A Quantum Bit Commitment Scheme Provably Unbreakable by both Parties

Gilles Brassard [*] *Université de Montréal* [†]

Claude Crépeau [‡] *École Normale Supérieure* [§]

Richard Jozsa *Université de Montréal* [†]

Denis Langlois [‡] *Université Paris–Sud* [¶]

Abstract

Assume that a party, $\mathcal{A}lice$, has a bit x in mind, to which she would like to be committed toward another party, $\mathcal{B}ob$. That is, $\mathcal{A}lice$ wishes, through a procedure *commit*(x), to provide $\mathcal{B}ob$ with a piece of evidence that she has a bit x in mind and that she cannot change it. Meanwhile, $\mathcal{B}ob$ should not be able to tell from that evidence what x is. At a later time, $\mathcal{A}lice$ can reveal, through a procedure *unveil*(x), the value of x and prove to $\mathcal{B}ob$ that the piece of evidence sent earlier really corresponded to that bit. Classical bit commitment schemes (by which $\mathcal{A}lice$'s piece of evidence is classical information such as a bit string) cannot be secure against unlimited computing power and none have been proven secure against algorithmic sophistication. Previous quantum bit commitment schemes (by which $\mathcal{A}lice$'s piece of evidence is quantum information such as a stream of polarized photons) were known to be invulnerable to unlimited computing power and algorithmic sophistication, but not to *arbitrary* measurements allowed by quantum physics: perhaps more sophisticated use of quantum physics could have defeated them.

We present a new quantum bit commitment scheme. The major contribution of this work is to provide the first *complete* proof that, according to the laws of quantum physics, neither participant in the protocol can cheat, except with arbitrarily small probability. In addition, the new protocol can be implemented with current technology.

* Supported in part by NSERC's E. W. R. Steacie Memorial Fellowship and Québec's FCAR.

† Département d'Informatique et R.O., Université de Montréal, C.P. 6128, succursale "A", Montréal (Québec), Canada H3C 3J7. e-mail: {brassard}{jozsa}@iro.umontreal.ca.

‡ Part of this work was performed while visiting †.

§ Laboratoire d'Informatique de l'École Normale Supérieure, (CNRS URA1327), 45 rue d'Ulm, 75230 Paris CEDEX 05, France. e-mail:crepeau@dmi.ens.fr.

¶ Labo. de Recherche en Informatique, Université Paris–Sud, Bâtiment 490, 91405 Orsay, France. e-mail: langlois@lri.lri.fr.

1 Introduction

Assume that a party, $\mathcal{A}lice$, has a bit x in mind, to which she would like to be committed toward another party, $\mathcal{B}ob$. That is, $\mathcal{A}lice$ wishes, through a procedure *commit*(x), to provide $\mathcal{B}ob$ with a piece of evidence that she has a bit x in mind and that she cannot change it. Meanwhile, $\mathcal{B}ob$ should not be able to tell from that evidence what x is. At a later time, $\mathcal{A}lice$ can reveal, through a procedure *unveil*(x), the value of x and prove to $\mathcal{B}ob$ that the piece of evidence sent earlier really corresponded to that bit.

Bit commitment schemes have several applications in the field of cryptographic protocols. In particular one can implement *zero-knowledge proofs* of a variety of statements using bit commitment schemes [GMR89, GMW91, BCC88]. The first implementations of bit commitment schemes were given in a computational complexity scenario [Blu82]. Unfortunately, proofs of their (computational) security have always required an unproved assumption since otherwise they would imply very strong results such as $\mathcal{P} \neq \mathcal{NP}$.

Over the last two decades a number of researchers have investigated the connection between cryptography and quantum physics, starting with the work of Wiesner in the late 1960's (though published much later [Wie83]), and continuing with the work of Bennett and Brassard [BBBW83, BB84, BB85, BBR88, BB89, BBBSS92] and later of Crépeau [CK88, Cré90, BC91, BBCS92, Cré93]. The security of these protocols would not be compromised if a cheater had unlimited computing power, but in essentially all cases it has not yet been ruled out that still more sophisticated use of quantum physics might defeat them.

The first quantum bit commitment scheme ever proposed is due to Bennett and Brassard [BB84] (actually, the protocol they describe is only claimed to implement coin tossing, but implicitly it implements bit commitment). Their scheme had two major flaws: it was impossible to use in practice because faint pulses of light would compromise the security of the scheme (it required individual photons to be transmitted), and the scheme could actually be cheated by *Alice* using the Einstein–Podolsky–Rosen effect [EPR35]. A later protocol of [BC91] did not suffer from these problems but was sensitive to transmission errors in the quantum channel and no formal proof of its security was, at the time, available.

The protocol we describe in this current paper is an improvement on the protocol of [BC91], which can deal with transmission errors and is more efficient. We also provide the first mathematical proof that the protocol is *perfectly* secure, in the sense that neither party can cheat without arbitrarily high probability of detection, according to the laws of quantum physics.

As a side benefit, the current result and proof provide the missing piece to the protocol of [BBCS92] for quantum oblivious transfer. Thus, it is now possible to prove the security of that scheme as well. In turn, this unconditionally secure quantum oblivious transfer protocol allows for provably unconditionally secure discreet two-party computation and decision making.

2 The New Quantum Bit Commitment Scheme

Let $\textcircled{c}^p$ denote the random variable that takes the binary value 0 with probability p and 1 with probability $1-p$. We often drop the index when $p=\frac{1}{2}$ and write $\textcircled{c}$ instead of $\textcircled{c}^{\frac{1}{2}}$. Also, denote by $[\,]_i$ the selection function such that $[a_0, a_1, ..., a_k]_i = a_i$. Let $x \odot y$ denote the Boolean scalar product, i.e. if x_i, y_i are the i^{th} bits of x and y we have $x \odot y = \bigoplus_{i=1}^{n} x_i \wedge y_i$.

Let $+ = (|\leftrightarrow\rangle, |\updownarrow\rangle)$ and $\times = (|\nearrow\rangle, |\nwarrow\rangle)$ denote respectively the bases of rectilinear and diagonal polarization in the quantum state space of a photon. Please consult the Appendix for an explanation of this notation and a summary of relevant basic quantum physics.

2.1 The formal protocols

Let ε be an upper bound on the error rate of the quantum channel, i.e. the probability that a $|\leftrightarrow\rangle$ polarized photon is detected as $|\updownarrow\rangle$. In order for *Alice* to commit to a bit x, she uses protocol *commit*(x) with *Bob*. (an informal description of the protocol follows in Subsection 2.2)

Protocol 2.1 (*commit*(x) **)**

1: *Bob* chooses a Boolean matrix G as the generating matrix of a binary linear (n,k,d)-code C such that the ratio $d/n > 10\varepsilon$ and the ratio $k/n = 0.52$ and announces it to *Alice*

2: *Alice* chooses a non-zero random n-bit string $r \leftarrow (\textcircled{c}_1 \textcircled{c}_2 ... \textcircled{c}_n)$ and announces it to *Bob*

3: *Alice* chooses a random n-bit codeword $c \leftarrow (\textcircled{c}_1 ... \textcircled{c}_k)G$ from C such that $c \odot r = x$

4: $\mathbf{DO}_{i=1}^{n}$

- *Alice* chooses a random bit $b_i \leftarrow \textcircled{c}$ and defines her transmission basis $(\varphi_i, \varphi_i^{\perp}) \leftarrow [+, \times]_{b_i}$
- *Alice* sends to *Bob* a photon π_i with polarization $[\varphi_i, \varphi_i^{\perp}]_{c_i}$

5: $\mathbf{DO}_{i=1}^{n}$

- *Bob* chooses a random bit $b'_i \leftarrow \textcircled{c}$ and measures photon π_i in basis $(\theta_i, \theta_i^{\perp}) \leftarrow [+, \times]_{b'_i}$
- *Bob* sets $c'_i \leftarrow \begin{cases} 0 & \text{if } \pi_i \text{ is seen as } \theta_i \\ 1 & \text{if } \pi_i \text{ is seen as } \theta_i^{\perp} \end{cases}$

Let c', b and b' be the vectors $c' = (c'_1 c'_2 \dots c'_n)$, $b = (b_1 b_2 \dots b_n)$, $b' = (b'_1 b'_2 \dots b'_n)$. *Alice* keeps x, c and b secret until (and if) unveiling takes place, whereas *Bob* keeps c' and b' secret forever. Theorem 3.4 shows that an honest *Alice* does not reveal much about her secret bit x by sending codeword c on the quantum channel.

If *Alice* subsequently decides to unveil her commitment, she initiates the following protocol with *Bob*.

Protocol 2.2 (*unveil*$((c,b,x),(c',b'))$ **)**

1: *Alice* reveals c, b and x to *Bob*

2: *Bob* sets $\delta \leftarrow \sum_{i \,|\, b'_i = b_i} \frac{c_i \oplus c'_i}{n/2}$

3: **if** $(\delta < 1.4\varepsilon)$, $(x = c \odot r)$ and (c is a codeword)
then *Bob* outputs "accept"
else *Bob* outputs "reject"

2.2 Intuition behind the protocols

Intuitively, $\mathcal{A}lice$ chooses a random vector r and a random codeword c such that $c \odot r = x$. She tells r to $\mathcal{B}ob$ in the clear, but she sends him c through the quantum channel. For this, she encodes each bit of c by a photon polarized in a randomly chosen basis (rectilinear or diagonal): bit $c_i = 0$ is thus encoded as $|\leftrightarrow\rangle$ if $b_i = 0$ or as $|\nearrow\rangle$ if $b_i = 1$, whereas bit 1 may be encoded either as $|\updownarrow\rangle$ or $|\nwarrow\rangle$. Since $\mathcal{B}ob$ does not know in which bases the photons are polarized, he measures them in randomly chosen bases. When he chooses the correct basis ($b'_i = b_i$), which happens with probability $\frac{1}{2}$, he obtains the correct bit ($c'_i = c_i$) except with error probability at most ε. On the other hand, when he chooses the wrong basis ($b'_i \neq b_i$), his bit is uncorrelated with $\mathcal{A}lice$'s bit ($c'_i = c_i$ with probability $\frac{1}{2}$). Therefore $\mathcal{B}ob$'s reading of $\mathcal{A}lice$'s word c is correct on roughly 75% of the bits. (We shall see later that a cheating $\mathcal{B}ob$ is able to get as much as about 85% of the bits correctly, and that this is the best possible.)

The binary linear code C is chosen so that there are exponentially many codewords around $\mathcal{B}ob$'s received c' that are at the same Hamming distance as $\mathcal{A}lice$'s transmitted c. For this, the minimum distance between codewords should not be too large. Because r is chosen randomly, knowledge of r and c' give $\mathcal{B}ob$ an exponentially small amount of expected Shannon information on $x = c \odot r$. (See Theorem 3.4.) On the other hand, the minimum distance between codewords must be sufficiently large to prevent $\mathcal{A}lice$ from finding two different codewords c^0 and c^1 (together with possibly fake sending bases b^0 and b^1) so that $\mathcal{B}ob$ would be willing to blame on transmission errors the differences between either codeword and his measured c'. (See Theorem 3.7.) Thus we see that a balancing act is needed in the choice of code C to prevent both $\mathcal{A}lice$ and $\mathcal{B}ob$ from cheating, thence the mysterious parameters 10ε and 0.52 in protocol *commit*.

3 Analysis

There are very few ways in which the above protocols might fail. This section is divided into four parts, each analysing one way in which failure could occur.

- $\mathcal{B}ob$ gets too much information about x
- $\mathcal{B}ob$ chooses an unsuitable G
- $\mathcal{A}lice$ changes x without detection
- $\mathcal{B}ob$ rejects a valid c

3.1 Analysis of $\mathcal{B}ob$'s information about x

Given the public parameters G and r, all the information available to $\mathcal{B}ob$ about x is provided to him through the quantum transmission of c in step 4. We identify the measurement that $\mathcal{B}ob$ may perform in step 5 which will maximize his information about c (and x). Let us first define two magic constants that will be used later in the analysis:

$$\kappa = \cos(\pi/8) \approx 0.9238795$$

and

$$\sigma = \sin(\pi/8) \approx 0.3826834.$$

Let $|\mathcal{B}\rangle$ denote the state midway between $|\leftrightarrow\rangle$ and $|\nearrow\rangle$, and let $|\mathcal{B}^\perp\rangle$ denote the state midway between $|\updownarrow\rangle$ and $|\nwarrow\rangle$, i.e.

$$|\mathcal{B}\rangle = \kappa|\leftrightarrow\rangle + \sigma|\updownarrow\rangle \text{ and } |\mathcal{B}^\perp\rangle = -\sigma|\leftrightarrow\rangle + \kappa|\updownarrow\rangle.$$

3.1.1 Optimal measurement

Theorem 3.1 *The quantum measurement that will maximize (cheating) $\mathcal{B}ob$'s information about x is the measurement of each photon π_i separately in the basis $(\theta_i, \theta_i^\perp) = (\mathcal{B}, \mathcal{B}^\perp)$.*

Proof. The density matrix ρ_0 (please consult the Appendix) describing the quantum mixture of states sent to $\mathcal{B}ob$ to represent a 0 in step 4 of the original protocol *commit* is given by

$$\rho_0 = \frac{1}{2}|\leftrightarrow\rangle\langle\leftrightarrow| + \frac{1}{2}|\nearrow\rangle\langle\nearrow|$$

$$= \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

and for a bit 1, the density matrix ρ_1 is given by

$$\rho_1 = \frac{1}{2}|\updownarrow\rangle\langle\updownarrow| + \frac{1}{2}|\nwarrow\rangle\langle\nwarrow| =$$

$$\frac{1}{2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

Consider a protocol *commit'* that is identical to *commit* except that step 4 becomes

> **4':** $\mathbf{DO}_{i=1}^{n}$
>
> - $\mathcal{A}lice$ chooses a random bit $b_i \leftarrow \copyright^{\kappa^2}$
> - $\mathcal{A}lice$ sends to $\mathcal{B}ob$ a photon π_i with polarization $[|\mathcal{B}\rangle, |\mathcal{B}^\perp\rangle]_{c_i \oplus b_i}$

For the modified protocol *commit'* the density matrices ρ'_0, ρ'_1 describing the quantum mixtures sent to $\mathcal{B}ob$ to represent a 0 and a 1 in step 4' are the same as in *commit*:

$$\rho'_0 = \kappa^2|\mathcal{B}\rangle\langle\mathcal{B}| + (1-\kappa^2)|\mathcal{B}^\perp\rangle\langle\mathcal{B}^\perp| =$$

$$\kappa^2\begin{pmatrix} \kappa^2 & \sigma\kappa \\ \sigma\kappa & \sigma^2 \end{pmatrix} + \sigma^2\begin{pmatrix} \sigma^2 & -\sigma\kappa \\ -\sigma\kappa & \kappa^2 \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

$$\rho'_1 = (1-\kappa^2)|\mathcal{B}\rangle\langle\mathcal{B}| + \kappa^2|\mathcal{B}^\perp\rangle\langle\mathcal{B}^\perp| =$$

$$\sigma^2\begin{pmatrix} \kappa^2 & \sigma\kappa \\ \sigma\kappa & \sigma^2 \end{pmatrix} + \kappa^2\begin{pmatrix} \sigma^2 & -\sigma\kappa \\ -\sigma\kappa & \kappa^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

Furthermore, if we call ρ_c the density matrix associated with the mixture of states used in *commit* to send c and similarly, ρ'_c for *commit'* we get

$$\rho_c = \rho_{c_1} \otimes \rho_{c_2} \otimes \cdots \otimes \rho_{c_n} = \rho'_{c_1} \otimes \rho'_{c_2} \otimes \cdots \otimes \rho'_{c_n} = \rho'_c$$

where the operation $\otimes$ is the tensor product. Finally, the density matrices $\varrho_0, \varrho_1, \varrho'_0, \varrho'_1$ describing the quantum mixtures of states sent to $\mathcal{B}ob$ to commit to a 0 (or a 1) with the protocols *commit* and *commit'* are given by

$$\varrho_0 = \sum_{\{c\in C|c\odot r=0\}} \frac{\rho_c}{2^{k-1}} = \sum_{\{c\in C|c\odot r=0\}} \frac{\rho'_c}{2^{k-1}} = \varrho'_0$$

and

$$\varrho_1 = \sum_{\{c\in C|c\odot r=1\}} \frac{\rho_c}{2^{k-1}} = \sum_{\{c\in |c\odot r=1\}} \frac{\rho'_c}{2^{k-1}} = \varrho'_1.$$

Thus $\mathcal{B}ob$ is able to get *exactly* the same information about c and x in protocol *commit* as in *commit'*. (This follows from a theorem of quantum physics stipulating that mixtures with identical density matrices cannot be distinguished by *any quantum measurement* whatsoever.) We also observe that in the protocol *commit'* the measurement that will optimize $\mathcal{B}ob$'s information about c (and thus x) consists of measuring every single photon π_i in basis $(\mathcal{B}, \mathcal{B}^\perp)$ since in that basis he gets *all the information available!* (All the photons sent are either polarized as $|\mathcal{B}\rangle$ or $|\mathcal{B}^\perp\rangle$.) We thus conclude that the optimal measurement for $\mathcal{B}ob$ in protocol *commit* is the same. In particular, any more general joint measurement on all the photons together will be of no advantage. (This was the main open question of [BC91].) $_{3.1}\blacksquare$

3.1.2 Further analysis

Despite the fact that the honest $\mathcal{B}ob$ is not expected to perform the above optimal measurement, we show that even if he did he would get very little information about x.

We start with a lemma stating that the vector c' received by $\mathcal{B}ob$ must be fairly far from the vector c sent by $\mathcal{A}lice$. Let $\gamma = H^{-1}(1/2) \approx 0.1100279$.

Lemma 3.2 *Even if (cheating) $\mathcal{B}ob$ performs the optimal measurement, there exists a positive constant $\alpha < 1$ such that he ends up with a vector c' at distance less than γn from c with probability at most α^n.*

Proof. Assume that the quantum channel is noiseless (in reality things are even worse for a cheating $\mathcal{B}ob$). When $\mathcal{B}ob$ performs the optimal measurement, his distribution on c' is ruled by the fact that for all i, $1 \le i \le n$, we have $[\text{Prob}(c_i = c'_i) = \kappa^2]$. Therefore the number of differences between c and c' is expected to be $\sigma^2 n \approx .1464466n$. We can use "Bernshtein's law of large numbers" [Kra86] to estimate the probability that the number of errors will be less than γn.

Theorem 3.3 (Bernshtein) *Let $x_1, x_2, ..., x_n$ be independent Bernoulli variables. If $\text{Prob}(x_i = 1) = p$ for $1 \le i \le n$ then for all $0 < \delta \le p(1-p)$ we have*

$$\text{Prob}(|\sum_{i=1}^{n} \frac{x_i}{n} - p| \ge \delta) \le 2e^{-n\delta^2}.$$

Let x_i be the indicator variable of c_i, c'_i, i.e. $x_i = c_i \oplus c'_i$. The number of differences between c and c' is given by $\sum_{i=1}^n x_i$ and $\text{Prob}(x_i = 1) = \sigma^2$. Therefore the probability that c' is at distance less than γn of c is given by

$$\text{Prob}(\sum_{i=1}^{n} \frac{x_i}{n} \le \gamma) \quad \le \quad \text{Prob}(|\sum_{i=1}^{n} \frac{x_i}{n} - \sigma^2| \ge \sigma^2 - \gamma)$$

$$\le 2e^{-n(\sigma^2-\gamma)^2} \approx 2e^{-0.00132632n}.$$

$_{3.2}\blacksquare$

We conclude that most of the time c' is at distance at least γn from c.

Theorem 3.4 *Even if $\mathcal{B}ob$ knew the exact Hamming distance $d \leftarrow d_H(c, c')$, he would have very little information about x, when $d > \gamma n$.*

Proof. The number of words at Hamming distance d from c' is $\binom{n}{d}$. Using the fact that $d > \gamma n$ and the standard approximation [MS77]

$$\frac{2^{H(\lambda)n}}{\sqrt{8n\lambda(1-\lambda)}} \leq \binom{n}{\lambda n} \leq \frac{2^{H(\lambda)n}}{\sqrt{2\pi n\lambda(1-\lambda)}},$$

we get the lower bound

$$\binom{n}{d} > \binom{n}{\gamma n} \geq \frac{2^{H(\gamma)n}}{\sqrt{n}} = \frac{2^{n/2}}{\sqrt{n}}$$

because $8\lambda(1-\lambda) \leq 1$ precisely when $\lambda \geq \kappa^2$ or $\lambda \leq 1-\kappa^2$, and also because $H(\gamma) = 1/2$. If we divide by 2^{n-k} (the number of symdroms of the code C) we get:

$$E(\text{number of codewords at distance } d) > \frac{2^{k-n/2}}{\sqrt{n}}$$

which is exponentially large in n as long as $\frac{k}{n} > \frac{1}{2}$. Indeed, we show that

Lemma 3.5 *The number of codewords at distance d from c' is at least $\frac{2^{k-n/2-\alpha n}}{\sqrt{n}}$ except with probability $2^{-\alpha n}$ for any $\alpha > 0$.*

Proof. Let S_w be the syndrome of a word w. Let $N_d(x,y)$ be the number of words with syndrome y at distance d from a fixed word w with syndrome x (this function is well defined because its value is independent of the specific choice of w; moreover, $N_d(x,y) = N_d(\vec{0}, x \oplus y) = N_d(y,x)$). We first show that $N_d(S_{c'}, S_c) \geq 2^{-r}\frac{\binom{n}{d}}{2^{n-k}}$ with probability at least $1-2^{-r}$ for any security parameter $r > 0$.

Starting from word c, each syndrome s occurs $N_d(S_c, s)$ times among the words at distance d from c. Therefore syndrome s has probability $N_d(S_c,s)/\binom{n}{d} = N_d(s,S_c)/\binom{n}{d}$ of being selected, i.e. of being that of the actual c'. Thus, any syndrome s for which $N_d(s,S_c) < 2^{-r}\frac{\binom{n}{d}}{2^{n-k}}$ (which would be bad because it would mean less uncertainty for $\mathcal{B}ob$) has probability of occurrence less than $(2^{-r}\frac{\binom{n}{d}}{2^{n-k}})/\binom{n}{d} = \frac{1}{2^{n-k}}2^{-r}$. Even if all but one syndrome were in that category, their collective probability would still be less than 2^{-r}. This establishes the claim that $N_d(S_{c'}, S_c) \geq 2^{-r}\frac{\binom{n}{d}}{2^{n-k}}$, except with probability less than 2^{-r}. Given that $\binom{n}{d} \geq \frac{2^{n/2}}{\sqrt{n}}$, setting $r = \alpha n$ leads to the result of the lemma. $_{3.5}\blacksquare$

From $\mathcal{B}ob$'s point of view, the codeword c is one of the, at least $\frac{2^{k-n/2-\alpha n}}{\sqrt{n}}$ many, equally likely codewords at distance d from c' forming a set E. The following lemma says how much information $\mathcal{B}ob$ will consequently have about $c \odot r$ for a random r.

Lemma 3.6 ([BBR88]) If $\mathcal{B}ob$ has narrowed down the value of c to a set E of equally likely candidates and if a random subset of the bits of c is chosen, the expected amount of Shannon information available to $\mathcal{B}ob$ about the parity of the bits in this subset is less than $2/|E|\ln 2$ bit.

In our case, this means that the number of bits of information $\mathcal{B}ob$ has after seeing c' is less than $\frac{2\sqrt{n}}{2^{k-n/2-\alpha n}\ln 2}$. This number of bits is exponentially small as soon as $k > n/2 + \alpha n$. Thus, if we pick $\alpha = 0.1$ we find that the number of bits of information is at most $2^{-0.1n}\sqrt{n}/\ln 2$ whenever $k/n > 0.51n$, even if he knew the exact number of errors d. $_{3.4}\blacksquare$

In reality, $\mathcal{B}ob$'s situation is much more difficult: he may not perform the optimal measurement, his measuring apparatus may be imperfect, and he does not even know the exact number of errors. Since his information about x would be very small even if he knew d and made no mistakes, his actual knowledge cannot be any better.

3.2 Analysis of $\mathcal{B}ob$'s probability of choosing an unsuitable G

It is a well known fact [MS77] that a random binary matrix G of size $k \times n$ defines a binary linear code with minimal distance at least d except with probability $2^{-\alpha n}$ as long as

$$k < n - H(d/n)n - \alpha n.$$

In particular if we set $\varepsilon \leq 1\%$ we find that a random binary matrix defines a binary linear code with minimal distance at least $10\varepsilon n$ except with probability $2^{-0.01n}$ as long as

$$k/n = 0.52 < 1 - H(0.1) - 0.01.$$

Therefore, $\mathcal{B}ob$ may choose G at random of size $0.52n \times n$ and only with probability $2^{-0.01n}$ will the code thus defined have minimal distance less than $10\varepsilon n$, again, as long as $\varepsilon \leq 1\%$.

3.3 Analysis of $\mathcal{A}lice$'s probability of changing x without detection

Although an honest $\mathcal{B}ob$ would not have as much information about x as the cheating $\mathcal{B}ob$ who reads each photon in basis $(|\mathcal{B}\rangle, |\mathcal{B}^{\perp}\rangle)$, he would have something more: the honest $\mathcal{B}ob$ has the ability to check that $\mathcal{A}lice$ is indeed committed to some bit.

Theorem 3.7 *There exists a positive constant $\alpha < 1$ with the following property: the probability that $\mathcal{A}lice$ is able to announce either pair (c^0, b^0) or pair (c^1, b^1) at her choosing in protocol unveil leading $\mathcal{B}ob$ to accept a 0 and a 1, is less than α^n.*

Proof. Let (c^0, b^0) and (c^1, b^1) be any pairs of n-bit strings such that $c^0 \odot r = 0$ and $c^1 \odot r = 1$. Since $c^0 \odot r \neq c^1 \odot r$, it must be that $c^0 \neq c^1$. By construction of the code C, any two codewords must be at distance at least $10\varepsilon n$ from each other. Let I be the set of indices on which c^0 and c^1 disagree: $I = \{i \mid c^0_i \neq c^1_i\}$. We show that whatever $\mathcal{A}lice$ does, with high probability, $I_0 \leftarrow \{i \in I \mid c^0_i \neq c'_i \wedge b^0_i = b'_i\}$ or $I_1 \leftarrow \{i \in I \mid c^1_i \neq c'_i \wedge b^1_i = b'_i\}$ has size more than $0.7\varepsilon n$. Since $I_0 \cap I_1 = \emptyset$, and thus $|I_0 \cup I_1| = |I_0| + |I_1|$, it suffices to show

Lemma 3.8 *Except with probability α^n for some constant $\alpha < 1$,*

$$I_0 \cup I_1 = \{i \in I \mid c^0_i \neq c'_i \wedge b^0_i = b'_i \vee c^1_i \neq c'_i \wedge b^1_i = b'_i\}$$

has size more than $1.4\varepsilon n$.

Proof. For each $i \in I$ consider

$$\text{Prob}(c^0_i \neq c'_i \wedge b^0_i = b'_i \vee c^1_i \neq c'_i \wedge b^1_i = b'_i).$$

The size of $I_0 \cup I_1$ can be estimated by a binomial distribution with respect to a lower bound for this probability. We start by giving such a lower bound.

Lemma 3.9 *For each $i \in I$,*

$$\text{Prob}(c^0_i \neq c'_i \wedge b^0_i = b'_i \vee c^1_i \neq c'_i \wedge b^1_i = b'_i) \geq \sigma^2$$

Proof. First notice that for $i \in I$, if $b^0_i = b^1_i$ then

$$\text{Prob}(c^0_i \neq c'_i \wedge b^0_i = b'_i \vee c^1_i \neq c'_i \wedge b^1_i = b'_i) = 1/2 \quad (1)$$

since $\mathcal{B}ob$ has probability $\frac{1}{2}$ of choosing the same basis $b^0_i = b'_i = b^1_i$.

The more complicated question is to determine this probability for $b^0_i \neq b^1_i$. Without loss of generality we may assume $b^1_i = c^0_i \neq c^1_i = b^0_i$, since the only other possibility $b^0_i = c^0_i \neq c^1_i = b^1_i$, is treated similarly. Then we get

$$\text{Prob}(c^0_i \neq c'_i \wedge b^0_i = b'_i \vee c^1_i \neq c'_i \wedge b^1_i = b'_i)$$

$$= \text{Prob}(c'_i = b'_i) \quad (2)$$

because

$$\begin{aligned}
&\text{Prob}(c^0_i \neq c'_i \wedge b^0_i = b'_i \vee c^1_i \neq c'_i \wedge b^1_i = b'_i) \\
&= \text{Prob}(b^1_i \neq c'_i \wedge b^0_i = b'_i \vee b^0_i \neq c'_i \wedge b^1_i = b'_i) \\
&= \text{Prob}(c'_i = b'_i = b^0_i \vee c'_i = b'_i = b^1_i) \\
&= \text{Prob}(c'_i = b'_i)
\end{aligned}$$

(The other case leads to $\text{Prob}(c'_i \neq b'_i)$)

But how small can $\mathcal{A}lice$ make these probabilities?

Assume first that $\mathcal{A}lice$ sends $\mathcal{B}ob$ a photon π_i polarized as a pure state $|\Psi\rangle$ at step i (please consult the Appendix for the notion of pure states).

$$\text{Prob}(c'_i = b'_i) = \tfrac{1}{2}\langle\leftrightarrow|\Psi\rangle^2 + \tfrac{1}{2}\langle\nwarrow\!\!\searrow|\Psi\rangle^2$$

$$= \frac{1}{2}\left[(1,0)\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}\right]^2 + \frac{1}{2}\left[(\tfrac{-1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}})\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}\right]^2$$

for the θ such that $|\Psi\rangle = (\cos\theta)|\leftrightarrow\rangle + (\sin\theta)|\updownarrow\rangle$. Therefore

$$\begin{aligned}
\text{Prob}(c'_i = b'_i) &= \frac{1}{2}\cos^2\theta + \frac{1}{2}\left(\frac{\sin\theta - \cos\theta}{\sqrt{2}}\right)^2 \\
&= \frac{\frac{1}{2} + \cos^2\theta - \sin\theta\cos\theta}{2}.
\end{aligned}$$

The minimum and maximum of this expression are σ^2 and κ^2. So, for any pure state $|\Psi\rangle$,

$$\sigma^2 \leq \text{Prob}(c'_i = b'_i) \leq \kappa^2 \quad (3)$$

(and similarly $\sigma^2 \leq \text{Prob}(c'_i \neq b'_i) \leq \kappa^2$).

It is a fact of quantum physics that the polarization of any photon can be described by a mixture of pure states (see the Appendix). Therefore, if π_i is not in a pure state, it may be represented as a mixture of pure states $|\Psi_1\rangle, |\Psi_2\rangle, ..., |\Psi_m\rangle$ with probabilities $p_1, p_2, ..., p_m$ such that $\sum p_j = 1$. We get the same result in this case since it holds for each $|\Psi_j\rangle$ individually. We therefore conclude from (1), (2) and (3) that in all cases, for each $i \in I$

$$\sigma^2 \leq \text{Prob}(c^0_i \neq c'_i \wedge b^0_i = b'_i \vee c^1_i \neq c'_i \wedge b^1_i = b'_i) \leq \kappa^2$$

3.9 ■

Now, since $|I| = 10\varepsilon n$, $|I_0 \cup I_1|$ will be given by a binomial distribution with mean at least $10\sigma^2\varepsilon n$ over $10\varepsilon n$ trials. Let u_i be the characteristic function of $I_0 \cup I_1$, i.e.

$$u_i \leftarrow \begin{cases} 1 & \text{if } c^0_i \neq c'_i \wedge b^0_i = b'_i \vee c^1_i \neq c'_i \wedge b^1_i = b'_i \\ 0 & \text{otherwise} \end{cases}$$

The probability that $|I_0 \cup I_1| < 1.4\varepsilon n$ is therefore going to be very small:

$$
\begin{aligned}
&\mathrm{Prob}(|I_0 \cup I_1| < 1.4\varepsilon n) \\
&= \mathrm{Prob}(\sum_{i\in I} u_i < 1.4\varepsilon n) \\
&= \mathrm{Prob}(\sum_{i\in I} \frac{u_i}{10\varepsilon n} < 0.14) \\
&= \mathrm{Prob}(\sum_{i\in I} \frac{u_i}{10\varepsilon n} - \sigma^2 < 0.14 - \sigma^2) \\
&\leq \mathrm{Prob}(|\sum_{i\in I} \frac{u_i}{10\varepsilon n} - \sigma^2| \geq \sigma^2 - 0.14) \\
&\leq 2e^{-10\varepsilon n(\sigma^2 - 0.14)^2} \\
&\approx 2e^{-0.000415587\varepsilon n}.
\end{aligned}
$$

3.8 $\blacksquare$

We conclude from the lemma, that except with probability $2e^{-0.000415587\varepsilon n}$, at least one of I_0 or I_1 must have size more than $0.7\varepsilon n$ and thus $\mathcal{B}ob$ would necessarily reject either (c^0, b^0) or (c^1, b^1) at step 3 of protocol *unveil*. 3.7 $\blacksquare$

3.4 Analysis of $\mathcal{B}ob$'s probability of rejecting a valid c

Despite the good will of $\mathcal{A}lice$, there is a small probability that $\mathcal{B}ob$ will reject the correct pair (c, b) because of unlikely extreme noise in the quantum channel. We finally show that this event occurs with exponentially small probability.

Theorem 3.10 *If $\mathcal{A}lice$ is honest, then there exists a constant $\alpha < 1$ such that an honest $\mathcal{B}ob$ rejects (c, b, x) with probability less than α^n.*

Proof. An error will be observed by $\mathcal{B}ob$ exactly if $b_i = b_i'$, while $c_i \neq c_i'$. The probability of such an event due to noise is less than $\frac{\varepsilon}{2}$, by definition of ε and because $\mathrm{Prob}(b_i' = b_i) = \frac{1}{2}$. What is therefore the probability of observing at least $0.7\varepsilon n$ errors?

Let

$$
v_i = \begin{cases} 1 & \text{if } b_i = b_i' \wedge c_i \neq c_i' \\ 0 & \text{otherwise} \end{cases}
$$

be the characteristic function of the observation of errors. The probability of observing more than $0.7\varepsilon n$ errors is given by $\mathrm{Prob}(\sum_{i=1}^n v_i > 0.7\varepsilon n)$ and is bounded as follows

$$
\begin{aligned}
&\mathrm{Prob}(\sum_{i=1}^n v_i > 0.7\varepsilon n) \\
&= \mathrm{Prob}(\sum_{i=1}^n \frac{v_i}{n} - 0.5\varepsilon > 0.2\varepsilon) \\
&\leq \mathrm{Prob}(|\sum_{i=1}^n \frac{v_i}{n} - 0.5\varepsilon| > 0.2\varepsilon) \\
&\leq 2e^{-n(0.2\varepsilon)^2} = 2e^{-0.04\varepsilon^2 n}.
\end{aligned}
$$

3.10 $\blacksquare$

4 Conclusion and Open Questions

We have described a complete protocol for Bit Commitment based on the transmission of polarized photons. We have shown that under the laws of quantum physics, this protocol cannot be cheated by either party except with exponentially small probability (exponential in the running time needed to implement the honest protocol).

A more thorough analysis is required to adjust all the constants used in this paper to get the best performance from our construction. Better performances may probably be achieved by using a third conjugate transmission-reception basis of circular polarization. This analysis will appear in the final version of this paper.

Acknowledgements

We wish to thank Charles H. Bennett, Joe Kilian, Asher Peres, Louis Salvail, and Miklòs Santhà, for support, suggestions and comments on this work. We are grateful to Joe Kilian for noticing that a single codeword is sufficient in protocol *commit*, and to Louis Salvail for asking “Who will choose the code C?”.

Finally, we thank God for playing dice.

References

[BBBSS92] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin. Experimental quantum cryptography. *Journal of Cryptology*, 5:3–28, 1992.

[BB84] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing*, pages 175–179. IEEE, 1984.

[BB85] C. H. Bennett and G. Brassard. An update on quantum cryptography. In *Advances in Cryptology: Proceedings of Crypto '84*, Lecture Notes in Computer Science, pages 475–480, Vol. 196. Springer-Verlag, 1985.

[BB89] C. H. Bennett and G. Brassard. The dawn of a new era for quantum cryptography: the experimental prototype is working! *SIGACT News*, 20(4):78–82, 1989.

[BBBW83] C. H. Bennett, G. Brassard, S. Breidbard, and S. Wiesner. Quantum cryptography, or unforgeable subway tokens. In *Advances in Cryptology: Proceedings of Crypto '82*, pages 267–275. Plenum Press, 1983.

[BBCS92] C. H. Bennett, G. Brassard, C. Crépeau, and M.-H. Skubiszewska. Practical quantum oblivious transfer. In *Advances in Cryptology: Proceedings of Crypto '91*, Lecture Notes in Computer Science, Vol. 576, pages 351–366. Springer-Verlag, 1992.

[BBR88] C. H. Bennett, G. Brassard, and J.–M. Robert. Privacy amplification by public discussion. *SIAM Journal on Computing*, 17(2):210–229, April 1988.

[Blu82] M. Blum. Coin flipping by telephone. In *Proceedings of IEEE Spring Computer Conference*, pages 133–137. IEEE, 1982.

[BCC88] G. Brassard, D. Chaum, and C. Crépeau. Minimum disclosure proofs of knowledge. *Journal of Computer and System Sciences*, 37:156–189, 1988.

[BC91] G. Brassard and C. Crépeau. Quantum bit commitment and coin tossing protocols. In *Advances in Cryptology: Proceedings of Crypto '90*, Lecture Notes in Computer Science, Vol. 537, pages 49–61. Springer-Verlag, 1991.

[Cré90] C. Crépeau. *Correct and Private Reductions among Oblivious Transfers.* PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1990. Supervised by Silvio Micali.

[Cré93] C. Crépeau. Cryptographic primitives and quantum theory. In *Proceedings of Workshop on Physics and Computation, PhysComp 92*, pages 200–204, 1993.

[CK88] C. Crépeau and J. Kilian. Achieving oblivious transfer using weakened security assumptions. In *29^{th} Symposium on Foundations of Computer Science*, pages 42–52. IEEE, 1988.

[EPR35] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47:777, 1935.

[GMW91] O. Goldreich, S. Micali, and A. Wigderson. Proofs that yield nothing but their validity, or All languages in $\mathcal{NP}$ have zero-knowledge proof systems. *Journal of the ACM*, 38:691–729, 1991.

[GMR89] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proof-systems. *SIAM Journal on Computing*, 18:186–208, 1989.

[Hel76] C. W. Helstrom. *Quantum Detection and Estimation Theory.* Academic Press, 1976.

[Kra86] E. Kranakis. Primality and cryptography. John Wiley and Sons, 1986.

[MS77] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes.* North-Holland, 1977.

[Sud86] A. Sudbery. *Quantum Mechanics and the Particles of Nature — an Outline for Mathematicians.* Cambridge University Press, 1986.

[Wie83] S. Wiesner. Conjugate coding. *SIGACT News*, 15(1):78–88, 1983. manuscript written *circa* 1970, unpublished until it appeared in SIGACT News.

Appendix: Outline of Some Relevant Quantum Theory

In quantum physics, the state space of a single photon is the collection of all unit vectors in a two dimensional complex Hilbert space $\mathcal{H}$. We use the Dirac bracket notation ([Sud86] chapter 2), commonly used in physics, to denote the states. In this notation the state vectors are written using right-handed pointed brackets. If $\binom{u}{v}$ is a state in $\mathcal{H}$ (given in terms of its components with respect to some basis) we write

$$|\psi\rangle = \begin{pmatrix} u \\ v \end{pmatrix}.$$

The corresponding left-handed bracket, enclosing the same symbol, denotes the complex conjugate transpose

$$\langle\psi| = (u^*, v^*)$$

and juxtaposition represents matrix multiplication. Thus if

$$|\psi_1\rangle = \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} \qquad |\psi_2\rangle = \begin{pmatrix} u_2 \\ v_2 \end{pmatrix}$$

then

$$\begin{aligned} \langle\psi_1|\psi_2\rangle &= (u_1^*, v_1^*) \begin{pmatrix} u_2 \\ v_2 \end{pmatrix} \\ &= u_1^* u_2 + v_1^* v_2 \end{aligned}$$

is a complex number giving the inner product of the states and

$$\begin{aligned} |\psi_1\rangle\langle\psi_2| &= \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} (u_2^*, v_2^*) \\ &= \begin{pmatrix} u_1 u_2^* & u_1 v_2^* \\ v_1 u_2^* & v_1 v_2^* \end{pmatrix} \end{aligned}$$

is an outer product giving a linear operation on $\mathcal{H}$, which maps a vector $|\xi\rangle$ to the vector $|\psi_1\rangle\langle\psi_2|\xi\rangle$. In particular $|\psi\rangle\langle\psi|$ is the operation of orthogonal projection into the one dimensional subspace of $\mathcal{H}$ in the direction of the unit vector $|\psi\rangle$.

The states of horizontal and vertical polarization, denoted $|\updownarrow\rangle$ and $|\leftrightarrow\rangle$, form an orthonormal basis of $\mathcal{H}$ (called the *rectilinear basis*) as do the states of diagonal polarization (at 45° and 135°) defined by

$$\begin{aligned} |\nearrow\rangle &= \Big(|\updownarrow\rangle + |\leftrightarrow\rangle\Big)/\sqrt{2} \\ |\searrow\rangle &= \Big(|\updownarrow\rangle - |\leftrightarrow\rangle\Big)/\sqrt{2}. \end{aligned}$$

Another important basis is $|\mathcal{B}\rangle, |\mathcal{B}^\perp\rangle$ defined in Section 3.1, which corresponds to linear polarization in directions midway between the above two bases. (In this paper we do not use photons with circular polarization, which correspond to linear combinations of $|\updownarrow\rangle$ and $|\leftrightarrow\rangle$ with complex, rather than real, coefficients.)

According to the formalism of quantum physics, any physical measurement on a photon is described in terms of a decomposition of the Hilbert space $\mathcal{H}$ into a family of orthogonal subspaces, one for each measurement outcome[1]. When a measurement is performed on a photon in state $|\psi\rangle$ each outcome may occur with probability given by the squared length of the projection of $|\psi\rangle$ into the corresponding subspace. As a result of the measurement, the original state $|\psi\rangle$ is obliterated and replaced by the projected vector (renormalized to unit length) corresponding to the seen outcome. Thus, we may associate a measurement to any orthogonal basis of $\mathcal{H}$. The rectilinear and diagonal bases have the following "conjugacy" property: if a measurement in one basis is carried out on either vector of the other basis, then the two outcomes always occur with probability $\frac{1}{2}$ and all information about the measured state is obliterated. Thus we get zero information about which of the two basis states was supplied. However if a basis vector is measured in the same basis, then it is identified with certainty and the state is left unchanged.

In certain situations, the state of a photon may not be describable by a "pure" state in $\mathcal{H}$ (that is, a vector in the Hilbert space). This occurs in two possible ways (see [Sud86] chapter 5) (a) the state is known only to the extent of being one of a "mixture" of states $|\psi_1\rangle, \ldots, |\psi_n\rangle$ with probabilities $p_1, \ldots, p_n$; (b) the photon is "entangled" with some other system and only the larger joint system has a description as a pure state (in a larger Hilbert space). The situation (a) occurs in step 4 of protocol *commit* — $\mathcal{A}lice$'s signal to $\mathcal{B}ob$ for the value $c_i = 0$ is one of two possible (non orthogonal) states randomly chosen with probability $\frac{1}{2}$ (and similarly for the value 1). To the general mixture in (a) above we associate the **density matrix**

$$\rho = \sum_i p_i \, |\psi_i\rangle\langle\psi_i|,$$

which is the average projection operator for the state distribution. For the special case of a pure state $|\psi\rangle$,

[1] Actually there is a more general notion of measurement, the so called POM or "Positive-Operator-Valued" measurements (see [Hel76, pp. 74–83] for details) which we omit for the sake of clarity. However it is straightforward to see that the argument in the proof of Theorem 3.1 also covers these more general measurements.

the density matrix is simply $|\psi\rangle\langle\psi|$. A straightforward extension of the measurement theory outlined above shows that the results of any physical measurement whatever on the mixture depend on the states and probabilities constituting the mixture **only** through the combination ρ. This is rather curious since the same density matrix may arise from very different mixtures of states. Thus any two such mixtures, having the same density matrix, cannot be distinguished by any physical measurement. We exploit this property of quantum measurement theory in Theorem 3.1, using the fact that a $(\frac{1}{2}, \frac{1}{2})$ mixture of the non orthogonal states $|\updownarrow\rangle$ and $|\searrow\rangle$ has the same density matrix as the $(1-\kappa^2, \kappa^2)$ mixture of the orthogonal states $|\mathcal{B}\rangle$ and $|\mathcal{B}^\perp\rangle$.

With regard to the situation (b), the most famous example of an entangled state is the joint state of two particles occurring in the Einstein–Podolsky–Rosen effect [EPR35]. If a photon is entangled with any other system then it can be shown ([Sud86] chapter 5) that the photon alone may always be described by a suitable density matrix, i.e. as far as measurements on the photon **alone** are concerned, it is physically indistinguishable from a suitable mixture of states, and follows the analysis of (a). This fact is relevant in the proof of Lemma 3.9 (where *Alice* may attempt to cheat by entangling her photons with each other or with some other system, which is precisely how the 1984 quantum bit commitment scheme of Bennett and Brassard could be broken [BB84]). Thus the density matrix formalism provides a uniform way of describing a single photon in the most general possible situation.