Improved Device Independent Weak Coin Flipping Protocols

5th August 2021

Abstract

[OUTDATED: Needs to be rewritten]
We report a device independent weak coin flipping protocol[1] with $P_A^* \leq \cos^2(\pi/8)$ and $P_B^* \leq 0.667...$, by making seemingly minor changes to the best known protocol due to SCAKPM'11 [10.1103/PhysRevLett.106.220501], with $P_A^* \leq \cos^2(\pi/8) \approx 0.85$ and $P_B^* \leq 3/4 = 0.75$. In terms of bias, we improve the SCAKPM'11 result from ≈ 0.336 to ≈ 0.3199. This improvement is due to two ingredients: a self-testing (of GHZ) step and an extra cheat detection step for Bob. We also introduce a new bias suppression technique that ekes out further security from the abort probability to obtain ... Note that the SCAKPM'11 result held for both strong and weak coin flipping; ours holds only for the latter. TODO: Fix me!

Contents

[1] which are analysed

1 Introduction

INTERNAL/Atul: Colour coding—Purple is for informal discussions, black is for formal statements and blue is for proofs. We can remove these from the final version; I put it to minimise verbiage.

1.1 About Weak Coin Flipping

Secure two-party computation is a cryptographic setting where two parties, conventionally called Alice and Bob, receive inputs x and y and their goal is to compute some function $f_A(x, y)$ and $f_B(x, y)$ respectively which depends on both their inputs. However, they do not wish to reveal their inputs. Coin flipping (CF) is a cryptographic primitive in this setting, i.e. a building block for constructing more applicable secure two-party cryptographic schemes, where Alice and Bob wish to exchange messages and agree on a random bit, without trusting each other. A protocol that implements coin flipping must protect an honest player from a malicious[2] player.

A weaker primitive, unsurprisingly, known as *weak coin flipping* (WCF) is where a zero corresponds to Alice winning and one corresponds to Bob winning. It is weaker because now the protocol has to protect Alice from a malicious Bob who tries to bias the outcome towards one (and not towards zero) and conversely, it must protect Bob from a malicious Alice who tries to bias the outcome towards zero (and not towards one). To emphasise the distinction, the former primitive is often termed *strong coin flipping* (SCF).

We primarily focus on WCF in this article and begin with introducing some notation. We denote by P_A^* the highest probability of a malicious Alice convincing an honest Bob that she won (i.e. in the WCF protocol, Alice uses her best cheating strategy against Bob who in turn is following the protocol as described, to convince him that the outcome is zero). Analogously, P_B^* is the highest probability of a malicious Bob convincing an honest Alice that he won. The bias of a WCF protocol is defined as $\epsilon := \max\left\{P_A^*, P_B^*\right\} - \frac{1}{2}$. A protocol that is completely secure, has $\epsilon = 0$ and one that is completely insecure has $\epsilon = \frac{1}{2}$.

Using a classical channel of communication between Alice and Bob, unless one makes further assumptions such as computational hardness of certain problems or relativistic assumptions,[3] coin flipping (even weak) is impossible to implement with any security, to wit: for all classical protocols at least one of the parties, viz. a malicious Alice or a malicious Bob, can win with certainty because one can show $\epsilon = \frac{1}{2}$ (viz. $\max\{P_A^*, P_B^*\} = 1$). Using a quantum channel of communication, it was shown that WCF can be implemented with vanishing bias. These works, however, do not account for noise in their implementation. One path towards more robust security is device independence wherein the players do not even trust their devices (recall, they already do not trust the other party). This is in contrast to the device independent setting considered in key distribution where the two parties trust each other but neither their devices nor the communication channel (TODO: is the classical communication channel trusted?).

1.2 Contributions

[TODO: fix it—this is outdated] In this work, we start with a device independent (DI) coin flipping (CF) protocol introduced[4] in [SCA+11] which has $P_A^* = \cos^2(\pi/8) \approx 0.854$ and $P_B^* = 3/4 = 0.75$. They then compose these protocols to give a balanced protocol, i.e. with $P_A^* = P_B^* \approx \frac{1}{2} + 0.336$. To the best of our knowledge, this DI CF protocol has the best security guarantee. While Kitaev's bound for CF rules out perfect DI CF, no lower bounds on the bias are known for DI WCF. In this work, however, we focus on improving the upper bound on the bias, viz. we give DI WCF protocols with biases ≈ 0.319.

We introduce two key new ideas which result in better protocols. The first, is the use of self-testing by one party before initiating the protocol and the second, is a more general technique to convert unbalanced protocols (i.e. ones in which the probability of maliciously winning for Alice and Bob are unequal) into balanced ones.

[2] (or cheating, we use these adjectives interchangeably)

[3] in terms of the spatial locations of the observers; not to be confused with the term *relativising* from computational complexity.

[4] In fact, they introduced a device independent bit commitment protocol which they in turn use to construct a strong coin flipping protocol with the same cheating probabilities for Alice and Bob, ≈ 0.854 and 0.75 respectively.

1.3 Proof Technique

Notation and Cheat Vectors

We introduce some notation to facilitate the discussion here. Denote the DI CF protocol introduced in [SCA+11] by $\mathcal{I}$ and let $p^*_A(\mathcal{I}) \approx 0.853\ldots$ denote the maximum probability with which a malicious Alice can win against honest Bob who is following the protocol $\mathcal{I}$ and similarly, let $p^*_B(\mathcal{I}) \approx 0.75$ denote the maximum probability with which a malicious Bob can win against an honest Alice who is following the protocol $\mathcal{I}$.

One of the key observations we make in this work is the use of what we call "cheat vectors"—it is any tuple of probabilities which can arise in a CF protocol when one player is honest. More precisely, suppose Alice is (possibly) malicious and Bob follows the protocol $\mathcal{I}$. Then, the cheat vectors for Alice constitute the set

$$\mathbb{C}_A(\mathcal{I}) := \{(\alpha, \beta, \gamma) : \exists \text{ a strategy for } A \text{ s.t. an honest } B \text{ outputs 0,1, and } \perp \text{ with probabilities } \alpha, \beta \text{ and } \gamma\}. \quad (1)$$

We analogously define $\mathbb{C}_B(\mathcal{I})$. Cheat vectors become useful when we try to compose protocols. The observation then, is that the abort event can be taken to abort the full protocol instead of being treated as the honest player winning. The latter gives the malicious player further opportunity to cheat and so preventing it improves the security.

Protocols

We introduce two variants of protocol $\mathcal{I}$, which we call $\mathcal{P}$ and $\mathcal{Q}$.

- $\mathcal{P}$ is essentially the same as $\mathcal{I}$ except that Alice self-tests her boxes before starting the protocol and performs an additional test to ensure Bob doesn't cheat. We show that $p^*_A(\mathcal{P}) \lesssim 0.853\ldots$ and $p^*_B(\mathcal{P}) \lesssim 0.667\ldots$. We also show that $\mathbb{C}_B(\mathcal{P})$ can be cast as an SDP.
- $\mathcal{Q}$ is also essentially the same as $\mathcal{I}$ except that Bob self-tests his boxes before starting the protocol. In this case, $p^*_X(\mathcal{Q}) = p^*_X(\mathcal{I})$ for both values of $X \in \{A, B\}$ so the advantage isn't manifest. However, now $\mathbb{C}_A(\mathcal{Q})$ can be cast as an SDP which, as we shall see, yields an advantage when $\mathcal{Q}$ is composed.

Compositions

As the protocols $\mathcal{X} \in \{\mathcal{I}, \mathcal{P}, \mathcal{Q}\}$ all have skewed security—either $p^*_A(\mathcal{X}) > p^*_B(\mathcal{X})$ or the other way—and therefore the bias is determined by $p^*_{\max}(\mathcal{X}) := \max\{p^*_A(\mathcal{X}), p^*_B(\mathcal{X})\}$. Note that, $p^*_{\max}(\mathcal{X}) = p^*_{\max}(\mathcal{Y})$ for all $\mathcal{X}, \mathcal{Y} \in \{\mathcal{I}, \mathcal{P}, \mathcal{Q}\}$, which means that we don't immediately obtain an advantage. However, the most obvious method of composing these protocols to obtain a new protocol, which we describe later, "balances" the advantage. After this composition procedure is applied to some protocol $\mathcal{X}$, we denote the resulting protocol by $C_{LL}(\mathcal{X})$. Applying this technique to $\mathcal{P}$, we already obtain a more secure protocol.

- For all $X \in \{A, B\}$ the cheating probabilities for protocol $\mathcal{I}$ under the standard composition is given by

  $$p^*_X(C^{LL}(\mathcal{I})) \approx \frac{1}{2} + 0.336\ldots$$

  while for the improved protocol $\mathcal{P}$, these are given by

  $$p^*_X(C^{LL}(\mathcal{P})) \approx \frac{1}{2} + 0.3199\ldots. \quad (2)$$

The standard composition technique doesn't yield any improvement for $\mathcal{Q}$ because the cheating probabilities are identical to those of $\mathcal{I}$. We can extract an advantage by using a composition technique that uses "cheat vectors" and the abort event. We describe it in detail later but for now, we simply denote the new protocol obtained using this improved "abort phobic" composition (of protocol $\mathcal{X}$) by $C_{\perp L}(\mathcal{X})$ or $C_{L\perp}(\mathcal{X})$.

- Applying the technique to $\mathcal{P}$, the cheating probabilities become

  $$p^*_X(C^{\perp L}(\mathcal{P})) \approx \frac{1}{2} + 0.3148\ldots$$

  which is a further improvement.

Protocol	Bias
$C^{LL}(\mathcal{W},\ldots,\mathcal{W})$	$0.336\ldots$
$C^{LL}(\mathcal{P},\ldots,\mathcal{P})$	$0.3199\ldots$
$C^{\perp L}(\mathcal{P},\ldots,\mathcal{P})$	$0.3148\ldots$
$C^{L\perp}(\mathcal{Q},\ldots,\mathcal{Q})$	$0.3226\ldots$
$C^{L\perp}(\mathcal{Q},\ldots,\mathcal{Q},\mathcal{P})$	$0.29104\ldots$

Table 1:

- Using this technique on $\mathcal{Q}$, the cheating probabilities become

$$p_X^*(C^{L\perp}(\mathcal{Q})) \approx \frac{1}{2} + 0.3226\ldots$$

for all $X \in \{A, B\}$, which is worse than even Equation (2).

- However, when we combine both these protocols to obtain (again, for all $X \in \{A, B\}$)

$$p_X^*(C^{L\perp}(\mathcal{Q},\mathcal{Q},\ldots,\mathcal{Q},\mathcal{P})) \approx \frac{1}{2} + 0.29104\ldots$$

where we use the same composition technique except that at the last "level" we use[5] $\mathcal{P}$ instead of $\mathcal{Q}$.

2 Device Independent Weak Coin Flipping protocols | State Of The Art

In the following, we first discuss how one can describe DI WCF protocols in terms of the players exchanging "boxes"—devices which take classical inputs and give classical outputs. Subsequently we recall the GHZ test and finally we use these to delineate the DI-CF due to [SCA+11].

2.1 Device Independence and the Box Paradigm

We describe device independent protocols as classical protocols with the one modification: we assume that the two parties can exchange boxes and that the parties can shield their boxes (from the other boxes i.e. the boxes can't communicate with each other once shielded).[6]

Definition 1 (Box)**.** A *box* is a device that takes an input $x \in \mathcal{X}$ and yields an outputs $a \in \mathcal{A}$ where $\mathcal{X}$ and $\mathcal{A}$ are finite sets. Typically, a set of n boxes, taking inputs $x_1, x_2, \ldots x_n$ and yielding outputs $a_1, a_2 \ldots a_n$ are *characterised* by a joint conditional probability distribution, denoted by

$$p(a_1, a_2 \ldots a_n | x_1, x_2 \ldots x_n).$$

Further, if $p(a_1, a_2 \ldots a_n|x_1, x_2 \ldots x_n) = \operatorname{tr}\left[M^1_{a_1|x_1} \otimes M^2_{a_2|x_2} \cdots \otimes M^n_{a_n|x_n} \rho\right]$ then we call the set of boxes, *quantum boxes*, where for a fixed x' $\{M^i_{a'|x'}\}_{a' \in \mathcal{A}_i}$ constitute a POVM for the ith subsystem, ρ is a density matrix and their dimensions are mutually consistent.

Henceforth, we restrict ourselves to quantum boxes.

Definition 2 (Protocol in the box formalism)**.** A generic two-party protocol in the box formalism has the following form:

[5] $C^{\perp L}(\mathcal{P},\mathcal{P},\ldots,\mathcal{P},\mathcal{Q})$ is strictly worse than considering $C^{\perp L}(\mathcal{P},\mathcal{P},\ldots,\mathcal{P},\mathcal{P})$; this should become evident later.

[6] TODO: Verify if this notion is in fact correct; I hope I'm not making a major mistake somehow. I should be able to take the POVMs as tensor products right, because I can change them at will, independent of the others (and ensuring that there's no communication between them; could they be somehow entangled, i.e. could it be that somehow the measurement operators are themselves quantum correlated?); I would like to reach the conclusion starting from the locality assumption.

1. Inputs:
   (a) Alice is given boxes $\square_1^A, \square_2^A \ldots \square_p^A$ and Bob is given boxes $\square_1^B, \square_2^B, \ldots \square_q^B$.
   (b) Alice is given a random string r^A and Bob is given a random string r^B (of arbitrary but finite length).
2. Structure: At each round of the protocol, the following is allowed.
   (a) Alice and Bob can locally perform arbitrary but finite time computations on a Turing Machine.
   (b) They can exchange classical strings/messages and boxes.

A protocol in the box formalism is readily expressed as a protocol which uses a (trusted) classical channel (i.e. they trust their classical devices to reliably send/receive messages), untrusted quantum devices and an untrusted quantum channel (i.e. a channel that can carry quantum states but may be controlled by the adversary).

Assumption 3 (Setup of Device Independent Two-Party Protocols)**.** *Alice and Bob*

1. *both have private sources of randomness,*
2. *can send and receive classical messages over a (trusted) classical channel,*
3. *can prevent parts of their untrusted quantum devices from communicating with each other, and*
4. *have access to an untrusted quantum channel.*

We restrict ourselves to a "measure and exchange" class of protocols—protocols where Alice and Bob start with some pre-prepared states and subsequently, only perform classical computation and quantum measurements locally in conjunction with exchanging classical and quantum messages. More precisely, we consider the following (likely restricted) class of device independent protocols.

Definition 4 (Measure and Exchange (Device Independent Two-Party) Protocols)**.** A *measure and exchange (device independent two-party) protocol* has the following form:

1. Inputs:
   (a) Alice is given quantum registers $A_1, A_2, \ldots A_p$ together with POVMs[7]
   $$\{M_{a|x}^{A_1}\}_a, \{M_{a|x}^{A_2}\}_a, \ldots \{M_{a|x}^{A_p}\}_a$$
   which act on them and Bob is, analogously, given quantum registers $B_1, B_2, \ldots B_q$ together with POVMs
   $$\{M_{b|y}^{B_1}\}_b, \{M_{b|y}^{B_2}\}_b, \ldots, \{M_{b|y}^{B_q}\}_b.$$
   Alice shields $A_1, A_2, \ldots A_p$ (and the POVMs) from each other and from Bob's lab. Bob similarly shields $B_1, B_2 \ldots B_q$ (and the POVMs) from each other and from Alice's lab.
   (b) Alice is given a random string r^A and Bob is given a random string r^B (of arbitrary but finite length).
2. Structure: At each round of the protocol, the following is allowed.
   (a) Alice and Bob can locally perform arbitrary but finite time computations on a Turing Machine.
   (b) They can exchange classical strings/messages.
   (c) Alice (for instance) can
      i. send a register A_l and the encoding of her POVMs $\{M_i^{A_l}\}_i$ to Bob, or
      ii. receive a register B_m and the encoding of the POVMs $\{M_i^{B_m}\}_i$.

   Analogously for Bob.

It is clear that a protocol in the box formalism (Definition 2) which uses only quantum boxes (Definition 1) can be implemented as a measure and exchange protocol (Definition 4).

[7] For concreteness, take the case of binary measurements. By $\{M_{a|x}^{A_1}\}_a$, for instance, we mean $\{M_{0|x}^{A_1}, M_{1|x}^{A_1}\}$ is a POVM for $x \in \{0, 1\}$.

2.2 The GHZ Test

Before we define the current best DI CF protocol, we briefly remind the reader of the GHZ test, upon which the aforementioned protocol depends, and set up some conventions.

Definition 5. Suppose we are given three boxes, $\square^A, \square^B$ and $\square^C$, which accept binary inputs $a, b, c \in \{0,1\}$ and produces binary output $x, y, z \in \{0,1\}$ respectively. The boxes pass the GHZ test if $a \oplus b \oplus c = xyz \oplus 1$, given the inputs satisfy $x \oplus y \oplus z = 1$.

Claim 6. Quantum boxes pass the GHZ test with certainty (even if they cannot communicate), for the state $|\psi\rangle_{ABC} = \frac{|000\rangle_{ABC}+|111\rangle_{ABC}}{\sqrt{2}}$, and measurement[8] $\frac{\sigma_x+\mathbb{I}}{2}$ for input 0 and $\frac{\sigma_y+\mathbb{I}}{2}$ for input 1 (in the notation introduced earlier, $M^A_{0|0} = |+\rangle\langle+|, M^A_{1|0} = |-\rangle\langle-|$ and so on, where $|\pm\rangle = \frac{|0\rangle\pm|1\rangle}{\sqrt{2}}$).[9]

The proof is easier to see in the case where the outcomes are ± 1; it follows from the observations that $\sigma_y \otimes \sigma_y \otimes \sigma_y |\psi\rangle = -|\psi\rangle$, $\sigma_x \otimes \sigma_x \otimes \sigma_x |\psi\rangle = |\psi\rangle$ and the anti-commutation of σ_x and σ_y matrices, i.e. $\sigma_x\sigma_y + \sigma_y\sigma_x = 0$.

2.3 The Protocol

The best DI CF protocol known is the one introduced in [SCA+11]. While this is a protocol for SCF, and so also works as a WCF protocol, we do not know of any better protocol for the latter.

Algorithm 7 (SCF, original)**.** *Alice has one box and Bob has two boxes (in the security analysis, we let the cheating player distribute the boxes). Each box takes one binary input and gives one binary output.*

1. *Alice chooses $x \in_R \{0,1\}$ and inputs it into her box to obtain a. She chooses $r \in_R \{0,1\}$ to compute $s = a \oplus x.r$ and sends s to Bob.*
2. *Bob chooses $g \in_R \{0,1\}$ (for "guess") and sends it to Alice.*
3. *Alice sends x and a to Bob. They both compute the output $x \oplus g$.*
4. *Test round*
   (a) *Bob tests if $s = a$ or $s = a \oplus x$. If the test fails, he aborts. Bob chooses $b, c \in_R \{0,1\}$ such that $a \oplus b \oplus c = 1$ and then performs a GHZ using a, b, c as the inputs and x, y, z as the output from the three boxes. He aborts if this test fails.*

From Claim 6, it is clear that when both players follow Algorithm 7 using GHZ boxes (Definition 5), Bob never aborts and they win with equal probabilities. The security of the protocol is summarised next.

Lemma 8 (Security of SCF)**.** *[SCA+11] Let $\mathcal{I}$ denote the protocol corresponding to Algorithm 7. Then, the success probability of cheating Bob, $p^*_B(\mathcal{I}) \leq \frac{3}{4}$ and that of cheating Alice, $p^*_A(\mathcal{I}) \leq \cos^2(\pi/8)$. Further, both bounds are saturated by a quantum strategy which uses a GHZ state and the honest player measures along the σ_x/σ_y basis corresponding to input $0/1$ into the box. Cheating Alice measures along $\sigma_{\hat{n}}$ for $\hat{n} = \frac{1}{\sqrt{2}}(\hat{x} + \hat{y})$ while cheating Bob measures his first box along σ_x and second along σ_y.*

Note that both players can cheat maximally assuming they share a GHZ state and the honest player measures along the associated basis. This entails that even though the cheating player could potentially tamper with the boxes before handing them to the honest player, surprisingly, exploiting this freedom does not offer any advantage to the cheating player.

[8] we added the identity so that the eigenvalues associated become $0, 1$ instead of $-1, 1$.
[9] TODO: Think: Should I add the classical value? This would require me to add what it means to have a classical box.

3 First Technique: Self-testing (single shot, unbalanced)

TODO: Assumption: No honest abort.

We make two observations.

First, in Algorithm 7 only Bob performs the test round. In WCF, there is a notion of Alice winning and Bob winning. Thus, if $x \oplus g = 0$, i.e. the outcome corresponding to "Alice wins", we can imagine that Bob continues to perform the test to ensure (at least to some extend) that Alice did not cheat. However, if $x \oplus g = 1$, i.e. the outcome corresponding to "Bob wins", we can require Alice to now complete the GHZ test to ensure that Bob did not cheat. It turns out that this does not lower p_B^*. Interestingly, the best cheating strategy deviates from the GHZ state and measurements for the honest player. We omit the details here (see TODO: write this down somewhere) but mention this to motivate the following.

Second, Alice (say) can harness the self-testing property of GHZ states and measurements to ensure that Bob has not tampered with her boxes. One way of proceeding is that N copies of the supposedly correct boxes are distributed. Alice now picks one out of these N boxes at random and asks Bob to send the associated two boxes to each $N-1$ box that Alice posses. Alice runs the GHZ test on each box and if even one test fails, she declares that Bob cheated. This way, for a large N, Alice can ensure with near certainty, that she has a box containing the correct state and (which performs the correct) measurements. Note that no such scheme can be concocted which simultaneously self-tests Alice and Bob's boxes. More precisely, no such procedure can ensure that Alice and Bob share a GHZ state (Alice one part, Bob the other two, for instance) because this would mean perfect (or near perfect) SCF is possible which is forbidden even in the device dependent case. Kitaev showed that for any SCF protocol, $\epsilon \geq \frac{1}{\sqrt{2}} - \frac{1}{2}$.

Combining these two observations, results in an improvement in the security for Alice. We obtain a protocol with $P_A^* \leq 3/4$, which is the same as before, but $P_B^* \lesssim 0.667....$

3.1 Cheat Vectors

As alluded to in Section 1.3, using cheat vectors, it is sometimes possible to compose protocols and obtain a lower bias compared to protocols which are composed without using cheat vectors. We describe such procedures in the next section, Section 4. Here, we simply define cheat vectors and show that self-testing allows one to express relevant optimisation problems over cheat vectors as semi definite programmes.

Definition 9 (Cheat Vectors)**.** Given a protocol $\mathcal{I}$, denote by $\mathbb{C}_B(\mathcal{I})$ the set of *cheat vectors* for Bob, which is defined as follows :

$$\mathbb{C}_B(\mathcal{I}) := \{(\alpha, \beta, \gamma) : \exists \text{ a strategy of } B \text{ s.t. an honest } A \text{ outputs 0,1, and } \perp \text{ with probabilities } \alpha, \beta \text{ and } \gamma\}$$

and analogously, denote by $\mathbb{C}_A(\mathcal{I})$ the set of cheat vectors for Alice (see Equation (1)).

3.2 Alice self-tests | Protocol $\mathcal{P}$

We begin with the case where Alice self-tests. In the honest implementation, the *trio* of boxes used in the following are characterised by the GHZ setup (see Claim 6).

Algorithm 10 (Alice self-tests her boxes)**.** *There are N trios of boxes; Alice has the first part and Bob has the remaining two parts, of each trio.*

1. *Alice selects a number $i \in_R \{1, 2 \dots N\}$ and sends it to Bob.*
2. *Bob sends his part of the trio of boxes corresponding to $\{1, 2 \dots N\} \backslash i$, i.e. he sends all the boxes, except the ones corresponding to the trio i.*
3. *Alice performs a GHZ test on all the trios labelled $\{1, 2 \dots N\} \backslash i$, i.e. all the trios except the ith.*

We restrict ourselves to the ith trio. Alice has one box and Bob has two boxes. Each box takes one binary input and gives one binary output.

1. *Alice chooses $x \in_R \{0, 1\}$ and inputs it into her box to obtain a. She chooses $r \in_R \{0, 1\}$ to compute $s = a \oplus x.r$ and sends s to Bob.*

2. *Bob chooses $g \in_R \{0,1\}$ (for "guess") and sends it to Alice.*

3. *Alice sends x to Bob. They both compute the output $x \oplus g$.*

4. *Test rounds:*

   (a) *If $x \oplus g = 0$:*
   Alice sends a to Bob.
   Bob tests if $s = a$ or $s = a \oplus x$. If the test fails, he aborts. Bob chooses $y, z \in_R \{0,1\}$ such that $x \oplus y \oplus z = 1$ and then performs a GHZ using x, y, z as the inputs and a, b, c as the output from the three boxes. He aborts if this test fails.

   (b) *Else, if $x \oplus g = 1$:*

      i. *Alice chooses $y, z \in_R \{0,1\}$ s.t. $x \oplus y \oplus z = 1$ and sends them to Bob.*

      ii. *Bob inputs y, z into his boxes, obtains and sends b, c to Alice.*

   Alice tests if x, y, z as inputs and a, b, c as outputs, satisfy the GHZ test. She aborts if this test fails.

Lemma 11. *Let $\mathcal{P}$ denote the protocol corresponding to Algorithm 10. Then Alice's cheating probability $p_A^*(\mathcal{P}) \leq \cos^2(\pi/8) \approx 0.852$. Further, let $c_0, c_1, c_\perp \in \mathbb{R}$, and $\mathbb{C}_B(\mathcal{P})$ be the set of cheat vectors for Bob. Then, as $N \to \infty$, the solution to the optimisation problem $\max(c_0\alpha + c_1\beta + c_\perp\gamma)$ over $\mathbb{C}_B(\mathcal{Q})$ approaches that of a semi definite programme. In particular, i.e. for $c_0 = c_\perp = 0$ and $c_1 = 1$, $p_B^*(\mathcal{P}) \lesssim 0.667...$ (in the limit).*

We defer the proof to Section 5.1.1. The value for $p_B^*(\mathcal{P})$ was obtained by numerically solving the corresponding semi definite programme while the analysis for cheating Alice is the same as that of the original protocol.

3.3 Bob self-tests | Protocol $\mathcal{Q}$

What if we modified the protocol and had Bob self-test his boxes? Does that yield a better protocol? We address the first question now and the second in the subsequent section.

Algorithm 12 (Bob self-tests his boxes)**.** *Proceed exactly as in Algorithm 10, except for the self-testing where the rolls of Alice and Bob are reversed. More explicitly, suppose there are N trios of boxes; Alice has the first part and Bob has the remaining two parts, of each trio.*

1. *Bob selects a number $i \in_R \{1, 2 \dots N\}$ and sends it to Alice.*

2. *Alice sends her part of the trio of boxes corresponding to $\{1, 2 \dots N\} \backslash i$, i.e. she sends all the boxes, except the ones corresponding to the trio i.*

3. *Bob performs a GHZ test on all the trios labelled $\{1, 2 \dots N\} \backslash i$, i.e. all the trios except the ith.*

Henceforth, proceed as in Algorithm 10 after the self-testing step.

As already indicated in Section 1.3, we don't expect the cheating probabilities to improve but we do expect an SDP characterisation of Alice's cheat vectors.

Lemma 13. *Let $\mathcal{Q}$ denote the protocol corresponding to Algorithm 12. Then, Alice's cheating probability, $p_A^*(\mathcal{Q}) \leq 3/4$ and Bob's cheating probability, $p_B^*(\mathcal{Q}) \leq \cos^2(\pi/8)$ (which are the same as those in Lemma 8). Further, let $c_0, c_1, c_\perp \in \mathbb{R}$, and $\mathbb{C}_A(\mathcal{Q})$ be the set of cheat vectors for Alice. Then, as $N \to \infty$, the solution to the optimisation problem $\max(c_0\alpha + c_1\beta + c_\perp\gamma)$ over $(\alpha, \beta, \gamma) \in \mathbb{C}_A(\mathcal{Q})$ approaches that of a semi definite programme.*

The proof is again deferred; see Section 5.1.2.

4 Second Technique: Bias Suppression

In this section, we use the convention that $\mathcal{I}, \mathcal{P}$ and $\mathcal{Q}$ correspond to the protocols described in Algorithm 7, Algorithm 10 and Algorithm 12, respectively. Notice that $p_A^*(\mathcal{X}) > p_B^*(\mathcal{X})$ where $\mathcal{X} \in \{\mathcal{I}, \mathcal{P}, \mathcal{Q}\}$. We call such protocols "unbalanced". In this section we start from unbalanced WCF protocols and compose them to construct balanced WCF protocols. To this end, we introduce some notation and the term "polarity", to capture which among A and B is favoured.

Definition 14 (Unbalanced protocols, Polarity)**.** Given a WCF protocol $\mathcal{X}$, we say that it is unbalanced if $p_A^*(\mathcal{X}) \neq p_B^*(\mathcal{X})$. We say that $\mathcal{X}$ has polarity A if $p_A^*(\mathcal{X}) > p_B^*(\mathcal{X})$ and polarity B if $p_A^*(\mathcal{X}) < p_B^*(\mathcal{X})$.

Finally, let $X, Y \in \{A, B\}$ be distinct and suppose that $\mathcal{R}$ is an unbalanced protocol. Then, we define $\mathcal{R}_X$ to be protocol $\mathcal{R}$ where Alice's and Bob's roles are possibly interchanged so that $\mathcal{R}_X$ has polarity X, i.e. $p_X^*(\mathcal{R}_X) > p_Y^*(\mathcal{R}_X)$. We refer to $\mathcal{R}_X$ as $\mathcal{R}$ polarised towards X.

We now describe how these protocols can be composed such that the "winner gets polarity".

4.1 Composition

Definition 15 ($C(.,.)$ and $C(.)$)**.** Given two unbalanced WCF protocols, $\mathcal{X}$ and $\mathcal{Y}$, let $\mathcal{X}_A, \mathcal{X}_B$ and $\mathcal{Y}_A, \mathcal{Y}_B$ be their polarisations (see Definition 14). Define $C(\mathcal{X}, \mathcal{Y})$ as follows:

1. Alice and Bob execute $\mathcal{X}_A$ and obtain outcome $X \in \{A, B, \perp\}$.
2. If
   (a) $X = A$, execute $\mathcal{Y}_A$ and obtain outcome $Y \in \{A, B, \perp\}$, else if
   (b) $X = B$, execute $\mathcal{Y}_B$ and obtain outcome $Y \in \{A, B, \perp\}$, and finally if
   (c) $X = \perp$, set $Y = \perp$.

   Output Y.

Let $\mathcal{Z}^{i+1} := C(\mathcal{X}, \mathcal{Z}^i)$ for $i \geq 1$, and $\mathcal{Z}^1 := \mathcal{X}$. Then, formally, define $C(\mathcal{X}) := \lim_{i \to \infty} \mathcal{Z}^i$.[10]

The study of such composed protocols is simplified by assuming that in an honest run, neither player outputs $\perp$ (abort), i.e. they either output A or B. We take a moment to explain this.

Consider any protocol $\mathcal{R}$ where, when both players are honest, the probability of abort is zero. The protocols we have described so far, satisfy this property, so long as we assume that honest players can prepare perfect GHZ boxes. Such protocols are readily converted into protocols where an honest player never outputs abort.

For instance, suppose that in the execution of the aforementioned protocol $\mathcal{R}$ (with no-honest-abort), Alice behaves honestly but Bob is malicious. Suppose after interacting with Bob, Alice reaches the conclusion that she must abort. Since she knows that if Bob was honest, the outcome abort could not have arisen, she concludes that Bob is cheating and declares herself the winner, i.e. she outputs A. Similarly, when Bob is honest and after the interaction, reaches the outcome abort, he knows Alice cheated and can therefore declare himself the winner, i.e. output B.

Whenever we modify a protocol so that an honest Alice (Bob) replaces the outcome abort with Alice (Bob) winning, we say Alice (Bob) is *lenient*. This is motivated by the fact that when we compose protocols, if Alice can conclude that Bob is cheating, and she still outputs A instead of aborting, she is giving Bob further opportunity to cheat—she is being lenient.

Definition 16 ($\mathcal{R}$ with lenient players)**.** Suppose $\mathcal{R}$ is a WCF protocol such that when both players are honest, the probability of outcome abort, $\perp$, is zero. Then by "$\mathcal{R}$ *with lenient Alice (Bob)*", which we denote by $\mathcal{R}^{L\perp}$ ($\mathcal{R}^{\perp L}$), we mean that Alice (Bob) follows $\mathcal{R}$ except that the outcome $\perp$ replaced with A (B). Finally, by "*lenient* $\mathcal{R}$", which we denote by $\mathcal{R}^{LL}$, we mean $\mathcal{R}$ with lenient Alice and Bob.

For clarity and conciseness, we define C^{LL} to be compositions with lenient variants of the given protocols. We work out some examples of such protocols and analyse their security in the following section. These can be improved by considering $C^{L\perp}$ and $C^{\perp L}$—compositions where only one player is lenient. We discuss those afterwards.

[10] This is just to facilitate notation. This way the cheating probabilities p_A^* and p_B^* converge and numerically this only takes a few compositions to reach in our case.

Definition 17 (C^{LL}, $C^{\perp L}$ and $C^{L\perp}$)**.** Suppose a WCF protocol $\mathcal{X}$ can be transformed into its *lenient* variants (see Definition 16). Then define

$$\begin{aligned} C^{LL}(\mathcal{X},\mathcal{Y}) &:= C(\mathcal{X}^{LL},\mathcal{Y}),\\ C^{\perp L}(\mathcal{X},\mathcal{Y}) &:= C(\mathcal{X}^{\perp L},\mathcal{Y}), \quad \text{and}\\ C^{L\perp}(\mathcal{X},\mathcal{Y}) &:= C(\mathcal{X}^{L\perp},\mathcal{Y}). \end{aligned}$$

In words, C^{LL} is referred to as a *standard* composition, while $C^{\perp L}$ and $C^{L\perp}$ are referred to as *abort-phobic* compositions.

4.2 Standard Composition | C^{LL}

We begin with the simplest case, standard composition, C^{LL}. Let us take an example. Consider protocol $\mathcal{P}$ (see Algorithm 10) and recall (see Lemma 11)

$$\begin{aligned} p^*_A(\mathcal{P}_A) &=: \alpha \approx 0.852\ldots,\\ p^*_B(\mathcal{P}_A) &=: \beta \approx 0.667\ldots. \end{aligned}$$

Note that therefore $p^*_A(\mathcal{P}_B) = \beta$ and $p^*_B(\mathcal{P}_B) = \alpha$. Further, let $\mathcal{P}' := C^{LL}(\mathcal{P},\mathcal{P})$, i.e. Alice and Bob (who are both lenient) first execute $\mathcal{P}_A$ and if the outcome is A, they execute $\mathcal{P}_A$, while if the outcome is B, they execute $\mathcal{P}_B$. This is illustrated in Figure 1 where note that the event abort doesn't appear due to the leniency assumption. This allows us to evaluate the cheating probabilities for the resulting protocol as

$$\begin{aligned} p^*_A(\mathcal{P}') &= \alpha\alpha + (1-\alpha)\beta =: \alpha^{(1)}, \quad \text{and}\\ p^*_B(\mathcal{P}') &= \beta\alpha + (1-\beta)\beta =: \beta^{(1)}. \end{aligned} \tag{3}$$

To see this, consider Equation (3). Alice knows that if she wins the first round, her probability of winning is $\alpha > \beta$. She knows that in the first round, she can force the outcome A with probability α. From leniency, she knows that Bob would output B with the remaining probability.[11]

A side remark: one consequence of this simplified analysis is that[12] $\alpha^{(1)} > \beta^{(1)}$. Intuitively, it means that plority is preserved by the composition procedure. Proceeding similarly, i.e. defining $\mathcal{P}'' := C^{LL}(\mathcal{P},\mathcal{P}')$, and repeating $k+1$ times overall, one obtains[13]

$$\begin{aligned} \alpha^{(k+1)} &= \alpha\alpha^{(k)} + (1-\alpha)\beta^{(k)}\\ \beta^{(k+1)} &= \beta\alpha^{(k)} + (1-\beta)\beta^{(k)}. \end{aligned}$$

In the limit of $k \to \infty$, one obtains

$$p^*_A(C^{LL}(\mathcal{P})) = p^*_B(C^{LL}(\mathcal{P})) = \lim_{k\to\infty}\alpha^{(k)} = \lim_{k\to\infty}\beta^{(k)} \approx 0.8199\ldots.$$

Proceeding similarly, one obtains for $X \in \{A,B\}$ and $\mathcal{X} \in \{\mathcal{I},\mathcal{Q}\}$,

$$p^*_X(C^{LL}(\mathcal{X})) \approx 0.836\ldots$$

We thus have the following.

Theorem 18. *Let $X \in \{A,B\}$ and $\mathcal{X} \in \{\mathcal{I},\mathcal{Q}\}$. Then $p^*_X(C^{LL}(\mathcal{P})) \approx 0.8199\ldots$ and $p^*_X(C^{LL}(\mathcal{X})) \approx 0.836\ldots$.*

[11] Without leniency, this probability could have been shared between the outcomes $\perp$ (abort) and B. Consequently, only upper bounds could be obtained on $p^*_A(\mathcal{P}')$ and $p^*_B(\mathcal{P}')$ using only α and β as security guarantees for $\mathcal{P}_A$. Upper bounds, however, would not be enough to determine the polarity of $\mathcal{P}'$ and an stymie unambiguous repetition of the composition procedure (at least as it is defined). One could nevertheless compose by hoping that the upper bounds can be used to accurately represent the polarity. This would still yield a protocol and the same calculation would yield correct bounds but the composition itself might be sub-optimal.

[12] $\alpha^{(1)} - \beta^{(1)} = (\alpha-\beta)\alpha - (\alpha-\beta)\beta = (\alpha-\beta)^2 > 0$

[13] Again, note that $\alpha^{(k+1)} - \beta^{(k+1)} = (\alpha^{(k)} - \beta^{(k)})(\alpha-\beta) > 0$.

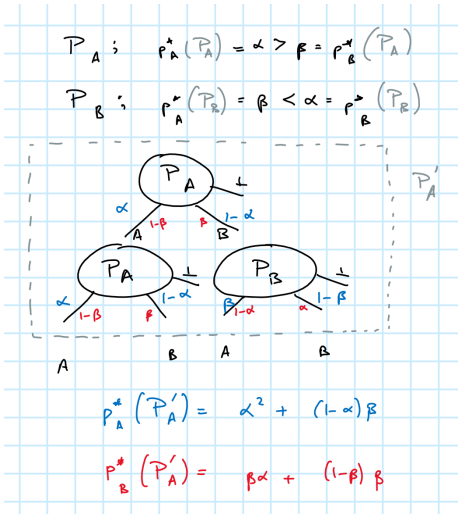

Figure 1: Standard analysis (TODO: remove the abort)

4.3 Abort Phobic Compositions | $C^{L\perp}, C^{\perp L}$

We now look at the case of abort phobic compositions, $C^{L\perp}$ and $C^{\perp L}$. We work through essentially the same example as above and see what changes in this setting. Consider protocol $\mathcal{P}$ (see ...) and recall that as before

$$p_A^*(\mathcal{P}_A) =: \alpha \approx 0.852\ldots,$$
$$p_B^*(\mathcal{P}_A) =: \beta \approx 0.667\ldots.$$

In addition, we know from Lemma 11 that cheat vectors for Bob, $(\alpha, \beta, \gamma) \in \mathbb{C}_B(\mathcal{P}_A)$ admit a nice characterisation courtesy of the self testing step. Let $\mathcal{P}' := C^{\perp L}(\mathcal{P}, \mathcal{P})$, i.e. Alice and Bob execute $\mathcal{P}_A$ and if the outcome is A, they execute $\mathcal{P}_A$ while if the outcome is B, they execute $\mathcal{P}_B$. Bob is assumed to be lenient so an honest Bob never outputs abort, $\perp$. However, an honest Alice can output abort, $\perp$ so we keep that output in the illustration, Lemma 11. Our goal is to find $p_A^*(\mathcal{P}')$ and $p_B^*(\mathcal{P}')$. The former is the same as before because Bob is lenient:

$$p_A^*(\mathcal{P}') = \alpha \cdot \alpha + (1-\alpha)\cdot\beta.$$

Clearly, $p_B^*(\mathcal{P}') \leq \beta\alpha + (1-\beta)\beta$ but this bound may not be tight because $(1-\beta)$ is the combined probability of Alice aborting and Alice outputting A. However, we can use cheat vectors to obtain

$$p_B^*(\mathcal{P}') = \max_{(v_A, v_B, v_\perp)\in\mathbb{C}_B} v_B\alpha + v_A\beta$$

which is an SDP one can solve numerically. Unlike the previous case, the polarity of the resulting protocol, $\mathcal{P}'$, might have flipped (compared to the polarity of $\mathcal{P}$).

Repeating this procedure, one can consider $\mathcal{P}'' := C^{\perp L}(\mathcal{P}, \mathcal{P}')$ and obtain $p_A^*(\mathcal{P}'')$ directly as illustrated above and numerically solve for $p_B^*(\mathcal{P}'')$ using the cheat vectors. Numerically, we found that ten-fifteen repetitions caused the cheating probabilities to converge to approximately 0.81459. We saw that the abort probabilities associated with $\mathcal{P}$ were quite small and therefore one could hope that $\mathcal{Q}$ fares better. Proceed analogously for protocol and considering $\mathcal{Q}' := C^{L\perp}(\mathcal{Q}, \mathcal{Q})$, $\mathcal{Q}'' := C^{L\perp}(\mathcal{Q}, \mathcal{Q}')$, etc., the cheating probabilities converge to approximately 0.822655.

Theorem 19. *Let $X \in \{A, B\}$. Then*

$$p_X^*(C^{\perp L}(\mathcal{P})) \approx 0.81459$$

and

$$p_X^*(C^{L\perp}(\mathcal{Q})) \approx 0.822655$$

where the latter holds assuming Conjecture ?? is true.

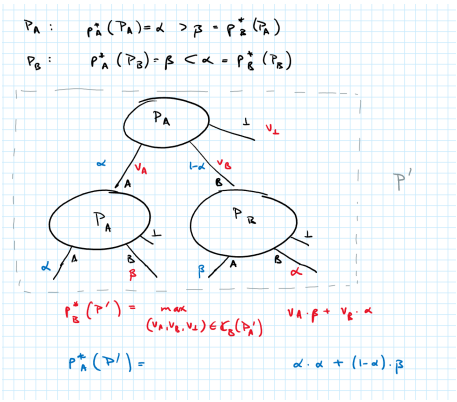

Figure 2: Cheat vector analysis. (TODO: improve the caption)$(v_A, v_B, v_\perp) \in \mathbb{C}_B$;

While by itself Q doesn't seem to help, one can suppress the bias further, by noting that at the very last step, only the cheating probabilities $p^*_A(Q)$ and $p^*_B(Q)$ played a role (i.e. the fact that the cheating vectors $\mathbb{C}_A$ for Q had an SDP characterisation was not used). Further, we know that $p^*_A(\mathcal{P}) = p^*_A(Q)$ but $p^*_B(\mathcal{P}) < p^*_B(Q)$, i.e. using $\mathcal{P}$ at the very last step will result in a strictly better protocol.

Theorem 20. *Let* $X \in \{A, B\}$,

$$\mathcal{Z}^1 := C(Q, \mathcal{P}), \quad \text{and}$$
$$\mathcal{Z}^{i+1} := C(Q, \mathcal{Z}^i) \quad i > 1.$$

Then

$$\lim_{i\to\infty} p^*_X(\mathcal{Z}^i) \approx 0.791044\ldots$$

assuming Conjecture ?? holds.

5 Security Proof | Asymptotic

In this section, we prove the security under the following assumption:

Assumption 21. *In protocol* $\mathcal{P}$ (Q)*, Alice (Bob) does not perform the box verification step and instead it is assumed that her box is (his boxes are) taken from a trio of boxes which win the GHZ game with certainty.*

Later, we drop the assumption and use the box verification step (see ..) to estimate the probability of winning the GHZ game. When the winning probability is exactly one, the states and measurements are the same as the GHZ state and σ_x, σ_y measurements, up to local isometries and this allows us to use semi definite programming.

Lemma 22. *Let* $a, b, c, x, y, z \in \{0, 1\}$*. Consider a trio of quantum boxes, specified by projectors* $\{M^A_{a|x}, M^B_{b|y}, M^C_{c|z}\}$ *acting on finite dimensional Hilbert spaces* $\mathcal{H}^A, \mathcal{H}^B$ *and* $\mathcal{H}^C$*, and* $|\psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B \otimes \mathcal{H}^C =: \mathcal{H}^{ABC}$*. If the trio pass the GHZ test with certainty, then there exists a local isometry*

$$\Phi = \Phi^A \otimes \Phi^B \otimes \Phi^C : \mathcal{H}^{ABC} \to \mathcal{H}^{ABC} \otimes \mathbb{C}^{2\times 3}$$

such that

$$\Phi\left(|\psi\rangle\right) = |\chi\rangle \otimes |\text{junk}\rangle\,,$$
$$\Phi\left(M^D_{d|t}\,|\psi\rangle\right) = \Pi^D_{d|t}\,|\text{GHZ}\rangle \otimes |\text{junk}\rangle \quad \forall D \in \{A, B, C\},\ \text{and}\ d, t \in \{0, 1\}$$

where $|\text{GHZ}\rangle = \frac{|000\rangle+|111\rangle}{\sqrt{2}} \in \mathbb{C}^{2\times 3}$, $|\text{junk}\rangle \in \mathcal{H}^{ABC}$ *is some arbitrary state and* $\{\Pi^A_{a|x}, \Pi^B_{b|y}, \Pi^C_{c|z}\}$ *are projectors corresponding to* σ_x *on the first, second and third qubit of* $|\text{GHZ}\rangle$ *respectively, for* $x = 0$ *and corresponding to* σ_y *for* $x = 1$*, as in Claim 6.*

INTERNAL; (TODO: remove): Isometries can only increase dimensions (they must be injective; that is to ensure they preserve inner products of vectors). Therefore the isometry can't get rid of the |junk⟩ part.

5.1 Cheat vectors optimisation using Semi Definite Programming

5.1.1 SDP when Alice self-tests

Asymptotic proof of Lemma 11. We prove Lemma 11 under Assumption 21. We begin by making two observations.

First, note that in the protocol, if Alice applies an isometry on her box *after* she has inputted x, obtained the outcome a (and has noted it somewhere), the security of the resulting protocol is unchanged because the rest of the protocol only depends on x and a, and Alice's isometry only amounts to relabelling of the post measurement state. This freedom allows us to simplify the analysis.

Second, in the analysis, we cannot model Alice's random choice, say for x, as a mixed state because Bob can always hold a purification and thus know x. Therefore, we model the randomness using pure states and measure them in the end.

Notation: Other than PQR, all other registers store qubits.

We proceed step by step.

1. We can model (justified below) Alice's act of inputting a random x and obtaining an outcome a from her box through the state

$$|\Psi_0\rangle := \frac{1}{2} \sum_{x,a\in\{0,1\}} |xa\rangle_{XA} |\Phi(x,a)\rangle_{IJ}$$

where X represents the random input and A the output. Here, $|\Phi(x,a)\rangle_{IJ}$ are Bell states (see Equation (5)) and the registers IJ are held by Bob. Alice's act of choosing r at random, computing $s = a \oplus x.r$ is modelled as

$$|\Psi_1\rangle := \frac{1}{2\sqrt{2}} \sum_{x,a,r\in\{0,1\}} |xa\rangle_{XA} |\Phi(x,a)\rangle_{IJ} |r\rangle_R |a \oplus x.r\rangle_S . \tag{4}$$

Finally, Alice's act of sending s is modelled as Alice starting with the state

$$\mathrm{tr}_{IJS} \left[|\Psi_1\rangle \langle\Psi_1| \right] \in XAR.$$

Justification for starting with $|\Psi_0\rangle$.
To see why we start with the state $|\Psi_0\rangle$, model Alice's choice of x as $|+\rangle_X$, suppose her measurement result is stored in $|0\rangle_A$, the state of the boxes before measurement is $|\psi\rangle_{PQR}$ and Alice holds P, i.e.

$$\left|\Psi_0'\right\rangle := |+\rangle_X |0\rangle_A |\psi\rangle_{PQR} .$$

Let $\{M^P_{a|x}\}$ be the measurement operators corresponding to Alice's box. The measurement process is unitarily modelled as

$$\left|\Psi_1'\right\rangle := U_{\text{measure}} \left|\Psi_0'\right\rangle = \frac{1}{\sqrt{2}} \sum_{x,a\in\{0,1\}} |x\rangle_X |a\rangle_A M^P_{a|x} |\psi\rangle_{PQR}$$

where

$$U_{\text{measure}} = \sum_{x\in\{0,1\}} |x\rangle\langle x|_X \otimes \left[\mathbb{I}_A \otimes M^P_{0|x} + X_X \otimes M^P_{1|x} \right] \otimes \mathbb{I}_{QR}.$$

Now we harness the freedom of applying an isometry to the post measured state (as observed above). We choose the local isometry in Lemma 22. Without loss of generality, we can assume that Bob had already applied his part of the isometry before sending the boxes (because he can always reverse it when it is his turn). We thus have,

$$\begin{aligned} \left|\Psi_2'\right\rangle := \Phi_{PQR} \left|\Psi_1'\right\rangle &= \frac{1}{\sqrt{2}} \sum_{x,a\in\{0,1\}} |x\rangle_X |a\rangle_A \Pi^H_{x|a} |\mathrm{GHZ}\rangle_{HIJ} \otimes |\mathrm{junk}\rangle_{PQR} \\ &= \frac{1}{2} \sum_{x,a\in\{0,1\}} |x\rangle_X |a\rangle_A U^H(x,a) |0\rangle_H |\Phi(x,a)\rangle_{IJ} \otimes |\mathrm{junk}\rangle_{PQR} \end{aligned}$$

where

$$|\Phi(x,a)\rangle_{IJ} = \frac{|00\rangle + (-1)^a (i)^x |11\rangle}{\sqrt{2}} \tag{5}$$

and $U^H(x,a)\,|0\rangle_H$ is $\frac{|0\rangle+(-1)^a(i)^x|1\rangle}{\sqrt{2}}$. Since the state of register H is completely determined by registers X and A, we can drop it from the analysis without loss of generality. Finally, since $|\text{junk}\rangle_{PQR}$ is completely tensored out, we can drop it too without affecting the security. Formally, we can assume that Alice gives Bob the register P at this point.

2. Bob sending g is modelled by introducing $\rho_2 \in XARG$ satisfying $\text{tr}_{IJS}\left[|\Psi_1\rangle\langle\Psi_1|\right] = \text{tr}_G(\rho_2)$.

3. At this point, either $x \oplus g$ is zero, in which case Alice's output is fixed or $x \oplus g$ is one, and in that case Bob will already know x because he knows g (he sent it) and Alice will proceed to testing Bob. Formally, therefore, we needn't do anything at this step.

4. Assuming $x \oplus g = 1$, Alice sends y, z to Bob such that $x \oplus y \oplus z = 1$. However, since Bob already knows x, he can deduce z from y. We thus only need to model Alice sending y and Bob responding with $d = b \oplus c$ (because Alice will only use $b \oplus c$ to test the GHZ game, so it suffices for Bob to send d). This amounts to introducing $\rho_3 \in XARGYD$ satisfying $\rho_2 \otimes \frac{\mathbb{I}_Y}{2} = \text{tr}_D(\rho_3)$.

5. Since we postponed the measurements to the end, we add this last step. Alice now measures ρ_3 to determine $x \oplus g$ and if it is one, whether the GHZ test passed. Let

$$\Pi_i := \sum_{x,y\in\{0,1\}:x\oplus g=i} |xg\rangle\langle xg|_{XG} \otimes \mathbb{I}_{ARYD},$$
$$\Pi^{\text{GHZ}} := \sum_{\substack{x,y\in\{0,1\},\\ a,d\in\{0,1\}:a\oplus d\oplus 1=xy\cdot(1\oplus x\oplus y)}} |xyad\rangle\langle xyad|_{XYAD} \otimes \mathbb{I}_{RG}. \tag{6}$$

   Then, we can write the cheat vector for Alice, i.e. the tuple of probabilities that Alice outputs 0, 1 and abort (see Definition 9), as

$$(\alpha,\beta,\gamma) = (\text{tr}(\Pi_0\rho_3), \text{tr}(\Pi_1\Pi^{\text{GHZ}}\rho_3), \text{tr}(\Pi_1\bar{\Pi}^{\text{GHZ}}\rho_3))$$

   where $\bar{\Pi} := \mathbb{I} - \Pi$.

To summarise, the final SDP is as follows: let $|\Psi_1\rangle \in XAIJRS$ be as given in Equation (4), $\rho_2 \in XARG$ and $\rho_3 \in XARGYD$

$$\max \quad \text{tr}([c_0\Pi_0 + \Pi_1(c_1\Pi^{\text{GHZ}} + c_\perp\bar{\Pi}^{\text{GHZ}})]\rho_3)$$

subject to

$$\text{tr}_{IJS}\left[|\Psi_1\rangle\langle\Psi_1|\right] = \text{tr}_G(\rho_2)$$
$$\rho_2 \otimes \frac{\mathbb{I}_Y}{2} = \text{tr}_D(\rho_3)$$

where the projectors are defined in Equation (6). □

5.1.2 SDP when Bob self-tests

Proof of Algorithm 12. Denote by $\mathcal{I}$ the protocol corresponding to Algorithm 7.

It is evident that $p_B^*(Q) \leq p_B^*(\mathcal{I})$ because compared to $\mathcal{I}$, in Q Alice performs an extra test. However, it is not hard to see that the inequality is saturated, i.e. $p_B^*(Q) = p_B^*(\mathcal{I})$. Consider ... (TODO: recall/re-construct the cheating strategy for Bob that lets him win with the same 3/4 probability).

From Lemma 8, it is also clear that $p_A^*(Q) = p_A^*(\mathcal{I})$ because the only difference between Bob's actions in Q and $\mathcal{I}$ is that Bob self-tests to ensure his boxes are indeed GHZ. However, the optimal cheating strategy for $\mathcal{I}$ can be implemented using GHZ boxes.

This establishes the first part of the lemma. For the second part, i.e. establishing that optimising $c_0\alpha + c_1\beta + c_\perp\gamma$ over $(\alpha,\beta,\gamma) \in \mathbb{C}_A$ is an SDP, we proceed as follows. Suppose Assumption 21 holds. Then we can assume that Bob starts with the state

$$\rho_0 := \text{tr}_H(|\text{GHZ}\rangle\langle\text{GHZ}|_{HIJ}) \tag{7}$$

and the effect of measuring the two boxes can be represented by the application of projectors of pauli operators X and Z.

The justification is similar to that given in the former proof. Suppose Bob holds registers QR of $|\psi\rangle_{PQR}$ which is the combined state of the three boxes. Suppose his measurement operators are $\{M^Q_{b|y}, M^R_{c|z}\}$. Then using the isometry in Lemma 22, Bob can relabel his state (and without loss of generality, we can suppose Alice also relabels according to the aforementioned isometry) to get $\Phi_{PQR}|\psi\rangle_{PQR} = |\text{GHZ}\rangle_{HIJ} \otimes |\text{junk}\rangle_{PQR}$. Further, since $\Phi_{PQR}(M^Q_{b|y} \otimes M^R_{c|z}|\psi\rangle_{PQR}) = \Pi^I_{b|y}\Pi^J_{c|z}|\text{GHZ}\rangle_{HIJ} \otimes |\text{junk}\rangle_{PQR}$ Bob's act of measurement, in the new labelling, corresponds to simply measuring the GHZ state in the appropriate Pauli basis.

1. Bob receiving s from Alice is modelled by introducing $\rho_1 \in SIJ$ satisfying $\text{tr}_S(\rho_1) = \rho_0$.
2. Bob sending $g \in_R \{0,1\}$ can be seen as appending a mixed state: $\rho_1 \otimes \frac{1}{2}\mathbb{I}_G$.
3. Alice sending x (and a) can be modelled as introducing $\rho_2 \in AXSIJG$ satisfying $\text{tr}_A(\rho_2) = \rho_1 \otimes \frac{\mathbb{I}_G}{2}$.
4. To model the GHZ test, introduce a register Y in the state $\frac{|0\rangle_Y + |1\rangle_Y}{\sqrt{2}}$. Recall that to perform the GHZ test, we need $x \oplus y \oplus z = 1$ i.e. $z = 1 \oplus y \oplus x$. Further introduce registers B and C to hold the measurement results, define
$$U := \sum_{y,x\in\{0,1\}} |yx\rangle\langle yx|_{YX} \otimes (\mathbb{I}_B \otimes \Pi^I_{0|y} + X_B \otimes \Pi^I_{1|y}) \otimes (\mathbb{I}_C \otimes \Pi^J_{0|(1\oplus y\oplus x)} + X_C \otimes \Pi^J_{1|(1\oplus y\oplus x)}) \otimes \mathbb{I}_{ASG}. \quad (8)$$
By construction, $\rho_3 := U(|+\rangle\langle+|_Y \otimes |00\rangle\langle 00|_{BC} \otimes \rho_2)U^\dagger \in YBCAXSIJG$ models the measurement process. (TODO: this equality would become approximately true...but perhaps the noise can be absorbed in ρ_0 with some argument)
5. Since we postponed the measurements to the end, we add this step. Define
$$\Pi_i := \sum_{x,g\in\{0,1\}:x\oplus g=i} |xg\rangle\langle xg|_{XG} \otimes \mathbb{I}_{YABSIJ}$$
to determine who won. Define
$$\Pi^{\text{sTest}} := \sum_{s,a,x\in\{0,1\}:s=a\vee s=a\oplus x} |sax\rangle\langle sax|_{SAX} \otimes \mathbb{I}_{GYBCIJ}$$
to model the first test, i.e. s should either be a or $a \oplus x$. Define
$$\Pi^{\text{GHZ}} := \sum_{\substack{x,y\in\{0,1\},\\ a,b,c\in\{0,1\}:a\oplus b\oplus c\oplus 1=xy\cdot(1\oplus x\oplus y)}} |xyabc\rangle\langle xyabc|_{XYABC} \otimes \mathbb{I}_{GSIJ}$$
to model the GHZ test. Let
$$\Pi^{\text{Test}} := \Pi^{\text{GHZ}}\Pi^{\text{sTest}}, \quad \bar{\Pi}^{\text{Test}} := \mathbb{I} - \Pi^{\text{Test}}. \quad (9)$$
One can then write the cheat vector for Bob, i.e. the tuple of probabilities that Bob outputs 0, 1 and abort (see Definition 9), as
$$(\alpha, \beta, \gamma) = (\text{tr}(\Pi_0\Pi^{\text{Test}}\rho_3), \text{tr}(\Pi_1\rho_3), \text{tr}(\Pi_0\bar{\Pi}^{\text{Test}}\rho_3)).$$

To summarise, the final SDP is as follows: let $\rho_0 \in IJ$ be as defined in Equation (7), $\rho_1 \in SIJ$ and $\rho_2 \in AXSIJG$. Then,
$$\max \quad \text{tr}\left([\Pi_0(c_0\Pi^{\text{Test}} + c_\perp\bar{\Pi}^{\text{Test}}) + c_1\Pi_1]U(|+00\rangle\langle+00|_{YBC} \otimes \rho_2)U^\dagger\right)$$
subject to
$$\text{tr}_S(\rho_1) = \rho_0$$
$$\text{tr}_A(\rho_2) = \frac{1}{2}\rho_1 \otimes \mathbb{I}_G$$
where U is as defined in Equation (8) and the projectors as in Equation (9).

□

6 Security Analysis | Finite n

6.1 Alice self tests

The basic idea here is to treat the state and measurements inside the boxes as variables which are optimised over, subject to the constraint that they are ϵ-close to the ideal GHZ state and measurements. This ceases to be an SDP so we relax the constraint that the post measurement states must arise from measuring some fixed state and let them be arbitrary states. The requirement that these are close to the ideal GHZ state and post measured states is still enforced. When $\epsilon = 0$, we recover the asymptotic SDP (and that is no longer a relaxation). Since we only change the constant ϵ, convergence of the objective value of these SDPs is easy to show [JAMIE].

aoeu

Proof. We first write exactly what is going on physically, except that we take the liberty of "renaming", i.e. applying global isometries. We treat the state $|\psi\rangle_{PQR}$ and the measurements $\{M^P_{a|x}\}$ as variables.

1. We begin as before with $\left|\Psi'_0\right\rangle$,
$$\left|\Psi'_0\right\rangle := |+\rangle_X |0\rangle_A |\psi\rangle_{PQR'}$$
and[14] obtain the "post measurement state" as
$$\left|\Psi'_1\right\rangle = \frac{1}{2} \sum_{x,a\in\{0,1\}} |x\rangle_X |a\rangle_A M^P_{a|x} |\psi\rangle_{PQR'} .$$
Since we are allowed to "rename" (without changing the value of the SDP), we have
$$\left|\Psi'_2\right\rangle = \frac{1}{2} \sum_{x,a\in\{0,1\}} |x\rangle_X |a\rangle_A \Phi_{PQR} M^P_{a|x} |\psi\rangle_{PQR'} . \tag{10}$$
At this point, in the asymptotic case, we could directly apply the self-testing result and replace $\Phi_{PQR'} M^P_{a|x} |\psi\rangle_{PQR'}$ with $\Pi^H_{x|a} |\mathrm{GHZ}\rangle_{HIJ} \otimes |\mathrm{junk}\rangle_{PQR'}$. Now, instead, we require
$$\left\| \Phi_{PQR'} M^P_{a|x} |\psi\rangle_{PQR'} - \Pi^H_{x|a} |\mathrm{GHZ}\rangle_{HIJ} \otimes |\mathrm{junk}\rangle_{PQR'} \right\| \leq \epsilon \quad \forall a, x \in \{0,1\}$$
[EDIT: here the norm is just the vector norm but we can impose it as density matrices; there we use the trace norm; the remark shows how to convert that into an SDP]where the norm[15] here is the trace norm $\|.\|$ (see Remark 23) and ϵ' is a function ϵ which vanishes as ϵ vanishes (ϵ comes from the self testing step). One could, henceforth, continue as in the asymptotic case. More precisely, one could start with $|\Psi_0\rangle := \left|\Psi'_2\right\rangle$, model the classical computation step as
$$\begin{aligned} |\Psi_1\rangle &= U_{\mathrm{comp}} |\Psi_0\rangle |00\rangle_{RS} \\ &= \frac{1}{2\sqrt{2}} \sum_{x,a,r\in\{0,1\}} |xa\rangle_{XA} |r\rangle_R |a \oplus x.r\rangle_S \Phi_{PQR} M^P_{a|x} |\psi\rangle_{PQR'} \end{aligned}$$
where U_{comp} is implicitly defined to yield the stated state. Then, the act of sending s (which is the first communication step) is modelled as
$$\mathrm{tr}_{IJSPQR'}(|\Psi_1\rangle \langle\Psi_1|) \in XARH.$$
2. The remaining steps are unchanged except that Alice additionally, always holds the register H now.

The final optimisation problem is defined on the variables $|\psi\rangle \in PQR'$, $M^P_{a|x}$ projectors (or POVMs) acting on PQR, $\Phi_{PQR'}$ a local isometry[16] from $PQR' \to HIJPQR'$, $|\mathrm{junk}\rangle \in PQR'$, $\rho_2 \in XARGH$ and $\rho_3 \in XARGYDH$. The problem is:
$$\max \quad \mathrm{tr}([c_0\Pi_0 + \Pi_1(c_1\Pi^{\mathrm{GHZ}} + c_\perp \bar{\Pi}^{\mathrm{GHZ}})]\rho_3)$$

[14] (we put R' because we already used R for the random register)
[15] We could have used other norms but they would be a relaxation of the constraints.
[16] by local we mean it has the form $\Phi_P \otimes \Phi_Q \otimes \Phi_R$ where, for instance, $\Phi_P : P \to HP$.

subject to

$$\left\| \Phi_{PQR'} M^P_{a|x} \left|\psi\right\rangle_{PQR'} - \Pi^H_{x|a} \left|\mathrm{GHZ}\right\rangle_{HIJ} \otimes \left|\mathrm{junk}\right\rangle_{PQR'} \right\| \leq \epsilon \quad \forall a, x \in \{0,1\} \tag{11}$$
$$\left|\Psi_1\right\rangle := U_{\mathrm{comp}} \left|\Psi_0\right\rangle \left|00\right\rangle_{RS}$$
$$\mathrm{tr}_{IJSPQR'} \left[\left|\Psi_1\right\rangle \left\langle\Psi_1\right|\right] = \mathrm{tr}_G(\rho_2)$$
$$\rho_2 \otimes \frac{\mathbb{I}_Y}{2} = \mathrm{tr}_D(\rho_3)$$

where $\left|\Psi_0\right\rangle$ is as defined above (see Equation (10) and recall that $\left|\Psi_0\right\rangle = \left|\Psi_2'\right\rangle$). This, as it is stated, is not an SDP. However, it is clear that when $\epsilon = 0$, we recover the asymptotic case (many variables can be dropped because they either are fixed (and no longer variable, e.g. $\left|\Psi_0\right\rangle$) or become redundant, e.g. register H). Let $v(\epsilon, d)$ be the value of the optimization program above where d encodes the dimension of systems PQR. We now relax the constraints to obtain an SDP. Let $v'(\epsilon, d)$ be its value. We want the relaxation to be such that $v'(0, d) = v(0, d)$. Additionally, because it is a relaxation, we know $v(\epsilon, d) \leq v'(\epsilon, d)$. It then suffices to show the continuity of the relaxation (the SDP), to establish the convergence of $v(\epsilon, d)$ to $v(0, d)$ as $\epsilon \to 0$ [JAMIE double-check!].

There are two steps to the relaxation. First, we relax the Equation (11) as in Remark 23. This is straightforward. Second, we remove the variables $\left|\psi\right\rangle, M^P_{a|x}$ and $\Phi_{PQR'}$ and instead introduce variables $\left|\xi^{a,x}\right\rangle \in HIJPQR'$ for $a, x \in \{0,1\}$. We substitute[17] $\left|\psi\right\rangle$ with $\left|\xi\right\rangle$ and $M^P_{a|x} \left|\psi\right\rangle$ with $\left|\xi^{a,x}\right\rangle$ in the definition of $\left|\Psi_0\right\rangle$ and in the constraint Equation (11). This is evidently a relaxation (because one can represent any choice of $\left|\psi\right\rangle, M^P_{a|x}$ and $\Phi_{PQR'}$ using $\left|\xi\right\rangle$ and $\left|\xi_{a,x}\right\rangle$ in the optimisation problem). Relaxing further to mixed states, the SDP is then defined on $\xi^{aa',xx'} \in \mathrm{L}(HIJPQR')$ for $a, a', x, x' \in \{0,1\}$, $\rho_{\mathrm{junk}} \in \mathrm{PSD}(PQR')$, $\rho_2 \in \mathrm{PSD}(XARGH)$ and $\rho_3 \in \mathrm{PSD}(XARGYDH)$ as

$$\max \quad \mathrm{tr}([c_0\Pi_0 + \Pi_1(c_1\Pi^{\mathrm{GHZ}} + c_\perp \bar{\Pi}^{\mathrm{GHZ}})]\rho_3)$$

subject to

$$\left\| \xi^{aa',xx'} - \Pi^H_{a|x} \left|\mathrm{GHZ}\right\rangle \left\langle\mathrm{GHZ}\right|_{HIJ} \Pi^H_{a'|x'} \otimes \rho_{\mathrm{junk}} \right\| \leq \epsilon'' \quad \forall a, a', x, x' \in \{0,1\}$$
$$\bar{\Psi}_1 := U_{\mathrm{comp}} \bar{\Psi}_0 \otimes \left|00\right\rangle \left\langle 00\right|_{RS} U^\dagger_{\mathrm{comp}}$$
$$\mathrm{tr}_{IJSPQR'} \left[\bar{\Psi}_1\right] = \mathrm{tr}_G(\rho_2)$$
$$\rho_2 \otimes \frac{\mathbb{I}_Y}{2} = \mathrm{tr}_D(\rho_3)$$

where

$$\bar{\Psi}_0 := \frac{1}{4} \sum_{x,x',a,a' \in \{0,1\}} \left|xa\right\rangle \left\langle x'a'\right|_{XA} \xi^{aa',xx'}_{HIJPQR'}$$

and ϵ'' is a function of ϵ which vanishes as ϵ vanishes. Clearly, when $\epsilon = 0$, we recover the asymptotic SDP and by construction, the SDP is a relaxation of the optimisation problem we started with. Recall that $v'(\epsilon, d)$ is the value of this SDP. It is easy [for JAMIE! please help] that $v'(\epsilon, d)$ is continuous as a function of ϵ (at least for small ϵ?); my guess would be that we are slowly enlarging the feasible region so won't expect any jumps. □

aoeu

Remark 23. It is straightforward to show that $\|\left|\rho\right\rangle - \left|\sigma\right\rangle\| \leq \epsilon \implies \mathrm{tr}\left|\rho - \sigma\right| \leq \epsilon'$ where $\epsilon' = 2\sqrt{1 - (1 - \epsilon)}$.

For many norms (including the trace norm), we have $\|X\| \leq \epsilon \implies |\lambda_{\max}(X)| \leq \epsilon'$ where ϵ' vanishes as ϵ vanishes. It is easy to bound $\lambda_{\max}(X) \leq \epsilon$ as

$$\begin{pmatrix} X - \epsilon\mathbb{I} & 0 \\ 0 & \epsilon\mathbb{I} - X \end{pmatrix} \geq 0.$$

This is an SDP constraint because we can define some $\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix} \geq 0$ and then set the linear constraint $Y_{11} = X - \epsilon\mathbb{I}$ and $Y_{22} = \epsilon\mathbb{I} - X$.

[17] we can drop the pure state requirement; we use it for notational simplicity

[EDIT When X is not Hermitian, we can relax it using Schur's complement as

$$\begin{pmatrix} \mathbb{I} & X \\ X^T & \epsilon''\mathbb{I} \end{pmatrix} \geq 0 \iff \epsilon''\mathbb{I} \geq X^T X$$

and if $\|X\| \leq \epsilon$, then there should be some function ϵ'' of ϵ that satisfies the above (with possibly a multiplicative factor of $\dim(X)$).]

6.2 Bob self tests

For Bob's case, we work out an example which is essentially the same as what we want to prove. In this case, we are unable to find a simple SDP relaxation as above and instead rely on the NPA hierarchy for the continuity result.

Example 24. We consider three optimisation problems. The first is supposed to be the "asymptotic version", the second is supposed to by a toy model of what happens in the lab with ϵ as a parameter, and finally the third is an SDP relaxation of the second, obtained using the NPA hierarchy.

First: Let $\rho_0 := \mathrm{tr}_{HI}\left[|\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|\right]_{HIJ}$. The variable is $\rho_1 \in ZJ$. The SDP program is

$$\max \quad \mathrm{tr}(\Pi_{\mathrm{obj}}\rho_2\Pi_{\mathrm{obj}})$$

subject to

$$\begin{aligned} \mathrm{tr}_Z(\rho_1) &= \rho_0 \\ \rho_2 &= \sum_{\substack{z,z' \\ c,c'}} |c\rangle\langle c'|_C \quad \otimes \quad \Pi^J_{c|z} \otimes \Pi^Z_z \quad \rho_1 \quad \Pi^J_{c'|z'} \otimes \Pi^Z_{z'} \end{aligned}$$

where Π_{obj} is an arbitrary but fixed projector which acts non-trivially on registers CZ and $\{\Pi^J_{c|z}\}$ constitute two sets of projective measurements, the setting indexed by z and outcome by c.

The main simplifications we make, compared to Bob's asymptotic SDP, are:
(1) we keep only the J register from the HIJ registers used in the GHZ test,
(2) we skipped the part where Alice first sends s, then Bob sends g and in turn Alice sends x and a which are finally used to do the test; we simply have her send z, the basis in which to measure,
(3) the action of the (appropriately adapted) unitary U is captured directly by defining ρ_3
(4) the final measurement operator is left arbitrary so long as it acts on "classical registers", CZS.

These simplifications can be undone with the main idea unchanged. We now proceed with defining the second variant which has the PQR registers as well.

Second: The variables are $\rho_0 \in HIJPQR$, $\rho_1 \in ZJR$, $\rho_{\mathrm{junk}} \in R$ and $\{M_{0|z}, M_{1|z}\}$ are projectors acting on JR, for $z \in \{0,1\}$. The optimisation problem is

$$\max \quad \mathrm{tr}(\Pi_{\mathrm{obj}}\rho_2\Pi_{\mathrm{obj}})$$

subject to

$$\left\|\rho_0 - |\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|_{HIJ} \otimes \rho_{\mathrm{junk}}\right\| \leq \epsilon_0 \tag{12}$$

$$\left\|M_{c|z}\rho_0 M_{c'|z'} - \Pi^J_{c|z}|\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|\Pi^J_{c'|z'} \otimes \rho_{\mathrm{junk}}\right\| \leq \epsilon_1 \quad \forall \quad c,c',z,z' \in \{0,1\}. \tag{13}$$

$$\begin{aligned} \mathrm{tr}_Z(\rho_1) &= \mathrm{tr}_{HIPQ}\rho_0 \\ \rho_2 &= \sum_{\substack{z,z' \\ c,c'}} |c\rangle\langle c'|_C \quad \otimes \quad M_{c|z} \otimes \Pi^Z_z \quad \rho_1 \quad M_{c'|z'} \otimes \Pi^Z_{z'} \end{aligned}$$

where ϵ_0 and ϵ_1 are functions of ϵ which vanish as ϵ vanishes.

We briefly justify why this optimization problem correctly captures the physical situation, modulo the simplifications listed above (which again, don't change the argument here). Let $|\psi\rangle \in HIJPQR$ be the state in the box and $M_{c|z}$ the measurement operators for the last box. Since we're allowing Bob to optimise over $|\psi\rangle$ and $M_{c|z}$ we don't quite need to worry about the isometry in the self-testing step. We suppress the c's and z's for the moment. The

self-testing statement says that $\||\psi\rangle - |\text{GHZ}\rangle \otimes |\text{junk}\rangle\| \leq \epsilon$ which entails Equation (12). The self-testing statement also says that $\|M|\psi\rangle - \Pi|\text{GHZ}\rangle \otimes |\text{junk}\rangle\| \leq \epsilon$ which implies Equation (13).

It is straightforward to see that for $\epsilon = 0$, this optimization problem reduces to the first one. The ρ_0 part is trivial and replacement of $M_{c|z}$ with $\Pi_{c|z}$ in ρ_2 can be made as in illustrated Example 25 below.

Third: Denote the value of the second program by $v(\epsilon, d)$. As argued, $v(0, d)$ is the value of the first program for all d (finite d). [EDIT: I realised even an NPA relaxation is not simple/obvious here] Let $w(\epsilon, d, k)$ denote the value of the NPA relaxation of the second program, to level k. The NPA hierarchy is well known and for our purposes here, it suffices to note two facts. First, the NPA relaxation is always an SDP and second, the NPA relaxation converges to, in this case, the second program as k tends to infinity. Since $v(\epsilon, d) \leq w(\epsilon, d, k)$ for all k and d, the continuity result follows [JAMIE: complete the argument?].

Example 25. Let ρ_{AB} be a density matrix, Π^B, Π'^B be projectors on B and M^B, M'^B be measurement (Kraus) operators on B. Suppose $M^B \rho_{AB} M'^B = \Pi^B \rho_{AB} \Pi'^B$. Suppose

$$M^B \rho_{AB} M'^B = \Pi^B \rho_{Ab} \Pi'^B. \tag{14}$$

If σ_{AB} is another density matrix such that $\text{tr}_A(\sigma_{AB}) = \text{tr}_B(\rho_{AB})$, then

$$M^B \sigma_{AB} M'^B = \Pi^B \sigma_{AB} \Pi'^B. \tag{15}$$

This follows from Uhlman's theorem which guarantees that there exists a U acting on system A such that $(U \otimes \mathbb{I}_B)\, \sigma_{AB} \left(U^\dagger \otimes \mathbb{I}_B\right) = \rho_{AB}$. Thus, conjugating Equation (14) with $U \otimes \mathbb{I}_B$, we obtain .

6.3 Bob Self-tests | simplified

Consider the following scenario:

- Bob has one box from a triple of GHZ boxes (assume he knows this box is ϵ close to a GHZ box, up to some local isomorphism Φ)
- He receives a bit z from Alice
- He inputs the bit z into the box, obtains an outcome c.
- Alice wants some function of z, c maximized. Call this *value* $\eta_\epsilon^{\text{lab}}$

Our objective is to show that as $\epsilon \to 0$, $\eta_\epsilon^{\text{lab}} = \eta$ where η is the value of an SDP where Bob's box is replaced with the GHZ state and measurement. We do this in three steps.

- We show $\eta_\epsilon^{\text{lab}} \leq \eta_\epsilon$ where η_ϵ does not depend explicitly depend on the isomorphism Φ.
- Then, we show that $\eta_\epsilon \leq \eta_\epsilon^{\text{Tr}}$ where $\eta_\epsilon^{\text{Tr}}$ does not involve the "junk" space.
  - This also means $\eta_\epsilon^{\text{Tr}}$ is continuous as a function of ϵ.
- And finally, we show $\eta_0^{\text{Tr}} = \eta$ (where η is as defined above).

In what follows, PQR are arbitrary fixed dimensional spaces while the remaining spaces denote qubits.

Definition 26 ($\eta_\epsilon^{\text{lab}}$)**.** Let $\rho_0 \in D(PQR), \rho_1 \in D(ZR), \rho_{\text{junk}} \in D(PQR)$ be density matrices, $M_{c|z} \in \text{Proj}(R)$ for $c, z \in \{0, 1\}$ and $\Pi_{\text{obj}} \in \text{Proj}(CZ)$ be projectors and finally, let $\Phi : PQR \to HIJPQR$ be a local isometry and let $U \otimes V \otimes W$ denote its action, where $U^\dagger U = \mathbb{I}_P$ and so on. Then, define

$$\eta_\epsilon^{\text{lab}} := \max \text{tr}(\Pi_{\text{obj}} \otimes \mathbb{I}_R \cdot \rho_2)$$

s.t.

$$\begin{aligned}
\left\|\Phi(\rho_0) - |\text{GHZ}\rangle\langle\text{GHZ}|_{HIJ} \otimes \rho_{\text{junk}}\right\| &\leq \epsilon \\
\left\|\Phi(M_{cz}\rho_0 M_{c'z'}) - \Pi^J_{cz}|\text{GHZ}\rangle\langle\text{GHZ}|_{HIJ}\Pi^J_{c'z'} \otimes \rho_{\text{junk}}\right\| &\leq \epsilon \qquad \forall c, c', z, z' \in \{0, 1\} \\
\text{tr}_Z(\rho_1) &= \text{tr}_{PQ}(\rho_0) \\
\rho_2 &= \sum_{z,z',c,c'} |c\rangle\langle c'|_C \otimes \left(M_{cz} \otimes \Pi^Z_z \cdot \rho_1 \cdot M_{c'z'} \otimes \Pi^Z_{z'}\right)
\end{aligned}$$

Definition 27 (η_ϵ)**.** Let $\sigma_0 \in D(HIJPQR), \sigma_1 \in D(ZJR), \sigma_{\text{junk}} \in D(PQR)$ be density matrices and $N_{c|z} \in \text{Proj}(JR)$. Let Π_{obj}, Φ, U, V, W be as in Definition 26. Then, define

$$\eta_\epsilon := \max \text{tr}(\Pi_{\text{obj}} \otimes \mathbb{I}_{JR} \cdot \sigma_2)$$

s.t.

$$\begin{aligned} \left\|\sigma_0 - |\text{GHZ}\rangle\langle\text{GHZ}|_{HIJ} \otimes \sigma_{\text{junk}}\right\| &\leq \epsilon \\ \left\|\sigma_0 - \Pi^J_{cz} |\text{GHZ}\rangle\langle\text{GHZ}|_{HIJ} \Pi^J_{c'z'} \otimes \sigma_{\text{junk}}\right\| &\leq \epsilon \qquad \forall c, c', z, z' \in \{0,1\} \\ \text{tr}_Z(\sigma_1) &= \text{tr}_{PQIJ}(\sigma_0) \\ \sigma_2 &= \sum_{z,z',c,c'} |c\rangle\langle c'|_C \otimes \left(N_{cz} \otimes \Pi^Z_z \cdot \sigma_1 \cdot N_{c'z'} \otimes \Pi^Z_{z'}\right). \end{aligned}$$

Lemma 28. *One has* $\eta^{\text{lab}}_\epsilon \leq \eta_\epsilon$.

Proof. [TODO: add details] Suppose ρ_0, ρ_1, ρ_2 and M_{cz} achive the maximum for $\eta^{\text{lab}}_\epsilon$. Then $\sigma_0 = V\rho_0 V^\dagger$, $\sigma_1 = V\rho_1 V^\dagger$, $\sigma_2 = V\rho_2 V^\dagger$ and $N_{cz} = VM_{cz}V^\dagger$, observe that the constraints of η_ϵ are satisfied. Further, observe that $\text{tr}(\Pi_{\text{obj}} \otimes \mathbb{I}_{IR}\sigma_2) = \text{tr}(\Pi_{\text{obj}} \otimes \mathbb{I}_R \rho_2)$. Hence, η_ϵ is at least as large as $\eta^{\text{lab}}_\epsilon$. □

Lemma 29. *JAMIE:* η_ϵ *is continuous as a function of* ϵ *as* $\epsilon \to 0$.

Definition 30 ($\eta^{\text{Tr}}_\epsilon$)**.** Let $\tau_0 \in D(HIJ), \tau_1 \in D(ZJ)$ be density matrices and $P_{c|z} \in \text{Proj}(J)$ [EDIT: or perhaps POVMs]. Let Π_{obj} be as in Definition 26. Then, define

$$\eta^{\text{Tr}}_\epsilon := \max \text{tr}(\Pi_{\text{obj}} \otimes \mathbb{I}_J \cdot \tau_2)$$

s.t.

$$\begin{aligned} \left\|\tau_0 - |\text{GHZ}\rangle\langle\text{GHZ}|_{HIJ}\right\| &\leq \epsilon \\ \left\|P_{cz}\tau_0 P_{c'z'} - \Pi^J_{cz} |\text{GHZ}\rangle\langle\text{GHZ}|_{HIJ} \Pi^J_{c'z'}\right\| &\leq \epsilon \qquad \forall c, c', z, z' \in \{0,1\} \\ \text{tr}_Z(\tau_1) &= \text{tr}_{IJ}(\tau_0) \\ \tau_2 &= \sum_{z,z',c,c'} |c\rangle\langle c'|_C \otimes \left(P_{cz} \otimes \Pi^Z_z \cdot \tau_1 \cdot P_{c'z'} \otimes \Pi^Z_{z'}\right). \end{aligned}$$

Lemma 31. *[may break] One has* $\eta_\epsilon \leq \eta^{\text{Tr}}_\epsilon$.

Proof. Suppose $\sigma_0, \sigma_1, \sigma_2$ and N_{cz} achive the maximum for $\eta^{\text{lab}}_\epsilon$. Let, $\tau_0 = \text{tr}_{PQR}(\sigma_0)$, $\tau_1 = \text{tr}_R(\sigma_1)$, $\tau_2 = \text{tr}_R(\sigma_2)$ and P_{cz} =? [TODO: figure out what happens to the measurements; perhaps I must relax them to POVM!]. Then, the inequality constraints of η_ϵ imply the inequality constraints of $\eta^{\text{Tr}}_\epsilon$ by the monotonicity of trace distance. The equality constraints of η_ϵ directly imply the equality constraints of $\eta^{\text{Tr}}_\epsilon$ as partial traces are a special case. Finally, ...[TODO: the measurement part] □

Definition 32 (η)**.** Let $\rho_0 = |\text{GHZ}\rangle\langle\text{GHZ}|_{HIJ}$, $\rho_1 \in D(ZJ)$ be density matrices and $\Pi^J_{cz} \in \text{Proj}(J)$ be GHZ scenario projectors (Pauli x and Pauli y projectors). Let Π_{obj} be as in Definition 26. Define

$$\eta := \max \text{tr}(\Pi_{\text{obj}} \otimes \mathbb{I}_J \cdot \rho_2)$$

s.t.

$$\begin{aligned} \text{tr}_Z(\rho_1) &= \text{tr}_{IJ}(\rho_0) \\ \rho_2 &= \sum_{z,z',c,c'} |c\rangle\langle c'|_C \otimes (\Pi^J_{cz} \otimes \Pi^Z_z \cdot \rho_1 \cdot \Pi^J_{c'z'} \otimes \Pi^Z_{z'}). \end{aligned}$$

Lemma 33. *One has for* $\epsilon = 0$, $\eta_\epsilon = \eta$.

Proof. Follows directly from setting $\epsilon = 0$ and using Example 34 below to argue (in η_ϵ) that P_{cz} can be replaced with Π_{cz} when they act on τ_1 even though $\epsilon = 0$ only yields this result for τ_0. □

Example 34. Let ρ_{AB} be a density matrix, Π^B, Π'^B be projectors on B and M^B, M'^B be measurement (Kraus) operators on B. Suppose $M^B \rho_{AB} M'^B = \Pi^B \rho_{AB} \Pi'^B$. Suppose

$$M^B \rho_{AB} M'^B = \Pi^B \rho_{Ab} \Pi'^B. \tag{16}$$

If σ_{AB} is another density matrix such that $\mathrm{tr}_A(\sigma_{AB}) = \mathrm{tr}_B(\rho_{AB})$, then

$$M^B \sigma_{AB} M'^B = \Pi^B \sigma_{AB} \Pi'^B. \tag{17}$$

This follows from Uhlman's theorem which guarantees that there exists a U acting on system A such that $(U \otimes \mathbb{I}_B)\, \sigma_{AB} \left(U^\dagger \otimes \mathbb{I}_B\right) = \rho_{AB}$. Thus, conjugating Equation (16) with $U \otimes \mathbb{I}_B$, we obtain .

6.4 The self-testing step [Discuss with Tom before writing]

Proposition 35. *For any implementation of the boxes and choice of $\delta > 0$, the joint probability that the test Ω passes and that the conclusion T is false is small, i.e. $\Pr[\Omega \cap \bar{T}] \leq \frac{1}{1-\delta+n\delta} \leq \frac{1}{n\delta}$ where the first upper-bound is tight.*

6.5 Robust Self Testing

Lemma 36. *Let $a, b, c, x, y, z \in \{0, 1\}$. Consider a trio of quantum boxes, specified by projectors $\{M^A_{a|x}, M^B_{b|y}, M^C_{c|z}\}$ acting on finite dimensional Hilbert spaces $\mathcal{H}^A, \mathcal{H}^B$ and $\mathcal{H}^C$, and $|\psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B \otimes \mathcal{H}^C =: \mathcal{H}^{ABC}$. If the trio pass the GHZ test with probability $1 - \epsilon$ (for $1 > \epsilon > 0$), then there exists a local isometry,*

$$\Phi = \Phi^A \otimes \Phi^B \otimes \Phi^C : \mathcal{H}^{ABC} \to \mathcal{H}^{ABC} \otimes \mathbb{C}^{2\times 3}$$

and a decreasing function of ϵ, $f(\epsilon)$ such that

$$\begin{aligned} \|\Phi(|\psi\rangle) - |\chi\rangle \otimes |\mathrm{junk}\rangle\| &\leq f(\epsilon), \\ \left\| \Phi\left(M^D_{d|t} |\psi\rangle\right) - \Pi^D_{d|t} |\mathrm{GHZ}\rangle \otimes |\mathrm{junk}\rangle \right\| &\leq f(\epsilon) \quad \forall D \in \{A, B, C\},\ \text{and } d, t \in \{0, 1\} \end{aligned}$$

where $|\mathrm{GHZ}\rangle = \frac{|000\rangle + |111\rangle}{\sqrt{2}} \in \mathbb{C}^{2\times 3}$, $|\mathrm{junk}\rangle \in \mathcal{H}^{ABC}$ is some arbitrary state and $\{\Pi^A_{a|x}, \Pi^B_{b|y}, \Pi^C_{c|z}\}$ are projectors corresponding to σ_x on the first, second and third qubit of $|\mathrm{GHZ}\rangle$ respectively, for $x = 0$ and corresponding to σ_y for $x = 1$, as in Claim 6.

6.6 The continuity argument [Enter Jamie]

References

[SCA+11] J. Silman, A. Chailloux, N. Aharon, I. Kerenidis, S. Pironio, and S. Massar, *Fully distrustful quantum bit commitment and coin flipping*, Physical Review Letters **106** (2011), no. 22.