Two Provers in Isolation*

Claude Crépeau[1,**], Louis Salvail[2], Jean-Raymond Simard[3], and Alain Tapp[2]

[1] School of Computer Science, McGill University,
Montréal, QC, Canada
crepeau@cs.mcgill.ca
[2] Département d'Informatique et R.O., Université de Montréal,
Montréal, QC, Canada
{salvail,tappa}@iro.umontreal.ca
[3] GIRO inc., Montréal, QC, Canada
Jean-Raymond.Simard@GIRO.ca

Abstract. We revisit the Two-Prover Bit Commitment Scheme of BenOr, Goldwasser, Kilian and Wigderson [BGKW88]. First, we introduce Two-Prover Bit Commitment Schemes similar to theirs and demonstrate that although they are classically secure using their proof technique, we also show that if the provers are allowed to share quantum entanglement, they are able to successfully break the binding condition. Secondly, we translate this result in a purely classical setting and investigate the possibility of using this Bit Commitment scheme in applications. We observe that the security claim of [BGKW88] based on the assumption that the provers cannot communicate is not a sufficient criteria to obtain soundness. We develop a set of conditions, called *isolation*, that must be satisfied by any third party interacting with the provers to guarantee the binding property of the Bit Commitment.

1 Introduction

The notion of Multi-Prover Interactive Proofs was introduced by BenOr, Goldwasser, Kilian and Wigderson [BGKW88]. In the Two-Prover scenario, we have two provers, Peggy and Patty, that are allowed to share arbitrary information before the proof, but they become physically separated from each other during the execution of the proof, in order to prevent them from communicating. It was demonstrated by Babai, Fortnow, and Lund [BFL91] that Two-Prover Interactive Proofs (with a polynomial-time verifier) exist for all languages in NEXP-time. A fully parallel amalog was achieved by Lapidot and Shamir [LS97].

A quantum mechanical version of this scenario was considered by Kobayashi, Matsumoto, Yamakami and Yao [KM03, KMY03, Yao03]. To this day, it is still

* An earlier version of this work was presented under the title "Classical and Quantum Strategies for Two-Prover Bit Commitments", at QIP '06, *The 9th Workshop on Quantum Information Processing*, January 16-20, 2006, Paris.

** Supported in part by CIFAR, NSERC, MITACS, QuantumWorks and FQRNT's INTRIQ.

D.H. Lee and X. Wang (Eds.): ASIACRYPT 2011, LNCS 7073, pp. 407–430, 2011.

an open problem to establish the exact power of Multi-Prover Quantum Interactive Proofs. A rather vast litterature now exists on this topic (see [BHOP08], [CSUU07], [DLTW08], [IKM09], [IKPSY08], [KKMV08], [Weh06]). However, it is still not even clear whether two provers are as powerful as more-than-two provers.

The Two-Prover Zero-Knowledge Interactive Proofs of [BGKW88] rely on the construction of a Bit Commitment scheme, information theoretically secure under the assumption that the provers cannot communicate. We refer the reader to their paper to understand the application of this Bit Commitment scheme to the construction of Two-Prover Zero-Knowledge Proofs. We solely focus on their Bit Commitment scheme for the rest of our work. In this paper, we consider several important questions regarding Two-Prover Bit Commitment schemes. We do not limit our interest of Two-Prover Bit Commitment to the context of Zero-Knowledge proofs; as already discussed in [BGKW88] similar techniques lead them to a secure Oblivious Transfer under the same assumption. Given that *any* two-party computation may be achieved from Oblivious Transfer [Kil88], we consider the security of such Bit Commitment scheme in a very general context. We discuss at length the security in a very general composability situation.

In order to argue the security of their Bit Commitment scheme, the authors of [BGKW88] asserted the following assumption:

```
"there is no communication between the two provers while
 interacting with the verifier".
```

The current paper is concerned with the sufficiency of this assertion. We show is Section 3.2 that, although this assumption *must be made*, it is however considerably too weak, because we exhibit variations of the scheme that are equally binding classically but that are not at all binding if the provers were allowed to share entanglement. It is however a very well known fact that entanglement does not allow communication. Although it is true that they can cheat if they can communicate, it is also true that they can cheat without communicating. Therefore the assumption that the provers cannot communicate is too weak.

This observation can be turned into a purely classical argument by exhibiting a black-box two-party computation, that does not allow them to communicate, but that allows them to cheat the binding condition of the Bit Commitment scheme. This peculiar source of randomness may replace the entanglement used by the attack. Furthermore, the above assertion of BGKW can be interpreted as a prescription to the verifier that he should make sure not to help the provers to communicate while interacting with him. Again, this prescription would not prevent him from acting like the black-box we exhibit. Thus, a stronger prescription is mandatory in order to assert security.

We carefully define a notion of *isolation* by which the two provers may not communicate nor perform any non-local sampling beyond what is possible via quantum mechanics. We finally formalize a set of conditions that any third party involved in a Two-Prover Bit Commitment scheme may satisfy to make sure he does not break the assumption that the provers are in isolation. In particular, we

make sure that if such a Bit Commitment scheme is used in another larger cryptographic protocol, its security properties will carry over to the larger context.

1.1 Related Work

The starting point of this research is clearly the Bit Commitment scheme introduced by BenOr, Goldwasser, Kilian and Wigderson [BGKW88]. The security of a Two-Prover Bit Commitment scheme against quantum adversaries has been considered in the past in the work of Brassard, Crépeau, Mayers and Salvail [BCMS98]. They showed that if such a Bit Commitment scheme is used in combination to the Quantum Oblivious Transfer protocol of [BCMS98] it is not sufficient to guarantee the security of the resulting QOT if the two provers can get back together at the end of the protocol. In the current work, we consider only the situation while the provers are isolated.

The research by Cleve, Høyer, Toner and Watrous [CHTW04] is the main inspiration of the current paper. They have established some relations between Two-Prover Interactive Proofs and so called "non-locality games". More precisely, they showed that certain languages have a classical Two-Prover Interactive Proof that looses soundness if the provers are allowed to share entanglement. Some of our results are very similar to this. However, our new contributions are numerous. While [CHTW04] focuses on languages, we focus on the tool known as Bit Commitment. This tool is used in many contexts other than proofs of membership to a language: proofs of knowledge, Oblivious Transfer, Zero-Knowledge proofs, general two-party computations. Moreover inspired by the observations of [CHTW04], we analyze the security of such Two-Prover tools in a completely classical situation. We conclude that proving security of such protocols is very subtle when used in combination with other such tools. We also argue that the claim of security of the protocols of [BGKW88] requires a lot more assumptions than the mere "no communication" assumption (even in the purely classical situation).

Despite the impossibility theorems of Mayers [May96] and Lo & Chau [LC97], the possibility of information theoretically secure Bit Commitment schemes in the Two-Prover model is *not* excluded in the classical and quantum models. Indeed, the computations sufficient to cheat the binding condition of a Quantum Bit Commitment scheme in the above "no-go" theorems cannot, in general, be performed by the two provers when they are isolated from each other. This is the reason why these theorems do not apply.

In a closely related piece of work, Kent [Ken05] showed how impossibility of communication, implemented through relativistic assumptions, may be used to obtain a Bit Commitment scheme similar to BGKW that can be constantly updated to avoid cheating. Kent proves the classical security of his scheme while remaining elusive about its quantum security. However, he claims security of one round (see [Ken05], Lemma 3, p. 329) of his protocol which is more or less the same as our Lemma 1. Unfortunately, his proof is incomplete as pointed out in our proof of the Lemma. But we clearly recongnized that he was first to address this question.

A very different set of results [BCU+06] relates non-locality boxes and two-party protocols such as Bit Commitment and Oblivious Transfer. These are only marginally connected to the current research. They showed how these cryptographic protocols may be securily implemented from those non-locality boxes. On the cotrary, we show how to break such protocols using non-locality boxes...

2 Preliminaries

2.1 Isolation

First let us define the condition imposed on the two provers: we use the word *isolation* to describe the relation between Peggy and Patty during the protocol. The intuitive meaning of this term is that Peggy and Patty cannot communicate with each other, since this condition is explicitly imposed by the Two-Prover model. However, we introduce this new terminology instead of the traditional "cannot communicate with one another" because we noticed that the meaning of "no-communication" is too weak and must be very clearly defined to produce valid security proofs. This *isolation* will be formally defined in Section 4. For now, the reader may follow his intuition and picture Peggy and Patty as restricted to compute their messages using only local variables.

2.2 Bit Commitment

The primitive known as "Bit Commitment" is a protocol in which a player Alice first sends some information to another player Bob, such that this information *binds* her to a particular bit value b. However, the information sent by Alice is not enough for Bob to learn b (b is *concealed*). At a later time, Alice sends the rest of the information to unveil the bit b, and she cannot change her mind to reveal $\bar{b}$ and convince Bob that this was the value to which she was committed in the first step. The following definitions will be used to characterize the security of a Bit Commitment scheme. Note that the function $\mu(n)$ always refers to a negligible function in n.

Definition 1. *A Bit Commitment scheme is* statistically concealing *if only a negligible amount of information on the committed bit can leak to the verifier before the unveiling stage.*

Definition 2. *A Bit Commitment scheme is statistically binding if, for $b \in \{0,1\}$, the probability p_b that Alice successfully unveils for b satisfies*

$$p_0 + p_1 \leq 1 + \mu(n). \tag{1}$$

This binding condition was first proposed by Dumais, Mayers, and Salvail [DMS00], as a weaker substitute to the traditional definition $p_b \leq \mu(n)$ for either $b = 0$ or 1. This definition has been henceforward used to show security of many Bit Commitment schemes against quantum adversaries in various models, e.g. [DMS00, CLS01, DFSS05].

More recent definitions have been introduced since then ([DFRSS07]) that appear to be better characterization of Bit Commitment security in a quantum setting. However, we have not been able, so far, to find protocols that satisfy these definitions. This, we hope, will be part of future work in this area.

3 Two-Prover Bit Commitment scheme

For simplicity reasons, we replace the original scheme of [BGKW88] by a far simpler and compact version, which we call "simplified-BGKW" (or sBGKW as a short-hand). Still, we strongly recommend the reader to [BGKW88] for the details of the original construction. For an n-bit string r and a bit b, we define the n-bit string $b \cdot r := b \wedge r_1 || b \wedge r_2 || \ldots || b \wedge r_n$. The scheme is as follows:

Peggy and Patty agree on a uniform n-bit string w and a random bit d. They are then isolated from one another.

Protocol 31 (sBGKW - **Commit to** b)

1: *Vic sends a random n-bit string r to Patty,*
2: *Patty replies with* $x := (d \cdot r) \oplus w$,
3: *Peggy announces* $z := b \oplus d$.

Protocol 32 (sBGKW - **Unveil** b)

1: *Peggy announces bit b and the n-bit string w,*
2: *Vic accepts iff* $w = ((b \oplus z) \cdot r) \oplus x$.

Note that at the unveiling stage, as in the original scheme it is not required that Peggy be the one announcing b. It is as good to let Vic deduce b: Vic computes $y := w \oplus x$, if $y = 0^n$ he sets $b := z$ and if $y = r$ he sets $b := \bar{z}$, and otherwise rejects. Indeed, Peggy may not even know b!

3.1 BGKW's Notion of Isolation

The assumption made in [BGKW88] is that Peggy and Patty are not allowed to communicate with each other. Based solely on that constraint, the following seems a "valid" security proof (it is more or less the same proof as in [BGKW88]).

Theorem 1. *Constraining the provers as in* [BGKW88]*, the* sBGKW *protocol is secure classically.*

Proof. Vic does not know w, and w is uniformly distributed among all possible n-bit strings for both values of z. It follows that the two strings w and $r \oplus w$ have the exact same uniform distribution and are perfectly indistinguishable from one another. We can say the same for the pairs (z, w) and $(z, r \oplus w)$. Hence sBGKW is concealing.

Now suppose that Peggy and Patty would like to be able to unveil a certain instance of b both as 0 and as 1. To do so, Peggy would like to announce $\widehat{w}_b$ such that $\widehat{w}_b = (b \cdot r) \oplus x$. We note that this models the two possible dishonest behaviors for Peggy and Patty: honestly commit to $\bar{b}$ and try to change to b afterwards, and commit to nothing by sending some x and decide which b they want to unveil *only* at the unveiling stage. It follows that in both scenarios, a successful cheating strategy would allow to produce the two strings $\widehat{w}_0$ and $\widehat{w}_1$, such that $\{\widehat{w}_0, \widehat{w}_1\} = \{x, r \oplus x\}$. However, the string $\widehat{w}_0 \oplus \widehat{w}_1 = x \oplus r \oplus x = r$ is completely unknown to Peggy by the no-communication assumption. Therefore, even using unlimited computational power, her probability of issuing a valid pair $\widehat{w}_0, \widehat{w}_1$ is at most $1/2^n$. Hence sBGKW is binding.

Nevertheless, this result is incomplete[1]! Indeed, we show next how a correlated random variable can be used to invalidate the result of Theorem 1 while not violating the "no-communication" assumption. This suggest that the conventional wording "no-communication" is insufficient as it is not explicit enough to cover any kind of cheating mechanism Peggy and Patty can employ.

3.2 Cheating sBGKW with an NL-box

An **NL**-box, short-hand for "Non-Locality box" introduced by Popescu and Rohrlich [PR94, PR97], is a device with two inputs s and t, and two output bits u and v such that u and v are individually uniformly distributed and satisfy the relation $f(s,t) = u \oplus v$ for some function f. The pair (s,u) is on Peggy's side while the pair (t,v) is on Patty's side. Because u and v are individually uniformly distributed, no **NL**-box allow Peggy and Patty to communicate, in either direction. The **NL**-boxes are usually assumed as asynchronous devices, that is, feeding in the input s is sufficient to obtain u even if t has not been input yet, and likewise for t. Such a particular box, known as the **PR**-box, is defined for $f(s,t) = s \wedge t$, where s and t are binary inputs. It is known that two classical players can simulate the **PR**-box with success probability[2] at most 75% for all s,t, while quantum players sharing an entangled state can achieve a success probability of $\cos^2(\pi/8) \approx 85\%$ (consult [CHTW04] for details).

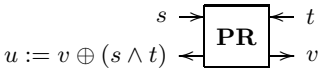

Fig. 1. the cheating **PR**-box

Let the two provers be given a black-box access to this **PR**-box. The following shows how this **PR**-box allows Peggy and Patty to unveil the bits committed

[1] The broad explanation is that we implicitly *assumed* the provers had only access to local variable. We'll see we need to guarantee this restriction for the proof to hold.

[2] This result is shown optimal by enumerating every possible classical strategies.

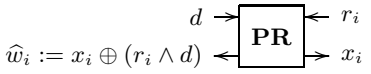

Fig. 2. Using the **PR**-box

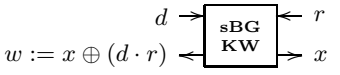

Fig. 3. The cheating **sBGKW**-box

through sBGKW in either way, at Peggy's will. For each position i, $1 \leq i \leq n$, Patty inputs in the **PR**-box the bit $s := r_i$ received from Vic and obtains output $x_i := u$ from the **PR**-box, which corresponds to the i-th bit of the commitment string. Patty sends x to Vic. Peggy discloses z a random bit to Vic. To unveil bit b, Peggy inputs $t := d := b \oplus z$ in the **PR**-box and obtains the output $\widehat{w}_i := v$ from the **PR**-box, which she sends to Vic together with b.

If $d = 0$ then $d \wedge r_i = 0$ and thus $\widehat{w}_i = x_i$ which is the right value she must disclose. If $d = 1$ then $d \wedge r_i = r_i$ and thus $\widehat{w}_i \oplus x_i = r_i$ or $\widehat{w}_i = x_i \oplus r_i$ which is again the right value she must disclose.

Indeed, we can view an arbitrary cheat on the sBGKW as a non-local computation between the provers as in Fig. 3. Essentially we have just demonstrated that an sBGKW-box can be emulated perfectly by perfect **PR**-boxes. However, a valid cheating strategy might not succeed 100% of the time, so an sBGKW-box that is correct 80% of the time, for instance, would be enough to break the binding property. It seems quite obvious, nevertheless, that a **PR**-box that is correct 80% of the time will not help implementing an sBGKW-box that is correct 80% of the time. For that matter, any **PR**-box that is correct a constant fraction $p < 1$ of the time will not help either...

It is not obvious that a **sBGKW**-box with error probability greater than zero is equivalent to the **PR**-box, but it would be very interesting to prove either way.

3.3 Quantumly Insecure - Two-Prover Bit Commitments

We exhibit an intermediate scheme to emphasize how shared entanglement can be used to cheat with probability almost one a classically "secure" Two-Prover Bit Commitment. The protocol is a weaker version of the sBGKW scheme, called wBGKW, where the acceptance criteria of the unveiling stage is loosen to tolerate some errors. A second protocol (available in Sub-Section 3.7) is also a modified version of the sBGKW scheme where the acceptance criteria is based on a game described later, called the Magic Square game.

A weaker acceptance criteria: the wBGKW scheme Consider a weaker acceptance criteria where the string $\widehat{w}$ sent by Peggy can differ in at most $n/5$ positions from what it should be. Formally the verifier Vic is to accept b if $d(\widehat{w}, ((b \oplus z) \cdot r) \oplus x) < n/5$, where $d(\cdot)$ is the binary Hamming distance. The

interest of such a modification is that now a cheating quantum pair Peggy and Patty can use the non-local property of entanglement to approximate the **PR**-box and successfully cheat wBGKW, while, as we show next, the Bit Commitment is "secure" classically. To facilitate notation we add an index b to the string $\widehat{w}$, since $\widehat{w}$ is different whether we unveil zero or one. Also, define as B the random variable corresponding to the value they unveil.

Theorem 2. *For any classical strategy, the probability that it outputs a string* $\widehat{w}_0$ *when* $B = 0$ *and* $\widehat{w}_1$ *when* $B = 1$ *s.t.* $E[d(\widehat{w}_b, ((b \oplus z) \cdot r) \oplus x)] < n/5$ *for both values of* b*, is exponentially small in* n*.*

Proof (of Theorem 2).

Wlog, we can assume the provers use a deterministic strategy that may produce such a $\widehat{w}_0$ when $B = 0$, and $\widehat{w}_1$ when $B = 1$, so they can in fact output *both* $\widehat{w}_0$ and $\widehat{w}_1$. Hence, Peggy can compute the string $\widehat{w}_0 \oplus \widehat{w}_1$. Recall that when $d(\widehat{w}_b, ((b \oplus z) \cdot r) \oplus x) = 0$ then $\widehat{w}_0 \oplus \widehat{w}_1 = r$. We want to determine the distance between $\widehat{w}_0 \oplus \widehat{w}_1$ and r in our situation. From the theorem's assumption, there exists a classical strategy that outputs $\widehat{w}_0$ *and* $\widehat{w}_1$ such that $E[d(\widehat{w}_b, ((b \oplus z) \cdot r) \oplus x)] < n/5$, for $b = 0, 1$. We easily obtain that for such a strategy, the expected distance from r is

$$E[d(\widehat{w}_0 \oplus \widehat{w}_1, r)] = E[d(\widehat{w}_0 \oplus \widehat{w}_1, x \oplus (x \oplus r))] \leq E[d(\widehat{w}_0, x)] + E[d(\widehat{w}_1, x \oplus r)] < 2n/5$$

by the triangular inequality. Using a standard Chernoff bound argument, and since r is absolutely unknown to Peggy, her probability of outputting a string $y = \widehat{w}_0 \oplus \widehat{w}_1$ such that $E[d(y, r)] < (1/2 - \epsilon) \cdot n$ is exponentially small in n for any $0 < \epsilon \leq 1/4$. Hence, because $1/4 < 2/5 < 1/2$, we conclude that such a strategy cannot exist except with exponentially small probability, and so unveiling *must* fail for one of the two possibilities.

Conversely, this scheme is almost totally insecure against quantum adversaries.

Theorem 3. *There exists a quantum strategy that successfully cheats the* wBGKW *scheme with probability* $1 - \mu(n)$*.*

Proof (of Theorem 3). We saw in Section 3.2 that the **PR**-box, taken as a black box, correctly produces the needed $\widehat{w}_b$ to unveil as b. Using the well-known result [e.g. [CHTW04]] that through entanglement, Peggy and Patty can optimally simulate the **PR**-box such that for each i taken independently, $1 \leq i \leq n$, the **PR**-box produces correlated outputs with probability $\cos^2(\pi/8) \approx 0.85$. Therefore, using the standard Chernoff bound, this independent quantum strategy yields that for both values of b,

$$E[d(\widehat{w_b}, ((b \oplus z) \cdot r) \oplus x)] = (1 - \cos^2(\pi/8)) \cdot n$$

with probability exponentially close to one. Having that $(1 - \cos^2(\pi/8)) \cdot n < 0.15 \cdot n < n/5$, we conclude that a pair of quantum provers defeats the binding condition of the scheme with probability $1 - \mu(n)$.

3.4 Discussion

The limitation of Theorem 1 (and Theorem 2) is that it claims that the following non-local computation, named **sBGKW2**-box (see Fig. 4) , is a communication

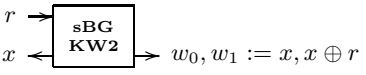

Fig. 4. the cheating **sBGKW2**-box

device (which is obvious) assuming that any implementation of an **sBGKW**-box is sufficient to implement it (which is false, since the **sBGKW**-box is *not* a communication device, it is impossible to implement any communication device from it).

However, these proofs are not *wrong* either since it is impossible to accomplish the **sBGKW**-box without some sort of communication, which also works for the **sBGKW2**-box. In particular, it means that this proof is seriously context-dependent. In a context where Patty and Peggy have access to a third party that scrupulously monitors that they are not communicating with each other, the proof does not hold anymore because using the third party as a **sBGKW**-box is not excluded.

The bottom line here is that this proof is valid *solely* in a stand-alone security model. As soon as one starts composing such protocols, one has to, not only, monitor that the actions of the third party do not allow communication but also do not constitute any form of correlation between Patty and Peggy.

This demonstrates that certain non-local correlations are enough to cheat Two-Prover Bit Commitment schemes while they are not enough to communicate. Thus we have to define the prover's isolation in terms of these non-local correlations and not only in terms of communication. This is the purpose of Section 4.

3.5 A Non-Local Box to Cheat the Original BGKW Scheme

Similarly to the sBGKW scheme, we can define an analogous cheating box for the original BGKW scheme with two binary inputs s, t, and two uniformly generated ternary outputs x, y.

The original protocol goes as follows:
Peggy and Patty agree on a uniform n-trit string w. They are then isolated from one another.

Protocol 33 (BGKW - **Commit to** b)

1: *Vic sends a random n-bit string r to Patty,*
2: *Patty replies with x such that for all k,* $x_k := \sigma_{r_k}(w_k) - b \bmod 3$.

Protocol 34 (BGKW - Unveil b)

1: *Peggy announces bit b and the string w,*
2: *Vic accepts iff w is such that for all k, $b = \sigma_{r_k}(w_k) - x_k \bmod 3$.*

Where the σ function of [BGKW88] can be re-written as the single expression: $\forall\, r \in \{0,1\}, w \in \{0,1,2\}$

$$\sigma_r(w) = (1+r)w \mod 3. \tag{2}$$

So using (2), we want from the cheating **NL**-box that $u := (s+1)v - t \mod 3$ for each s, t, and uniformly chosen v. Because for any binary s, t we can easily define the inverse permutation over trits to be $v := (t+u)(s+1) \mod 3$, the following **PR3**-box does not allow to communicate since individually u and v are uniformly distributed.

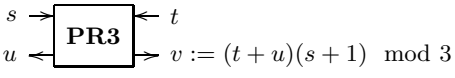

Fig. 5. A non-local box to cheat BGKW

It is not hard to verify that the **PR3**-box that implements this non-local computation from s, t is exactly the one needed to cheat the original BGKW scheme. As with the **PR**-box, for each round i, Peggy inputs in the box $s := r_i$ and obtains the trit $x_i := u$, which she sends to Vic. If Patty wants to unveil for b, she inputs $t := b$ in the **PR3**-box, which correctly outputs $\widehat{w}_i := v$. Clearly, they successfully cheat since

$$\begin{aligned} \forall\, i \quad (1+r_i)\widehat{w_i} - x_i \mod 3 &= (1+r_i)(b+x_i)(1+r_i) - x_i \mod 3 \\ &= (1+r_i)^2(b+x_i) - x_i \mod 3 \\ &= (b+x_i) - x_i \mod 3 \\ &= b. \end{aligned}$$

We can also demonstrate that the **PR3**-box is as powerful as the **PR**-box. It is straightforward to check that the outputs x' and y' depicted in Figure 6 are indeed the correct outputs to cheat the sBGKW scheme.

3.6 Magic Square Non-locality Game

A square is a 3×3 matrix whose entries are in $\{0,1\}$. A row is said to be *correct* if its parity is even, and a column is said to be *correct* if its parity is odd. We use the following definition of the Magic Square game (from [CHTW04]), which slightly differs from the original game due to Aravind [Ara02]. The verifier Vic picks at random a row or column, say column c_i, and a position x^i_j on c_i, $i, j \in \{1,2,3\}$.

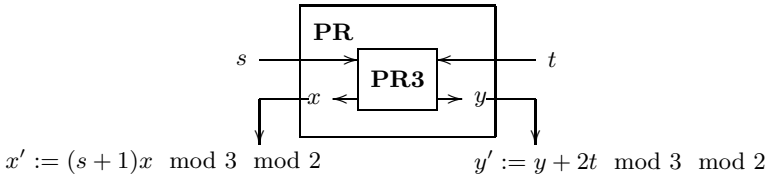

Fig. 6. Reduction from the **PR**-box to the **PR3**-box.

He then asks the entries of column c_i to Peggy, and the value in position x_j^i to Patty. The two provers win if the parity of c_i is odd (more generally, if the row or column asked for is *correct*), and if the value returned by Patty matches the value at position x_j^i in Peggy's answer. The following defines the *validity* of a square.

Definition 3. *A* (3×3) *matrix* S is valid *for zero if all rows of* S *xor to 0, and* S is valid *for one when all columns of* S *xor to 1.*

For instance the following matrix S_0 is valid for zero while S_1 is valid for one:

$$S_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{3}$$

Any classical strategy successfully wins this Magic Square game with probability at most $\left(\frac{17}{18}\right)$. Remarkably, there exists a quantum strategy that allows Peggy and Patty to successfully win this game *every* time, see [CHTW04, Ara02] for details.

3.7 Magic Square Bit Commitment

It is not hard to exploit the Magic Square game to build another Bit Commitment scheme. This scheme is particularly relevant in our study of Bit Commitments in the Two-Prover model as it is perfectly secure classically but can easily be cheated with probability one using a quantum strategy. The scheme is as follows:

Peggy and Patty agree on a random bit v and n random squares S_i such that S_i is valid for v. They are then isolated.

Protocol 35 (MSBC - **Commit to** b)

1: *Peggy computes* $x := v \oplus b$ *and sends* x *to Vic.*
2: *Vic picks a pair of random trits* r_i *and* c_i *and asks Peggy for* $S_i(r_i, c_i)$.

Protocol 36 (MSBC - **Unveil** b)

1: *Peggy sends* b *to Vic,*
2: *Vic asks Patty for row number* r_i *of* S_i *if* $b = x$*, or column number* c_i *of* S_i *if* $b = \bar{x}$.
3: *Vic accepts* b *if, for each* i*, the row or column that should xor to* b *does, and if the entry returned by Peggy matches with Patty's answer. Vic rejects otherwise.*

Theorem 4. *Any classical strategy successfully cheats the binding property of the* MSBC *scheme with probability at most* $\left(\frac{8}{9}\right)^{n/6}$*, except with exponentially small probability.*

Proof (of Theorem 4).

Wlog, it is sufficient to consider deterministic strategies. Consider the strategy where only the entry $(2,2)$ is used to make the square S_i correct for $\widehat{w}_i$. When $t_i = 0$ or 1, Peggy answers the line or column of S_i as is. However, when $t_i = 2$, she sets the entry $(2,2)$ to the correct value such that a line xores to 0 or a column xores to 1. On query (y_i, z_i), Patty answers the entry (y_i, z_i) of S_i if $(y_i, z_i) \neq (2,2)$, otherwise she answers 0. It is not hard to show that this strategy is optimal, since Peggy knows all the information (the S_i's, x, and r), and Patty knows nothing about x and r.

The problem for the provers is that whenever $b \cdot r_i = 1$, they succeed for at most only one of $b \in \{0,1\}$. This is because the square S_i they share cannot be correct for both x_i and $\overline{x_i}$. Since r is uniformly distributed, by a Chernoff argument, r contains at least $n/3$ 1's. Thus, there is at least one of $b \in \{0,1\}$ for which in at least $n/6$ challenges the provers will answer correctly with probability at most $8/9$ (the sum of the challenges where she succeeds with probability at most $8/9$ for 0, and those where she succeeds with probability at most $8/9$ for 1, adds up to $n/3$). Therefore, their probability of successfully cheating is at most $\left(\frac{8}{9}\right)^{n/6}$ for any classical strategy, except with exponentially small probability.

However, there exists a quantum strategy that allows Peggy and Patty to successfully break the binding condition with probability 1 by winning the Magic Square game *every* time.

Theorem 5. *There exists a quantum strategy that successfully cheats* MSBC *with probability 1.*

4 Defining and Checking Isolation

The existence of such an inputs-correlated[3] random variable, which does not allow communication but allows cheating of the sBGKW Two-Prover Bit Commitment scheme sheds some light on the limitations of the original assumption of [BGKW88].

Indeed, the assumption of [BGKW88] is necessary but not sufficient to guarantee the binding property of the Bit Commitment scheme. Among its weakness, we note that it does not *explicitly* force any cheating strategy to be repeatable. The **PR**-box not being a repeatable process[4] gives a first understanding why

[3] We emphasize that at least one of the "inputs" to the random variable needs to be obtained once the provers are isolated, otherwise such a random variable can be shared while the provers are together, and is thus useless to cheat the sBGKW scheme.

[4] The **PR**-box cannot be repeated to generate two *valid* strings $\widehat{w}_0$ and $\widehat{w}_1$.

we can still cheat the sBGKW scheme despite the result of Theorem 1, which implicitly assumed repeatability of the cheating strategy.

Clearly, to achieve the binding condition, a stronger assumption is needed. One could require that once the provers are isolated, there exists no mechanism by which they may sample a joint random variable which is dependent on the inputs they provide. We note that, among other things, this new condition excludes communication between the two provers, as desired. However, it excludes a lot more, such as shared entanglement! This is simply too strong; we need to be more subtle in the way we define this *"mechanism to sample a joint random variable".*

It seems reasonable to believe that nature does not allow the existence of a **PR**-box (consult [CHTW04]). So why even ask for a stronger assumption than the no-communication assumption of [BGKW88]? Part of the answer is that Vic can play the role of the **PR**-box, or any other third party. In no circumstances can we ignore the fact that both Peggy and Patty individually talk to Vic. Definitely, we need to consider this aspect of the protocol with great care. For instance, consider the scenario where r is sent to Peggy but unveiling is not done immediately after committing, but rather once Vic and the two provers have been involved in other, unrelated, interactive protocols. It is perfectly conceivable that within those protocols, for each i, Peggy and Patty succeed in sending r_i and b to Vic, and then in a completely different context (or a moment of unawareness) Vic performs the required computation and output x_i and $\widehat{w}_i$, which are then sent respectively to Peggy and Patty. It is obvious that if such a computation, or any alike, can take place with enough probability then Peggy and Patty would succeed in cheating the sBGKW protocol!

More generally, we must not only consider Vic but any other third party, call it Ted, to which Peggy and Patty might have access to obtain correlated information. The previous situation highlights the fact that there is a whole class of functions with inputs coming from Peggy and Patty for which Ted must not send the outputs. Intuitively, each time Ted sends a message to either Peggy or Patty, he must ensure that the message does not outperform what Peggy and Patty can achieve using local variables in the sense of quantum mechanics. We propose two different approaches to formulate that statement as a criteria. The first considers the practical flavor of the problem, when Ted is working with instances of variables. The second approach is based on an information theoretic argument. At this point, we will not consider the scenario where the players can share quantum resources.

Let Peggy be identified by P_0 and Patty by P_1. The variable $D \in \{0,1\}$ is a reference to player P_D, and $T \in \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$ is a tag appended to each message that indicates to Ted the player(s) that is (are) eligible for receiving this message, where $T = \{0,1\}$ means by both players and $T = \emptyset$ means by none of them. The message about to be sent from Ted to prover P_D is represented by $(m,T)_D$. We formalize Ted's behavior as follows.

Definition 4 (Practical criteria). *Ted is said to be a "secure third party" if* $\forall D \in \{0,1\}$*, Ted follows these points.*

1. *A message received from player P_D is tagged with $T := \{D\}$.*
2. *A message generated without involving any of the previous messages, e.g. picking a random string, is tagged with $T := \{0, 1\}$.*
3. *A message obtained from a computation involving previous messages is tagged with the intersection of the tags of* all *the messages involved in that computation.*
4. *A message $(m, T)_D$ is sent to player P_D only if $D \in T$.*

Note: It is important that the communication pattern between Ted and the isolated provers be specified ahead of time, otherwise the traffic pattern (not only the message contents) may leak information.

We now explain why Ted will not send a message that allows P_0 and P_1 to communicate or establish non-local correlations. Let $(m, T)_D$ be the message Ted is about to send to player P_D. From the fourth point of Definition 4, Ted will send $(m, T)_D$ only if it is tagged $T = \{D\}$ or $\{0, 1\}$. Looking at the message's tag assignment rule number 3, this happens only if there is absolutely no message tagged $\{1 - D\}$ or $\emptyset$ used in the computation of $(m, T)_D$. Using an induction argument, it is not hard to see that this happens only when all the variables involved in the computation of $(m, T)_D$ are independent of the information of P_{1-D}, that is, they have been themselves generated using variables tagged $\{D\}$ or $\{0, 1\}$. Thus, such a message $(m, T)_D$ is also independent of the information known only to P_{1-D}. Therefore, the messages sent by Ted do not let the two players communicate.

The case of non-locality is slightly more subtle, yet pretty straightforward. Recall that in a general non-local process, both players use a message each and receive a message uniformly distributed, from their point of view, such that the four messages satisfy a certain relation. The received message does not allow to communicate with the other player. Suppose P_{1-D} receives his message first. Since from his point of view, this message is uniformly distributed, Ted *can* in fact generate a uniformly distributed message, tag it with $T := \{0, 1\}$ and send it to P_{1-D}. At this point, this behavior does not violate anything because non-locality has not been created yet. Then, Ted computes the message for P_D. Because this message needs to satisfy the relation that binds together the four messages, at least a message tagged with $T \neq \{D\}$ and one tagged with $T \neq \{1 - D\}$ are used in its computation (it can be the same message), so the resulting message $(m, T)_D$ will be assigned a tag $T := \emptyset$ because the intersection does not contain $\{D\}$ nor $\{1-D\}$. This message $(m, \emptyset)_D$ is the one creating the non-local relation. However, from point 4 of Definition 4, since $D \notin \emptyset$, Ted will never send $(m, \emptyset)_D$.

As mentioned before the previous definition, we can alternatively formalize Ted's behavior in terms of entropy. The advantage of doing so is to enable analysis of existing protocols. To satisfy the above practical criteria, the wrapping protocol must be designed in a rather restricted way. To consider general protocols, we offer this alternate definition.

Let the message about to be sent from Ted to prover P_D be represented by the variable $(M, T)_D$. The set of variables $S_{D,T}$ represents all the variables (messages) with tag T sent by prover P_D to Ted, and the set of variables $R_{D,T}$ all the variables (messages) with tag T sent by Ted to prover P_D before $(M, T)_D$.

Definition 5 (Information based criteria). *Ted is said to be a "secure third party" if* $\forall D \in \{0,1\}$*, Ted follows these points.*

1. *An information received from player* P_D *is tagged with* $T := \{D\}$[5].
2. *A variable* M *to be sent to* P_D *is tagged with the less restrictive tag* $T \in \{\emptyset, \{D\}, \{0,1\}\}$ *that satisfies the following relation*[6]*. Note that the calligraphic tag* $\mathcal{T}'$ *stands for the tag* $\{0,1\}/\big(T \cap \{D\}\big)$ *and the calligraphic tag* $\mathcal{T}''$ *stands for the tag* $\{D\} \cup \big(T \cap \{1-D\}\big)$.

$$\begin{aligned} H((M,T)_D | S_{D,\{D\}}, R_{D,\{D\}}, R_{D,\{0,1\}}, S_{1-D,\mathcal{T}'}, R_{1-D,\mathcal{T}'}, R_{1-D,\{0,1\}}) \\ = \quad H((M,T)_D | S_{D,\mathcal{T}''}, R_{D,\mathcal{T}''}, R_{D,\{0,1\}}, R_{1-D,\{0,1\}}) \end{aligned} \quad (4)$$

3. *A variable* $(M,T)_D$ *is sent to player* P_D *only if* $D \in T$.

We warn the reader that the tags and players' variables D and $1-D$ do not play any role in the computation of the entropies; they are only present to discriminate the variables and determine which ones to include in the conditional part of the entropies. Notice also that, contrary to Definition 4, a variable's tag is set only when Ted considers sending it to a player, except for incoming variables. This relaxation will turn out to be the key point to explain why this generalized definition is not stronger than local variables on the players' side.

The process of determining which tag to assign can be broken into two steps. We start with the empty tag $\emptyset$. The first step is to decide whether we can add $\{D\}$ to the tag, or not. Notice that the *right*-hand side of equation (4) is the same for $T \in \{\emptyset, \{D\}\}$. This results from the calligraphic tag $\mathcal{T}''$, which is equivalent to $\{D\}$ in this case. On the other hand, the calligraphic tag $\mathcal{T}'$ introduces the terms $S_{1-D,\{1-D\}}$ and $R_{1-D,\{1-D\}}$ in the *left*-hand side of equation (4) when $T = \{D\}$. Thus, if the result of this first step is that the tag is at least $\{D\}$, then it means that the message to be sent is independent of the private information held by P_{1-D}. However, if we find that the tag is not even $\{D\}$, then it means that the message to be sent has some dependencies with the private information of P_{1-D}, and therefore the message should not be sent.

If the first step terminates with a tag containing $\{D\}$, then we can move on to determine whether we can add $\{1-D\}$ to the tag, or not. We note that $\mathcal{T}'$ won't change for $T \in \{\{D\}, \{0,1\}\}$, so the *left*-hand side is invariant. However, the calligraphic tag $\mathcal{T}''$ will remove the terms $S_{D,\{D\}}$ and $R_{D,\{D\}}$ from the *right*-hand side if we consider the tag $T = \{0,1\}$. Hence, if equation (4) is satisfied with $T = \{0,1\}$, it means that the message to be sent is not only independent of the private information of P_{1-D} (from the first step), but also of the private information of P_D. It follows naturally that this message be eligible for distribution to both players.

[5] This implies that the sets $S_{D,\{0,1\}}$ and $S_{1-D,\{0,1\}}$ are always empty. Therefore we do not include them in equation (4), but a formal expression should include them in the conditional part on both sides of the equality.

[6] In order to write a clear equation, we had to specify to which player the message is intended. As a result, we did not include $\{1-D\}$ in the set of possible tags. It turns out that the empty set tag is sufficient to cover both communication and correlation.

The interest of Definition 5 is that it is more flexible in the tag assignation than the practical Definition 4 (and thus more general). Indeed, whenever Ted deliberately randomizes a message with new [uniformly distributed] information, the information-based criteria concludes that there is no problem to send to P_D a message that would have been tagged with $T = \{1 - D\}$ or $\emptyset$ by the practical definition. The reason is that by randomizing completely all the [private] variables related to P_{1-D}, Ted is reducing the message he sends to P_D to what P_D can exactly achieve using local variables. That is to say, P_D already has (using local variables) a random view of P_{1-D}'s variables (and so of the global message), so there is no problem for Ted to first randomize P_{1-D}'s variables and then send this message to P_D. Note however that the variables used to randomize will never be sent to P_D since they now carry the sensible information. We give two examples of these particular cases in the Appendix A.

Henceforth, the Two-Prover model's assumption is based on this refined definition of isolation.

Definition 6. *We say that Peggy and Patty are* isolated *from one another if they cannot communicate with one another, and if they only have access as external resource to secure third parties.*

Using this new definition of isolation, we are now *guaranteed* that any strategy that Peggy and Patty try to perform through a third party can be achieved using *only* local variables on each side. Using this fact together with the general assumption that the cheating strategy is deterministic[7], it is straightforward to fix the proof of Theorem 1 by arguing that their classical strategy can be run on each copy of the information to output *both* $\widehat{w}_0$ and $\widehat{w}_1$.

5 Quantum Secure Bit Commitment in the Two-Prover Model

We now present the modified version of the sBGKW scheme, called the mBGKW scheme, and prove its security against quantum adversaries. Although the two schemes are almost identical, it turns out the proof against quantum provers is easier with the latter. The security of the sBGKW and BGKW schemes will follow as corollaries of mBGKW's security. The scheme is as follows:

Peggy and Patty agree on an n-bit string w. They are then isolated as in Definition 6.

Protocol 51 (mBGKW **- Commit to** b)

1: *Vic sends two random n-bit strings r_0, r_1 to Peggy.*
2: *Peggy replies with $x := r_b \oplus w$.*

[7] A probabilistic strategy can be made deterministic by fixing the randomness to the best sequence.

Protocol 52 (mBGKW - **Unveil** b)

1: *Patty announces an n-bit string* $\widehat{w}$
2: *Vic computes* $r := \widehat{w} \oplus x$. *He accepts iff* $r \in \{r_0, r_1\}$ *and deduces* b *from* $r = r_b$.

We want to show that the mBGKW scheme is secure against a quantum adversary. Clearly the commitment is concealing because Vic does not know w. This means that there exists w and w' such that $x = r_0 \oplus w = r_1 \oplus w'$, and Vic cannot determine which one has been used.

To prove that the binding property holds according to Definition 2, we again use the crucial observation that if Patty could simultaneously compute $(\widehat{w}_0, \widehat{w}_1)$, then she would learn $r_0 \oplus r_1 = \widehat{w}_0 \oplus \widehat{w}_1$. Let $p_\oplus := \Pr[\text{Patty determines } r_0 \oplus r_1]$. The next lemma relates $p_\oplus$ to $p_0 + p_1$ in the desired way. Notice however that because quantum information is involved this statement is much less straightforward than the classical analog: p_0 and p_1 still correspond to running the attack twice on the same data but an attacker cannot do both.

Lemma 1. *Assume Patty has probability* p_b *to unveil bit* b *successfully, for both values of* b, *and such that* $p_0 + p_1 \geq 1 + \varepsilon$ *for* $\varepsilon > 0$. *Then, Patty can guess* $r_0 \oplus r_1$ *with probability* $p_\oplus \geq \varepsilon^2/4$.

Proof (of Lemma 1).

Assume without loss of generality that when the unveiling phase of mBGKW starts, Patty holds the pure state $|\psi\rangle \in \mathcal{H}^N$ of dimension $N \geq 2^n$. Note that we do not need to consider the whole bipartite state between Peggy and Patty since when the unveiling phase starts, Peggy does no longer play an active role in the protocol and no communication is allowed between the two; hence her system can be traced-out of the global Hilbert space. Moreover, by linearity, the proof also holds if $|\psi\rangle$ is replaced by a mixed state. Notice also that, from the new model's assumption, Peggy and Patty cannot do better using a third party than what they can achieve with entanglement.

Generally speaking, Patty has two possible strategies depending upon the bit b she wants to unveil. When $B = 0$, she applies a unitary transform U_0 to $|\psi\rangle$ in order to get the state $|\psi_0\rangle := U_0|\psi\rangle$ that she measures in the computational basis $\{|w\rangle\langle w|\}_{w \in \{0,1\}^n}$ applied to the first n qubits of $|\psi_0\rangle$. When $B = 1$, she proceeds similarly with unitary transform U_1 allowing to prepare the state $|\psi_1\rangle := U_1|\psi\rangle$. She then measures $|\psi_1\rangle$ using the same measurement as for $B = 0$. All general measurement can be realized in this fashion, this is thus a general strategy for Patty. Notice that in the proof of Kent [Ken05], the use of unitary transformations U_0 and U_1 is obscured by the fact that he works with projective measurements. Notice also that the measurement on the first n qubits of $|\psi_b\rangle$ can alternatively be expressed by the measurement operators $\{|w\rangle\langle w| \otimes I_M\}_{w \in \{0,1\}^n}$ on the whole state $|\psi_b\rangle$, where I_M is the identity matrix on the system of dimension $M = N/2^n$.

From the values $r_0, r_1, x \in \{0,1\}^n$ announced by Vic and Peggy during the committing phase, we define $\widehat{w}_b := r_b \oplus x$ as the string Patty has to announce in order to open b with success. We have,

$$p_b = \langle \psi_b | \widehat{w}_b \rangle \langle \widehat{w}_b | \psi_b \rangle, \tag{5}$$

which by assumption satisfies

$$p_0 + p_1 \geq 1 + \varepsilon,\ \varepsilon > 0. \tag{6}$$

Notice that $\langle \psi_b | \widehat{w}_b \rangle$ is a generalized inner product[8] since $|\widehat{w}_b\rangle$ lives in a subspace of dimension 2^n in $\mathcal{H}^N$. Therefore when $\widehat{w}_b$ is obtained, there is some state left in $\mathcal{H}^{\mathcal{N}}$ of dimension $N/2^n$ which we label as $|\widehat{v}_b\rangle$ (i.e. $|\psi_b\rangle$ has not been completely collapsed by the measurement). Thus, using (5) we can write $|\psi_b\rangle$ as

$$|\psi_b\rangle = \sqrt{p_b}|\widehat{w}_b\rangle|\widehat{v}_b\rangle + \sqrt{1-p_b}|\widehat{w}_b^{\perp}\rangle, \tag{7}$$

where $\|\langle \widehat{v}_b | \langle \widehat{w}_b | \widehat{w}_b^{\perp} \rangle\|^2 = 0$. Note that the "state" $|\widehat{w}_b^{\perp}\rangle$ has not necessarily a physical signification. It is simply a mathematical tool that allows us to conveniently carry the statistics.

We want to determine a lower bound for the probability $p_{\oplus}$. One possible way for Patty to compute $r_0 \oplus r_1$ is to obtain $\widehat{w}_0$ and $\widehat{w}_1$ individually. Again, one possible way to do this is to use the following strategy:

1. Patty applies the strategy allowing to open $B = 0$ from $|\psi_0\rangle = U_0|\psi\rangle$ resulting in the state $|\tilde{\psi}_0\rangle$ after the measurement in the computational basis $\{|w\rangle\langle w|\}_{w\in\{0,1\}^n}$ has been performed on the first n qubits, and
2. Patty prepares $|\tilde{\psi}_1\rangle := U_1 U_0^{\dagger}|\tilde{\psi}_0\rangle$ before applying again the measurement in the computational basis $\{|w\rangle\langle w|\}_{w\in\{0,1\}^n}$ on the first n qubits.

Note that when preparing $|\tilde{\psi}_1\rangle$, we applied $U_0^{\dagger}$ before U_1. This is to put back the state $|\tilde{\psi}_0\rangle$ as close as possible as the original state $|\psi\rangle$. From (6) and for N big enough, the probability to measure $\widehat{w}_0$ in the first step is not too small and so, by applying the inverse of all the unitary transformations generated by U_0, the state $|\tilde{\psi}\rangle$ we get before applying U_1 is a good enough approximation of the original $|\psi\rangle$. Similarly we can say that the fidelity $F(|\tilde{\psi}\rangle, |\psi\rangle)$ is large enough. By invariance under unitary transformation, it follows that $|\tilde{\psi}_1\rangle$ approximates $|\psi_1\rangle$ with the same fidelity $F(|\tilde{\psi}\rangle, |\psi\rangle)$.

In the strategy described above, the probability to determine $r_0 \oplus r_1$ is

$$p_0 \cdot p_{\widehat{w}_1|\widehat{w}_0}\ .$$

As we said earlier, this is only *one of the* possible strategies to determine $r_0 \oplus r_1$, thus

$$p_{\oplus} \geq p_0 \cdot p_{\widehat{w}_1|\widehat{w}_0}\ .$$

[8] If $|w\rangle \in \mathcal{H}^M$ and $|\psi\rangle \in \mathcal{H}^N$ then for $|\psi\rangle^N = \sum_i \alpha_i |a_i\rangle^M \otimes |b_i\rangle^{N/M}$ we define $\langle w|\psi\rangle = \sum_i \alpha_i \langle w|a_i\rangle |b_i\rangle$.

Let us first find a lower bound on the probability $p_{\widehat{w}_1|\widehat{w}_0}$ to produce $\widehat{w}_1$ given that $\widehat{w}_0$ has already been produced after step 1. Since $\widehat{w}_0$ was obtained, the state $|\tilde{\psi}_0\rangle$ is equal to $|\widehat{w}_0\rangle|\widehat{v}_0\rangle$. We have,

$$\begin{aligned}
|\tilde{\psi}_1\rangle &= U_1 U_0^\dagger |\tilde{\psi}_0\rangle \\
&= U_1 U_0^\dagger |\widehat{w}_0\rangle|\widehat{v}_0\rangle \\
&= U_1 \left(U_0^\dagger \frac{|\psi_0\rangle}{\sqrt{p_0}} - U_0^\dagger \sqrt{\frac{1-p_0}{p_0}} |\widehat{w}_0^\perp\rangle \right) && (8) \\
&= U_1 \frac{|\psi\rangle}{\sqrt{p_0}} - U_1 U_0^\dagger \sqrt{\frac{1-p_0}{p_0}} |\widehat{w}_0^\perp\rangle && (9) \\
&= \frac{|\psi_1\rangle}{\sqrt{p_0}} - U_1 U_0^\dagger \sqrt{\frac{1-p_0}{p_0}} |\widehat{w}_0^\perp\rangle && (10) \\
&= \frac{1}{\sqrt{p_0}} \left(\sqrt{p_1}|\widehat{w}_1\rangle|\widehat{v}_1\rangle + \sqrt{1-p_1}|\widehat{w}_1^\perp\rangle - U_1 U_0^\dagger \sqrt{1-p_0}|\widehat{w}_0^\perp\rangle \right), && (11)
\end{aligned}$$

where (8) follows from isolating $|\widehat{w}_0\rangle|\widehat{v}_0\rangle$ in (7), (9) and (10) are obtained by definition of U_0 and U_1 respectively, and (11) also follows from (7). At this point, Patty applies the measurement in the computational basis in order to obtain $\widehat{w}_1$. Since we are interested only in finding a lower bound, the probability to obtain $\widehat{w}_1$ is minimized when $U_1 U_0^\dagger |\widehat{w}_0^\perp\rangle = |\widehat{w}_1\rangle|\widehat{v}_1\rangle$. It easily follows that,

$$\begin{aligned}
p_{\widehat{w}_1|\widehat{w}_0} &= \langle\tilde{\psi}_1|\widehat{w}_1\rangle\langle\widehat{w}_1|\tilde{\psi}_1\rangle \\
&\geq \frac{1}{p_0}\left(\sqrt{p_1} - \sqrt{1-p_0}\right)^2 && (12) \\
&\geq \frac{1}{p_0}\left(\sqrt{p_1} - \sqrt{p_1-\varepsilon}\right)^2 && (13) \\
&\geq \frac{\varepsilon^2}{4p_0}, && (14)
\end{aligned}$$

where (12) follows from (11), (13) is obtained from (6), and (14) follows from a Taylor expansion. Finally, (14) gives the desired result since

$$p_\oplus \geq p_0 \cdot p_{\widehat{w}_1|\widehat{w}_0} \geq \frac{\varepsilon^2}{4}.$$

Theorem 6. *If there exists an algorithm A that can cheat the* mBGKW *Bit Commitment scheme with probabilities $p_0 + p_1 > 1 + 2/\sqrt{2^n}$ then there exists an algorithm A' that can predict an unknown n-bit string $(r_0 \oplus r_1)$ with probabilities better than $1/2^n$, which is impossible.*

Proof (of Theorem 6). From the isolation assumption, we have

$$p_\oplus = \frac{1}{2^n}\,.$$

Using the result from Lemma 1,

$$\frac{1}{2^n} \geq \frac{\varepsilon^2}{4} \quad \Longrightarrow \quad \varepsilon \leq \frac{1}{\sqrt{2^{n-2}}}\,. \tag{15}$$

It follows that the binding condition is satisfied: plugging (15) in Lemma 1, we get for any cheating strategies

$$p_0 + p_1 \leq 1 + \frac{1}{\sqrt{2^{n-2}}} .$$

Notice that the proof presented in Lemma 1 can easily be generalized to a whole class of Bit Commitment schemes with the properties that information unknown to Patty is sent to Peggy to commit, and an *exact* answer is needed from Patty to unveil successfully the committed bit. Theorem 6 therefore holds for a whole class of Bit Commitment schemes in the Two-Prover model.

Note that sBGKW is the same as mBGKW where $r_0 := 000...0$ is the all-zero string all the time. The statement and proof of Lemma 1 is equally valid for any fixed choice of either (but not both)$r_0$or$r_1$because the probability to predict $r_0 \oplus r_1$ remains exponentially small. Hence using only the model's assumption we get:

Corollary 1. *If there exists an algorithm A that can cheat the* sBGKW *Bit Commitment scheme with probabilities* $p_0 + p_1 > 1 + 2/\sqrt{2^n}$ *then there exists an algorithm* A' *that can predict an unknown n-bit string r with probabilities better than* $1/2^n$*, which is impossible.*

However, as previously, this proof is valid *solely* in a stand-alone security model. As soon as one starts composing such protocols, this proof is not necessarily valid anymore.

6 Conclusion and Open Problems

This paper contained several results. It showed that Two-Prover Bit Commitment schemes may or not be secure quantumly when they are classically. It also considered for the first time ever the exact conditions that the provers and verifier must satisfy to obtain security proofs of such Bit Commitment schemes both classically and quantumly.

A natural question would be to determine if the binding condition of ALL Two-Prover Quantum Bit Commitment schemes can be broken by a non-local computation that does not allow to communicate. This would imply that the no-communication assumption is NEVER sufficient to asses security of such schemes. A hierarchy of non-local correlations may be imagined with higher up correlations simulating lower down correlations, but not the opposite. What is the Bit Commitment scheme that can be broken only by a very highest correlation ?

In our definition of Bit Commitment, we assessed that cheating meant $p_0 + p_1 > 1+\epsilon$ for non-negligible ϵ. However, recently more precise binding conditions have been introduced [DFRSS07]. The results of this paper should be extended to suit this newer definition.

The last natural question that results from our work is to find the complexity class corresponding to Quantum Two-Prover Zero-Knowledge Interactive Proofs (and similarly for $k > 2$ provers). Remember that these questions are not even settled for Quantum Two-Prover Interactive Proofs alone. As soon as the verifier is also quantum it is not clear how Bit Commitments may be used to "encrypt" the verifier's computations, thus the classical methodologies fall apart.

Acknowledgements. We are thankful the anonymous referees (of several conferences) for their comments and numerous suggestions.

References

[Ara02] Aravind, P.K.: Bell's theorem without inequalities and only two distant observers. Foundation of Physics Letters, 397–405 (2002)

[BCMS98] Brassard, G., Crépeau, C., Mayers, D., Salvail, L.: Defeating classical bit commitment schemes with a quantum computer. ArXiv Quantum Physics e-prints (1998)

[BCU+06] Buhrman, H., Christandl, M., Unger, F., Wehner, S., Winter, A.: Implications of superstrong nonlocality for cryptography. Proceedings of The Royal Society A 462(2071), 1919–1932 (2006)

[BFL91] Babai, L., Fortnow, L., Lund, C.: Non-deterministic exponential time has two-prover interactive protocols. Computational Complexity 1, 3–40 (1991)

[BGKW88] BenOr, M., Goldwasser, S., Kilian, J., Widgerson, A.: Multi-prover interactive proofs: how to remove intractability. In: STOC 1988: Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing, pp. 113–131. ACM Press, New York (1988)

[BHOP08] Ben-Or, M., Hassidim, A., Pilpel, H.: Quantum Multi Prover Interactive Proofs with Communicating Provers. In: 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008), pp. 467–476. IEEE Computer Society (2008)

[CHTW04] Cleve, R., Hoyer, P., Toner, B., Watrous, J.: Consequences and limits of nonlocal strategies. In: CCC 2004: Proceedings of the 19th IEEE Annual Conference on Computational Complexity, pp. 236–249. IEEE Computer Society, Washington, DC, USA (2004)

[CLS01] Crépeau, C., Légaré, F., Salvail, L.: How to Convert the Flavor of a Quantum Bit Commitment. In: Pfitzmann, B. (ed.) EUROCRYPT 2001. LNCS, vol. 2045, pp. 60–77. Springer, Heidelberg (2001)

[CSUU07] Cleve, R., Slofstra, W., Unger, F., Upadhyay, S.: Perfect Parallel Repetition Theorem for Quantum XOR Proof Systems. In: CCC 2007: Proceedings of the 2007 IEEE 22nd Annual Conference on Computational Complexity, pp. 109–114. IEEE Computer Society, Los Alamitos (2007)

[DFSS05] Damgård, I., Fehr, S., Salvail, L., Schaffner, C.: Cryptography in the bounded quantum-storage model. In: 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), pp. 449–458. IEEE Computer Society (2005)

[DFRSS07] Damgård, I., Fehr, S., Renner, R., Salvail, L., Schaffner, C.: A Tight High-Order Entropic Quantum Uncertainty Relation with Applications. In: Menezes, A. (ed.) CRYPTO 2007. LNCS, vol. 4622, pp. 360–378. Springer, Heidelberg (2007)

[DLTW08] Doherty, A.C., Liang, Y.-C., Toner, B., Wehner, S.: The Quantum Moment Problem and Bounds on Entangled Multi-prover Games. In: CCC 2008: Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity, pp. 199–210. IEEE Computer Society, Washington, DC, USA (2008)

[DMS00] Dumais, P., Mayers, D., Salvail, L.: Perfectly concealing quantum bit commitment from any quantum one-way permutation, pp. 300–315 (2000)

[IKM09] Ito, T., Kobayashi, H., Matsumoto, K.: Oracularization and Two-Prover One-Round Interactive Proofs against Nonlocal Strategies. In: CCC 2009: Proceedings of the 2009 IEEE 24th Annual Conference on Computational Complexity, pp. 217–228. IEEE Computer Society, Los Alamitos (2009)

[IKO03] Ibaraki, T., Katoh, N., Ono, H. (eds.): ISAAC 2003. LNCS, vol. 2906. Springer, Heidelberg (2003)

[IKPSY08] Ito, T., Kobayashi, H., Preda, D., Sun, X., Yao, A.C.-C.: Generalized Tsirelson Inequalities, Commuting-Operator Provers, and Multi-prover Interactive Proof Systems. In: CCC 2008: Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity, pp. 187–198. IEEE Computer Society, Washington, DC, USA (2008)

[Ken05] Kent, A.: Secure classical bit commitment using fixed capacity communication channels. J. Cryptology 18(4), 313–335 (2005)

[Kil88] Kilian, J.: Founding cryptography on oblivious transfer. In: Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing, pp. 20–31 (1988)

[KKMV08] Kempe, J., Kobayashi, H., Matsumoto, K., Vidick, T.: Using Entanglement in Quantum Multi-prover Interactive Proofs. In: CCC 2008: Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity, pp. 211–222. IEEE Computer Society, Washington, DC, USA (2008)

[KM03] Kobayashi, H., Matsumoto, K.: Quantum multi-prover interactive proof systems with limited prior entanglement. J. Comput. Syst. Sci. 66(3), 429–450 (2003)

[KMY03] Kobayashi, H., Matsumoto, K., Yamakami, T.: Quantum merlin-arthur proof systems: Are multiple merlins more helpful to arthur? In: Ibaraki, et al. (eds.) [IKO03], pp. 189–198 (2003)

[LC97] Lo, H.-K., Chau, H.F.: Is quantum bit commitment really possible? Phys. Rev. Lett. 78(17), 3410–3413 (1997)

[LS97] Lapidot, D., Shamir, A.: Fully parallelized multi-prover protocols for nexptime. J. Comput. Syst. Sci. 54(2), 215–220 (1997)

[May96] Mayers, D.: Unconditionally secure quantum bit commitment is impossible (November 1996)

[PR94] Popescu, S., Rohrlich, D.: Nonlocality as an axiom. Foundations of Physics 24, 379 (1994)

[PR97] Popescu, S., Rohrlich, D.: Causality and nonlocality as axioms for quantum mechanics. In: Symposium on Causality and Locality in Modern Physics and Astronomy (1997)

[Weh06] Wehner, S.: Entanglement in Interactive Proof Systems with Binary Answers. In: Durand, B., Thomas, W. (eds.) STACS 2006. LNCS, vol. 3884, pp. 162–171. Springer, Heidelberg (2006)

[Yao03] Yao, A.C.-C.: Interactive proofs for quantum computation. In: Ibaraki, et al. (eds.) [IKO03], p. 1 (2003)

A Isolation Examples

Example 1:
Let P_0 send to Ted a message represented by $(X, \{0\})_0$ (the variable X is tagged with $\{0\}$ and comes from P_0). Then Ted generates a uniform random variable $(W, T)_D$ (its tag and receiver have not been set yet) and produces the message $M = X \oplus W$ for P_1. Checking with equation (4) we see there is no problem setting M's tag to $\{1\}$, as

$$H((M, \{1\})_1 | (X, \{0\})_0) = H((W, T)_D) = H((M, \{1\})_1).$$

This is satisfied since $(W, T)_D$ is uniform and has never been sent. However, the practical definition would have assigned the tag $T := \{0\}$ since W's tag would have been $\{0, 1\}$ (by the second rule) and $\{0\} = \{0\} \cap \{0, 1\}$. Let Ted send $(M, \{1\})_1$. We now get that for *both* $D = 0$ and 1, if $T = \{D\}$ or $\{0, 1\}$ then the left-hand side of equation (4) for W is

$$H((W, T)_D | (X, \{0\})_0, (M, \{1\})_1) = 0,$$

and the right-hand side is respectively

$$\begin{aligned} H((W, \{0\})_0 | (X, \{0\})_0) &= H((W, \{0\})_0) = 1, \\ H((W, \{1\})_1 | (M, \{1\})_1) &= H((X, \{0\})_0) = 1, \\ H((W, \{0, 1\})_D) &= 1. \end{aligned}$$

Because equation (4) is not satisfied for both $T = \{D\}$ and $\{0, 1\}$, W's tag is set to $T := \emptyset$, and Ted should not send $(W, \emptyset)_D$ to neither of P_D, for $D = 0, 1$.

Example 2:
Similarly, we can send to P_1 a message M that would have been tagged $\emptyset$ by the practical definition. We take the **PR**-box relation as example. Suppose the variables $(X, \{0\})_0$ and $(Y, \{1\})_1$ have already been sent to Ted by the players (and tagged accordingly), and $(U, \{0, 1\})_0$ [9] has been sent by Ted to P_0. Let $(W, T)_D$ be a uniformly distributed random variable chosen by Ted, with $D \in \{0, 1\}$. Consider the following variable for P_1,

$$V = U \oplus \big(W \oplus X\big) \wedge T,$$

that is, we randomized the variable tagged $\{0\}$ (i.e. X) in the **PR**-box relation. In the practical definition, because W is chosen uniformly and independently of previous variables, the second rule would have assigned a tag $\{0, 1\}$ to it, and so V's tag would have been set to $\emptyset = \{0, 1\} \cap \{0, 1\} \cap \{0\} \cap \{1\}$. However, checking with equation (4), because W has not been sent yet, we get that there is no problem setting V's tag to $\{1\}$, as

$$H((V, \{1\})_1 | (Y, \{1\})_1, (X, \{0\})_0, (U, \{0, 1\})_0) = \frac{1}{2} = H((V, \{1\})_1 | (Y, \{1\})_1, (U, \{0, 1\})_0).$$

[9] It is straightforward to verify that this is the less restrictive tag.

So Ted would send this message $(V, \{1\})_1$ to P_1. Is this a problem? No, because the classical limitations of non-locality have not been violated yet! The reason is simple: by randomizing completely all the [private] variables related to P_0, Ted is reducing the message he sends to P_1 to what P_1 can exactly achieve using local variables. That is to say, P_1 already has a random view of P_0's variables, so there is no problem for Ted to first randomize P_0's variables and then send this message to P_1. If we make the calculations, we see that indeed, for the variable V sent, the relation

$$V = U \oplus X \wedge Y$$

holds with probability 75%, just as in the classical scenario, and no W will never let us beat that. Of course, as in the previous example, the variable $(W, T)_D$ used to randomize can never be disclosed to *any* of the two players, and equation (4) agrees with that (W's tag will be set to $T := \emptyset$ for both D).