secure as long as communication is *never* possible at all (see e.g. [20] for further background). Similar protocols were derived using relativistic assumptions [15, 16], whose first round is closely related to [26]. The classical protocol of [26] is secure against all quantum attacks ([16]; see also [27]), essentially because a successful quantum attack would imply the possibility of communication between the separate sites. However, these classical protocols would not be secure even against classical attacks if communication were possible during the commit phase, and so the two site quantum protocol implemented here would offer a theoretical advantage in such a setting.

It is a very interesting open question, whether there exists an natural experimentally implementable physical scenario in which communication during the commit phase is allowed, but communication during the wait and open phase is forbidden.

Looking at the protocol of [26] one could imagine that supplementing special relativity with the assumption that only a finite amount of communication is possible within a given time frame, and that an infinitesimally small point in space can, by the holographic principle, not store an arbitrary amount of randomness—required to execute the protocol of [26]—can possibly lead to such a scenario. However, these are only clues we can derive from a particular classical protocol [26] and they do not by itself answer this interesting open question.

## Appendix F: Extension for three agents

Suppose that Bob sites his third agent,  $B_0$ , equidistant from  $B_1$  and  $B_2$  on the shorter arc of a great circle joining them and preshare  $b$  and  $r^{(b)}$  among all three agents, while Alice locates her third agent,  $A_0$ , right next to  $B_0$ . Then, instead of  $B_1$  and  $B_2$  committing the bit  $a = b \oplus b'$ ,  $B_0$  can do so by sending  $b'$  to  $A_0$  at  $t_c$ , while  $B_1$  and  $B_2$  unveil by sending  $b$  and  $r^{(b)}$  at  $t_c + d/(2c)$ . This allows  $B_0$  to commit to a bit as soon as he learns it, without presharing it with  $B_1$  and  $B_2$ , while Alice is guaranteed that Bob's agents have collectively committed him from  $t_c$  until  $t_c + d/(2c)$ .

This freedom to commit bits in real time could be made more practical if the  $B_i$  preshare a large number of quantumly generated random bits  $b_j$  and commitment data  $r^{(b_j)}$  and if  $B_1$  and  $B_2$  release these in rapid sequence at coordinated times pre-agreed with  $B_0$ .  $B_0$  can then commit a bit  $a = b_j \oplus b'$  at any time of his choice, by sending  $A_0$  both  $b'$  and the value  $j$  (chosen so that the unveiling times of  $r^{(b_j)}$  guarantee commitment for roughly  $t_c + (d/(2c))$ ). Adding a third site adds logistical complexity but no significant technological challenge beyond those addressed in our implementation, which thus demonstrates the practicality of secure real time relativistic quantum commitment of bits acquired at a single location.