

# On the Composition of Two-Prover Commitments, and Applications to Multi-Round Relativistic Commitments

Serge Fehr and Max Fillinger\*

Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands  
{serge.fehr,max.fillinger}@cwi.nl

**Abstract.** We consider the related notions of *two-prover* and of *relativistic* commitment schemes. In recent work, Lunghi *et al.* proposed a new relativistic commitment scheme with a *multi-round sustain phase* that keeps the binding property alive as long as the sustain phase is running. They prove security of their scheme against classical attacks; however, the proven bound on the error parameter is very weak: it blows up *double exponentially* in the number of rounds. In this work, we give a new analysis of the multi-round scheme of Lunghi *et al.*, and we show a *linear* growth of the error parameter instead (also considering classical attacks only). Our analysis is based on a new *composition theorem* for two-prover commitment schemes. The proof of our composition theorem is based on a better understanding of the binding property of two-prover commitments that we provide in the form of new definitions and relations among them. These new insights are certainly of independent interest and are likely to be useful in other contexts as well. Finally, our work gives rise to several interesting open problems, for instance extending our results to the quantum setting, where the dishonest provers are allowed to perform measurements on an entangled quantum state in order to try to break the binding property.

## 1 Introduction

**TWO-PROVER COMMITMENT SCHEMES.** We consider the notion of *2-prover commitment schemes*, as originally introduced by Ben-Or, Goldwasser, Kilian and Wigderson in their seminal paper [BGKW88]. In a 2-prover commitment scheme, the prover (i.e., the entity that is responsible for preparing and opening the commitment) consists of two agents,  $P$  and  $Q$ , and it is assumed that these two agents cannot communicate with each other during the execution of the protocol. With this approach, the classical and quantum impossibility results for unconditionally secure commitment schemes [May97,LC97] can be circumvented.

A simple 2-prover bit commitment scheme is the scheme proposed by Crépeau *et al.* [CSST11], which works as follows. The verifier  $V$  chooses a uniformly random  $a \in \{0,1\}^n$  and sends it to  $P$ , who replies with  $x := y + a \cdot b$ , where  $b$  is the bit to commit to, and  $y \in \{0,1\}^n$  is a uniformly random string known (only) to  $P$  and  $Q$ . Furthermore, “+” is bit-wise XOR, and “ $\cdot$ ” is scalar multiplication (of the scalar  $b$  with the vector  $a$ ). In order to open the commitment (to  $b$ ),  $Q$  sends  $y$  to  $V$ , and  $V$  checks if  $x + y = a \cdot b$ . It is clear that this scheme is hiding: the commitment  $x = y + a \cdot b$  is uniformly random and independent of  $a$  no matter what  $b$  is. On the other hand, the binding property follows from the observation that in order to open the commitment to  $b = 0$ ,  $Q$  needs to announce  $y = x$ , and in order to open to  $b = 1$ , he needs to announce  $y = x + a$ . Thus, in order to open to *both*, he must know  $x$  and  $x + a$ , and thus  $a$ , which is a contradiction to the no-communication assumption, because  $a$  was sent to  $P$  only.

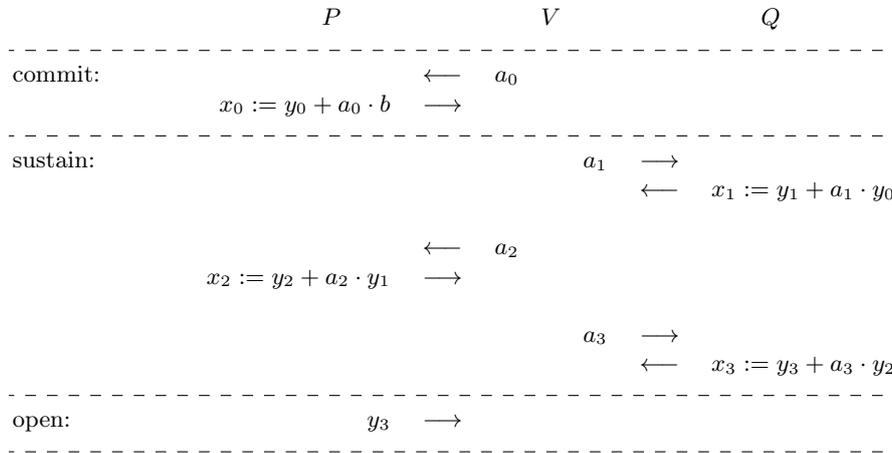
In the quantum setting, where the dishonest provers are allowed to share an entangled quantum state and can produce  $x$  and  $y$  by means of performing measurements on their respective parts of the state, the above reasoning for the binding property does not work anymore. Nevertheless, as shown in [CSST11], the binding property still holds (though with a weaker parameter).

**RELATIVISTIC COMMITMENT SCHEMES.** The idea of *relativistic commitment schemes*, as introduced by Kent [Ken99], is to take a 2-prover commitment scheme as above and enforce the no-communication assumption by means of relativistic effects: place  $P$  and  $Q$  spatially far apart, and execute the scheme fast enough, so that there is not enough time for them to communicate. The obvious downside of such a relativistic commitment scheme is that the binding property stays alive only for a very short time: the opening has to take place almost immediately after the committing, before the provers have the chance to exchange information. This limitation can be circumvented by considering *multi-round* schemes, where

\* Supported by the *NWO Free Competition* grant 617.001.203.

after the actual commit phase there is a *sustain phase*, during which the provers and the verifier keep exchanging messages, and as long as this sustain phase is running, the commitment stays binding (and hiding), until the commitment is finally opened. Such schemes were proposed in [Ken99] and [Ken05], but they are rather inefficient, and the security analyses are somewhat informal (e.g., with no formal security definitions) and of asymptotic nature.

More recently, Lunghi *et al.* [LKB<sup>+</sup>15] proposed a new and simple multi-round relativistic commitment scheme, and provided a rigorous security analysis. Their scheme works as follows. The actual commit protocol is the commit protocol from the Crépeau *et al.* scheme:  $V$  sends a uniformly random string  $a_0 \in \{0,1\}^n$  to  $P$ , who returns  $x_0 := y_0 + a_0 \cdot b$ . Then, to sustain the commitment, before  $P$  has the chance to tell  $a_0$  to  $Q$ ,  $V$  sends a new uniformly random string  $a_1 \in \{0,1\}^n$  to  $Q$  who replies with  $x_1 := y_1 + a_1 \cdot y_0$ , where  $y_1 \in \{0,1\}^n$  is another random string shared between  $P$  and  $Q$ , and the multiplication  $a_1 \cdot y_0$  is in a suitable finite field. Then, to further sustain the commitment,  $V$  sends a new uniformly random string  $a_2 \in \{0,1\}^n$  to  $P$  who replies with  $x_2 := y_2 + a_2 \cdot y_1$ , etc. Finally, after the last sustain round where  $x_m := y_m + a_m \cdot y_{m-1}$  has been sent to  $V$ , in order to finally open the commitment,  $y_m$  is sent to  $V$  (by the other prover). See Figure 1. In order to verify the opening,  $V$  computes  $y_{m-1}, y_{m-2}, \dots, y_0$  inductively in the obvious way, and checks if  $x_0 + y_0 = a_0 \cdot b$ .



**Fig. 1.** The Lunghi *et al.* multi-round scheme (for  $m = 3$ ).

What is crucial is that in round  $i$  (say for odd  $i$ ), when preparing  $x_i$ , the prover  $Q$  must not know  $a_{i-1}$ , but he is allowed to know  $a_1, \dots, a_{i-2}$ . Thus, the execution must be timed in such a way that between subsequent rounds there is not enough time for the provers to communicate, but they may communicate over multiple rounds.

As for the security of this scheme, it is obvious that the hiding property stays satisfied up to the open phase: every single message  $V$  receives is one-time-pad encrypted. As for the binding property, Lunghi *et al.* prove that the scheme with a  $m$ -round sustain phase is  $\varepsilon_m$ -binding against classical attacks, where  $\varepsilon_m$  satisfies  $\varepsilon_0 = 2^{-n}$  (this is just the standard Crépeau *et al.* scheme) and  $\varepsilon_m \leq 2^{-n-1} + \sqrt{\varepsilon_{m-1}}$  for  $m \geq 1$ . Thus, even when reading this recursive formula liberally by ignoring the  $2^{-n-1}$  term, we obtain

$$\varepsilon_m \lesssim \sqrt[m]{\varepsilon_0} = 2^{-\frac{n}{m}},$$

i.e., the error parameter blows up *double exponentially* in  $m$ .<sup>1</sup> In other words, in order to have a non-trivial  $\varepsilon_m$  we need that  $n$ , the size of the strings that are communicated, is *exponential* in  $m$ . This means that Lunghi *et al.* can only afford a very small number of rounds. For instance, in their implementation where they can manage  $n = 512$  (beyond that, the local computation takes too long), asking for an

<sup>1</sup> Lunghi *et al.* also provide a more complicated recursive formula for  $\varepsilon_m$  that is slightly better, but the resulting blow-up is still double exponential.

error parameter  $\varepsilon_m$  of approximately  $2^{-32}$ , they can do  $m = 4$  rounds.<sup>2</sup> This allows them to keep a commitment alive for 2 ms.

**OUR RESULTS.** Our main goal is to improve the bound on the binding parameter of the above multi-round scheme. Indeed, our results show that the binding parameter blows up only *linearly* in  $m$ , rather than double exponentially. Explicitly, our results show that (for classical attacks)

$$\varepsilon_m \leq (m + 1) \cdot 2^{-\frac{n}{2} + 2}.$$

Using the same  $n$  and error parameter as in the implementation of Lunghi *et al.*, we can now afford approximately  $m = 2^{224}$  rounds. Scaling up the 2 ms from the Lunghi *et al.* experiment for 4 rounds gives us a time that is in the order of  $10^{56}$  years.

We use the following strategy to obtain our improved bound on  $\varepsilon_m$ . We observe that the first sustain round can be understood as committing on the opening information  $y_0$  of the actual commitment, using an extended version of the Crépeau *et al.* scheme that commits to a *string* rather than to a bit. Similarly, the second sustain round can be understood as committing on the opening information  $y_1$  of that commitment from the first sustain round, etc. Thus, thinking of the  $m = 1$  version of the scheme, what we have to prove is that if we have two commitment schemes  $\mathcal{S}$  and  $\mathcal{S}'$ , and we modify the opening phase of  $\mathcal{S}$  in that we first commit to the opening information (using  $\mathcal{S}'$ ) and then open that commitment, then the resulting commitment scheme is still binding; note that, intuitively, this is what one would indeed expect. Given such a composition theorem, we can then apply it inductively and conclude security (i.e. the binding property) of the Lunghi *et al.* multi-round scheme.

Our main result is such a general composition theorem, which shows that if  $\mathcal{S}$  and  $\mathcal{S}'$  are respectively  $\varepsilon$ - and  $\delta$ -binding (against classical attacks) then the composed scheme is  $(\varepsilon + \delta)$ -binding (against classical attacks), under some mild assumptions on  $\mathcal{S}$  and  $\mathcal{S}'$ . Hence, the error parameters simply add up; this is what gives us the linear growth. The proof of our composition theorem crucially relies on new definitions of the binding property of 2-prover commitment schemes, which seem to be handier to work with than the  $p_0 + p_1 \leq 1 + \varepsilon$  definition as for instance used by Lunghi *et al.* Our definitions formalize the following intuitive requirement: After the commit phase, even if the provers are dishonest, there should exist some bit  $\hat{b}$  such that opening the commitment to any other bit fails (with high probability). We show that one of our new definitions is equivalent to the  $p_0 + p_1$ -definition, while the other one is strictly stronger. Our result holds for both definitions, so we not only obtain a better parameter than Lunghi *et al.* but also with respect to a stronger definition, and thus we improve the result also in that direction.

One subtle issue is that the extended version of the Crépeau *et al.* scheme to strings, as it is used in the sustain phase, is not a fully secure string commitment scheme. The reason is that for *any*  $y$  that may be announced in the opening phase, there exists a string  $s$  such that  $x + y = a \cdot s$ ; as such, the provers can commit to some fixed string, and then can still decide to either open the commitment to that string (by running the opening phase honestly), or to open it to a random string that is out of their control (by announcing a random  $y$ ). We deal with this by also introducing a *relaxed* version (which we call *fairly-binding*) of the binding property, which captures this limited freedom for the provers, and we show that it is satisfied by the (extended version of the) Crépeau *et al.* scheme and that our composition theorem holds for this relaxed version; finally, we observe that the composed fairly-binding string commitment scheme is a binding *bit* commitment scheme when restricting the domain to a bit.

As such, we feel that our techniques and insights not only give rise to an improved analysis of the Lunghi *et al.* multi-round scheme, but they significantly improve our understanding of the security of 2-prover commitment schemes, and as such are likely to find further applications.

**OPEN PROBLEMS.** Our work gives rise to a list of interesting and challenging open problems. For instance, our composition theorem only applies to pairs  $\mathcal{S}, \mathcal{S}'$  of commitment schemes of a certain restricted form, e.g., only one prover should be involved in the commit phase (as it is the case in the Crépeau *et al.* scheme). Our proof crucially relies on this, but there seems to be no fundamental reason for such a restriction. Thus, we wonder if it is possible to generalize our composition theorem to a larger class of pairs of schemes, or, ultimately, to *all* pairs of schemes (that “fit together”).

<sup>2</sup> Note that [LKB<sup>+</sup>15] mentions  $\varepsilon_m \approx 10^{-5} \approx 2^{-16}$ , but this is an error, as communicated to us by the authors, and as can easily be verified. Also, [LKB<sup>+</sup>15] mentions  $m = 5$  rounds, but this is because they include the commit round in their counting, and we do not.

In another direction, some of our observations and results generalize immediately to the quantum setting, where the two dishonest provers are allowed to compute their messages by performing measurements on an entangled quantum state, but in particular our main result, the composition theorem, does not generalize. Also here, there seems to be no fundamental reason, and thus, generalizing our composition theorem to the quantum setting is an interesting open problem. Finally, in order to obtain security of the Lunghi *et al.* multi-round scheme against quantum attacks, beyond a quantum version of the composition theorem, one also needs to prove security against quantum attacks of the (extended version of the) original Crépeau *et al.* scheme as a (fairly-binding) *string* commitment scheme.

**CONCURRENT WORK.** In independent and concurrent work, Chakraborty, Chailloux and Leverrier [CCL15] showed (almost) the same linear bound for the Lunghi *et al.* scheme, but with respect to the original—and thus weaker—notation of security. Their approach is more direct and tailored to the specific scheme; our approach is more abstract and provides more insight, and our result applies much more generally.

## 2 Preliminaries

### 2.1 Basic Notation

**PROBABILITY DISTRIBUTIONS.** For the purpose of this work, a (*probability*) *distribution* is a function  $p : \mathcal{X} \rightarrow [0, 1]$ ,  $x \mapsto p(x)$ , where  $\mathcal{X}$  is a finite non-empty set, with the property that  $\sum_{x \in \mathcal{X}} p(x) = 1$ . For specific choices  $x_\circ \in \mathcal{X}$ , we tend to write  $p(x = x_\circ)$  instead of  $p(x_\circ)$ . For any subset  $A \subset \mathcal{X}$ , called an *event*, the probability  $p(A)$  is naturally defined as  $p(A) = \sum_{x \in A} p(x)$ , and it holds that

$$p(A) + p(\Gamma) = p(A \cup \Gamma) + p(A \cap \Gamma) \leq 1 + p(A \cap \Gamma) \quad (1)$$

for all  $A, \Gamma \subset \mathcal{X}$ , and, more generally, that

$$\sum_{i=1}^k p(A_i) \leq p(A_1 \cup \dots \cup A_k) + \sum_{i < j} p(A_i \cap A_j) \leq 1 + \sum_{i < j} p(A_i \cap A_j) \quad (2)$$

for all  $A_1, \dots, A_k \subset \mathcal{X}$ . For a distribution  $p : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  on two (or more) variables, probabilities like  $p(x = y)$ ,  $p(x = f(y))$ ,  $p(x \neq y)$  etc. are naturally understood as

$$p(x = y) = p(\{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid x = y\}) = \sum_{\substack{x \in \mathcal{X}, y \in \mathcal{Y} \\ \text{s.t. } x=y}} p(x, y)$$

etc., and the *marginals*  $p(x)$  and  $p(y)$  are given by  $p(x) = \sum_y p(x, y)$  and  $p(y) = \sum_x p(x, y)$ , respectively. Vice versa, given two distributions  $p(x)$  and  $p(y)$ , we say that a distribution  $p(x, y)$  on two variables is a *consistent joint distribution* if the two marginals of  $p(x, y)$  coincide with  $p(x)$  and  $p(y)$ , respectively. We will make use of the following property on the existence of a consistent joint distribution that maximizes the probability that  $x = y$ ; the proof is given in the appendix.

**Lemma 2.1.** *Let  $p(x)$  and  $p(y)$  be two distributions on a common set  $\mathcal{X}$ . Then there exists a consistent joint distribution  $p(x, y)$  such that  $p(x = y = x_\circ) = \min\{p(x = x_\circ), p(y = x_\circ)\}$  for all choices of  $x_\circ \in \mathcal{X}$ . Additionally,  $p(x, y)$  satisfies  $p(x, y \mid x \neq y) = p(x \mid x \neq y) \cdot p(y \mid x \neq y)$ .*

**PROTOCOLS.** In this work, we will consider 3-party (interactive) *protocols*, where the parties are named  $P$ ,  $Q$  and  $V$  (the two “provers” and the “verifier”). Such a protocol  $\text{prot}_{PQV}$  consists of a triple  $(\text{prot}_P, \text{prot}_Q, \text{prot}_V)$  of  $L$ -round *interactive algorithms* for some  $L \in \mathbb{N}$ . Each interactive algorithm takes an input, and for every round  $\ell \leq L$  computes the messages to be sent to the other algorithms/parties in that round as deterministic functions of its input, the messages received in the previous rounds, and the local randomness. In the same way, the algorithms produce their respective outputs after the last round. We write

$$(\text{out}_P \parallel \text{out}_Q \parallel \text{out}_V) \leftarrow (\text{prot}_P(\text{in}_P) \parallel \text{prot}_Q(\text{in}_Q) \parallel \text{prot}_V(\text{in}_V))$$

to denote the execution of the protocol  $\text{prot}_{PQV}$  on the respective inputs  $\text{in}_P, \text{in}_Q$  and  $\text{in}_V$ , and that the respective outputs  $\text{out}_P, \text{out}_Q$  and  $\text{out}_V$  are produced. Clearly, for any protocol  $\text{prot}_{PQV}$  and any input  $\text{in}_P, \text{in}_Q, \text{in}_V$ , the probability distribution  $p(\text{out}_P, \text{out}_Q, \text{out}_V)$  of the output is naturally well defined.

If we want to make the local randomness explicit, we write  $\text{prot}_P[\xi_P](in_P)$  etc., and understand that  $\xi_P$  is correctly sampled — without loss of generality, we may assume it to be a uniformly random bit string of sufficient length. Furthermore, we write  $\text{prot}_P[\xi_{PQ}](in_P)$  and  $\text{prot}_Q[\xi_{PQ}](in_Q)$  to express that  $\text{prot}_P$  and  $\text{prot}_Q$  use *the same* randomness, in which case we speak of *joint randomness*.

We can *compose* two interactive algorithms  $\text{prot}_P$  and  $\text{prot}'_P$  in the obvious way, by applying  $\text{prot}'_P$  to the output of  $\text{prot}_P$ . The resulting interactive algorithm is denoted as  $\text{prot}'_P \circ \text{prot}_P$ . Composing the respective algorithms of two protocols  $\text{prot}_{PQV} = (\text{prot}_P, \text{prot}_Q, \text{prot}_V)$  and  $\text{prot}'_{PQV} = (\text{prot}'_P, \text{prot}'_Q, \text{prot}'_V)$  results in the composed protocol  $\text{prot}'_{PQV} \circ \text{prot}_{PQV}$ . If  $\text{prot}_P$  is a *non-interactive* algorithm, then  $\text{prot}'_{PQV} \circ \text{prot}_P$  is naturally understood as the protocol  $\text{prot}'_{PQV} \circ \text{prot}_P = (\text{prot}'_P \circ \text{prot}_P, \text{prot}'_Q, \text{prot}'_V)$ , and similarly  $\text{prot}'_{PQV} \circ \text{prot}_{QV}$  in case  $\text{prot}_{QV}$  is a protocol among  $Q$  and  $V$  only.

## 2.2 2-Prover Commitment Schemes

**Definition 2.2.** A 2-prover (string) commitment scheme  $\mathcal{S}$  consists of two interactive protocols  $\text{com}_{PQV} = (\text{com}_P, \text{com}_Q, \text{com}_V)$  and  $\text{open}_{PQV} = (\text{open}_P, \text{open}_Q, \text{open}_V)$  between the two provers  $P$  and  $Q$  and the verifier  $V$ , with the following syntactics. The commit protocol  $\text{com}_{PQV}$  uses joint randomness  $\xi_{PQ}$  for  $P$  and  $Q$  and takes a string  $s \in \{0, 1\}^n$  as input for  $P$  and  $Q$  (and independent randomness and no input for  $V$ ), and it outputs a commitment  $com$  to  $V$  and some state information to  $P$  and  $Q$ :

$$(\text{state}_P \| \text{state}_Q \| c) \leftarrow (\text{com}_P[\xi_{PQ}](s) \| \text{com}_Q[\xi_{PQ}](s) \| \text{com}_V(\emptyset)).$$

The opening protocol  $\text{open}_{PQV}$  uses joint randomness  $\eta_{PQ}$  and outputs a string or a rejection symbol to  $V$ , and nothing to  $P$  and  $Q$ :

$$(\emptyset \| \emptyset \| s) \leftarrow (\text{open}_P[\eta_{PQ}](\text{state}_P) \| \text{open}_Q[\eta_{PQ}](\text{state}_Q) \| \text{open}_V(c))$$

with  $s \in \{0, 1\}^n \cup \{\perp\}$ . The set  $\{0, 1\}^n$  is called the domain of  $\mathcal{S}$ ; if  $n = 1$  then we refer to  $\mathcal{S}$  as a bit commitment scheme instead, and we tend to use  $b$  rather than  $s$  to denote the committed bit.

*Remark 2.3.* By convention, we assume throughout the paper that the commitment  $c$  output by  $V$  equals the *communication* that takes place between  $V$  and the provers during the commit phase. This is without loss of generality since, in general,  $c$  is computed as a (possibly randomized) function of the communication, which  $V$  just as well can apply in the opening phase.

*Remark 2.4.* Note that we specify that  $P$  and  $Q$  use *fresh* joint randomness  $\eta_{PQ}$  in the opening phase, and, if necessary, the randomness  $\xi_{PQ}$  from the commit phase can be “handed over” to the opening phase via  $\text{state}_P$  and  $\text{state}_Q$ ; this will be convenient later on. Alternatively, one could declare that  $P$  and  $Q$  *re-use* the joint randomness from the commit phase.

Whenever we refer to such a 2-prover commitment scheme, we take it as understood that the scheme is complete and hiding, as defined below, for “small” values of  $\gamma$  and  $\delta$ . Since our focus will be on the binding property, we typically do not make the parameters  $\eta$  and  $\delta$  explicit.

**Definition 2.5.** A 2-prover commitment scheme is  $\gamma$ -complete if in an honest execution  $V$ ’s output  $s$  of  $\text{open}_{PQV}$  equals  $P$  and  $Q$ ’s input  $s$  to  $\text{com}_{PQV}$  except with probability  $\eta$ , for any choice of  $P$  and  $Q$ ’s input  $s \in \{0, 1\}^n$ .

The standard definition for the hiding property is as follows:

**Definition 2.6.** A 2-prover commitment scheme is  $\delta$ -hiding if for any commit strategy  $\overline{\text{com}}_V$  and any two strings  $s_0$  and  $s_1$ , the distribution of the commitments  $c_0, c_1$ , produced as

$$(\text{state}_P \| \text{state}_Q \| c_b) \leftarrow (\text{com}_P[\xi_{PQ}](s_b) \| \text{com}_Q[\xi_{PQ}](s_b) \| \overline{\text{com}}_V(\emptyset)), b = 0, 1$$

have statistical distance at most  $\delta$ . A 0-hiding scheme is also called perfectly hiding.

Defining the binding property is more subtle. First, note that an attack against the binding property consists of an “allowed” commit strategy  $\overline{\text{com}}_{PQ} = (\overline{\text{com}}_P, \overline{\text{com}}_Q)$  and an “allowed” opening strategy  $\overline{\text{open}}_{PQ} = (\overline{\text{open}}_P, \overline{\text{open}}_Q)$  for  $P$  and  $Q$ . Any such attack fixes  $p(s)$ , the distribution of  $s \in \{0, 1\}^n \cup \{\perp\}$  that is output by  $V$  after the opening phase, in the obvious way.

What exactly “allowed” means may depend on the scheme and needs to be specified. Typically, in the 2-prover setting, we only allow strategies  $\overline{\text{com}}_{PQ}$  and  $\overline{\text{open}}_{PQ}$  with *no communication* at all between the two provers, but we may also be more liberal and allow some *well-controlled* communication, as in the Lunghi *et al.* multi-round scheme. Furthermore, in this work, we focus on *classical* attacks, where  $\overline{\text{com}}_P, \overline{\text{com}}_Q, \overline{\text{open}}_P$  and  $\overline{\text{open}}_Q$  are classical interactive algorithms as specified in the previous section, with access to joint randomness. But one could also consider *quantum* attacks, in which the provers can perform measurements on an entangled quantum state. Our main result holds for classical attacks only, and so the unfamiliar reader can safely ignore the possibility of quantum attacks, but some of our insights also apply to quantum attacks.

A somewhat accepted definition for the binding property of a 2-prover *bit* commitment scheme, as it is for instance used in [CSST11, LKB<sup>+</sup>15, FF15] (up to the factor 2 in the error parameter), is as follows. Here, we assume it has been specified which attacks are *allowed*, e.g., those where  $P$  and  $Q$  do not communicate during the course of the scheme.

**Definition 2.7.** *A 2-prover bit commitment scheme is  $\varepsilon$ -binding in the sense of  $p_0 + p_1 \leq 1 + 2\varepsilon$  if for every allowed commit strategy  $\overline{\text{com}}_{PQ}$ , and for every pair of allowed opening strategies  $\overline{\text{open}}_{PQ}^0$  and  $\overline{\text{open}}_{PQ}^1$ , which fix distributions  $p(b_0)$  and  $p(b_1)$  for  $V$ 's respective outputs, it holds that*

$$p(b_0=0) + p(b_1=1) \leq 1 + 2\varepsilon.$$

In the literature (see e.g. [CSST11] or [LKB<sup>+</sup>15]), the two probabilities  $p(b_0=0)$  and  $p(b_1=1)$  above are usually referred to as  $p_0$  and  $p_1$ , respectively.

### 2.3 The $\mathcal{CHSH}^n$ Scheme

Our main example is the bit commitment scheme by Crépeau *et al.* [CSST11] we mentioned in the introduction, and which works as follows. The commit phase  $\text{com}_{PQV}$  instructs  $V$  to sample and send to  $P$  a uniformly random  $a \in \{0, 1\}^n$ , and it instructs  $P$  to return  $x := r + a \cdot b$  to  $V$ , where  $r$  is the joint randomness, uniformly distributed in  $\{0, 1\}^n$ , and  $b$  is the bit to commit to, and the opening phase  $\text{open}_{PQV}$  instructs  $Q$  to send  $y := r$  to  $V$ , and  $V$  outputs the (smaller) bit  $b$  that satisfies  $x + y = a \cdot b$ , and  $b := \perp$  in case no such bit exists. Note that the provers in this scheme use the same randomness in the commit and opening phase; thus, formally,  $Q$  needs to output the shared randomness  $r \leftarrow \xi_{PQ}$  as  $\text{state}_Q$ . The opening phase uses no fresh randomness.

It is easy to see that this scheme is  $2^{-n}$ -complete and perfectly hiding (completeness fails in case  $a = 0$ ). For *classical* provers that do not communicate at all, the scheme is  $2^{-n-1}$ -binding in the sense of  $p_0 + p_1 \leq 1 + 2^{-n}$ , i.e. according to Definition 2.7. As for *quantum* provers, Crépeau *et al.* showed that the scheme is  $2^{-n/2}$ -binding; this was recently minorly improved to  $2^{-(n+1)/2}$  by Sikora, Chailloux and Kerenidis [SCK14].

We also want to consider an extended version of the scheme, where the bit  $b$  is replaced by a string  $s \in \{0, 1\}^n$  in the obvious way (where the multiplication  $a \cdot s$  is then understood in a suitable finite field), and we want to appreciate this extension as a 2-prover *string* commitment scheme. However, it is a priori not clear what is a suitable definition for the binding property, especially because for this particular scheme, the dishonest provers can always honestly commit to a string  $s$ , and can then decide to correctly open the commitment to  $s$  by announcing  $y := r$ , or open to a *random* string by announcing a randomly chosen  $y$  — any  $y$  satisfies  $x + y = a \cdot s$  for *some*  $s$  (unless  $a = 0$ , which almost never happens).<sup>3</sup>

Due to its close relation to the CHSH game [CHSH69], in particular to the arbitrary-finite-field version considered in [BS15], we will refer to this *string* commitment scheme as  $\mathcal{CHSH}^n$ .

## 3 On the Binding Property of 2-Prover Commitment Schemes

We introduce new definitions for the binding property of 2-prover commitment schemes. In the case of *bit* commitment schemes, they imply Definition 2.7, as we will show. Although not necessarily simpler, we feel that our definitions are closer to the intuition of what is expected from a commitment scheme, and as such easier to work with. Indeed, the proofs of our composition results are heavily based on our

<sup>3</sup> This could easily be prevented by requiring  $Q$  to announce  $s$  (rather than letting  $V$  compute it), but we want the information announced during the opening phase to fit into the domain of the commitment scheme.

new definitions. Also, our new notions are more flexible in terms of tweaking it; for instance, we modify them to obtain a *relaxed* notion for the binding property, which captures the binding property that is satisfied by the string commitment scheme  $\mathcal{CHSH}^n$ .

Throughout this section, when quantifying over attacks against (the binding property of) a scheme, it is always understood that there is a notion of *allowed* attacks for that scheme (e.g., all attacks for which  $P$  and  $Q$  do not communicate), and that the quantification is over all such allowed attacks. Also, even though our focus is on classical attacks, Proposition 3.10 and Theorem 3.11 also apply to quantum attacks.

### 3.1 Defining The Binding Property

Intuitively, we say that a scheme is binding if after the commit phase there exists a string  $\hat{s}$  so that no matter what the provers do in the opening phase, the verifier will output either  $s = \hat{s}$  or  $s = \perp$  (except with small probability). We consider two definitions of the binding property which interpret this intuitive requirement in two different ways. In the first definition, which we introduce in this section,  $\hat{s}$  is a function of the provers' (combined) view immediately after the commit phase. In the second one, which we introduce in Section 3.2,  $\hat{s}$  is specified by its distribution only. Both of these definitions admit a composition theorem.

**Definition 3.1 (Binding property).** *A 2-prover commitment scheme  $\mathcal{S}$  is  $\varepsilon$ -binding if for every commit strategy  $\overline{\text{com}}_{PQ}[\bar{\xi}_{PQ}]$  there exists a function  $\hat{s}(\bar{\xi}_{PQ}, c)$  of the joint randomness  $\bar{\xi}_{PQ}$  and the commitment<sup>4</sup>  $c$  such that for every opening strategy  $\overline{\text{open}}_{PQ}$  it holds that  $p(s \neq \hat{s}(\bar{\xi}_{PQ}, c) \wedge s \neq \perp) \leq \varepsilon$ . In short:*

$$\forall \overline{\text{com}}_{PQ} \exists \hat{s}(\bar{\xi}_{PQ}, c) \forall \overline{\text{open}}_{PQ} : p(s \neq \hat{s} \wedge s \neq \perp) \leq \varepsilon. \quad (3)$$

The string commitment scheme  $\mathcal{CHSH}^n$  does *not* satisfy this definition (the bit commitment version does, as we will show): after the commit phase, the provers can still decide to open the commitment to a *fixed* string, chosen before the commit phase, or to a *random* string that is out of their control. We capture this by the following relaxed version of the binding property: we allow  $V$ 's output  $s$  to be different from  $\hat{s}$  and  $\perp$ , but in this case the provers should have little control over  $s$ : for any fixed *target string*  $s_o$ , it should be unlikely that  $s = s_o$ . Formally, this is captured as follows; we will show in Section 3.4 that  $\mathcal{CHSH}^n$  is fairly-binding in this sense.

**Definition 3.2 (Fairly binding property).** *A 2-prover commitment scheme  $\mathcal{S}$  is  $\varepsilon$ -fairly-binding if for every commit strategy  $\overline{\text{com}}_{PQ}[\bar{\xi}_{PQ}]$  there exists a function  $\hat{s}(\bar{\xi}_{PQ}, c)$  such that for every opening strategy  $\overline{\text{open}}_{PQ}[\bar{\eta}_{PQ}]$  and all functions  $s_o(\bar{\xi}_{PQ}, \bar{\eta}_{PQ})$  it holds that  $p(s \neq \hat{s}(\bar{\xi}_{PQ}, c) \wedge s = s_o(\bar{\xi}_{PQ}, \bar{\eta}_{PQ})) \leq \varepsilon$ . In short:*

$$\forall \overline{\text{com}}_{PQ} \exists \hat{s}(\bar{\xi}_{PQ}, c) \forall \overline{\text{open}}_{PQ} \forall s_o(\bar{\xi}_{PQ}, \bar{\eta}_{PQ}) : p(s \neq \hat{s} \wedge s = s_o) \leq \varepsilon. \quad (4)$$

*Remark 3.3.* By means of standard techniques, one can easily show that it is sufficient for the (fairly) binding property to consider *deterministic* provers. In this case,  $\hat{s}$  is a function of  $c$  only, and, in the case of fairly-binding,  $s_o$  runs over all *fixed* strings.

*Remark 3.4.* Clearly, the binding property implies the fairly binding property. Furthermore, in the case of *bit* commitment schemes it obviously holds that  $p(b \neq \hat{b} \wedge b \neq \perp) = p(b \neq \hat{b} \wedge b = 0) + p(b \neq \hat{b} \wedge b = 1)$ , and thus the fairly-binding property implies the binding property with a factor-2 loss in the parameter. Furthermore, every fairly-binding *string* commitment scheme gives rise to a binding *bit* commitment scheme in a natural way, as shown by the following proposition.

**Proposition 3.5.** *Let  $\mathcal{S}$  be a  $\varepsilon$ -fairly-binding string commitment scheme. Fix any two distinct strings  $s_0, s_1 \in \{0, 1\}^n$  and consider the bit-commitment scheme  $\mathcal{S}'$  obtained as follows. To commit to  $b \in \{0, 1\}$ , the provers commit to  $s_b$  using  $\mathcal{S}$ , and in the opening phase  $V$  checks if  $s = s_b$  for some bit  $b \in \{0, 1\}$  and outputs this bit if it exists and else outputs  $b = \perp$ . Then,  $\mathcal{S}'$  is a  $2\varepsilon$ -binding bit commitment scheme.*

<sup>4</sup> Recall that by convention (Remark 2.3),  $c$  equals the communication between  $V$  and the provers during the commit phase.

*Proof.* Fix some commit strategy  $\overline{\text{com}}_{PQ}$  for  $\mathcal{S}'$  and note that it can also be used to attack  $\mathcal{S}$ . Thus, there exists a function  $\hat{s}(\bar{\xi}_{PQ}, c)$  as in Definition 3.2. We define

$$\hat{b}(\bar{\xi}_{PQ}, c) = \begin{cases} 0 & \text{if } \hat{s}(\bar{\xi}_{PQ}, c) = s_0 \\ 1 & \text{otherwise} \end{cases}$$

Now fix an opening strategy  $\overline{\text{open}}_{PQ}$  for  $\mathcal{S}'$ , which again is also a strategy against  $\mathcal{S}$ . Thus, we have  $p(\hat{s} \neq s = s_o) \leq \varepsilon$  for any  $s_o$  (and in particular  $s_o = s_0$  or  $s_1$ ). This gives us

$$\begin{aligned} p(\hat{b} \neq b \neq \perp) &= p(\hat{b} = 1 \wedge b = 0) + p(\hat{b} = 0 \wedge b = 1) \\ &= p(\hat{s} \neq s_0 \wedge s = s_0) + p(\hat{s} = s_0 \wedge s = s_1) \\ &\leq p(\hat{s} \neq s_0 \wedge s = s_0) + p(\hat{s} \neq s_1 \wedge s = s_1), \\ &\leq 2\varepsilon \end{aligned}$$

and thus  $\mathcal{S}'$  is a  $2\varepsilon$ -binding bit-commitment scheme.  $\square$

*Remark 3.6.* The proof of Proposition 3.5 generalizes in a straightforward way to  $k$ -bit string commitment schemes: given a  $\varepsilon$ -fairly-binding  $n$ -bit string commitment scheme  $\mathcal{S}$ , for  $k < n$ , we define a  $k$ -bit string commitment scheme  $\mathcal{S}_k$  as follows: to commit to a  $k$ -bit string, the provers pad the string with  $n - k$  zeros and then commit to the padded string using  $\mathcal{S}$ . In the opening phase, the verifier outputs the first  $k$  bits of  $s$  if the remaining bits in  $s$  are all zeros, and  $\perp$  otherwise. Then,  $\mathcal{S}'$  is  $2^k\varepsilon$ -binding.

### 3.2 The Weak Binding Property

Here, we introduce yet another definition for the binding property. It is similar in spirit to Definition 3.1, but weaker. One advantage of this weaker notion is that it is also meaningful when considering quantum attacks, whereas Definition 3.1 is not. In the subsequent section, we will see that for *bit* commitment schemes, this weaker notion of the binding property is equivalent to Definition 2.7.

**Definition 3.7 (Weak binding property).** *A 2-prover commitment scheme  $\mathcal{S}$  is  $\varepsilon$ -weak-binding if for all commit strategies  $\overline{\text{com}}_{PQ}$  there exists a distribution  $p(\hat{s})$  such that for every opening strategy  $\overline{\text{open}}_{PQ}$  (which then fixes the distribution  $p(s)$  of  $V$ 's output  $s$ ) there is a consistent joint distribution  $p(\hat{s}, s)$  such that  $p(s \neq \hat{s} \wedge s \neq \perp) \leq \varepsilon$ . In short:*

$$\forall \overline{\text{com}}_{PQ} \exists p(\hat{s}) \forall \overline{\text{open}}_{PQ} \exists p(\hat{s}, s) : p(s \neq \hat{s} \wedge s \neq \perp) \leq \varepsilon. \quad (5)$$

We also consider a related, i.e., ‘‘fairly’’, version of this binding property, similar to Definition 3.2.

**Definition 3.8 (Fairly weak binding property).** *A 2-prover commitment scheme  $\mathcal{S}$  is  $\varepsilon$ -fairly-weak-binding if for all commit strategies  $\overline{\text{com}}_{PQ}$  there exists a distribution  $p(\hat{s})$  such that for every opening strategy  $\overline{\text{open}}_{PQ}$  (which then fixes the distribution  $p(s)$  of  $V$ 's output  $s$ ) there is a consistent joint distribution  $p(\hat{s}, s)$  so that for all  $s_o \in \{0, 1\}^n$  it holds that  $p(s \neq \hat{s} \wedge s = s_o) \leq \varepsilon$ . In short:*

$$\forall \overline{\text{com}}_{PQ} \exists p(\hat{s}) \forall \overline{\text{open}}_{PQ} \exists p(\hat{s}, s) \forall s_o : p(s \neq \hat{s} \wedge s = s_o) \leq \varepsilon. \quad (6)$$

*Remark 3.9.* Remarks 3.3 and 3.4 also hold for the weak binding properties. Furthermore, it is easy to see that the binding and fairly-binding properties imply their weak counterparts.

**Proposition 3.10.** *Let  $\mathcal{S}$  be a  $\varepsilon$ -fairly-weak-binding string commitment scheme and define  $\mathcal{S}'$  as in Proposition 3.5. Then,  $\mathcal{S}'$  is a  $2\varepsilon$ -weak-binding bit commitment scheme.*

*Proof.* The proof of Proposition 3.5 can be easily adapted: let  $p(\hat{s})$  be as required by Definition 3.8. We define  $p(\hat{b})$  by taking the marginal of  $p(\hat{s}, \hat{b})$  where  $\hat{b} = 0$  if  $\hat{s} = s_0$ , and  $\hat{b} = 1$  otherwise. An opening strategy  $\overline{\text{open}}_{PQ}$  for  $\mathcal{S}'$  can also be viewed as a strategy for  $\mathcal{S}$ . As such, there is a joint distribution  $p(\hat{s}, s)$  as required by Definition 3.7 which we can extend to  $p(\hat{s}, s, b)$  by setting  $b = 0$  if  $s = s_0$ ,  $b = 1$  if  $s = s_1$  and  $b = \perp$  otherwise. We define  $p(\hat{b}, b) := \sum_{\hat{s}, s} p(\hat{b}, \hat{s}) \cdot p(s, b|\hat{s})$ . As in the proof of Proposition 3.5, one can easily check that  $p(\hat{b} \neq b \neq \perp) \leq 2\varepsilon$  holds.

### 3.3 Relations Between The Definitions

Here, we show that in case of *bit* commitment schemes, the weak binding property as introduced in Definition 3.7 above is actually *equivalent* to the  $(p_0 + p_1)$ -definition. Even though our focus is on classical attacks, the proof immediately carries over to quantum attacks as well.

**Theorem 3.11.** *A 2-prover bit-commitment scheme is  $\varepsilon$ -binding in the sense of  $p_0 + p_1 \leq 1 + 2\varepsilon$  if and only if it is  $\varepsilon$ -weak-binding.*

*Proof.* First, consider a scheme that is  $\varepsilon$ -binding according to Definition 2.7. Fix a commit strategy  $\overline{\text{com}}_{PQ}$  and opening strategies  $\overline{\text{open}}_{PQ}^0$  and  $\overline{\text{open}}_{PQ}^1$  so that  $p_0 = p(b_0 = 0)$  and  $p_1 = p(b_1 = 1)$  are maximized, where  $b_i \in \{0, 1, \perp\}$  is  $V$ 's output when the dishonest provers use opening strategy  $\overline{\text{open}}_{PQ}^i$ . Let  $p_0 + p_1 = 1 + 2\varepsilon'$ . Since the scheme is  $\varepsilon$ -binding, we have  $\varepsilon' \leq \varepsilon$ . We define the distribution  $p(\hat{b})$  as  $p(\hat{b} = 0) := p_0 - \varepsilon'$  and  $p(\hat{b} = 1) := p_1 - \varepsilon'$ . To see that this is indeed a probability distribution, note that  $p_0, p_1 \geq 2\varepsilon'$  (otherwise, we would have  $p_0 > 1$  or  $p_1 > 1$ ) and that  $p(\hat{b} = 0) + p(\hat{b} = 1) = p_0 + p_1 - 2\varepsilon' = 1$ . Now we consider an arbitrary opening strategy  $\overline{\text{open}}_{PQ}$  which fixes a distribution  $p(b)$ . By definition of  $p_0$  and  $p_1$ , we have  $p(b = i) \leq p_i$  and thus  $p(b = i) \leq p(\hat{b} = i) + \varepsilon' \leq p(\hat{b} = i) + \varepsilon$ . By Lemma 2.1, there exists a consistent joint distribution  $p(\hat{b}, b)$  with the property that  $p(\hat{b} = b = i) = \min\{p(b = i), p(\hat{b} = i)\}$ . We wish to bound  $p(\hat{b} \neq b \wedge b \neq \perp) = p(\hat{b} = 0 \wedge b = 1) + p(\hat{b} = 1 \wedge b = 0)$ . For  $i \in \{0, 1\}$ , it holds that

$$\begin{aligned} p(\hat{b} = 1 - i \wedge b = i) &= p(b = i) - p(\hat{b} = b = i) \\ &= p(b = i) - \min\{p(\hat{b} = i), p(b = i)\} \\ &= \max\{0, p(b = i) - p(\hat{b} = i)\} \\ &\leq \varepsilon \end{aligned}$$

and furthermore, there is at most *one*  $i \in \{0, 1\}$  such that  $p(b = i) > p(\hat{b} = i)$ , for if  $p(b = i) > p(\hat{b} = i)$  for both  $i = 0$  and  $i = 1$ , then  $p(b = 0) + p(b = 1) > p(\hat{b} = 0) + p(\hat{b} = 1) = 1$  which is a contradiction. Thus, we have  $p(\hat{b} \neq b \wedge b \neq \perp) \leq \varepsilon$ . This proves one direction of our claim.

For the other direction, consider a scheme that is  $\varepsilon$ -binding. Fix  $\overline{\text{com}}_{PQ}$  and let  $p(\hat{b})$  be a distribution such that for every opening strategy  $\overline{\text{open}}_{PQ}$ , there is a joint distribution  $p(\hat{b}, b)$  with  $p(\hat{b} \neq b \neq \perp) \leq \varepsilon$ . Now consider two opening strategies  $\overline{\text{open}}_{PQ}^0$  and  $\overline{\text{open}}_{PQ}^1$  which give distributions  $p(b_0)$  and  $p(b_1)$ . We need to bound  $p(b_0 = 0) + p(b_1 = 1)$ . There is a joint distribution  $p(\hat{b}, b_0)$  such that  $p(\hat{b} \neq b_0 \neq \perp) \leq \varepsilon$  and likewise for  $b_1$ . Thus,

$$\begin{aligned} p(b_0 = 0) + p(b_1 = 1) &= p(\hat{b} = 0, b_0 = 0) + p(\hat{b} = 1, b_0 = 0) + p(\hat{b} = 0, b_1 = 1) + p(\hat{b} = 1, b_1 = 1) \\ &\leq p(\hat{b} = 0) + p(\hat{b} = 1) + p(\hat{b} \neq b_0 \neq \perp) + p(\hat{b} \neq b_1 \neq \perp) \\ &\leq 1 + 2\varepsilon \end{aligned}$$

which proves the other direction. □

*Remark 3.12.* By Remark 3.9, it follows that Definition 3.1 also implies the  $p_0 + p_1$ -definition. In fact, Definition 3.1 is *strictly* stronger (and hence, also strictly stronger than the weak-binding definition). Consider the following (artificial and very non-complete) scheme: in the commit phase,  $V$  chooses a uniformly random bit and sends it to the provers, and then accepts everything or rejects everything during the opening phase, depending on that bit. Then,  $p_0 + p_1 = 1$ , yet a commitment can be opened to  $1 - \hat{b}$  (no matter how  $\hat{b}$  is defined) with probability  $\frac{1}{2}$ .

Since a non-complete separation example may not be fully satisfying, we note that it can be converted into a complete (but even more artificial) scheme. Fix a “good” (i.e., complete, hiding and binding with low parameters) scheme and call our example scheme above the “bad” scheme. We define a *combined* scheme as follows: at the start, the first prover can request either the “good” or “bad” scheme to be used. The honest prover is instructed to choose the former, guaranteeing completeness. The dishonest prover may choose the latter, so the combined scheme inherits the binding properties of the “bad” scheme: it is binding according to the  $(p_0 + p_1)$ -definition, but not according to Definition 3.1.

### 3.4 Security of $\mathcal{CHSH}^n$

In this section, we show that  $\mathcal{CHSH}^n$  is a fairly-binding string commitment scheme.<sup>5</sup> To this end, we introduce yet another version of the binding property and show that  $\mathcal{CHSH}^n$  satisfies this property. Then we show that this version of the binding property implies the fairly-binding property (up to some loss in the parameter, and some mild restrictions on the scheme).

This new binding property is based on the intuition that it should not be possible to open a commitment to two different values *simultaneously* (except with small probability). For this, we observe that (for *classical* attacks), when considering a commit strategy  $\overline{\text{com}}_{PQ}$ , as well as *two* opening strategies  $\overline{\text{open}}_{PQ}$  and  $\overline{\text{open}}'_{PQ}$ , we can run both opening strategies *simultaneously* on the produced commitment with two (independent) copies of  $\text{open}_V$ , by applying  $\overline{\text{open}}_{PQ}$  and  $\overline{\text{open}}'_{PQ}$  to two copies of the respective internal states of  $P$  and  $Q$ ). This gives rise to a *joint* distribution  $p(s, s')$  of the respective outputs  $s$  and  $s'$  of the two copies of  $\text{open}_V$ .

**Definition 3.13 (Simultaneous opening).** *A 2-prover commitment scheme  $\mathcal{S}$  is  $\varepsilon$ -fairly-binding in the sense of simultaneous opening<sup>6</sup> if for all  $\overline{\text{com}}_{PQ}$ , all pairs of opening strategies  $\overline{\text{open}}_{PQ}$  and  $\overline{\text{open}}'_{PQ}$ , and all pairs  $s_o, s'_o$  of distinct strings, we have  $p(s = s_o \wedge s' = s'_o) \leq \varepsilon$ .*

*Remark 3.14.* Also for this notion of fairly-binding, it is sufficient to consider *deterministic* strategies, as can easily be seen.

**Proposition 3.15.** *The string commitment scheme  $\mathcal{CHSH}^n$  is  $2^{-n}$ -fairly-binding in the sense of simultaneous opening.*

*Proof.* By Remark 3.14, it suffices to consider deterministic attack strategies. Fix a deterministic strategy  $\overline{\text{com}}_{PQ}$  and two deterministic opening strategies  $\overline{\text{open}}_{PQ}$  and  $\overline{\text{open}}'_{PQ}$ . The strategy  $\overline{\text{com}}_{PQ}$  specifies  $P$ 's output  $x$  as a function  $f(a)$  of the verifier's message  $a$ . The opening strategies are described by constants  $y$  and  $y'$ . By definition of  $\mathcal{CHSH}^n$ ,  $s = s_o$  implies  $f(a) + y = a \cdot s_o$  and likewise,  $s' = s'_o$  implies  $f(a) + y' = a \cdot s'_o$ . Therefore,  $s = s_o \wedge s' = s'_o$  implies  $a = (y - y') / (s_o - s'_o)$ . It thus holds that  $p(s = s_o \wedge s' = s'_o) \leq p(a = (y - y') / (s_o - s'_o)) \leq \frac{1}{2^n}$ , which proves our claim.  $\square$

*Remark 3.16.* It follows directly from (1) that every *bit* commitment scheme that is  $\varepsilon$ -fairly-binding in the sense of simultaneous opening (against classical attacks) is  $\varepsilon/2$ -binding in the sense of  $p_0 + p_1 \leq 1 + \varepsilon$  (and thus also according to Definitions 3.7). The converse is not true though: the schemes from Remark 3.12 again serve as counterexamples.

**Theorem 3.17.** *Let  $\mathcal{S} = (\text{com}_{PQV}, \text{open}_{PQV})$  be a 2-prover commitment scheme. If  $\mathcal{S}$  is  $\varepsilon$ -fairly-binding in the sense of simultaneous opening and  $\text{open}_V$  is deterministic, then  $\mathcal{S}$  is  $2\sqrt{\varepsilon}$ -fairly-binding.*

*Proof.* By Remark 3.3, it suffices to consider deterministic strategies for the provers. We fix some deterministic commit strategy  $\overline{\text{com}}_{PQ}$  and an enumeration  $\{\overline{\text{open}}^i_{PQ}\}_{i=1}^N$  of all deterministic opening strategies. Since we assume that  $\text{open}_V$  is deterministic, for any fixed opening strategy for the provers, the verifier's output  $s$  is a *function* of the commitment  $c$ . Thus, for each opening strategy  $\overline{\text{open}}^i_{PQ}$  there is a function  $f_i$  such that the verifier's output is  $s = f_i(c)$ . We will now define the function  $\hat{s}(c)$  that satisfies the properties required by Definition 3.2. Our definition depends on a parameter  $\alpha > 0$  which we fix later. To define  $\hat{s}$ , we partition the set  $C$  of all possible commitments into *disjoint* sets  $C = R \cup \bigcup_{s,i} C_{s,i}$  that satisfy the following three properties for every  $i$  and every  $s$ :

$$C_{s,i} \subseteq f_i^{-1}(\{s\}), \quad p(c \in C_{s,i}) \geq \alpha \text{ or } C_{s,i} = \emptyset, \quad \text{and } p(c \in R \wedge f_i(c) = s) < \alpha.$$

The second property implies that there are at most  $\alpha^{-1}$  non-empty sets  $C_{s,i}$ . It is easy to see that such a partitioning exists: start with  $R = C$  and while there exist  $s$  and  $i$  with  $p(c \in R \wedge f_i(c) = s) \geq \alpha$ , let  $C_{s,i} = \{c \in R \mid f_i(c) = s\}$  and remove the elements of  $C_{s,i}$  from  $R$ . For any  $c \in C$ , we now define  $\hat{s}(c)$  as follows. We set  $\hat{s}(c) = s$  for  $c \in C_{s,i}$  and  $\hat{s}(c) = 0$  for  $c \in R$ .

Now fix some opening strategy  $\overline{\text{open}}^i_{PQ}$  and a string  $s_o$ , and write  $s_i$  for the verifier's output. Using  $C_{\neq s_o}$  as a shorthand for  $\bigcup_{s \neq s_o} \bigcup_j C_{s,j}$ , we note that if  $\hat{s}(c) \neq s_o$  then  $c \in R \cup C_{\neq s_o}$ . Thus, it follows that

$$p(s_i \neq \hat{s}(c) \wedge s_i = s_o) = p(\hat{s}(c) \neq s_o \wedge s_i = s_o)$$

<sup>5</sup> It is understood that the allowed attacks against  $\mathcal{CHSH}^n$  are those where the provers do not communicate.

<sup>6</sup> We use "fairly" here to distinguish the notion from a "non-fairly" version with  $p(\perp \neq s \neq s' \neq \perp) \leq \varepsilon$ ; however, we do not consider this latter version any further here.

$$\begin{aligned}
&\leq p(c \in (R \cup C_{\neq s_o}) \wedge f_i(c) = s_o) \\
&= p(c \in R \wedge f_i(c) = s_o) + \sum_{s \neq s_o, j} p(c \in C_{s,j} \wedge f_i(c) = s_o) \\
&\leq p(c \in R \wedge f_i(c) = s_o) + \sum_{\substack{s \neq s_o, j \\ \text{s.t. } C_{s,j} \neq \emptyset}} p(f_j(c) = s \wedge f_i(c) = s_o) \\
&< \alpha + \alpha^{-1} \cdot \varepsilon
\end{aligned}$$

where the final inequality holds because  $p(c \in R \wedge f_i(c) = s_o) < \alpha$  by the choice of  $R$ , because  $p(f_j(c) = s \wedge f_i(c) = s_o) \leq \varepsilon$  by the assumed binding property, and because the number of non-empty  $C_{s,j}$  is at most  $1/\alpha$ . It is easy to see that the upper bound  $\alpha + \alpha^{-1} \cdot \varepsilon$  is minimized by setting  $\alpha = \sqrt{\varepsilon}$ . We conclude that  $p(s_i \neq \hat{s}(c) \wedge s_i = s_o) < 2\sqrt{\varepsilon}$ .  $\square$

Combining Proposition 3.15 and Theorem 3.17, we obtain the following statement for the (fairly-)binding property of the 2-prover string commitment scheme  $\mathcal{CHSH}^n$ .

**Corollary 3.18.**  *$\mathcal{CHSH}^n$  is  $2^{-\frac{n}{2}+1}$ -fairly-binding.*

For the fairly-weak-binding property, we can get a slightly better parameter. Note that we do not require  $\text{open}_V$  to be deterministic here. The proof of the theorem below is given in Appendix C.

**Theorem 3.19.** *Every 2-prover commitment scheme  $\mathcal{S}$  that is  $\varepsilon$ -fairly-binding in the sense of simultaneous opening (against classical attacks) is  $\sqrt{2\varepsilon}$ -fairly-weak-binding (against classical attacks).*

**Corollary 3.20.**  *$\mathcal{CHSH}^n$  is  $2^{-\frac{n-1}{2}}$ -fairly-weak-binding.*

*Remark 3.21.* It is not too hard to see that Corollary 3.20 above implies an upper bound on the classical value  $\omega$  of the game  $\text{CHSH}_{2^n}$  considered in [BS15] of  $\omega(\text{CHSH}_{2^n}) \leq 2^{-\frac{n-1}{2}} + 2^{-n}$ . As such, Theorem 1.3 in [BS15] implies that the above  $\varepsilon$  is asymptotically optimal for odd  $n$ , i.e., the square root loss to the binding property of the *bit* commitment version is unavoidable (for odd  $n$ ).

As for security against quantum attacks, we point out that [BS15, RAM15] provide an upper bound on the quantum value  $\omega^*(\text{CHSH}_q)$  of general finite-field CHSH; however, this does not directly imply security against quantum attacks of  $\mathcal{CHSH}^n$  as a (fairly-weak-binding) string commitment scheme.

## 4 Composing Commitment Schemes

### 4.1 The Composition Operation

We consider two 2-prover commitment schemes  $\mathcal{S}$  and  $\mathcal{S}'$  of a restricted form, and we compose them to a new 2-prover commitment scheme  $\mathcal{S}'' = \mathcal{S} \star \mathcal{S}'$  in a well-defined way; our composition theorem then shows that  $\mathcal{S}''$  is secure (against classical attacks) if  $\mathcal{S}$  and  $\mathcal{S}'$  are. We start by specifying the restriction to  $\mathcal{S}$  and  $\mathcal{S}'$  that we impose.

**Definition 4.1.** *Let  $\mathcal{S}$  and  $\mathcal{S}'$  be two 2-prover string commitment schemes. We call the pair  $(\mathcal{S}, \mathcal{S}')$  eligible if the following three properties hold, or they hold with the roles of  $P$  and  $Q$  exchanged.*

1. *The commit phase of  $\mathcal{S}$  is a protocol  $\text{com}_{PV} = (\text{com}_P, \text{com}_V)$  between  $P$  and  $V$  only, and the opening phase of  $\mathcal{S}$  is a protocol  $\text{open}_{QV} = (\text{open}_Q, \text{open}_V)$  between  $Q$  and  $V$  only. In other words,  $\text{com}_Q$  and  $\text{open}_P$  are both trivial and do nothing.<sup>7</sup> Similarly, the commit phase of  $\mathcal{S}'$  is a protocol  $\text{com}'_{QV}$  between  $Q$  and  $V$  only (but both provers may be active in the opening phase).*
2. *The opening phase  $\text{open}_{QV}$  of  $\mathcal{S}$  is of the following simple form:  $Q$  sends a bit string  $y \in \{0, 1\}^m$  to  $V$ , and  $V$  computes  $s$  deterministically as  $s = \text{Extr}(y, c)$ , where  $c$  is the commitment.<sup>8</sup>*
3. *The domain of  $\mathcal{S}'$  contains (or equals)  $\{0, 1\}^m$ .*

<sup>7</sup> Except that  $\text{com}_Q$  may output state information to the opening protocol  $\text{open}_Q$ , e.g., in order to pass on the commit phase randomness.

<sup>8</sup> Our composition theorem also works for a randomized  $\text{Extr}$ , but for simplicity, we restrict to the deterministic case.

Furthermore, we specify that the allowed attacks on  $\mathcal{S}$  are so that  $P$  and  $Q$  do not communicate during the course of the entire scheme, and the allowed attacks on  $\mathcal{S}'$  are so that  $P$  and  $Q$  do not communicate during the course of the commit phase but there may be limited communication during the opening phase.

An example of an eligible pair of 2-prover commitments is the pair  $(\mathcal{CHSH}^n, \mathcal{XCHSH}^n)$ , where  $\mathcal{XCHSH}^n$  coincides with scheme  $\mathcal{CHSH}^n$  except that the roles of  $P$  and  $Q$  are exchanged.

*Remark 4.2.* For an eligible pair  $(\mathcal{S}, \mathcal{S}')$ , it will be convenient to understand  $\text{open}_Q$  and  $\text{open}_V$  as *non-interactive* algorithms, where  $\text{open}_Q$  produces  $y$  as its *output*, and  $\text{open}_V$  takes  $y$  as additional *input* (rather than viewing the pair as a protocol with a single one-way communication round).

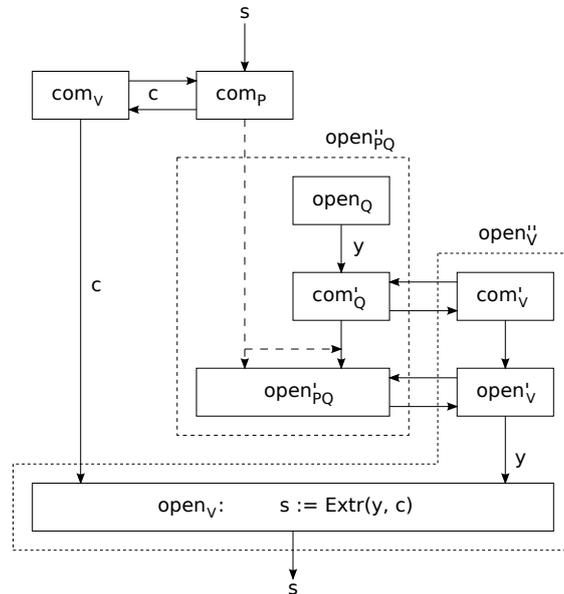
We now define the composition operation. Informally, committing is done by means of committing using  $\mathcal{S}$ , and to open the commitment,  $Q$  uses  $\text{open}_Q$  to locally compute the opening information  $y$  and he commits to  $y$  with respect to the scheme  $\mathcal{S}'$ , and then this commitment is opened (to  $y$ ), and  $V$  computes and outputs  $s = \text{Extr}(y, c)$ . Formally, this is captured as follows (see also Figure 2).

**Definition 4.3.** Let  $\mathcal{S} = (\text{com}_{PV}, \text{open}_{QV})$  and  $\mathcal{S}' = (\text{com}'_{QV}, \text{open}'_{PQV})$  be an eligible pair of 2-prover commitment schemes. Then, their composition  $\mathcal{S} \star \mathcal{S}'$  is defined as the 2-prover commitment scheme consisting of  $\text{com}_{PV} = (\text{com}_P[\xi_{PQ}], \text{com}_V)$  and

$$\text{open}'_{PQV} = (\text{open}'_P, \text{open}'_Q \circ \text{com}'_Q \circ \text{open}_Q, \text{open}_V \circ \text{open}'_V \circ \text{com}'_V),$$

where we make it explicit that  $\text{com}_P$  and  $\text{open}_Q$  use joint randomness, and so do  $\text{com}'_Q$  and  $\text{open}'_P$ .

When considering attacks against the binding property of the composed scheme  $\mathcal{S} \star \mathcal{S}'$ , we declare that the allowed deterministic attacks<sup>9</sup> are those of the form  $(\overline{\text{com}}_P, \overline{\text{open}}'_{PQ} \circ \text{ptoq}_{PQ} \circ \overline{\text{com}}'_Q)$ , where  $\overline{\text{com}}_P$  is an allowed deterministic commit strategy for  $\mathcal{S}$ ,  $\overline{\text{com}}'_Q$  and  $\overline{\text{open}}'_{PQ}$  are allowed deterministic commit and opening strategies for  $\mathcal{S}'$ , and  $\text{ptoq}_{PQ}$  is the one-way communication protocol that communicates  $P$ 's input to  $Q$  (see also Figure 3).<sup>10</sup>



**Fig. 2.** The composition of  $\mathcal{S}$  and  $\mathcal{S}'$  (assuming single-round commit phases). The dotted arrows indicate communication allowed to the dishonest provers.

<sup>9</sup> The allowed *randomized* attacks are then naturally given as those that pick one of the deterministic attacks according to some distribution.

<sup>10</sup> This one-way communication models that in the relativistic setting, sufficient time has passed at this point for  $P$  to inform  $Q$  about what happened during  $\text{com}_P$ .

*Remark 4.4.* It is immediate that  $\mathcal{S} \star \mathcal{S}'$  is a commitment scheme in the sense of Definition 2.2, and that it is complete if  $\mathcal{S}$  and  $\mathcal{S}'$  are, with the error parameters adding up. Also, the hiding property is obviously inherited from  $\mathcal{S}$ ; however, the point of the composition is to keep the hiding property alive for longer, namely up to before the last round of the opening phase—recall that, using the terminology used in context of relativistic commitments, these rounds of the opening phase up to before the last would then be referred to as the *sustain phase*. We show in Appendix D that  $\mathcal{S} \star \mathcal{S}'$  is hiding up to before the last round, with the error parameters adding up.

It is intuitively clear that  $\mathcal{S} \star \mathcal{S}'$  *should be* binding if  $\mathcal{S}$  and  $\mathcal{S}'$  are: committing to the opening information  $y$  and then opening the commitment allows the provers to *delay* the announcement of  $y$  (which is the whole point of the exercise), but it does not allow them to *change*  $y$ , by the binding property of  $\mathcal{S}'$ ; thus,  $\mathcal{S} \star \mathcal{S}'$  should be (almost) as binding as  $\mathcal{S}$ . This intuition is confirmed by our composition theorem below.

*Remark 4.5.* We point out that the composition  $\mathcal{S} \star \mathcal{S}'$  can be naturally defined for a *larger* class of pairs of schemes (e.g. where *both* provers are active in the commit phase of both schemes), and the above intuition still holds. However, our proof only works for this restricted class of (pairs of) schemes. Extending the composition result in that direction is an open problem.

*Remark 4.6.* We observe that if  $(\mathcal{S}, \mathcal{X}\mathcal{S})$  is an eligible pair, where  $\mathcal{X}\mathcal{S}$  coincides with  $\mathcal{S}$  except that the roles of  $P$  and  $Q$  are exchanged, then so is  $(\mathcal{X}\mathcal{S}, \mathcal{S} \star \mathcal{X}\mathcal{S})$ . As such, we can then compose  $\mathcal{X}\mathcal{S}$  with  $\mathcal{S} \star \mathcal{X}\mathcal{S}$ , and obtain yet another eligible pair  $(\mathcal{S}, \mathcal{X}\mathcal{S} \star \mathcal{S} \star \mathcal{X}\mathcal{S})$ , etc. Applying this to the schemes  $\mathcal{S} = \mathcal{CHSH}^n$ , we obtain the multi-round scheme from Lunghi *et al.* [LKB<sup>+</sup>15]. As such, our composition theorem below implies security of their scheme—with a *linear* blow-up of the error term (instead of double exponential).

We point out that formally we obtain security of the Lunghi *et al.* scheme as a *2-prover commitment scheme* under an *abstract restriction* on the provers' communication: in every round, the active prover cannot access the message that the other prover received in the previous round. As such, when the rounds of the protocol are executed fast enough so that it is ensured that there is no time for the provers to communicate between subsequent rounds, then security as a *relativistic commitment scheme* follows immediately.

Before stating and proving the composition theorem, we need to single out one more relevant parameter.

**Definition 4.7.** *Let  $(\mathcal{S}, \mathcal{S}')$  be an eligible pair, which in particular means that  $V$ 's action in the opening phase of  $\mathcal{S}$  is determined by a function  $\text{Extr}$ . We define  $k(\mathcal{S}) := \max_{c,s} |\{y \mid \text{Extr}(y, c) = s\}|$ .*

I.e.,  $k(\mathcal{S})$  counts the number of  $y$ 's that are consistent with a given string  $s$  (in the worst case). Note that  $k(\mathcal{CHSH}^n) = 1$ : for every  $a, x, s \in \{0, 1\}^n$  there is at most one  $y \in \{0, 1\}^n$  such that  $x + y = a \cdot s$ .

## 4.2 The Composition Theorems

In the following composition theorems, we take it as understood that the assumed respective binding properties of  $\mathcal{S}$  and  $\mathcal{S}'$  hold with respect to a well-defined respective classes of allowed attacks. We start with the composition theorem for the fairly-binding property, which is easier to prove than the one for the fairly-weak-binding property.

**Theorem 4.8.** *Let  $(\mathcal{S}, \mathcal{S}')$  be an eligible pair of 2-prover commitment schemes, and assume that  $\mathcal{S}$  and  $\mathcal{S}'$  are respectively  $\varepsilon$ -fairly-binding and  $\delta$ -fairly-binding. Then, their composition  $\mathcal{S}'' = \mathcal{S} \star \mathcal{S}'$  is  $(\varepsilon + k(\mathcal{S}) \cdot \delta)$ -fairly-binding.*

*Proof.* We first consider the case  $k(\mathcal{S}) = 1$ . We fix an attack  $(\overline{\text{com}}_P, \overline{\text{open}}''_{PQ})$  against  $\mathcal{S}''$ . Without loss of generality, the attack is deterministic, so  $\overline{\text{open}}''_{PQ}$  is of the form  $\overline{\text{open}}''_{PQ} = \overline{\text{open}}'_{PQ} \circ \text{ptoq}_{PQ} \circ \overline{\text{com}}'_Q$ .

Note that  $\overline{\text{com}}_P$  is also a commit strategy for  $\mathcal{S}$ . As such, by the fairly-binding property of  $\mathcal{S}$ , there exists a function  $\hat{s}(c)$ , only depending on  $\overline{\text{com}}_P$ , so that the property specified in Definition 3.2 is satisfied for every opening strategy  $\overline{\text{open}}_Q$  for  $\mathcal{S}$ . We will show that it is also satisfied for the (arbitrary) opening strategy  $\overline{\text{open}}''_{PQ}$  for  $\mathcal{S}''$ , except for a small increase in  $\varepsilon$ : we will show that  $p(\hat{s}(c) \neq s \wedge s = s_o) \leq \varepsilon + \delta$  for every fixed target string  $s_o$ . This then proves the claim.

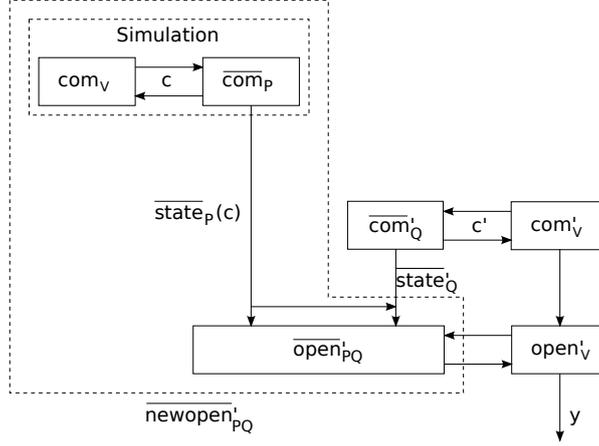
To show this property on  $\hat{s}(c)$ , we “decompose and reassemble” the attack strategy  $(\overline{\text{com}}_P, \overline{\text{open}}''_{PQ} \circ \text{ptoq}_{PQ} \circ \overline{\text{com}}'_Q)$  for  $\mathcal{S}''$  into an attack strategy  $(\overline{\text{com}}'_Q, \overline{\text{newopen}}'_{PQ})$  for  $\mathcal{S}'$  with  $\overline{\text{newopen}}'_{PQ}$  formally defined as

$$\overline{\text{newopen}}'_{PQ}[c](\overline{\text{state}}'_Q) := \overline{\text{open}}'_{PQ}(\overline{\text{state}}_P(c) \parallel (\overline{\text{state}}_P(c), \overline{\text{state}}'_Q))$$

where

$$(\overline{\text{state}}_P(c) || c) \leftarrow (\overline{\text{com}}_P || \text{com}_V).$$

Informally, this means that ahead of time,  $P$  and  $Q$  *simulate* an execution of  $(\overline{\text{com}}_P(\emptyset) || \text{com}_V(\emptyset))$  and take the resulting communication/commitment<sup>11</sup>  $c$  as shared randomness, and then  $\overline{\text{newopen}}'_{PQ}$  computes  $\overline{\text{state}}_P$  from  $c$  as does  $\overline{\text{com}}_P$ , and runs  $\overline{\text{open}}'_{PQ}$  (see Figure 3).<sup>12</sup> It follows from the fairly-binding property that there is a function  $\hat{y}(c')$  of the commitment  $c'$  so that  $p(\hat{y}(c') \neq y \wedge y = y_o(c)) \leq \delta$  for every function  $y_o(c)$ .



**Fig. 3.** Constructing the opening strategy  $\overline{\text{newopen}}'_{PQ}$  against  $\mathcal{S}'$ .

The existence of  $\hat{y}$  now gives rise to an opening strategy  $\overline{\text{open}}_Q$  for  $\mathcal{S}$ ; namely, simulate the commit phase of  $\mathcal{S}'$  to obtain the commitment  $c'$ , and output  $\hat{y}(c')$ . By Definition 3.2, for  $\tilde{s} := \text{Extr}(\hat{y}(c'), c)$  and every  $s_o$ ,  $p(\hat{s}(c) \neq \tilde{s} \wedge \tilde{s} = s_o) \leq \varepsilon$ .

We are now ready to put things together. Fix an arbitrary target string  $s_o$ . For any  $c$  we let  $y_o(c)$  be the unique string such that  $\text{Extr}(y_o(c), c) = s_o$  (and some default string if no such string exists); recall, we assume for the moment that  $k(\mathcal{S}) = 1$ . Omitting the arguments in  $\hat{s}(c)$ ,  $\hat{y}(c')$  and  $y_o(c)$ , it follows that

$$\begin{aligned} p(\hat{s} \neq s \wedge s = s_o) &\leq p(\hat{s} \neq s \wedge s = s_o \wedge s = \tilde{s}) + p(s = s_o \wedge s \neq \tilde{s}) \\ &\leq p(\hat{s} \neq \tilde{s} \wedge \tilde{s} = s_o) + p(\text{Extr}(y, c) \neq \text{Extr}(\hat{y}, c) \wedge \text{Extr}(y, c) = s_o) \\ &\leq p(\hat{s} \neq \tilde{s} \wedge \tilde{s} = s_o) + p(y \neq \hat{y} \wedge y = y_o) \\ &\leq \varepsilon + \delta. \end{aligned}$$

Thus,  $\hat{s}$  is as required.

For the general case where  $k(\mathcal{S}) > 1$ , we can reason similarly, except that we then list the  $k \leq k(\mathcal{S})$  possibilities  $y_o^1(c), \dots, y_o^k(c)$  for  $y_o(c)$ , and conclude that  $p(s \neq \tilde{s} \wedge s = s_o) \leq \sum_i p(y \neq \hat{y} \wedge y = y_o^i) \leq k(\mathcal{S}) \cdot \delta$ , which then results in the claimed bound.  $\square$

*Remark 4.9.* Putting things together, we can now conclude the security (i.e., the binding property) of the Lunghi *et al.* multi-round commitment scheme. Corollary 3.18 ensures the fairly-binding property of  $\mathcal{CHSH}^n$ , i.e., the Crépeau *et al.* scheme as a string commitment scheme, with parameter  $2^{-n/2+1}$ . The composition theorem (Theorem 4.8) then guarantees the fairly-binding property of the  $m$ -fold composition as a string commitment scheme, with parameter  $(m+1) \cdot 2^{-n/2+1}$ . Finally, Proposition 3.5 implies that the  $m$ -fold composition of  $\mathcal{CHSH}^n$  with itself is a  $\varepsilon_m$ -binding bit commitment scheme with error parameter  $\varepsilon_m = (m+1) \cdot 2^{-n/2+2}$  as claimed in the introduction, or, more generally, and by taking Remark 3.6 into account, a  $(m+1) \cdot 2^{-n/2+k+1}$ -binding  $k$ -bit-string commitment scheme.

<sup>11</sup> Recall that by convention (Remark 2.3), the commitment  $c$  equals the communication between  $V$  and, here,  $P$ .

<sup>12</sup> We are using here that  $Q$  is inactive during  $\overline{\text{com}}_{PQ}$  and  $P$  during  $\overline{\text{com}}'_{PQ}$ , and thus the two “commute”.

For completeness, we also show the composition theorem for the weak version of the binding property. Since this notion makes sense also against quantum attacks, we emphasize the restriction to classical attacks — extending the theorem to quantum attacks is an open problem.

**Theorem 4.10.** *Let  $(\mathcal{S}, \mathcal{S}')$  be an eligible pair of 2-prover commitment schemes, and assume that  $\mathcal{S}$  and  $\mathcal{S}'$  are respectively  $\varepsilon$ -fairly-weak-binding and  $\delta$ -fairly-weak-binding against classical attacks. Then, their composition  $\mathcal{S}'' = \mathcal{S} \star \mathcal{S}'$  is a  $(\varepsilon + k(\mathcal{S}) \cdot \delta)$ -fairly-weak-binding 2-prover commitment scheme against classical attacks.*

*Proof.* We first consider the case  $k(\mathcal{S}) = 1$ . We fix an arbitrary deterministic attack  $(\overline{\text{com}}_P, \overline{\text{open}}''_{PQ})$  against  $\mathcal{S}''$ , where  $\overline{\text{open}}''_{PQ}$  is of the form  $\overline{\text{open}}''_{PQ} = \overline{\text{open}}'_{PQ} \circ \text{ptoq}_{PQ} \circ \overline{\text{com}}'_Q$ . Let  $a$  be  $V$ 's randomness in  $\text{com}_V$ . Then,  $c$  is a function  $c(a)$  of  $a$ , and the distribution  $p(a, y)$  is well defined. Since  $\overline{\text{com}}_P$  is also an attack strategy against  $\mathcal{S}$ , there exists a distribution  $p(\hat{s})$  (only depending on  $\overline{\text{com}}_P$ ) such that Definition 3.8 is satisfied for every opening strategy  $\overline{\text{open}}_Q$  for  $\mathcal{S}$ .

Similar to the proof of Theorem 4.8, we reassemble the attack strategy  $(\overline{\text{com}}_P, \overline{\text{open}}'_{PQ} \circ \text{ptoq}_{PQ} \circ \overline{\text{com}}'_Q)$  for  $\mathcal{S}''$  into an attack strategy  $(\overline{\text{com}}'_Q, \overline{\text{newopen}}'_{PQ})$  for  $\mathcal{S}'$ . Concretely, for every fixed choice of  $a$ , we obtain a deterministic opening strategy  $\overline{\text{newopen}}'_{PQ,a}$  given by

$$\overline{\text{newopen}}'_{PQ,a}(\overline{\text{state}}'_Q) := \overline{\text{open}}'_{PQ}(\overline{\text{state}}_P(c(a)) \| (\overline{\text{state}}_P(c(a)), \overline{\text{state}}'_Q)),$$

and the distribution of the verifier's output  $y$  when the provers use  $\overline{\text{newopen}}'_{PQ,a}$  is  $p(y|a)$ . It follows from the fairly-weak-binding property of  $\mathcal{S}'$  that there exists a distribution  $p(\hat{y})$ , only depending on  $\overline{\text{com}}'_Q$ , so that for every choice of  $a$  there exists a consistent joint distribution  $p(\hat{y}, y|a)$  so that  $p(\hat{y} \neq y \wedge y = y_\circ | a) \leq \delta$  for every fixed target string  $y_\circ$ . Note that here, consistency in particular means that  $p(\hat{y}|a) = p(\hat{y})$ . This joint conditional distribution  $p(\hat{y}, y|a)$  together with the distribution  $p(a)$  of  $a$  then naturally defines the distribution  $p(a, \hat{y}, y)$ , which is consistent with  $p(a, y)$  considered above.

The existence of  $p(\hat{y})$  now gives rise to an opening strategy  $\overline{\text{open}}_Q$  for  $\mathcal{S}$ ; namely, sample  $\hat{y}$  according to  $p(\hat{y})$  and output  $\hat{y}$ . Note that the joint distribution of  $a$  and  $\hat{y}$  in this “experiment” is given by

$$p(a) \cdot p(\hat{y}) = p(a) \cdot p(\hat{y}|a) = p(a, \hat{y}),$$

i.e., is consistent with the distribution  $p(a, \hat{y}, y)$  above. By Definition 3.8, we know there exists a joint distribution  $p(\hat{s}, \tilde{s})$ , consistent with  $p(\hat{s})$  fixed above and with  $p(\tilde{s})$  determined by  $\tilde{s} := \text{Extr}(\hat{y}, c(a))$ , and such that  $p(\hat{s} \neq \tilde{s} \wedge \tilde{s} = s_\circ) \leq \varepsilon$  for every  $s_\circ$ . We can now “glue together”  $p(\hat{s}, \tilde{s})$  and  $p(c, \hat{y}, y, \tilde{s})$ , i.e., find a joint distribution that is consistent with both, by setting

$$p(a, \hat{y}, y, \tilde{s}, \hat{s}) := p(a, \hat{y}, y, \tilde{s}) \cdot p(\hat{s}|\tilde{s}).$$

We now fix an arbitrary target string  $s_\circ$ . Furthermore, for any  $a$  we let  $y_\circ(a)$  be the unique string such that  $\text{Extr}(y_\circ(a), c(a)) = s_\circ$  (and to some default string if no such string exists); recall, we assume for the moment that  $k(\mathcal{S}) = 1$ . With respect to the above joint distribution, it then holds that

$$\begin{aligned} p(\hat{s} \neq s \wedge s = s_\circ) &= p(\hat{s} \neq s \wedge s = s_\circ \wedge s = \tilde{s}) + p(s = s_\circ \wedge s = s_\circ \wedge s \neq \tilde{s}) \\ &\leq p(\hat{s} \neq \tilde{s} \wedge s = s_\circ \wedge s = \tilde{s}) + p(s \neq \tilde{s} \wedge s = s_\circ) \\ &\leq p(\hat{s} \neq \tilde{s} \wedge \tilde{s} = s_\circ) + p(\text{Extr}(y, c(a)) \neq \text{Extr}(\hat{y}, c(a)) \wedge \text{Extr}(y, c(a)) = s_\circ) \\ &\leq p(\hat{s} \neq \tilde{s} \wedge \tilde{s} = s_\circ) + p(y \neq \hat{y} \wedge y = y_\circ(a)) \\ &\leq p(\hat{s} \neq \tilde{s} \wedge \tilde{s} = s_\circ) + \sum_a p(a) \cdot p(y \neq \hat{y} \wedge y = y_\circ(a)|a) \\ &\leq \varepsilon + \delta. \end{aligned}$$

Thus, the distribution  $p(\hat{s}, s)$  is as required.

For the case where  $k(\mathcal{S}) > 1$ , we can reason similarly, except that we then list the  $k \leq k(\mathcal{S})$  possibilities  $y_\circ^1(a), \dots, y_\circ^k(a)$  for  $y_\circ(a)$ , and conclude that  $p(s \neq \tilde{s} \wedge s = s_\circ) \leq \sum_i p(y \neq \hat{y} \wedge y = y_\circ^i(a)) \leq k(\mathcal{S}) \cdot \delta$ , which then results in the claimed bound.  $\square$

*Remark 4.11.* Analogously to Remark 4.9, we can conclude from Corollary 3.20 and Theorem 4.10 that  $\mathcal{CHSH}^n$  is  $(m+1) \cdot 2^{-(n-1)/2}$ -fairly-weak-binding. It follows from Proposition 3.10 that  $\mathcal{CHSH}^n$  is a  $(m+1) \cdot 2^{-(n+1)/2}$ -weak-binding bit-commitment scheme. More generally, we can conclude that for any  $k < n$ , it is a  $(m+1) \cdot 2^{-(n-1)/2+k}$ -weak-binding  $k$ -bit string commitment scheme. Below, we show how to avoid the factor 2 introduced by invoking Proposition 3.10.

### 4.3 Variations

In this section, we show two variants of the composition theorems. The first one says that if we compose a weak-binding with a fairly-weak-binding scheme, we obtain a weak-binding scheme. This allows us to slightly improve the parameter in Remark 4.11. The proof crucially relies on the fact that, in the weak definition, there is some freedom in “gluing together” the distributions  $p(s)$  and  $p(\hat{s})$ . The second variant says that composing two binding (or weak-binding) schemes yields a binding (or weak-binding, respectively) scheme.

We start by proving the following two properties for fairly-weak-binding commitment schemes. The first property shows that one may assume the joint distribution  $p(\hat{s}, s)$  to be such that  $s$  and  $\hat{s}$  are independent conditioned on  $s \neq \hat{s}$ .

**Lemma 4.12.** *Let  $\mathcal{S}$  be a  $\varepsilon$ -fairly-weak-binding commitment scheme. Then, for any  $\overline{\text{com}}_{PQ}$  and  $\overline{\text{open}}_{PQ}$  there exists a joint distribution  $p(\hat{s}, s)$  as required by Definition 3.8, but with the additional property that*

$$p(\hat{s}, s | s \neq \hat{s}) = p(\hat{s} | s \neq \hat{s}) \cdot p(s | s \neq \hat{s}).$$

*Proof.* Since the scheme is  $\varepsilon$ -fairly-weak-binding, it follows that there exists a consistent joint distribution  $p(\hat{s}, s)$  such that  $p(s \neq \hat{s} \wedge s = s_o) \leq \varepsilon$  for every  $s_o$ . Because of this, we have

$$p(s = s_o) = p(s = s_o \wedge \hat{s} = s_o) + p(s = s_o \wedge \hat{s} \neq s_o) = p(s = s_o \wedge \hat{s} = s_o) + p(s \neq \hat{s} \wedge s = s_o) \leq p(\hat{s} = s_o) + \varepsilon.$$

We apply Lemma 2.1 to the marginal distributions  $p(\hat{s})$  and  $p(s)$ . The resulting joint distribution  $\tilde{p}(\hat{s}, s)$  satisfies  $\tilde{p}(\hat{s} = s_o \wedge s = s_o | s = \hat{s}) = \min\{p(s = s_o), p(\hat{s} = s_o)\}$  and  $\tilde{p}(\hat{s}, s | s \neq \hat{s}) = \tilde{p}(\hat{s} | s \neq \hat{s}) \cdot \tilde{p}(s | s \neq \hat{s})$ . It remains to show that  $\tilde{p}(s \neq \hat{s} \wedge s = s_o) \leq \varepsilon$  for all  $s_o$ . Indeed, we have

$$\begin{aligned} \tilde{p}(s \neq \hat{s} \wedge s = s_o) &= \tilde{p}(s = s_o) - \tilde{p}(s = \hat{s} \wedge s = s_o) \\ &= \tilde{p}(s = s_o) - \tilde{p}(\hat{s} = s_o \wedge s = s_o) \\ &= p(s = s_o) - \min\{p(\hat{s} = s_o), p(s = s_o)\} \\ &\leq p(s = s_o) - (p(s = s_o) - \varepsilon) \\ &= \varepsilon \end{aligned}$$

as claimed.  $\square$

The second property shows that the quantification over all *fixed*  $s_o$  in Definition 3.8 of the fairly-weak-binding property can be relaxed to  $s_o$  that may depend on  $\hat{s}$ , but only on  $\hat{s}$ . Note that we can obviously not allow  $s_o$  to depend (arbitrarily) on  $s$ , since then one could choose  $s_o = s$ .

**Proposition 4.13.** *Let  $\mathcal{S}$  be a  $\varepsilon$ -fairly-weak-binding commitment scheme. Then*

$$\forall \overline{\text{com}}_{PQ} \exists p(\hat{s}) \forall \overline{\text{open}}_{PQ} \exists p(\hat{s}, s) \forall p(s_o | \hat{s}) : p(s \neq \hat{s} \wedge s = s_o) \leq \varepsilon,$$

where it is understood that  $p(\hat{s}, s, s_o) := p(\hat{s}, s) \cdot p(s_o | \hat{s})$ . Thus, the joint distribution  $p(\hat{s}, s)$  is such that  $p(s \neq \hat{s} \wedge s = s_o) \leq \varepsilon$  holds in particular for any function  $s_o = f(\hat{s})$  of  $\hat{s}$ .

*Proof.* For given  $\overline{\text{com}}_{PQ}$  and  $\overline{\text{open}}_{PQ}$ , let  $p(\hat{s}, s)$  be as guaranteed by the fairly-weak-binding property. By Lemma 4.12, we may assume without loss of generality that  $p(\hat{s}, s | s \neq \hat{s}) = p(\hat{s} | s \neq \hat{s}) p(s | s \neq \hat{s})$ . Then, by Lemma B.1, we also have that  $p(s, s_o | s \neq \hat{s}) = p(s | s \neq \hat{s}) p(s_o | s \neq \hat{s})$ . It follows that

$$\begin{aligned} p(s \neq \hat{s} \wedge s = s_o) &= p(s \neq \hat{s}) \cdot p(s = s_o | s \neq \hat{s}) \\ &= p(s \neq \hat{s}) \sum_{s_o^*} p(s = s_o^* \wedge s_o = s_o^* | s \neq \hat{s}) \\ &= p(s \neq \hat{s}) \sum_{s_o^*} p(s = s_o^* | s \neq \hat{s}) \cdot p(s_o = s_o^* | s \neq \hat{s}) \\ &= \sum_{s_o^*} p(s \neq \hat{s} \wedge s = s_o^*) \cdot p(s_o = s_o^* | s \neq \hat{s}) \\ &\leq \varepsilon \cdot \sum_{s_o^*} p(s_o = s_o^* | s \neq \hat{s}) \\ &= \varepsilon \end{aligned}$$

where the inequality follows from the fact that  $p(s \neq \hat{s} \wedge s = s_o^*) \leq \varepsilon$  for every *fixed*  $s_o^*$ .  $\square$

For the rest of the section, we take it as understood that we only consider classical attacks.

**Theorem 4.14.** *Let  $(\mathcal{S}, \mathcal{S}')$  be an eligible pair of 2-prover commitment schemes, where  $\mathcal{S}$  is  $\varepsilon$ -weak-binding and  $\mathcal{S}'$  is  $\delta$ -fairly-weak-binding, and let  $\{0, 1\}^m$  be the domain of  $\mathcal{S}$ . Then, the composition  $S \star \mathcal{S}'$  is a  $(\varepsilon + (2^m - 1) \cdot k(\mathcal{S}) \cdot \delta)$ -weak-binding commitment scheme.*

*In particular, if  $\mathcal{S}$  is a bit commitment scheme then  $S \star \mathcal{S}'$  is a  $(\varepsilon + k(\mathcal{S}) \cdot \delta)$ -weak-binding.*

*Proof.* We follow the proof of Theorem 4.10, up to when it comes to choosing  $y_\circ$ . Let us first consider the case  $m = 1$ , i.e.,  $\mathcal{S}$  is a *bit* commitment scheme. In that case, and assuming for the moment that  $k(\mathcal{S}) = 1$ , we let  $y_\circ$  be the unique string that satisfies  $\text{Extr}(y_\circ, c) = s_\circ$ , but where now  $s_\circ := 1 - \tilde{s}$ . We emphasize that for a fixed  $c$ , this choice of  $y_\circ$  is *not* fixed anymore (in contrast to the choice in the proof of Theorem 4.10); namely, it is a function of  $\tilde{s} = \text{Extr}(\hat{y}, c)$ , which in turn is a function of  $\hat{y}$ . Therefore, by Proposition 4.13, it still holds that  $p(y \neq \hat{y} \wedge y = y_\circ | a) \leq \delta$ , and we can conclude that

$$\begin{aligned} p(\hat{s} \neq s \wedge s \neq \perp) &\leq p(\hat{s} \neq s \wedge s \neq \perp \wedge s = \tilde{s}) + p(s \neq \tilde{s} \wedge s \neq \perp) \\ &= p(\hat{s} \neq \tilde{s} \wedge s \neq \perp \wedge s = \tilde{s}) + p(s \neq \tilde{s} \wedge s = 1 - \tilde{s}) \\ &\leq p(\hat{s} \neq \tilde{s} \wedge \tilde{s} \neq \perp) + p(y \neq \hat{y} \wedge y = y_\circ) \\ &\leq p(\hat{s} \neq \tilde{s} \wedge \tilde{s} \neq \perp) + \sum_a p(a) p(y \neq \hat{y} \wedge y = y_\circ | a) \\ &\leq \varepsilon + \sum_a p(a) \delta \\ &= \varepsilon + \delta. \end{aligned}$$

In the case that  $k(\mathcal{S}) > 1$ , we instead randomly select one of the at most  $k(\mathcal{S})$  strings  $y_\circ$  that satisfy  $\text{Extr}(y_\circ, c) = s_\circ = 1 - \tilde{s}$ . Then, conditioned on  $a$ ,  $y_\circ$  is still independent of  $y$  given  $\hat{y}$ , so that Proposition 4.13 still applies, and we can argue as above, except that we get a factor  $k(\mathcal{S})$  blow-up from  $p(s \neq \tilde{s} \wedge s = 1 - \tilde{s}) \leq k(\mathcal{S}) \cdot p(y \neq \hat{y} \wedge y = y_\circ)$ .

Finally, for the case  $m > 1$ , we first pick a random  $s_\circ \in \{0, 1\}^m \setminus \{\tilde{s}\}$ , and then choose  $y_\circ$  such that  $\text{Extr}(y_\circ, c) = s_\circ$ , uniquely or at random, depending of  $k(\mathcal{S})$ . Conditioned on  $a$ ,  $y_\circ$  is still independent of  $y$  given  $\hat{y}$ , and therefore Proposition 4.13 still applies, but now we get an additional factor  $(2^m - 1)$  blow-up from  $p(s \neq \tilde{s} \wedge s \neq \perp) \leq (2^m - 1) p(s \neq \tilde{s} \wedge s = s_\circ)$ .  $\square$

*Remark 4.15.* Theorem 4.14 allows us to slightly improve the bound we obtain in Remark 4.11 on the Lunghi *et al.* multi-round commitment scheme. By Theorem 4.10, we can compose  $m$  instances of  $\mathcal{CHSH}^n$  to obtain a  $m \cdot 2^{-(n-1)/2}$ -fairly-weak-binding string commitment scheme. Then, we can compose the Crépeau *et al.* bit commitment scheme (i.e., the bit commitment version of  $\mathcal{CHSH}^n$ ), which is  $2^{-(n-1)}$ -weak-binding, with this fairly-weak-binding string commitment scheme; by Theorem 4.14, this composition, which is the Lunghi *et al.* multi-round bit commitment scheme, is  $(m \cdot 2^{-(n-1)/2} + 2^{-(n-1)})$ -weak-binding.

Finally, for completeness, we point out that the composition theorem also applies to two ordinary binding or weak-binding commitment schemes.

**Theorem 4.16.** *Let  $(\mathcal{S}, \mathcal{S}')$  be an eligible pair of 2-prover commitment schemes, where  $\mathcal{S}$  is  $\varepsilon$ -binding and  $\mathcal{S}'$  is  $\delta$ -binding. Then, the composition  $S \star \mathcal{S}'$  is  $(\varepsilon + \delta)$ -binding. The same holds for the weak-binding property.*

*Proof.* The proof is almost the same as in Theorem 4.8 or Theorem 4.10, respectively, except that now there are no  $s_\circ$  and  $y_\circ$ , and in the end we can simply conclude that

$$\begin{aligned} p(s \neq \hat{s} \wedge s \neq \perp) &\leq p(s \neq \hat{s} \wedge s \neq \perp \wedge s = \tilde{s}) + p(s \neq \tilde{s} \wedge s \neq \perp) \\ &\leq p(\tilde{s} \neq \hat{s} \wedge \tilde{s} \neq \perp) + p(y \neq \hat{y} \wedge y \neq \perp) \\ &\leq \varepsilon + \delta, \end{aligned}$$

where the second inequality holds since  $y = \perp$  implies that  $s = \text{Extr}(y, c) = \perp$ .  $\square$

## Acknowledgments

We would like to thank Jędrzej Kaniewski for helpful discussions regarding [LKB<sup>+</sup>15], and for commenting on an earlier version of our work.

## References

- BGKW88. Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Multi-Prover Interactive Proofs: How to Remove Intractability Assumptions. In Janos Simon, editor, *STOC 1988*, pages 113–131. ACM, 1988.
- BS15. Mohammad Bavarian and Peter W. Shor. Information Causality, Szemerédi-Trotter and Algebraic Variants of CHSH. In Tim Roughgarden, editor, *ITCS 2015*, pages 123–132. ACM, 2015.
- CCL15. Kaushik Chakraborty, André Chailloux, and Anthony Leverrier. Arbitrarily Long Relativistic Bit Commitment. *ArXiv e-prints*, 2015. <http://arxiv.org/abs/1507.00239>.
- CHSH69. John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed Experiment to Test Local Hidden-Variable Theories. *Physical Review Letters*, 23:880–884, 1969.
- CSST11. Claude Crépeau, Louis Salvail, Jean-Raymond Simard, and Alain Tapp. Two Provers in Isolation. In Dong Hoon Lee and Xiaoyun Wang, editors, *ASIACRYPT 2011*, volume 7073 of *Lecture Notes in Computer Science*, pages 407–430. Springer, 2011.
- FF15. Serge Fehr and Max Fillinger. Multi-Prover Commitments Against Non-Signaling Attacks. In Rosario Genaro and Matthew Robshaw, editors, *Advances in Cryptology - CRYPTO 2015, part II*, volume 9216 of *Lecture Notes in Computer Science*, pages 403–421, 2015.
- Ken99. Adrian Kent. Unconditionally Secure Bit Commitment. *Physical Review Letters*, 83(7):1447–1450, 1999.
- Ken05. Adrian Kent. Secure Classical Bit Commitment Using Fixed Capacity Communication Channels. *Journal of Cryptology*, 18(4):313–335, 2005.
- LC97. Hoi-Kwong Lo and Hoi Fung Chau. Is quantum bit commitment really possible? *Physical Review Letters*, 78:3410–3413, 1997.
- LKB<sup>+</sup>15. Tommaso Lunghi, Jędrzej Kaniewski, Felix Bussi eres, Raphael Houlmann, Marco Tomamichel, Stephanie Wehner, and Hugo Zbinden. Practical Relativistic Bit Commitment. *Physical Review Letters*, 115, 2015.
- May97. Dominic Mayers. Unconditionally Secure Quantum Bit Commitment is Impossible. *Physical Review Letters*, 18:3414–3417, 1997.
- RAM15. Ravishankar Ramanathan, Remigiusz Augusiak, and Gl acia Murta. XOR Games with  $d$  Outcomes and the Task of Non-Local Computation. *ArXiv e-prints*, 2015. <http://arxiv.org/abs/1502.02974>.
- SCK14. Jamie Sikora, Andr e Chailloux, and Iordanis Kerenidis. Strong Connections Between Quantum Encodings, Non-Locality and Quantum Cryptography. *Physical Review A*, page 9, 2014.

## A Proof of Lemma 2.1

We first extend the respective probability spaces given by the distributions  $p(x)$  and  $p(y)$  by introducing an event  $\Delta$  and declaring that

$$p(x = x_o \wedge \Delta) = \min\{p(x = x_o), p(y = x_o)\} = p(y = x_o \wedge \Delta)$$

for every  $x_o \in \mathcal{X}$ . Note that  $p(\Delta)$  is well defined (by summing over all  $x_o$ ). As we will see below,  $\Delta$  will become the event  $x = y$ . In order to find a consistent joint distribution  $p(x, y)$ , it suffices to find a consistent joint distribution  $p(x, y|\Delta)$  for  $p(x|\Delta)$  and  $p(y|\Delta)$ , and a consistent joint distribution  $p(x, y|\neg\Delta)$  for  $p(x|\neg\Delta)$  and  $p(y|\neg\Delta)$ . The former, we choose as

$$p(x = x_o \wedge y = x_o|\Delta) := \min\{p(x = x_o), p(y = x_o)\}/p(\Delta)$$

for all  $x_o \in \mathcal{X}$ , and  $p(x = x_o \wedge y = y_o|\Delta) := 0$  for all  $x_o \neq y_o \in \mathcal{X}$ , and the latter we choose as

$$p(x = x_o \wedge y = y_o|\neg\Delta) := p(x = x_o|\neg\Delta) \cdot p(y = y_o|\neg\Delta)$$

for all  $x_o, y_o \in \mathcal{X}$ . It is straightforward to verify that these are indeed *consistent* joint distributions, as required, so that  $p(x, y) = p(x, y|\Delta) \cdot p(\Delta) + p(x, y|\neg\Delta) \cdot p(\neg\Delta)$  is also consistent. Furthermore, note that  $p(x = y|\Delta) = 1$  and  $p(x = y|\neg\Delta) = 0$ ; the latter holds because we have  $p(x = x_o \wedge \Delta) = p(x = x_o)$  or  $p(y = x_o \wedge \Delta) = p(y = x_o)$  for each  $x_o \in \mathcal{X}$ , and thus  $p(x = x_o \wedge \neg\Delta) = 0$  or  $p(y = x_o \wedge \neg\Delta) = 0$ . As such,  $\Delta$  is the event  $x = y$ , and therefore  $p(x = y = x_o) = p(x = x_o \wedge \Delta) = \min\{p(x = x_o), p(y = x_o)\}$  for every  $x_o \in \mathcal{X}$  as required. Finally, the claim regarding  $p(x, y|x \neq y)$  holds by construction.  $\square$

## B A Property for Conditionally Independent Random Variables

Let  $p(x, y, z)$  be a distribution, and let  $A \subset \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$  be an event. Then, we write  $x \rightarrow y \rightarrow z$  to express that  $p(x, z|y) = p(x|y)p(z|y)$ , and  $x \rightarrow A \rightarrow y$  to express that  $p(x, y|A) = p(x|A)p(y|A)$ , etc.

**Lemma B.1.** *If  $x \rightarrow y \rightarrow z$  and  $x \rightarrow x \neq y \rightarrow y$ , then  $x \rightarrow x \neq y \rightarrow z$ .*

*Proof.* We assume that  $x \rightarrow y \rightarrow z$  and  $x \rightarrow x \neq y \rightarrow y$ . We first observe that

$$p(x, x \neq y, z|y) = p(x, x \neq y|y)p(z|x, y, x \neq y) = p(x, x \neq y|y)p(z|x, y) = p(x, x \neq y|y)p(z|y),$$

which means that  $(x, x \neq y) \rightarrow y \rightarrow z$ , and, by summing over  $x$ , implies  $x \neq y \rightarrow y \rightarrow z$ . It follows that

$$p(z|x, y, x \neq y) = p(z|y) = p(z|y, x \neq y),$$

which actually means that  $x \rightarrow (y, x \neq y) \rightarrow z$ . Therefore,

$$\begin{aligned} p(x, z|x \neq y) &= \sum_y p(x, y, z|x \neq y) = \sum_y p(x, y|x \neq y)p(z|x, y, x \neq y) \\ &= p(x|x \neq y) \sum_y p(y|x \neq y)p(z|y, x \neq y) \\ &= p(x|x \neq y) \sum_y p(y, z|x \neq y) \\ &= p(x|x \neq y)p(z|x \neq y), \end{aligned}$$

which was to be proven.  $\square$

## C Proof of Theorem 3.19

Fix a commit strategy  $\overline{\text{com}}_{PQ}$  against  $\mathcal{S}$ . Enumerate all strings in the domain  $\{0, 1\}^n$  of  $\mathcal{S}$  as  $s_1, \dots, s_{2^n}$ , and for every  $i \in \{1, \dots, 2^n\}$  let  $\overline{\text{open}}^i_{PQ}$  be an opening strategy maximizing  $p_i := p(s = s_i)$ , where  $s$  is the output of the verifier when  $P$  and  $Q$  use this strategy. We assume without loss of generality that the  $p_i$ s are in descending order. We define  $p(\hat{s})$  as follows. Let  $N \geq 2$  be an integer which we will fix later. By Definition 3.13 and inequality (2), it holds that

$$\sum_{i=1}^N p_i \leq 1 + \binom{N}{2} \cdot \varepsilon = 1 + \frac{N(N-1)}{2} \cdot \varepsilon$$

where we let  $p_i = 0$  for  $i > 2^n$  in case  $N > 2^n$ . We would like to define  $p(\hat{s})$  as  $p(\hat{s} = s_i) := p_i - (N-1)\varepsilon/2$  for all  $i \leq N, 2^n$ ; however, this is not always possible because  $p_i - (N-1)\varepsilon/2$  may be negative. To deal with this, let  $N'$  be the largest integer such that  $N' \leq N$  and  $p_1, \dots, p_{N'} \geq (N-1)\varepsilon/2$ . (We take  $N = 0$  if  $p_1 < (N-1)\varepsilon/2$ .) It follows that

$$\sum_{i=1}^{N'} p_i \leq 1 + \frac{N'(N'-1)}{2} \cdot \varepsilon \leq 1 + \frac{N'(N-1)}{2} \cdot \varepsilon \quad \text{and thus} \quad \sum_{i=1}^{N'} p_i = 1 + \frac{N'(N-1)}{2} \cdot \tilde{\varepsilon}$$

for some  $\tilde{\varepsilon} \leq \varepsilon$ . We now set  $p(\hat{s})$  to be  $p(\hat{s} = s_i) := p_i - (N-1)\tilde{\varepsilon}/2 \geq p_i - (N-1)\varepsilon/2 \geq 0$  for all  $i \leq N'$ . Now consider an opening strategy  $\overline{\text{open}}_{PQ}$  and let  $p(s)$  be the resulting output distribution. By definition of the  $p_i$ , it follows that  $p(s = s_i) \leq p_i$  for all  $i \leq 2^n$ , and  $p_i \leq p(\hat{s} = s_i) + (N-1)\varepsilon/2$  for all  $i \leq N'$ . By Lemma 2.1, we can conclude that there exists a consistent joint distribution  $p(\hat{s}, s)$  with  $p(\hat{s} = s = s_i) = \min\{p(s = s_i), p(\hat{s} = s_i)\} \geq p(s = s_i) - (N-1)\varepsilon/2$  for all  $i \leq N'$ , and thus  $p(\hat{s} \neq s = s_i) = p(s = s_i) - p(\hat{s} = s = s_i) \leq (N-1)\varepsilon/2$  for all  $i \leq N'$ . Furthermore, when  $N' < i \leq N$ , we have  $p(\hat{s} \neq s = s_i) = p(s = s_i) \leq p_i < (N-1)\varepsilon/2$  by definition of  $N'$ . Since the  $p_i$  are sorted in descending order, it follows that for all  $i > N$

$$p(\hat{s} \neq s = s_i) = p(s = s_i) \leq p_i \leq p_N \leq \frac{1}{N} \sum_{i=1}^N p_i \leq \frac{1}{N} + \frac{N-1}{2} \cdot \varepsilon$$

and thus, we have shown for all  $s_o \in \{0, 1\}^n$  that

$$p(\hat{s} \neq s = s_o) \leq \frac{1}{N} + \frac{N-1}{2} \cdot \varepsilon.$$

We now select  $N$  so that this value is minimized: it is easy to verify that the function  $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ ,  $x \mapsto 1/x + (x-1)\varepsilon/2$  has its global minimum in  $\sqrt{2/\varepsilon}$ ; thus, we pick  $N := \lceil \sqrt{2/\varepsilon} \rceil$ , which gives us

$$p(\hat{s} \neq s = s_o) \leq \frac{1}{N} + \frac{N-1}{2} \cdot \varepsilon \leq \frac{1}{\sqrt{2/\varepsilon}} + \frac{\sqrt{2/\varepsilon}}{2} \cdot \varepsilon = \sqrt{2\varepsilon}$$

for any  $s_o \in \{0, 1\}^n$ , as claimed.  $\square$

## D The Hiding Property of Composed Schemes

We already mentioned that the standard hiding property is not good enough for multi-round bit commitment schemes: The standard definition is not violated if the verifier learns the string  $s$  immediately after the commit phase. However, the purpose of multi-round schemes is to maintain the commitment over a longer period of time in the relativistic setting, without *disclosing* the string  $s$  until the very end. In this appendix, we define a hiding property that captures this requirement, and we prove that a composed scheme  $\mathcal{S}'' = \mathcal{S} \star \mathcal{S}'$  is hiding if both  $\mathcal{S}$  and  $\mathcal{S}'$  are hiding (with the error parameters adding up).

**Definition D.1.** Let  $\mathcal{S} = (\text{com}_{PQV}, \text{open}_{PQV})$  be a commitment scheme. We write  $v$  for the verifier's view immediately before the last round of communication in  $\text{open}_{PQV}$ . We say that a scheme is  $\varepsilon$ -hiding until the last round if for any (possibly dishonest) verifier  $V$  and any two inputs  $s_0$  and  $s_1$  to the honest provers, we have  $d(p(v|s_0), p(v|s_1)) \leq \varepsilon$ .

**Theorem D.2.** Let  $\mathcal{S}$  be a  $\varepsilon$ -hiding commitment scheme and  $\mathcal{S}'$  a scheme that is  $\delta$ -hiding until the last round. If  $(\mathcal{S}, \mathcal{S}')$  is eligible, then the composed scheme  $\mathcal{S}'' = \mathcal{S} \star \mathcal{S}'$  is  $(\varepsilon + \delta)$ -hiding until the last round.

*Proof.* Fix a strategy against the hiding-until-the-last-round property of  $\mathcal{S}''$ . We consider the distribution  $p(v, y, v'|s)$  where  $s$  is the string that the provers commit to,  $v$  the verifier's view after  $\overline{\text{com}}_{PQV}$  has been executed,  $y$  the opening information to which  $Q$  commits using the scheme  $\mathcal{S}'$ , and  $v'$  the verifier's view immediately before the last round of communication. We need to show that  $d(p(v'|s_0), p(v'|s_1)) \leq \varepsilon + \delta$  for any  $s_0$  and  $s_1$ .

First, note that  $p(v'|v, y, s_b) = p(v'|v, y)$  since  $v'$  is produced by  $P$ ,  $Q$  and  $V$  acting on  $y$  and  $v$  only. From any strategy against  $\mathcal{S}''$ , we can obtain a strategy against  $\mathcal{S}'$  by fixing  $v$ . Thus, by the hiding property of  $\mathcal{S}'$ , for any  $y_0$  and  $y_1$ , we have  $d(p(v'|v, y = y_0), p(v'|v, y = y_1)) \leq \delta$  and it follows by the convexity of the statistical distance in both arguments that

$$p(v'|v, s_0) = \sum_y p(y|v, s_0)p(v'|v, y) \approx_\delta \sum_y p(y|v, s_1)p(v'|v, y) = p(v'|v, s_1)$$

where we use  $\approx_\delta$  to indicate that the two distributions have statistical distance at most  $\delta$ . Since we have  $d(p(v|s_0), p(v|s_1)) \leq \varepsilon$  by the hiding property of  $\mathcal{S}$ , it follows that

$$p(v'|s_0) = p(v, v'|s_0) = p(v|s_0)p(v'|v, s_0) \approx_\delta p(v|s_0)p(v'|v, s_1) \approx_\varepsilon p(v|s_1)p(v'|v, s_1) = p(v, v'|s_1) = p(v'|s_1)$$

where the first and last equality hold because  $v'$  contains  $v$  since  $v'$  is the view of  $V$  at a later point in time.  $\square$