QUANTUM BIT COMMITMENT: A COMPLETE CLASSIFICATION OF PROTOCOLS

GIACOMO MAURO D'ARIANO

Quantum Optics & Information Group *of the INFM*
Dipartimento di Fisica "A. Volta", via Bassi 6, I-27100 Pavia, Italy

Department of Electrical and Computer Engineering,
Northwestern University, Evanston, IL 60208

This paper addresses the controversy between Mayers, Lo and Chau[1] on one side, and Yuen[2] on the opposite side, as to whether or not unconditionally secure protocols exist. For such purpose, a complete classification of all possible bit-commitment protocols is given, including all possible cheating attacks. For the simplest class of protocols (non-aborting and with complete and perfect verification), it is shown how a game-theoretical situation naturally arises. For these protocols, bounds for the cheating probabilities are derived, which turn out to be different from those given in the impossibility proof.[1] The whole classification and analysis has been carried out using a *finite open-system* approach. The discrepancy with the impossibility proof is explained on the basis of the implicit adoption of a *closed-system approach*—equivalent to modeling the commitment as being performed by two fixed machines interacting unitarily in an overall *closed system*. However, it is shown that the closed-system approach for the classification of commitment protocols unavoidably leads to infinite dimensions, which then invalidates the continuity argument at the heart of the impossibility proof.

1 Introduction

It is of practical relevance to establish if there exist secure quantum bit commitment protocols, since quantum bit commitment is a crucial element for building up more sophisticated protocols, such as remote quantum gambling, coin tossing, and unconditionally secure two-party quantum computation.

In bit commitment Alice provides Bob with a piece of evidence that she has chosen a bit $b = 0, 1$ which she commits to him. Later, Alice will open the commitment, revealing the bit b to Bob, and proving that it is indeed the committed bit with the evidence in Bob's possession. Therefore, Alice and Bob should agree on a protocol which satisfies simultaneously the following three requirements: (1) it must be *concealing*, namely Bob should not be able to retrieve b before the opening; (2) it must be *binding*, namely Alice should not be able to change b after the commitment; (3) it must be *verifiable*, namely Bob must be able to check b against the evidence in his possession, according to the rules of the protocol. In an in-principle proof of security of the commitment it is supposed that both parties possess unlimited technology, e.g., computational power, space, time, etc., and the protocol is said to be *unconditionally secure* if neither Alice nor Bob can cheat with significant probability of success as a consequence of physical laws.

In 1993, a quantum-mechanical protocol was proposed,[3] and the uncon-

ditional security of this protocol was generally accepted for long time. The insecurity of this protocol was shown by Mayers, Lo and Chau[1] in 1997, by recognizing the possibility for Alice to cheat by entangling the committed evidence with a quantum system in her possession, and it was argued that no unconditionally secure protocol is possible. Finally, after 2000, Yuen[2] presented some protocols which challenged the previous impossibility proof, mostly on the basis of the possibility of encoding the bit on an *anonymous state* given to Alice by Bob and known only to him, and suggesting the use of *decoy systems* that make the protocol concealing in the limit of infinitely-many systems, with the possibility for Bob to perform his quantum measurement before Alice opening, whence disputing the general availability of EPR cheating for Alice.

In this paper, in order to clarify the controversy, we will present a classification of all possible bit-commitment protocols based on a single commitment step, analyzing the main cheating strategies for both parties (a full derivation of the classification, the reduction of multi-step commitments to a single step, and a more exhaustive analysis of cheating attacks can be found in Ref. 4, of which the present paper is a much shorter version). For the simplest class of protocols (non-aborting, with complete and perfect verification) we will show how a game-theoretical situation naturally arises. Bounds for the cheating probabilities of these protocols are presented, which are different from those given in the impossibility proof.[1] In the final discussion we will see how the discrepancy between the two opposite analysis arises, due to the implicit adoption in the impossibility proof of a *closed-system approach*, equivalent to modeling the commitment as being performed by two fixed machines interacting unitarily in an overall *closed system*. However, it is shown that such modeling, along with the requirement of unlimited technology, necessarily leads to infinite dimensions, which invalidates the continuity argument at the heart of the impossibility proof.

2 The classification of protocols

The most general bit-commitment scheme with a single step is of the form: (1) Bob prepares the Hilbert space H with the *anonymous state* $|\varphi\rangle \in \mathsf{H}$, and sends H to Alice; (2) Alice *modulates* the value $b = 0, 1$ of the committed bit on the anonymous state $|\varphi\rangle$ and sends the output back to Bob. The *bit modulation* is a quantum operation (QO) $\mathrm{M}^{(b)}$ parametrized by b. Such scheme contains all possibilities, including Yuen's protocols,[2] and the protocols considered by Mayers, Lo and Chau,[1] which correspond to *openly known* $|\varphi\rangle$. In general, the output Hilbert space K of the QO will be different from H, since Alice can send back to Bob a quantum system different from what he sent to her.

In Ref. 4 a complete classification of all possible protocols is derived, on the basis of the fact that since Alice has unlimited technology, she can always achieve the encoding QO's $\mathrm{M}^{(b)}$ of the committed bit value b via a

perfect pure measurement. For *non-aborting protocols*, this corresponds to the following QO's

$$\mathrm{M}^{(b)}(|\varphi\rangle\langle\varphi|) = \mathrm{Tr}_{\mathsf{F}\otimes\mathsf{P}}[U^{(b)}(|\varphi\rangle\langle\varphi| \otimes |\omega\rangle\langle\omega|_{\mathsf{A}} \otimes \rho_{\mathsf{P}})U^{(b)\dagger}], \tag{2.1}$$

where A is the preparation ancilla/decoy Hilbert space prepared in the state $|\omega\rangle$; F is the measurement ancilla Hilbert space on which Alice performs a complete von Neumann measurement, and we have that $\mathsf{K} \otimes \mathsf{F} \simeq \mathsf{H} \otimes \mathsf{A}$; P is the space of the *secret parameter*, say j, which is needed in order to make the protocol *concealing* and at the same time *verifiable* (so that the modulation is actually a choice between two *ensembles* of QO's $\{\mathrm{M}_j^{(b)}\}$ for $b = 0, 1$). Therefore, the best option for Alice is to achieve the encoding QO by preparing the ancilla/decoy state $|\omega\rangle_{\mathsf{A}} \in \mathsf{A}$, performing the unitary transformation $U^{(b)}$ on $\mathsf{H}\otimes\mathsf{A}$, making a complete von Neumann measurement on F, with outcome say i, and finally send K to Bob. The partial trace on $\mathsf{F}\otimes\mathsf{P}$ on the basis $\{|i\rangle \otimes |j\rangle\}$, which describes Alice's measurement, corresponds to the Kraus decompositions $\mathrm{M}^{(b)} = \sum_{ji} p_j E_{ji}^{(b)} \cdot E_{ji}^{(b)\dagger}$, where j is the *secret parameter* and i is the *secret outcome*, and the probabilities $p_j = \langle j|\rho_{\mathsf{P}}|j\rangle$ for j will depend on the preparation ρ_{P}. In a protocol which is completely and perfectly verifiable Alice tells b, j and i to Bob, who verifies the state $E_{ji}^{(b)}|\varphi\rangle$. Since the local QO's on K and $\mathsf{F} \otimes \mathsf{P}$ commute, Alice has the possibility of: (1) first sending K to Bob; (2) then performing the measurement on $\mathsf{F} \otimes \mathsf{P}$ at the very last moment of the opening. As we will see, this is the basis for Alice's EPR cheating attacks. Notice that strictly trace-decreasing QO's, i.e., aborting protocols, pose limitations to Alice's EPR cheating. In fact, Alice cannot delay the abortion of the protocol at the opening, and must declare it at the commitment. Since both secret parameters j and i can be conveniently measured by Alice, they can be treated on equal footings as a single parameter $J \equiv (j, i)$. With the notation $E_J^{(b)} \doteq \sqrt{p_j} E_{ji}^{(b)} \in \mathsf{B}(\mathsf{H}, \mathsf{K})$, the maps write

$$\mathrm{M}^{(b)}(|\varphi\rangle\langle\varphi|) = \sum_j p_j \mathrm{M}_j^{(b)}(|\varphi\rangle\langle\varphi|) = \sum_J E_J^{(b)}|\varphi\rangle\langle\varphi|E_J^{(b)\dagger}. \tag{2.2}$$

3 Cheating

For a discussion of all possibilities for cheating see Ref. 4. Here we analyze the only the attacks that are useful for both Alice and Bob.

Alice cheating. After the commitment and before the opening Alice can try to cheat by performing a unitary transformation V on $\mathsf{F} \otimes \mathsf{P}$: this is the so-called EPR attack. Without changing the QO's $\mathrm{M}^{(b)}$, the maneuver will change their Kraus decompositions—which are relevant at the opening—as $\{E_J^{(b)}\} \to \{E_J^{(b)}(V)\}$, keeping the cardinality, in the following way

$$E_J^{(b)}(V) = \sum_L E_L^{(b)} V_{JL}, \qquad V_{JL} = \langle J|V|L\rangle. \tag{3.3}$$

The probability that Alice can cheat successfully in pretending to have committed, say, $b = 1$, whereas she committed $b = 0$ instead, is given by

$$P_c^A(V,\varphi) = \sum_J \frac{|\langle\varphi|E_J^{(0)\dagger}(V)E_J^{(1)}|\varphi\rangle|^2}{\|E_J^{(1)}\varphi\|^2}, \tag{3.4}$$

which depends on the anonymous state $|\varphi\rangle$ and on the cheating transformation V. Without any knowledge of $|\varphi\rangle$, the best that Alice can do is to adopt a conservative strategy, by maximizing her probability of cheating in the worst case, corresponding to the *minimax* choice of V

$$(P_c^A)_\mu \doteq \max_V \min_\varphi P_c^A(V,\varphi). \tag{3.5}$$

It is evident that in this way a game-theoretical situation arises, in which Bob chooses $|\varphi\rangle$ and Alice chooses V, with the probability $P(V,\varphi)$ playing the role of a *payoff matrix*. The actual game situation is more complicated—due for example to Bob cheating—and will be analyzed elsewhere.

Bob cheating. Bob can try to cheat by making the *best discrimination* between the two maps $\mathrm{M}^{(b)} = \sum_j p_j \mathrm{M}_j^{(b)}$. However, since he doesn't know the probabilities p_j actually used by Alice, his strategy will be suboptimal, and his actual cheating probability P_c^B will be lower than the probability $(P_c^B)_{\mathrm{opt}}$ corresponding to the optimal strategy with the right probabilities p_j. Since map-discrimination is generally more reliable with the map acting locally on an entangled state,[5] instead of preparing $|\varphi\rangle \in \mathsf{H}$ Bob prepares an entangled state on $\mathsf{H}\otimes\mathsf{R}$ and sends only H to Alice. Therefore, for equally probable bit values $b = 0, 1$, Bob's optimal probability of cheating is bounded as follows[4]

$$P_c^B \le (P_c^B)_{\mathrm{opt}} = \frac{1}{2} + \frac{1}{4}\left\|\mathrm{M}^{(1)} - \mathrm{M}^{(0)}\right\|_{cb}, \tag{3.6}$$

where $\|\cdot\|_{cb}$ denotes the completely-bounded (CB) norm.

Bounds for cheating probabilities. If the protocol is perfectly concealing the CB-norm in Eq. (3.6) is zero, and the two maps are the same, whence the their Kraus decompositions are connected via a unitary transformation V on $\mathsf{F}\otimes\mathsf{P}$, and Alice can cheat with probability one. Let's consider now the case in which Bob's optimal probability of cheating $(P_c^B)_{\mathrm{opt}}$ is infinitesimally close to $\frac{1}{2}$, namely $\|\mathrm{M}^{(1)} - \mathrm{M}^{(0)}\|_{cb} = \varepsilon$. Notice that generally ε is vanishing for increasing dimension of K (such as when the approximately concealing condition is achieved for an increasingly large number of decoy systems[2]), and no obvious continuity argument can be invoked to assert that Alice's cheating probability will approach unity for vanishing ε. More precisely, in the present context the continuity argument of Ref. 1 would imply that

$$1 - (P_c^A)_\mu = \omega\left(\left\|\mathrm{M}^{(1)} - \mathrm{M}^{(0)}\right\|_{cb}\right), \quad \lim_{\varepsilon\to 0}\omega(\varepsilon) = 0, \tag{3.7}$$

with the function $\omega(\varepsilon)$ independent of the dimension of K. However, using anonymous states such assertion may turn out to be false. In fact, it is obvious that if there is an alternative Kraus decomposition $\{E_J^{(0)}(V)\}$ for the map $\mathrm{M}^{(0)}$ such that the two Kraus decompositions $\{E_J^{(0)}(V)\}$ and $\{E_J^{(1)}\}$ are close, then the protocol is approximately concealing and not binding, since[4]

$$(P_c^B)_{opt} - \frac{1}{2} = \frac{1}{4}\left\|\mathrm{M}^{(1)} - \mathrm{M}^{(0)}\right\|_{cb} \leq \frac{1}{2}\sqrt{\left\|\sum_J \left|E_J^{(0)}(V) - E_J^{(1)}\right|^2\right\|}, \quad (3.8)$$

$$P_c^A(V,\varphi) \geq \left[1 - \frac{1}{2}\left\|\sum_J \left|E_J^{(0)}(V) - E_J^{(1)}\right|^2\right\|\right]^2, \quad (3.9)$$

where for any operator A we use the customary abbreviation $|A|^2 \doteq A^\dagger A$. However, the impossibility proof would be true if a bound of the form (3.8) were satisfied in the reverse direction, in which case one would have

$$1 - (P_c^A)_{\mu,av} \leq \min_V \left\|\sum_J \left|E_J^{(0)}(V) - E_J^{(1)}\right|^2\right\| \leq \omega\left(\left\|\mathrm{M}^{(1)} - \mathrm{M}^{(0)}\right\|_{cb}\right), \quad (3.10)$$

which would correspond to the following *continuity argument*: if two CP maps are close in CB-norm, then for a given fixed Kraus decomposition for one of the two maps, there is always an alternative Kraus decomposition for the other map such that the two are close. Since, as regards the cheating probabilities, we have considered only the case of non-aborting protocols with perfect-verification, proving the continuity argument (3.10) or directly the bound (3.7) would mean that a secure protocol could be sought outside such class of protocols. On the other hand, finding a counterexample to Eq. (3.7) would provide a perfectly verifiable and unconditionally secure protocol.

4 Discussion

The discrepancy between the previous analysis and the analysis beneath the impossibility proof[1] is essentially due to the fact that the latter is based on the assumption that the starting state of the commitment protocol is openly known, in the sense that the probability distribution of the state is given, and then the corresponding mixed state can be purified. The general underlying idea is that the protocol should be processed by *machines*, and therefore all probability distributions are defined, and purified inside the machines. However, such an assumption is certainly not realistic for a cryptographic protocol, in which each party has the freedom of changing or tuning the machine, namely choosing any desired probability distribution. One can continue to argue on this line, asserting that changing the machine is equivalent to use of a larger machine. However, this will be equivalent to considering *infinite machines*, corresponding to uniform probabilities on infinite sets, and this would invalidate an impossibility proof based on an unproven continuity argument.

The above ill-posed mathematical framework arises from the Bayesian approach to secret parameters, dictated from *closed-system* modeling with fixed machines and purification of probabilities. As an alternative to this closed-system approach, we have the realistic *finite open-system* approach, in which unknown parameters are treated as such, without the need of any a priori probability distribution, and in which we can address the problem for finite dimension with the parameter ε depending on it. Then, if one proceeds by treating unknown parameters as such, no openly -known state can be assumed, and the anonymous state encoding of Yuen[2] leads to the present classification of protocols. Notice that if the initial state $|\varphi\rangle$ is openly known, then for that given fixed state all QO's can be regarded as random unitary transformations (since all states are connected by unitary transformations), and this leads to the simple form of Alice's cheating probability in terms of fidelities,[1] whereas in the present context the probability of cheating has the more involved form (3.4), due to the fact that the state $|\varphi\rangle$ is unknown, and that there are QO's that don't admit random unitary Kraus decompositions.

Finally, regarding the possibility of aborting protocols, one could always reasonably adopt equivalent protocols which don't abort, since the repeated commitment will eventually be successful. However, such kind of protocols will necessarily be infinite convex-combinations of protocols on infinite-dimensional anonymous spaces H, and again the *closed-system* approach would necessarily lead to infinite dimensions.

Acknowledgments. This work has been jointly founded by the EC under the program ATESIT (Contract No. IST-2000-29681) and by the USA Army Research Office under MURI Grant No. DAAD19-00-1-0177. Extensive discussions with H. P. Yuen, who motivated this work, are acknowledged.

References

1. D. Mayers, Phys. Rev. Lett. **78**, 3414 (1997); H. K. Lo and H. F. Chau, Phys. Rev. Lett. **78**, 3410 (1997); H. K. Lo, Phys. Rev. A **56**, 1154 (1997).
2. H. P. Yuen, in *Quantum Communications and Measurements II*, P. Kumar, G. D'Ariano and O. Hirota, eds., Kluwer Academic/Plenum (New York 2000) p. 399; quant-ph/0006109, 0009113, 0106001, 0109055, 0207089. See also the paper in these proceedings.
3. G. Brassard, C. Crèpeau, R. Jozsa, and D. Langlois, in *Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science*, 1993 (IEEE, Los Alamitos, 1993), p. 362.
4. G. M. D'Ariano, quant-ph/0209149.
5. G. M. D'Ariano and P. Lo Presti and M. G. A. Paris, Phys. Rev. Lett. **87**, 270404 (2001).