

Bit-commitment-based quantum coin flipping

Ashwin Nayak*

Computer Science Department, and Institute for Quantum Information, California Institute of Technology, Mail Code 256-80,
Pasadena, California 91125

Peter Shor†

AT&T Research Laboratories, 180 Park Avenue, Florham Park, New Jersey 07932

(Received 14 April 2002; published 8 January 2003)

In this paper we focus on a special framework for quantum coin-flipping protocols, *bit-commitment-based protocols*, within which almost all known protocols fit. We show a lower bound of $1/16$ for the bias in any such protocol. We also analyze a sequence of multiround protocols that tries to overcome the drawbacks of the previously proposed protocols in order to lower the bias. We show an intricate cheating strategy for this sequence, which leads to a bias of $1/4$. This indicates that a bias of $1/4$ might be optimal in such protocols, and also demonstrates that a more clever proof technique may be required to show this optimality.

DOI: 10.1103/PhysRevA.67.012304

PACS number(s): 03.67.Dd

I. QUANTUM COIN FLIPPING

Coin flipping is the communication problem in which two distrustful parties wish to agree on a common random bit, by “talking over the phone” [1]. When the two parties follow a protocol honestly, the bit they agree on is required to be 0 or 1 with equal probability. Ideally, they would also like that if any (dishonest) party deviates from the protocol, they do not agree on any particular outcome with probability more than $1/2$. It is known that ideal coin flipping is impossible in both the classical and the quantum setting [2,3]. In fact, in any classical protocol, one of the two parties can force the outcome of the protocol to a value of her choice with probability 1. In [4], Aharonov, Ta-Shma, Vazirani, and Yao showed that it is possible to design a quantum coin-flipping protocol in which no player can force the outcome of the protocol with probability more than a constant $1/2 + \epsilon$, with bias a constant $\epsilon < 1/2$. In other words, any cheating player in such protocols is detected with constant probability. Later, Ambainis [5] gave an improved protocol with bias at most $1/4$. A similar protocol with bias $1/4$ was independently discovered by Spekkens and Rudolph [6].

Formally, a quantum coin-flipping protocol with bias ϵ is a two-party communication game in the style of [7], in which the players start with no input and compute values $c_A, c_B \in \{0,1\}$, respectively (or declare that the other player is cheating). The protocol satisfies the following additional properties.

(1) If both players are honest (i.e., follow the protocol), then they agree on the outcome of the protocol: $c_A = c_B$, and the outcome is 0 or 1 with equal probability: $\Pr(c_A = c_B = b) = 1/2$, for $b \in \{0,1\}$.

(2) If one of the players is honest (i.e., the other player may deviate arbitrarily from the protocol in his or her local computation), then the outcome of the protocol has bias at most ϵ : for any $b \in \{0,1\}$, $\Pr(c_A = c_B = b) \leq 1/2 + \epsilon$.

Almost all quantum coin-flipping protocols with a bias that is provably smaller than a half [4–6] are based on the notion of bit commitment. In other words, they have the following form, when the parties flipping the coin, Alice and Bob, are honest.¹

Protocol scheme II

The following steps describe the protocol scheme II.

(1) First, Alice and Bob each pick a random bit, a and b , respectively, and privately construct states ρ_a and σ_b . The states are over three sets of qubits, a *commitment* part, a *revelation* part, and a *verification* part. The revelation part consists of one qubit that contains the value of the bit picked.

(2) Next, they *commit* to their respective bits a and b by sending each other the commitment part of their states ρ_a and σ_b . They may do this over several rounds of communication, in which they send messages alternately.

(3) Then, they *reveal* to the other party the bits a, b they picked (in some order), and follow that up by sending the rest of the states ρ_a, σ_b (the verification part). This may again be over several rounds of communication. This allows each party to check via suitable measurements that the state with which the other, Alice (or Bob), committed to her (his) bit is indeed consistent with a (b).

The result of the protocol is $c = a \oplus b$, if neither player is detected cheating during the third (verification) stage.

For example, in the case of the protocol in [5], Alice uses the right half of the following state to commit to her bit a , and the left half to help Bob check her commitment:

$$\rho_a = \sum_{s=0,1} \frac{1}{2} |a,s\rangle \langle a,s| \otimes |\psi_{a,s}\rangle \langle \psi_{a,s}|,$$

where

¹We were recently informed [8] of a protocol of a different kind that also achieves a bias of $1/4$.

*Email address: nayak@cs.caltech.edu

†Email address: shor@research.att.com

$$|\psi_{a,s}\rangle = \frac{1}{\sqrt{2}}[|0\rangle + (-1)^s|a+1\rangle].$$

Bob skips the commit stage, and directly reveals his bit b . In the final stage, Bob checks that the state of the qutrit Alice sent in the first round is indeed consistent with a,s , by measuring it in a basis containing $|\psi_{a,s}\rangle$.

Protocols of the form described above may be recast in the following terms. First, Alice and Bob each pick a random bit. Then, they successively send each other qubits which do not depend on the qubits sent by the other party in the previous rounds. The qubits sent by Alice represent a commitment to her bit a along with auxiliary information required by Bob to check if she is cheating. The same is true for the qubits sent by Bob. Thus, after all communication is over, the states Alice and Bob sent to each other for 0 and 1 are perfectly distinguishable. They measure the states received from the other party (possibly with some ancilla) according to a von Neumann measurement to determine the bits a,b or to detect cheating. The outcome of the coin flip is $a \oplus b$ if no cheating is detected. In this description we have assumed, without loss of generality [9], that all measurements are done at the end. Note that the description is also slightly more general in that the players may not explicitly reveal the bits they intend to commit to, and the commitment and the verification stages may be interleaved. We will henceforth refer to such protocols as bit-commitment-based protocols.

In this paper we study coin-flipping protocols that fall into the special framework described above, that of bit-commitment-based protocols. We show a lower bound of $1/16$ for the bias in any such protocol (Theorem 2.2). This provides a single proof that these protocols, including the one proposed in [10], cannot lead to arbitrarily small bias. Tokunaga [11] has independently attempted, albeit unsuccessfully, to show a lower bound of $1/16$ for arbitrary protocols using the same cheating strategy that leads to our lower bound. Spekkens and Rudolph (Corollary 1 in [6]) have shown a lower bound of $0.19\dots$ for protocols as in scheme II above, in which only one party commits to a bit—the other directly sends his bit in the revelation phase of the protocol. Our lower bound thus applies to a more general kind of protocol.

Next, we analyze a sequence of protocols that tries to overcome the drawbacks of the previously proposed protocols, and also tries to circumvent the cheating strategy that leads to this above lower bound. We show an intricate cheating strategy for the sequence of protocols, which leads to a bias of $1/4$. This indicates that a bias of $1/4$ might be optimal in such protocols, and also demonstrates that a cleverer proof technique than the one used in Theorem 2.2 is required to show this optimality.

Recently, Kitaev has shown a lower bound of $(1/\sqrt{2}) - (1/2) = 0.2071\dots$ for the bias in arbitrary quantum coin-flipping protocols [12]. This is bigger than the bias of $1/16 = 0.0625$ that we show (and it applies to any coin-flipping protocol). In more detail, Kitaev's result states that the product of the optimum cheating probabilities for the two parties is at least $1/2$. We leave it as an open question whether these

optima sum to at least $3/2$, as is the case in all known protocols, including the ones considered in this paper.

Kitaev's $(1/\sqrt{2}) - (1/2)$ lower bound does not seem to apply to quantum games in which the two parties involved compete to “win” by getting a particular value of the coin as the outcome (say Alice wins if the outcome is 0, and Bob wins if the outcome is 1). This is also known as weak coin flipping in the literature. Protocols for weak coin flipping with bias less than $1/4$ have been discovered. Kerenidis and Nayak [13] have shown a protocol with cheating probability at most $0.739\dots$. Ambainis [14] and Spekkens and Rudolph [15] have shown protocols with an even lower cheating probability of at most $1/\sqrt{2} = 0.707\dots$. No lower bound is known for the bias in weak coin flipping; arbitrarily small bias may still be possible.

II. A LOWER BOUND ON THE BIAS

We first show that any bit-commitment-based protocol may be reduced to an extremely simple protocol of the same type, with bias at most that in the original protocol, and by increasing the number of rounds by at most 1.

Lemma 2.1. For any bit-commitment-based coin-flipping protocol \mathcal{P} (of the form II, or more generally, as described in Sec. I), there is another such protocol \mathcal{P}' so that we have the following.

(1) The states ρ_a, σ_b are pure: $\rho_a = |\psi_a\rangle\langle\psi_a|$ and $\sigma_b = |\phi_b\rangle\langle\phi_b|$, where $|\psi_0\rangle \perp |\psi_1\rangle$ and $|\phi_0\rangle \perp |\phi_1\rangle$.

(2) Alice measures the state she received from Bob according to the measurement given by the operators $P_0 = |\phi_0\rangle\langle\phi_0|$, $P_1 = |\phi_1\rangle\langle\phi_1|$, and $P_c = I - P_0 - P_1$, to determine Bob's bit or to check if he is cheating. Bob does an analogous measurement given by Q_0, Q_1, Q_c on the state he receives from Alice.

(3) The bias is at most the bias of \mathcal{P} .

Proof. The protocol \mathcal{P}' is obtained by stipulating that the players use a fixed purification $|\psi_a\rangle, |\phi_b\rangle$ of the states ρ_a, σ_b used in the original protocol \mathcal{P} . Since the states ρ_0, ρ_1 are perfectly distinguishable, their purifications are orthogonal. Similarly with ϕ_0, ϕ_1 . All but the last two rounds of \mathcal{P}' are as in the original one. We stipulate that the players send the entire (purified) state ψ_a or ϕ_b in \mathcal{P}' . Thus, the last player to send a message in \mathcal{P} sends the qubits used in the purification in the penultimate round of \mathcal{P}' . In the final round, the other player sends the qubits used in purifying her state. We also alter the measurement to the ones mentioned in the lemma.

We now show that this modification of the protocol results in bias at most that in the original one. We do this by showing that any cheating strategy of a player in the modified protocol \mathcal{P}' that achieves a bias of ϵ leads to a cheating strategy in the original protocol \mathcal{P} with at least the same bias.

For concreteness, we consider a cheating strategy for Alice in the protocol \mathcal{P}' . (The argument for the case of Bob is similar.) In her strategy in \mathcal{P} , Alice acts exactly as in the original strategy, except that she is not required to send the “purification qubits” meant for her last message in \mathcal{P}' . We need only show that the probability with which the measurement in \mathcal{P} yields 0 or 1 dominates the same probability for \mathcal{P}' .

Suppose that Bob uses a von Neumann measurement

given by the projection operators R_0, R_1, R_c in \mathcal{P} . We concentrate on the probability that Alice can convince Bob that she had picked $a=0$. The other probability may be bounded similarly. Since $R_0\rho_0R_0=\rho_0$, the purification $|\psi_0\rangle$ lies in the range of $I\otimes R_0$. The states of the qubits sent to Bob by Alice and her private qubits in \mathcal{P}' are together given by some mixed state $\sum_j \mu_j |\xi_j\rangle\langle\xi_j|$ (where the states $|\xi_j\rangle$ are over the space of ρ_a and the purification space). It thus suffices to show that for any state $|\xi\rangle \in \{|\xi_j\rangle\}$,

$$\|I\otimes R_0|\xi\rangle\|^2 \geq \|\langle\psi_0|\xi\rangle\|^2.$$

Note that the left-hand side (LHS) is the squared norm of the projection of $|\xi\rangle$ onto the range of $I\otimes R_0$, and the right-hand side (RHS) is the squared norm of the projection of the same vector onto a *subspace* of that range, the one-dimensional space spanned by $|\psi_0\rangle$. The inequality is then immediate. This shows that Alice can achieve at least the same bias as in \mathcal{P}' . ■

This simple characterization of bit-commitment-based protocols proves useful in the analysis of the smallest bias achievable with such protocols. Using this, we show that coin-flipping protocols based on bit-commitment cannot achieve arbitrarily small bias.

Theorem 2.2. In any quantum coin-flipping protocol based on bit commitment, one of the parties can achieve probability of cheating at least $9/16$.

Proof. As shown in Lemma 2.1, any such protocol between honest parties may be viewed as follows: first, Alice and Bob construct the states $|\psi_a\rangle$ and $|\phi_b\rangle$, respectively, corresponding to the random bits a and b . Then, they send each other a part of the states $|\psi_a\rangle, |\phi_b\rangle$ a few qubits at a time. Finally, they measure the qubits received from each other using projections P_0, P_1, P_c and Q_0, Q_1, Q_c .

Let $\rho_{a,i} = \text{Tr}_{A_i}(|\psi_a\rangle\langle\psi_a|)$ be the state sent to Bob by Alice by round i (so A_i are the qubits of $|\psi_a\rangle$ still with Alice after the i th round). Let $\sigma_{b,i}$ be the corresponding state sent to Alice by Bob.

Let there be n rounds in all. Let $F_{A,i} = F(\rho_{0,i}, \rho_{1,i})$ and similarly $F_{B,i} = F(\sigma_{0,i}, \sigma_{1,i})$. Here, $F(\cdot, \cdot)$ is the *fidelity* function as defined in [16]. So $F_{A,0} = F_{B,0} = 1$. Note that $Q_a\rho_{a,n}Q_a = \rho_{a,n}$, and similarly $P_b\sigma_{b,n}P_b = \sigma_{b,n}$, so that $F_{A,n} = F_{B,n} = 0$.

Lemma 2.3. Consider a protocol with honest players. For any constant $0 \leq \alpha \leq 1$, there is a player, say Alice, and a round $k \geq 0$ such that the states she sends to Bob by the k th round on $a=0$ and 1 have fidelity at least α , and the fidelity of the states she receives from Bob by the next round have fidelity at most α . In other words, $F_{A,k} \geq \alpha$ and $F_{B,k+1} \leq \alpha$.

Proof. The case $\alpha=1$ is trivial. Let $\alpha < 1$. Note that both $F_{A,i}, F_{B,i}$ decrease from 1 to 0 through the course of the protocol. Consider the first round $i \geq 1$ such that one of these, say $F_{B,i}$, becomes $\leq \alpha$. Round $k = i - 1$ satisfies the property we seek. ■

We will devise a cheating strategy for a player as given by Lemma 2.3 above for $\alpha = 1/4$. Say this player is Alice, and the round identified in the lemma is k . There is a unitary

transformation on the qubits A_k which achieves maximum fidelity between $\rho_{0,k}, \rho_{1,k}$ [16], i.e.,

$$|\langle\psi_0|U|\psi_1\rangle|^2 = F_{A,k} = F(\rho_{0,k}, \rho_{1,k}) \geq \frac{1}{4}.$$

Moreover, we may assume that $\langle\psi_0|U|\psi_1\rangle$ is real and non-negative.

Alice may cheat as follows. She constructs the state

$$|\xi\rangle = \frac{|\psi_0\rangle + U|\psi_1\rangle}{\|\psi_0 + U\psi_1\|}$$

and uses this state in the protocol until the $(k+1)$ th round. After this round, she makes the best possible measurement to distinguish $\sigma_{0,k+1}$ and $\sigma_{1,k+1}$ to guess the value of b . If her guess g is 0, she proceeds with the rest of the protocol. Otherwise, she applies U^\dagger to her part of $|\xi\rangle$ (the qubits in A_k) and then completes the protocol.

Lemma 2.4. $\Pr(c=0|g=b) \geq (1 + \sqrt{F_{A,k}})/2$.

Proof. First, note that

$$\|\psi_0 + U\psi_1\|^2 = \|\psi_0\|^2 + \|\psi_1\|^2 + 2\langle\psi_0|U|\psi_1\rangle = 2(1 + \sqrt{F_{A,k}}).$$

Suppose $b=0$. (The other case is similar.) Note that the probability that Alice succeeds in getting the outcome $c=0$ given that she guesses the value of b correctly is

$$\begin{aligned} \|\mathcal{Q}_0\xi\|^2 &= \frac{\|\mathcal{Q}_0\psi_0 + \mathcal{Q}_0U\psi_1\|^2}{\|\psi_0 + U\psi_1\|^2} \\ &= \frac{\|\mathcal{Q}_0\psi_0\|^2 + \|\mathcal{Q}_0U\psi_1\|^2 + 2\langle\psi_0|\mathcal{Q}_0U|\psi_1\rangle}{2(1 + \sqrt{F_{A,k}})}. \end{aligned}$$

Now $\mathcal{Q}_0|\psi_0\rangle = |\psi_0\rangle$, since $|\psi_0\rangle$ is the state Alice would have used if she were honest. So the first term in the numerator above is 1, and the last term is $2\sqrt{F_{A,k}}$. The second term may be bounded from below by noting that since $|\psi_0\rangle$ belongs to the range of \mathcal{Q}_0 , the square norm of the projection $\mathcal{Q}_0U|\psi_1\rangle$ is at least $\langle\psi_0|U|\psi_1\rangle^2 = F_{A,k}$. Thus, the probability of cheating is at least

$$\frac{1 + F_{A,k} + 2\sqrt{F_{A,k}}}{2(1 + \sqrt{F_{A,k}})} = \frac{1 + \sqrt{F_{A,k}}}{2},$$

which is the bound claimed. ■

The probability that Alice correctly guesses b is (using Bayes's strategy)

$$\Pr(g=b) = \frac{1}{2} + \frac{\|\sigma_{0,k+1} - \sigma_{1,k+1}\|_{\text{tr}}}{4}.$$

By a result of Fuchs and van de Graaf [17],

$$\|\sigma_{0,k+1} - \sigma_{1,k+1}\|_{\text{tr}} \geq 2(1 - \sqrt{F_{B,k+1}}) \geq 2(1 - \frac{1}{2}) = 1.$$

The net probability that Alice succeeds in biasing the coin towards 0 is therefore

$$\Pr(c=0) \geq \Pr(c=0|g=b)\Pr(g=b) \geq \frac{3}{4} \frac{3}{4} = \frac{9}{16}.$$

This proves the theorem. ■

III. A SEQUENCE OF HIGHLY INTERACTIVE PROTOCOLS

In this section we look at a sequence of bit-commitment-based protocols in which Alice and Bob very gradually send each other information about their bits in the commit stage. Intuitively, such protocols seem to be good candidates for achieving bias much smaller than 1/4, since a dishonest player does not get much information about the other’s bit until a significant number of rounds have elapsed, and he would have heavily committed to some bit by then. However, this intuition appears to be mistaken, and we give intricate cheating strategies for each of the players with which at least one of them can achieve bias at least as high as 1/4. This suggests that the optimal bias for this kind of protocol might be 1/4, and also that proving this optimality might require ideas more sophisticated than those in Theorem 2.2.

Define, for $x, s \in \{0,1\}$,

$$|\psi(x,s)\rangle = \sqrt{1-\epsilon}|0\rangle + (-1)^s \sqrt{\epsilon}|x+1\rangle.$$

These states provide the best trade-off between how much information they reveal and how much cheating in commitment they allow.

The protocol \mathcal{P}_n , $n \in \{1,2,3,\dots\}$, goes as follows. Alice picks $a \in_R \{0,1\}$, and Bob picks $b \in_R \{0,1\}$. Then they alternately send each other, for a total of n rounds, the states $|\psi(a,s)\rangle$ and $|\psi(b,s)\rangle$, respectively, for independently chosen random sign s , starting with Alice. The last player to receive such a state then reveals the bit he/she chose, followed by the other player. Then, they reveal the signs used in their states in the opposite order, and the other party checks the state he/she received against the claimed bit and sign. If no cheating is detected, the players declare the $c = a \oplus b$ as the result of the protocol.

More formally,

(1) For $i = 1, 2, 3, \dots, n$, if i is odd, Alice picks $s_i \in_R \{0,1\}$ and sends $|\psi(a, s_i)\rangle$ to Bob, or if i is even, Bob picks $s_i \in_R \{0,1\}$ and sends $|\psi(b, s_i)\rangle$ to Alice.

(2) If n is even, Alice sends a to Bob, and then Bob sends b to Alice. Otherwise, if n is even they reveal their bits in the opposite order.

(3) For $i = n, n-1, n-2, \dots, 1$, the player that picked s_i reveals it to the other player. In other words, the “signs” used in the states are revealed in the opposite order: If i is odd, Alice sends s_i to Bob, else Bob sends s_i to Alice. The player that receives this bit checks via a measurement that the state sent to her/him in the i th round of the protocol is indeed consistent with the bit and the sign that the other player sent. Note that the order of revealing signs is designed so that the naïve strategy of reusing a state that a player got in a previous round does not work.

(4) If all the checks are passed, the outcome of the protocol is $c = a \oplus b$.

Before we give cheating strategies for these protocols, we analyze general versions of \mathcal{P}_1 and \mathcal{P}_2 . This illustrates the main approach taken in the strategies for \mathcal{P}_n , $n \geq 3$.

A. The three-round version

We start by analyzing the protocol \mathcal{P}_1 with one round of commitment. This happens to be a parametrized version of the three-round protocol due to Ambainis [5]. We prove a property of this protocol that helps us analyze protocols with more rounds.

The protocol may be described with a parameter $\alpha \in [0, \pi]$ [such that $\tan(\alpha/2) = \sqrt{\epsilon/(1-\epsilon)}$] as follows:

(1) Alice picks $a, s \in_R \{0,1\}$, and sends Bob the state $|\psi_{a,s}\rangle$, where the state is defined as follows:

$$|\psi_{a,s}\rangle = \cos \frac{\alpha}{2} |0\rangle + (-1)^s \sin \frac{\alpha}{2} |a+1\rangle. \tag{1}$$

(2) Bob picks $b \in_R \{0,1\}$ and sends it to Alice.

(3) Alice then reveals the bits a, s to Bob, who checks for consistency with the state initially sent by Alice.

The output of the protocol is given by $c = a \oplus b$, if all the checks are passed.

Lemma 3.1. If Bob is honest, then $\Pr(c=0) \leq \frac{1}{4}(3 + \cos \alpha)$.

Proof. A cheating Alice creates a qutrit in some state (to be sent in the first round), and retains some qubits entangled with this qutrit. Without loss of generality we may assume that in the third round, Alice sends $a = b$, which is the bit that Bob sends in the second round. This only increases her chance of getting outcome 0. The following lemma now bounds the probability that Alice succeeds in cheating Bob, given the bit b .

Lemma 3.2. Suppose that Alice and Bob share a joint state $\sum_{i=0}^2 |\phi_i\rangle |i\rangle$. Alice now does some local computation and sends Bob a sign bit s , and Bob does a measurement to check if the state of his qutrit is $|\psi_{a,s}\rangle$. The probability with Bob’s test is passed is at most $(\|\phi_0\| \sqrt{1-\delta} + \|\phi_{a+1}\| \sqrt{\delta})^2$, where $\delta = \sin^2(\alpha/2)$. In other words, this probability is at most the fidelity $F(\rho, \rho_a)$, where ρ is the density matrix of Bob’s part of the joint state, if measured in the $|0\rangle, |1\rangle, |2\rangle$ basis, and ρ_a is the density matrix he would have if Alice were honest, i.e., $\rho_a = (1/2)(|\psi_{a,0}\rangle\langle\psi_{a,0}| + |\psi_{a,1}\rangle\langle\psi_{a,1}|)$.

Proof. We prove the statement for $a = 0$; the other case is similar. Note that in this case, Bob rejects with probability 1 the part of his state where his qutrit is $|2\rangle$. Thus we may restrict ourselves to the state $|\phi_0\rangle|0\rangle + |\phi_1\rangle|1\rangle$.

Suppose Alice applies a transformation U on an ancilla $|\bar{0}\rangle$ and her part of the shared state, and that $U|\phi_i\rangle = |\phi_{i0}\rangle|0\rangle + |\phi_{i1}\rangle|1\rangle$, where the trailing qubit contains the sign that Alice sends to Bob. When the sign Alice sends is s , the (unnormalized) density matrix of the Bob’s state is $\sum_{x,y} |x\rangle\langle y| \langle\phi_{xs}|\phi_{ys}\rangle$. Now the probability that this state passes Bob’s test may be seen to be

$$(1-\delta)\|\phi_{0s}\|^2 + \delta\|\phi_{1s}\|^2 + (-1)^s 2\sqrt{\delta(1-\delta)} \operatorname{Re}\langle\phi_{0s}|\phi_{1s}\rangle.$$

Summing up over $s = 0, 1$, and using the fact that $\|\phi_{i0}\|^2 + \|\phi_{i1}\|^2 = \|\phi_i\|^2$, we see that the net probability is

$$(1-\delta)\|\phi_0\|^2 + \delta\|\phi_1\|^2 + 2\sqrt{\delta(1-\delta)}\operatorname{Re}[\langle\phi_{00}|\phi_{10}\rangle - \langle\phi_{01}|\phi_{11}\rangle].$$

Now,

$$\begin{aligned} & |\operatorname{Re}[\langle\phi_{00}|\phi_{10}\rangle - \langle\phi_{01}|\phi_{11}\rangle]| \\ & \leq \|\phi_{00}\| \cdot \|\phi_{10}\| + \|\phi_{01}\| \cdot \|\phi_{11}\| \\ & \leq (\|\phi_{00}\|^2 + \|\phi_{01}\|^2)^{1/2} (\|\phi_{10}\|^2 + \|\phi_{11}\|^2)^{1/2} \\ & = \|\phi_0\| \cdot \|\phi_1\|. \end{aligned}$$

Therefore, the net probability of passing Bob's test is at most

$$(\sqrt{1-\delta}\|\phi_0\| + \sqrt{\delta}\|\phi_1\|)^2,$$

as claimed. \blacksquare

Averaging over the two choices for the bit b which Bob picks randomly, we see that Alice's cheating probability is

$$\Pr(c=0) \leq \frac{F(\rho_0, \rho) + F(\rho_1, \rho)}{2}, \quad (2)$$

where ρ is the state of Bob after the first round, and ρ_0, ρ_1 are the analogous states corresponding to $b=0,1$ respectively, if Alice were honest:

$$\rho_a = \frac{1}{2}(|\psi_{a,0}\rangle\langle\psi_{a,0}| + |\psi_{a,1}\rangle\langle\psi_{a,1}|),$$

and $F(\sigma_0, \sigma_1) = \|\sqrt{\sigma_0}\sqrt{\sigma_1}\|_{\text{tr}}^2$ denotes the fidelity of two density matrices [16].

We show in Lemma 3.3 below that the expression in Eq. (2) is bounded above by

$$\frac{1}{2}[1 + F(\rho_0, \rho_1)^{1/2}].$$

Since $F(\rho_0, \rho_1) = \cos^4(\alpha/2)$, the bound on Alice's cheating probability follows. \blacksquare

We now prove the lemma mentioned above. The lemma has also been proved by Spekkens and Rudolph (Lemma 2 in [6]).

Lemma 3.3. For any two density matrices σ_0, σ_1 ,

$$\max_{\sigma} F(\sigma_0, \sigma) + F(\sigma_1, \sigma) = 1 + F(\sigma_0, \sigma_1)^{1/2}.$$

Proof. Let σ be the density matrix that achieves the maximum, and let $|\phi_x\rangle$ be a purification of σ_x , for $x=0,1$. Let $|\xi_x\rangle$ be the purification of σ that achieves maximum fidelity with σ_x [16]:

$$F(\sigma_0, \sigma) = |\langle\phi_x|\xi_x\rangle|^2.$$

Since $|\xi_0\rangle, |\xi_1\rangle$ are the purifications of the same density matrix σ , there is a local unitary operator U such that $|\xi_1\rangle = U|\xi_0\rangle$. Now,

$$\begin{aligned} F(\sigma_0, \sigma) + F(\sigma_1, \sigma) &= |\langle\phi_0|\xi_0\rangle|^2 + |\langle\phi_1|\xi_1\rangle|^2 \\ &= |\langle\phi_0|\xi_0\rangle|^2 + |\langle\phi_1|U|\xi_0\rangle|^2 \\ &\leq \max_{|\xi\rangle} \{|\langle\phi_0|\xi\rangle|^2 + |\langle\phi_1|U|\xi\rangle|^2\} \\ &= 1 + |\langle\phi_1|U|\phi_0\rangle| \end{aligned} \quad (3)$$

$$\begin{aligned} &\leq 1 + \max_{\text{local } U} |\langle\phi_1|U|\phi_0\rangle| \\ &= 1 + F(\sigma_0, \sigma_1)^{1/2}. \end{aligned} \quad (4)$$

Equation (3) above follows by noticing that the state $|\xi\rangle$ that achieves the maximum is the vector that bisects the angle between $|\phi_0\rangle$ and $U|\phi_1\rangle$. Another way of getting the bound is by noticing that the expression is the maximum eigenvalue of the matrix $|\phi_0\rangle\langle\phi_0| + U|\phi_1\rangle\langle\phi_1|U^\dagger$. The last step—Eq. (4)—follows from a characterization of fidelity due to Jozsa [16].

To see that the above inequality is tight, we may construct the density matrix σ above as follows. Consider any two purifications $|\zeta_x\rangle$ ($x=0,1$) of σ_x that achieve the fidelity of the density matrices: $|\langle\zeta_0|\zeta_1\rangle| = F(\sigma_0, \sigma_1)^{1/2}$. Let

$$|\zeta\rangle = \frac{|\zeta_0\rangle + |\zeta_1\rangle}{\|\zeta_0 + \zeta_1\|}.$$

It is now not hard to see that $\|\zeta_0 + \zeta_1\|^2 = 2[1 + \sqrt{F(\sigma_0, \sigma_1)}]$, and that $|\langle\zeta|\zeta_x\rangle| = 1 + \sqrt{F(\sigma_0, \sigma_1)}$. The state σ obtained by tracing out the purification space from $|\zeta\rangle$ thus achieves the maximum. \blacksquare

B. The five-round protocol

Next, we give a tight analysis for a general five-round version of the protocol \mathcal{P}_2 described above. This version of the protocol still does not improve over the bias of 1/4 achieved by the three-round protocol of [5]. However, it suggests a better cheating strategy for the many-rounds version than the one given in the proof of Theorem 2.2.

The version of the protocol \mathcal{P}_2 we consider has the following five rounds with honest players.

(1) Alice picks $a, s \in_R \{0,1\}$ and sends Bob the state $|\psi_{a,s}\rangle$, where the state is defined as follows:

$$|\psi_{a,s}\rangle = \cos\frac{\alpha}{2}|0\rangle + (-1)^s \sin\frac{\alpha}{2}|a+1\rangle \quad (5)$$

for an angle α to be specified later.

(2) Similarly, Bob picks $b, s' \in_R \{0,1\}$ and sends Alice the state $|\phi_{b,s'}\rangle$, where the state is defined as follows:

$$|\phi_{b,s'}\rangle = \cos\frac{\beta}{2}|0\rangle + (-1)^{s'} \sin\frac{\beta}{2}|b+1\rangle, \quad (6)$$

for an angle β to be specified later.

(3) Alice then reveals the bit a to Bob.

(4) Bob reveals both b and s' to Alice, and Alice verifies, by an appropriate measurement, that the state sent by Bob is consistent with b, s' .

(5) Alice now discloses s as well, and Bob verifies that the state sent by Alice in the first round is consistent with a, s .

The output of the protocol, the coin flip c , is given by the exclusive-or of two bits, $c = a \oplus b$ (provided no cheating is detected).

In the case where the players are honest, $\Pr(c=0) = \Pr(c=1) = 1/2$. Below we prove an upper bound on the probability that any player can achieve (by deviating from the protocol) for an outcome of their choice. In the following discussion, we will assume, without loss of generality, that a dishonest player prefers the outcome $c = 0$. First, we prove a bound on the probability that a dishonest Bob can achieve, provided Alice is honest.

Lemma 3.4. If Alice is honest, then $\Pr(c=0) \leq 1 - \frac{1}{2} \cos^2(\alpha/2) \sin^2(\beta/2)$.

Proof. We claim that Bob's optimal cheating strategy is to measure the state received from Alice in the standard basis, and then commit to a state according to the outcome. If he observes $|1\rangle$ or $|2\rangle$, he can cheat with probability 1 in the rest of the protocol. In case he observes $|0\rangle$, his probability of cheating successfully is bounded by a constant less than one. This gives us the bound, as explained below.

Note that the last round is of no consequence to Bob's cheating strategy, and we may trace the sign bit s out after Alice sends the first message (and eliminate the last round). The protocol then becomes equivalent to one in which the first round takes the following form: Alice picks $b \in_R \{0,1\}$, and sends $|0\rangle$ with probability $\cos^2(\alpha/2)$, and $|b+1\rangle$ with probability $\sin^2(\alpha/2)$. This reduces the protocol to a convex combination of two protocols, with weights $\cos^2(\alpha/2)$ and $\sin^2(\alpha/2)$. In the first protocol, the first message is the same regardless of the value of b , and in the second, the first message reveals b entirely.

In the first case, the protocol reduces to a three-round protocol of the type studied in Sec. III A, with the role of Alice and Bob reversed. We may now use Lemma 3.1 to bound the probability that Bob can cheat by $(3 + \cos \beta)/4$.

In the second case, the protocol becomes trivial, and Bob can cheat with probability 1.

The probability of convincing Alice that $c=0$ is thus bounded by

$$\frac{3 + \cos \beta}{4} \cos^2 \frac{\alpha}{2} + \sin^2 \frac{\alpha}{2},$$

which reduces to the bound we seek. ■

Bob can easily achieve the probability bound stated in the lemma by following a cheating strategy as in Ref. [5], Lemma 10. This shows that the analysis in our proof is optimal.

We now turn to the case where Alice is dishonest. The following lemma bounds Alice's cheating probability.

Lemma 3.5. If Bob is honest, then $\Pr(c=0) \leq \frac{1}{2} [1 + \cos^2(\alpha/2) \sin^2(\beta/2)]$.

Proof. Suppose that Alice commits with one-half of the following state in the first round:

$$|\xi_0\rangle|0\rangle + |\xi_1\rangle|1\rangle + |\xi_2\rangle|2\rangle,$$

where the states $|\xi_i\rangle$ are held by Alice, and are such that $\sum_i \|\xi_i\|^2 = 1$.

When she receives a qutrit from Bob in the second round, Alice applies a unitary transformation U to some ancilla, her part of the above shared state, and the qutrit sent by Bob. This computes a bit a to send to Bob, and represents the bit she claims to have committed to in the first round. Let U have the following behavior:

$$|\bar{0}\rangle|\xi_i\rangle|j\rangle \mapsto |\xi_{ij0}\rangle|0\rangle + |\xi_{ij1}\rangle|1\rangle.$$

Suppose Bob had sent her a qutrit in state $|\phi_{b,s'}\rangle$ corresponding to the random bits b, s' . Their joint state after the third round is then

$$\sum_{i=0}^2 \sum_{a=0,1} (\sqrt{1-\epsilon}|\xi_{i0a}\rangle + (-1)^{s'}\sqrt{\epsilon}|\xi_{i,b+1,a}\rangle) \otimes |a,i\rangle_B,$$

where $\epsilon = \sin^2(\beta/2)$. The part of the joint state corresponding to the final outcome 0 is (when $a=b$):

$$\sum_{i=0}^2 (\sqrt{1-\epsilon}|\xi_{i0b}\rangle + (-1)^{s'}\sqrt{\epsilon}|\xi_{i,b+1,b}\rangle) \otimes |b,i\rangle_B.$$

From Lemma 3.2, the probability that Alice now passes Bob's final check is bounded by

$$\begin{aligned} & (\sqrt{1-\delta}\|\zeta_0(s')\| + \sqrt{\delta}\|\zeta_{b+1}(s')\|)^2 \\ &= (1-\delta)\|\zeta_0(s')\|^2 + \delta\|\zeta_{b+1}(s')\|^2 \\ & \quad + 2\sqrt{\delta(1-\delta)}\|\zeta_0(s')\| \cdot \|\zeta_{b+1}(s')\|, \end{aligned}$$

where

$$|\zeta_i(s')\rangle = \sqrt{1-\epsilon}|\xi_{i0b}\rangle + (-1)^{s'}\sqrt{\epsilon}|\xi_{i,b+1,b}\rangle.$$

Averaging it over the choices of s' that Bob makes, this probability may be bounded as

$$(1-\delta)\mu_0 + \delta\mu_{b+1} + 2\sqrt{\delta(1-\delta)}\mu_0\mu_{b+1},$$

where

$$\mu_i = (1-\epsilon)\|\xi_{i0b}\|^2 + \epsilon\|\xi_{i,b+1,b}\|^2.$$

Here, we have used the parallelogram law, i.e., $\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2)$, and the Cauchy-Schwartz inequality. Note that we may bound $\|\xi_{b+1,j,b}\|$ by $\|\xi_{b+1}\|$ and $\|\xi_{0,b+1,b}\|$ by $\|\xi_0\|$. So the probability above may be bounded as

$$\begin{aligned} & (1-\delta)[(1-\epsilon)\|\xi_{00b}\|^2 + \epsilon\|\xi_0\|^2] + \delta\|\xi_{b+1}\|^2 \\ & \quad + 2\sqrt{\delta(1-\delta)}\|\xi_{b+1}\|[(1-\epsilon)\|\xi_{00b}\|^2 + \epsilon\|\xi_0\|^2]^{1/2}. \end{aligned}$$

Averaging it over the choice of b that Bob makes, and using the Cauchy-Schwartz inequality once again, we see that this probability is bounded by

$$\frac{1}{2}[(1-\delta)(1+\epsilon)\|\xi_0\|^2 + \delta(1-\|\xi_0\|^2)] \\ + 2\sqrt{\delta(1-\delta)(1+\epsilon)}\|\xi_0\|\{1-\|\xi_0\|^2\}^{1/2}.$$

The maximum value of this expression as $\|\xi_0\|$ varies between 0 and 1 is

$$\frac{1}{2}[1 + \epsilon(1-\delta)] = \frac{1}{2}\left(1 + \sin^2\frac{\beta}{2}\cos^2\frac{\alpha}{2}\right).$$

This completes the proof of the lemma. \blacksquare

Next, we describe a cheating strategy for Alice that achieves the outcome of her choice with probability as high as in the upper bound above, showing that our analysis is optimal.

Lemma 3.6. If Bob is honest, Alice can achieve $\Pr(c=0) \geq \frac{1}{2}[1 + \cos^2(\alpha/2)\sin^2(\beta/2)]$.

Proof. Alice constructs the following entangled state and sends one half of it to Bob in the first round:

$$|\xi\rangle = \sqrt{1-\lambda}|0,0\rangle + \sqrt{\lambda}|1,1\rangle + \sqrt{\frac{\lambda}{2}}|2,2\rangle,$$

where

$$\lambda = \frac{\sin^2\frac{\alpha}{2}}{\left(1 + \sin^2\frac{\beta}{2}\right)\cos^2\frac{\alpha}{2} + \sin^2\frac{\alpha}{2}},$$

and so the density matrix of the qutrit with Bob is $\rho = \text{diag}(1-\lambda, \lambda/2, \lambda/2)$ in the 0,1,2 basis.

After Alice receives a qutrit from Bob in the second round, she applies a unitary transformation to “guess” a value for b to maximize her chances of getting $c=0$. This transformation acts on the qutrit she received from Bob, her entangled qutrit from the first round, and an ancilla qubit. It can be written as

$$|0\rangle\langle 0| \otimes (|0\rangle\langle 0| \otimes H + |1\rangle\langle 1| \otimes I + |2\rangle\langle 2| \otimes \sigma_x) \\ + |1\rangle\langle 1| \otimes I \otimes I + |2\rangle\langle 2| \otimes I \otimes \sigma_x.$$

where H is the Hadamard transform, and σ_x is the Pauli “bit flip” matrix. In other words, Alice guesses $a=b$, if the qutrit from Bob reveals the identity of b . Otherwise, if she committed with a $|d\rangle$ ($d=1,2$) she says $a=d-1$, since this commitment is irrevokable. If she committed with $|0\rangle$, she says $a=0,1$ with equal amplitude. It is crucial that she does this *in superposition*.

In the fourth round, Bob reveals the state of the qutrit he sent. If Alice sees that the bit Bob picked is different from her guess, i.e., $b \neq a$, then she sends, say, $s=0$ (an arbitrary value, since she has lost the game). Otherwise, she tries to

pick the best s possible to maximize her chance of passing Bob’s check. We describe her actions when $b=0, s'=0$; the other cases are similar.

The unnormalized density matrix of the qutrit she sent in the first round, when $b=0, s'=0$, conditioned on her reply being $a=0$ is

$$\cos^2\frac{\beta}{2}\left(\frac{1-\lambda}{2}|0\rangle\langle 0| + \frac{\lambda}{2}|1\rangle\langle 1|\right) + \sin^2\frac{\beta}{2}\rho.$$

Note that she knows the state of all their qubits, given b, s', a . She can thus transform her part of the state so that their joint state looks like

$$\sqrt{\frac{1-\lambda}{2}}\left(1 + \sin^2\frac{\beta}{2}\right)|00\rangle + \sqrt{\frac{\lambda}{2}}|11\rangle + \sqrt{\frac{\lambda}{2}}\sin\frac{\beta}{2}|22\rangle.$$

She sends $s=0$ if the entangled bit is 2. If her entangled qutrit is not 2, she does a Hadamard transform on her entangled qubit, and sends that to Bob as s . The probability with which Bob accepts $a=b, s$ is then

$$\left(\sqrt{\frac{1-\lambda}{2}}\left(1 + \sin^2\frac{\beta}{2}\right)\cos\frac{\alpha}{2} + \sqrt{\frac{\lambda}{2}}\sin\frac{\alpha}{2}\right)^2,$$

which evaluates to the expression stated in the lemma. She can achieve the same probability of success for all other values of b, s' as well. Thus, the overall chance of her succeeding in cheating is also given by this expression. \blacksquare

The properties we established above show that this protocol still has a bias of 1/4: the cheating probability for Alice and Bob are of the form $\frac{1}{2} + \delta$ and $1 - \delta$, respectively [with $\delta = \frac{1}{2}\cos^2(\alpha/2)\sin^2(\beta/2)$], and their maximum is minimized when $\delta=1/4$.

C. Cheating strategies for \mathcal{P}_n

Let $A_n(\epsilon), B_n(\epsilon)$ be the maximum value of Alice and Bob’s cheating probability in the protocol \mathcal{P}_n , when the other party is honest.

Bob’s optimal strategy may be reduced to a strategy for Alice as in the proof of Lemma 3.4. Alice always starts the protocol, and also sends the last message. Since the last message does not affect Bob’s cheating strategy, we may trace it out of the protocol when analyzing his optimal strategy. The protocol \mathcal{P}_n then reduces to a mixture of protocols where Alice sends $|0\rangle$ in the first round with probability $1 - \epsilon$ (i.e., does not reveal any information about the bit a), and sends $|a+1\rangle$ with probability ϵ (i.e., completely reveals the bit a). The rest of the protocol is the same as \mathcal{P}_{n-1} with the roles of Alice and Bob reversed. Thus,

$$B_n(\epsilon) = \epsilon + (1 - \epsilon)A_{n-1}(\epsilon) \quad \text{for } n \geq 2. \quad (7)$$

We already know from [5] that $B_1(\epsilon) = (1 + \epsilon)/2$. It thus suffices to analyze Alice’s cheating probability in all the protocols \mathcal{P}_n .

From our analysis of the three and five round protocols, we also know that

$$A_1(\epsilon) = 1 - \frac{\epsilon}{2},$$

$$B_2(\epsilon) = 1 - \frac{\epsilon}{2} + \frac{\epsilon^2}{2},$$

$$A_2(\epsilon) = \frac{1}{2} + \frac{\epsilon}{2} - \frac{\epsilon^2}{2}.$$

We now give a lower bound for the probability $A_n(\epsilon)$ by describing a cheating strategy for Alice that generalizes the strategy that we saw in the five round protocol.

First, we assume $n = 2k - 1 (k \geq 1)$ is odd, so that Alice has k rounds of commitment, and Bob has $k - 1$. The case of n even is addressed later. In the first round, Alice sends one-half of the state

$$\sqrt{1 - \lambda_1} |00\rangle + \sqrt{\frac{\lambda_1}{2}} |11\rangle + \sqrt{\frac{\lambda_1}{2}} |22\rangle$$

to Bob, and retains the other half. The half she keeps is referred to as an “entangled qutrit” below. All the parameters λ_j will be specified later. If one of the entangled qutrits she retains from any previous round is in state $x + 1$, in all the commitment rounds that follow, she sends the right half of the state

$$\frac{1}{\sqrt{2}} |0\rangle |\psi(x,0)\rangle + \frac{1}{\sqrt{2}} |1\rangle |\psi(x,1)\rangle$$

and keeps the first qubit (called a “sign qubit” below). If the entangled qutrits are all 0, and at least one of the qutrits she received in earlier rounds from Bob is in state $b + 1$, then she sends the right half of the state

$$\frac{1}{\sqrt{2}} |0\rangle |\psi(b,0)\rangle + \frac{1}{\sqrt{2}} |1\rangle |\psi(b,1)\rangle$$

to Bob. Otherwise, if none of the above two events occurs, she sends one half of the state

$$\sqrt{1 - \lambda_j} |00\rangle + \sqrt{\frac{\lambda_j}{2}} |11\rangle + \sqrt{\frac{\lambda_j}{2}} |22\rangle$$

in the j th commitment round. It is important that she do all this “in superposition,” i.e., via unitary operations controlled by her entangled qutrits and the qutrits she receives from Bob. She performs no measurements in the process.

Formally, for each qutrit that Alice is supposed to send, she has a qutrit-qubit pair. The first serves as the “entangled qutrit,” and the second serves as the “sign qubit.” They are all initialized to the 0 state. She prepares appropriate states over these according to the above rules, as the protocol proceeds. The joint state of both parties together after the n commitment rounds then looks as given below, for an arbitrary choice of b and the signs $\{s_{2j}\}$ picked by Bob. Here, the qutrits sent by Alice to Bob are underlined. The entangled qutrits and sign bits can be identified from the context. The first sum corresponds to the part of the state where Alice commits to a bit x by sending $|x + 1\rangle$, before she can identify which bit b Bob has picked. The next sum has the part of the state where Alice has not committed to any bit, and sees a $|b + 1\rangle$ in one of the qutrits Bob sent. The last term is the remaining part of the state. Note that Alice can differentiate between these three parts by examining her entangled qutrits and the qutrits Bob sent her. The odd lines contain the portion of the state constructed by Alice, and the even lines contain the portion prepared by Bob (and sent to Alice),

$$\sum_{x=0,1} \sum_{j=1}^{k-1} \left[\otimes_{l=1}^{j-1} \sqrt{1 - \lambda_l} |00\rangle \right] \otimes \sqrt{\frac{\lambda_j}{2}} |x+1, x+1\rangle \otimes \left[\frac{1}{\sqrt{2}} [|0\rangle |\psi(x,0)\rangle + |1\rangle |\psi(x,1)\rangle] \right]^{k-j}$$

$$\otimes (\sqrt{1 - \epsilon} |0\rangle)^{j-1} \otimes \left[\otimes_{l=j}^{k-1} |\psi(b, s_{2l})\rangle \right] + \sum_{j=1}^{k-1} \left[\otimes_{l=1}^j \sqrt{1 - \lambda_l} |00\rangle \right] \otimes \left[\frac{1}{\sqrt{2}} [|0\rangle |\psi(b,0)\rangle + |1\rangle |\psi(b,1)\rangle] \right]^{k-j}$$

$$\otimes (\sqrt{1 - \epsilon} |0\rangle)^{j-1} \otimes [(-1)^{s_{2j}} \sqrt{\epsilon} |b + 1\rangle] \otimes \left[\otimes_{l=j+1}^{k-1} |\psi(b, s_{2l})\rangle \right] + \left[\otimes_{l=1}^{k-1} \sqrt{1 - \lambda_l} |00\rangle \right] \otimes \left(\sqrt{1 - \lambda_k} |00\rangle \right.$$

$$\left. + \sqrt{\frac{\lambda_k}{2}} |11\rangle + \sqrt{\frac{\lambda_k}{2}} |22\rangle \right) \otimes (\sqrt{1 - \epsilon} |0\rangle)^{k-1}. \tag{8}$$

Since $n = 2k - 1$ is odd, Bob reveals his bit b first. Without loss of generality, we may assume that Alice would like to bias the coin towards 0. She therefore sends $a = b$ in the part of her state where her entangled qutrit is not equal to $\bar{b} + 1$ (which corresponds to a commitment which she cannot change). We will consider the residual state after Alice has sent back $k - i$ signs, in reverse order. This is the unnormalized part of the state (8) that has not been rejected by Bob. We will prove by induction that Alice can locally transform the residual state to a state $|\phi_i\rangle$ after every two rounds of sign exchange. This state is similar in form to the joint state (8) above, except that the first part is projected onto the space where $x = b$, and there is a factor of μ_i in the last term. The state $|\phi_i\rangle$ is displayed as

$$\begin{aligned}
& \sum_{j=1}^{i-1} \left[\bigotimes_{l=1}^{j-1} \sqrt{1-\lambda_l} |00\rangle \right] \otimes \sqrt{\frac{\lambda_j}{2}} |b+1, \underline{b+1}\rangle \otimes \left[\frac{1}{\sqrt{2}} (|0\rangle |\underline{\psi(b,0)}\rangle + |1\rangle |\underline{\psi(b,1)}\rangle) \right]^{i-j} \\
& \quad \otimes (\sqrt{1-\epsilon}|0\rangle)^{j-1} \otimes \left[\bigotimes_{l=j}^{i-1} |\psi(b, s_{2l})\rangle \right] \\
& + \sum_{j=1}^{i-1} \left[\bigotimes_{l=1}^j \sqrt{1-\lambda_l} |00\rangle \right] \otimes \left[\frac{1}{\sqrt{2}} (|0\rangle |\underline{\psi(b,0)}\rangle + |1\rangle |\underline{\psi(b,1)}\rangle) \right]^{i-j} \\
& \quad \otimes (\sqrt{1-\epsilon}|0\rangle)^{j-1} \otimes [(-1)^{s_{2j}} \sqrt{\epsilon} |b+1\rangle] \otimes \left[\bigotimes_{l=j+1}^{i-1} |\psi(b, s_{2l})\rangle \right] \\
& + \left[\bigotimes_{l=1}^{i-1} \sqrt{1-\lambda_l} |00\rangle \right] \otimes \left(\sqrt{\mu_i(1-\lambda_i)} |00\rangle + \sqrt{\frac{\lambda_i}{2}} |b+1, \underline{b+1}\rangle \right) \otimes (\sqrt{1-\epsilon}|0\rangle)^{i-1}, \tag{9}
\end{aligned}$$

where the numbers μ_i are as follows:

$$\begin{aligned}
\mu_k &= 1, \\
\mu_{i-1} &= (1-\epsilon)^2 \mu_i + \frac{\epsilon}{2} (3-\epsilon). \tag{10}
\end{aligned}$$

We can now also specify the parameters λ_i :

$$\lambda_i = \frac{\epsilon/2}{\mu_i(1-\epsilon) + \epsilon/2}. \tag{11}$$

Clearly, the state $|\phi_k\rangle$, when none of the signs has been revealed by Alice, is of this form with $\mu_k = 1$. Assume that this is also the case for some $i \leq k$. We will show by induction that the state after Alice has revealed $k-i+1$ sign bits $s_{2k-1}, s_{2k-3}, \dots, s_{2i-1}$ may be transformed to Eq. (9) and that Eq. (10) holds.

To send the sign s_{2i-1} , Alice does the following. The sign in part of the state in the first two summations of the state $|\phi_i\rangle$ in Eq. (9) is ‘‘precomputed’’ (in the sign qubit). To compute the sign in the last term, Alice first sets $b+1$ to 1 in the i th entangled qutrit, does a Hadamard transform, and exchanges that entangled qubit with the i th sign qubit. She then measures the i th sign qubit, and sends it across. It is easily seen that the unnormalized state that remains after Bob has checked the i th qutrit sent by Alice is as follows.² Here, we have written the last terms of the first two summations in Eq. (9) separately to facilitate the rest of the proof,

$$\begin{aligned}
& \sum_{j=1}^{i-2} \left[\bigotimes_{l=1}^{j-1} \sqrt{1-\lambda_l} |00\rangle \right] \otimes \sqrt{\frac{\lambda_j}{2}} |b+1, \underline{b+1}\rangle \otimes \left[\frac{1}{\sqrt{2}} [|0\rangle |\underline{\psi(b,0)}\rangle + |1\rangle |\underline{\psi(b,1)}\rangle] \right]^{i-1-j} \\
& \quad \otimes (\sqrt{1-\epsilon}|0\rangle)^{j-1} \otimes \left[\bigotimes_{l=j}^{i-1} |\psi(b, s_{2l})\rangle \right] + \sum_{j=1}^{i-2} \left[\bigotimes_{l=1}^j \sqrt{1-\lambda_l} |00\rangle \right] \otimes \left[\frac{1}{\sqrt{2}} [|0\rangle |\underline{\psi(b,0)}\rangle + |1\rangle |\underline{\psi(b,1)}\rangle] \right]^{i-1-j} \\
& \quad \otimes (\sqrt{1-\epsilon}|0\rangle)^{j-1} \otimes [(-1)^{s_{2j}} \sqrt{\epsilon} |b+1\rangle] \otimes \left[\bigotimes_{l=j+1}^{i-1} |\psi(b, s_{2l})\rangle \right] \\
& + \left[\bigotimes_{l=1}^{i-2} \sqrt{1-\lambda_l} |00\rangle \right] \otimes \left(\sqrt{\frac{\lambda_{i-1}}{2}} |b+1, \underline{b+1}\rangle \right) \otimes (\sqrt{1-\epsilon}|0\rangle)^{i-2} \otimes |\psi(b, s_{2(i-1)})\rangle \\
& + \left[\bigotimes_{l=1}^{i-2} \sqrt{1-\lambda_l} |00\rangle \right] \otimes (\sqrt{1-\lambda_{i-1}} |00\rangle)
\end{aligned}$$

²Actually, the state is a mixture of two states which are both $1/\sqrt{2}$ times the state given. The mixture arises because of the two possible values of the sign bit Alice sends for s_{2i-1} . The mixture is of course equivalent to the single state shown.

$$\begin{aligned} & \otimes (\sqrt{1-\epsilon}|0\rangle)^{i-2} \otimes [(-1)^{s_{2(i-1)}}\sqrt{\epsilon}|b+1\rangle] + \left[\otimes_{l=1}^{i-2} \sqrt{1-\lambda_l}|00\rangle \right] \otimes (\sqrt{1-\lambda_{i-1}}|00\rangle) \\ & \otimes (\sqrt{1-\epsilon}|0\rangle)^{i-2} \otimes \sqrt{1-\epsilon}|0\rangle [\mu_i(1-\epsilon) + \epsilon/2]^{1/2}. \end{aligned}$$

Now, when Bob sends the sign $s_{2(i-1)}$ to Alice, she rotates the $(i-1)$ th qutrit that Bob sent her in all but the last two terms in the sum above, to $|0\rangle$. She also rotates that qutrit in the last two terms from

$$[\mu_i(1-\epsilon) + \epsilon/2]^{1/2} \sqrt{1-\epsilon}|0\rangle + (-1)^{s_{2(i-1)}} \sqrt{\epsilon}|b+1\rangle$$

to $\sqrt{\mu_{i-1}}|0\rangle$, whereby $\mu_{i-1} = \mu_i(1-\epsilon)^2 + \epsilon(1-\epsilon)/2 + \epsilon$. This proves the induction step.

At the final round of the protocol \mathcal{P}_n , the state that they are left with is

$$\sqrt{\mu_1(1-\lambda_1)}|00\rangle + \sqrt{\frac{\lambda_1}{2}}|b+1, \underline{b+1}\rangle.$$

Following Alice's strategy for computing the sign as above, we see that the probability with which Alice succeeds in passing Bob's checks is [using Eq. (11)]

$$[\sqrt{\mu_1(1-\lambda_1)(1-\epsilon)} + \sqrt{\lambda_1\epsilon/2}]^2 = \mu_1(1-\epsilon) + \epsilon/2. \tag{12}$$

Solving the recurrence for μ_i given in Eq. (10), we get

$$\mu_i = (1-\epsilon)^{2(k-i)} + \frac{(3-\epsilon)}{2(2-\epsilon)} [1 - (1-\epsilon)^{2(k-i)}],$$

and so that from Eq. (12), for n odd,

$$\begin{aligned} A_n(\epsilon) & \geq \frac{\epsilon}{2} + (1-\epsilon)^n + \frac{(3-\epsilon)(1-\epsilon)}{2(2-\epsilon)} [1 - (1-\epsilon)^{n-1}] \\ & = \frac{1}{2(2-\epsilon)} [3 - 2\epsilon + (1-\epsilon)^{n+1}]. \end{aligned} \tag{13}$$

The analysis in the case that $n=2k$ is even is similar, except for the rule Alice uses to compute the bit she sends to Bob (since she is supposed to reveal her bit before Bob reveals his bit). In this case, she sends the bit x if any of her entangled qutrits is in state $|x+1\rangle$ or, if any of Bob's qutrits are in state $b+1$, she sends b . In the remaining case, she sends 0,1 with equal amplitude $1/\sqrt{2}$. This leads to a state similar to $|\phi_k\rangle$ above after Bob reveals his bit, and the last sign he used, except that here $\mu_k = (1+\epsilon)/2$. So, we get for n even,

$$\begin{aligned} A_n(\epsilon) & \geq \frac{\epsilon}{2} + (1-\epsilon)^{n-1}(1+\epsilon)/2 \\ & \quad + \frac{(3-\epsilon)(1-\epsilon)}{2(2-\epsilon)} [1 - (1-\epsilon)^{n-2}] \\ & = \frac{1}{2(2-\epsilon)} [3 - 2\epsilon - (1-\epsilon)^{n+1}]. \end{aligned} \tag{14}$$

From Eqs. (13), (14), and (7), we can deduce lower bounds for $B_n(\epsilon)$ as well, for $n \geq 2$. This expression matches the one for $B_1(\epsilon)$. Thus, for all n ,

$$B_n(\epsilon) \geq \frac{1}{2(2-\epsilon)} [3 - \epsilon + (-1)^n(1-\epsilon)^{n+1}]. \tag{15}$$

To determine the bias achieved, we examine the maximum of the cheating probabilities attained by Alice and Bob. Note that $A_n(\epsilon) + B_n(\epsilon) \geq 3/2$ for all n, ϵ . Thus, the bias of the protocol \mathcal{P}_n is at least $3/4$ for any n and ϵ . Since we would like this bias to be as small as possible, we optimize the maximum cheating probability with respect to ϵ .

For odd n , $A_n(0) = B_n(1) = 1$, and $A_n(1) = B(0) = 1/2$, A_n is monotonically decreasing, and B_n is monotonically increasing with respect to $\epsilon \in [0,1]$. Thus, the maximum bias achievable is minimized when $A_n(\epsilon) = B_n(\epsilon)$. This condition is satisfied when $\epsilon = \epsilon_0$ such that

$$(1-\epsilon_0)^{n+1} = \frac{\epsilon_0}{2}, \tag{16}$$

that is,

$$\epsilon_0 = \frac{1}{n+1} [\ln n - \ln \ln n + \Theta(1)].$$

Using Eq. (16) we can verify that the lower bound on $A_n(\epsilon_0)$ is $3/4$.

For even n , $A_n(0) = A_n(1) = 1/2$, and $B(0) = B_n(1) = 1$, A_n is concave, B_n is convex, and $B_n(\epsilon) \geq A_n(\epsilon)$ for $\epsilon \in [0,1]$. Thus, the bias achievable is minimized when B_n is, i.e., for $\epsilon = \epsilon_0$ such that

$$(1-\epsilon_0)^n = \frac{1}{(2-\epsilon_0)n+1}, \tag{17}$$

that is,

$$\epsilon_0 = \frac{1}{n} \ln[2n - \ln n + \Theta(1)].$$

The expression for the cheating probability $B_n(\epsilon_0)$ then evaluates to at least $\frac{3}{4}$, as may be seen by using Eq. (17).

This completes the analysis of the cheating strategies devised above.

ACKNOWLEDGMENTS

We thank Rob Spekkens and Terry Rudolph for helpful feedback on the paper. A.N. was supported by the Charles Lee Powell Foundation and NSF Grant Nos. CCR 0049092 and EIA 0086038. Part of this work was done while A.N. was at DIMACS Center, Rutgers University, and AT&T Laboratories, and was supported by NSF Grant Nos. STC 91-19999, CCR 99-06105, and EIA 00-80234.

-
- [1] M. Blum, in *Advances in Cryptology: A Report on CRYPTO'81*, IEEE Workshop on Communications Security, Santa Barbara, 1981, University of California, Santa Barbara, ECE Report No. 82-04 (unpublished), pp. 11–15.
- [2] H. Lo and H. Chau, *Physica D* **120**, 177 (1998). See also e-print quant-ph/9711065.
- [3] D. Mayers, *Phys. Rev. Lett.* **78**, 3414 (1997).
- [4] D. Aharonov, A. Ta-Shma, U. Vazirani, and A. Yao, in *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing* (ACM Press, New York, 2000), pp. 705–714.
- [5] A. Ambainis, in *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing* (ACM Press, New York, 2001), pp. 134–142.
- [6] R. W. Spekkens and T. Rudolph, *Phys. Rev. A* **65**, 012310 (2002).
- [7] A. C.-C. Yao, in *Proceedings of the 34th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society Press, Los Alamitos, CA, 1993), pp. 352–361.
- [8] L. Salvail (private communication).
- [9] E. Bernstein and U. Vazirani, *SIAM J. Comput.* **26**, 1411 (1997).
- [10] D. Mayers, L. Salvail, and Y. Chiba-Kohno, e-print quant-ph/9904078.
- [11] Y. Tokunaga, e-print quant-ph/0108026.
- [12] A. Kitaev (personal communication).
- [13] I. Kerenidis and A. Nayak, e-print quant-ph/020612.
- [14] A. Ambainis (personal communication).
- [15] R. W. Spekkens and T. Rudolph, *Phys. Rev. Lett.* **89**, 227901 (2002).
- [16] R. Jozsa, *J. Mod. Opt.* **41**, 2315 (1994).
- [17] C. A. Fuchs and J. van de Graaf, *IEEE Trans. Inf. Theory* **45**, 1216 (1999).