

# Verifier-on-a-Leash: new schemes for verifiable delegated quantum computation, with quasilinear resources

Andrea Coladangelo<sup>\*</sup>   Alex Grilo<sup>†</sup>   Stacey Jeffery<sup>‡</sup>   Thomas Vidick<sup>§</sup>

## Abstract

The problem of reliably certifying the outcome of a computation performed by a quantum device is rapidly gaining relevance. We present two protocols for a classical verifier to verifiably delegate a quantum computation to two non-communicating but entangled quantum provers. Our protocols have near-optimal complexity in terms of the total resources employed by the verifier and the honest provers, with the total number of operations of each party, including the number of entangled pairs of qubits required of the honest provers, scaling as  $O(g \log g)$  for delegating a circuit of size  $g$ . This is in contrast to previous protocols, which all require a prohibitively large polynomial overhead. Our first protocol requires a number of rounds that is linear in the depth of the circuit being delegated, and is blind, meaning neither prover can learn the circuit being delegated. The second protocol is not blind, but requires only a constant number of rounds of interaction.

Our main technical innovation is an efficient rigidity theorem which allows a verifier to test that two entangled provers perform measurements specified by an arbitrary  $m$ -qubit tensor product of single-qubit Clifford observables on their respective halves of  $m$  shared EPR pairs, with a robustness that is independent of  $m$ . Our two-prover classical-verifier delegation protocols are obtained by combining this rigidity theorem with a single-prover quantum-verifier protocol for the verifiable delegation of a quantum computation, introduced by Broadbent.

---

<sup>\*</sup>Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, USA. email: acoladan@caltech.edu. Supported by AFOSR YIP award number FA9550-16-1-0495.

<sup>†</sup>IRIF, CNRS/Université Paris Diderot, Paris, France. email: abgrilo@irif.fr. Supported by ERC QCC.

<sup>‡</sup>QuSoft and CWI, Amsterdam, the Netherlands. email: jeffery@cw.nl. Supported by an NWO WISE Grant.

<sup>§</sup>Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, USA. email: vidick@cms.caltech.edu. Supported by NSF CAREER Grant CCF-1553477, AFOSR YIP award number FA9550-16-1-0495, and the IQIM, an NSF Physics Frontiers Center (NSF Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028).

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Preliminaries</b>	<b>7</b>
2.1	Notation . . . . .	7
2.2	Quantum circuits . . . . .	8
2.3	Broadbent's EPR Protocol . . . . .	8
<b>3</b>	<b>Rigidity</b>	<b>13</b>
3.1	Testing . . . . .	14
3.2	The conjugation test . . . . .	16
3.3	Testing Clifford unitaries . . . . .	18
3.4	Tensor products of single-qubit Clifford observables . . . . .	21
3.5	Post-measurement states . . . . .	24
3.6	Tomography . . . . .	26
<b>4</b>	<b>The Verifier-on-a-Leash Protocol</b>	<b>27</b>
4.1	Protocol and statement of results . . . . .	27
4.2	Completeness . . . . .	33
4.3	Soundness . . . . .	33
4.4	Blindness . . . . .	35
<b>5</b>	<b>The Dog-Walker Protocol</b>	<b>37</b>
5.1	Protocol and statement of results . . . . .	37
5.2	Completeness . . . . .	40
5.3	Soundness . . . . .	40
<b>6</b>	<b>Running our protocols in sequence</b>	<b>45</b>
6.1	Sequential version of our protocols . . . . .	46
<b>A</b>	<b>Some simple tests</b>	<b>47</b>
A.1	The Magic Square game . . . . .	47
A.2	Elementary tests . . . . .	48
A.3	The Bell basis . . . . .	50
<b>B</b>	<b>The <math>m</math>-qubit Pauli group</b>	<b>52</b>
B.1	The $m$ -qubit Weyl-Heisenberg group . . . . .	52
B.2	The $m$ -qubit Pauli group . . . . .	54

# 1 Introduction

Quantum computers hold the potential to speed up a wide range of computational tasks (see, for example, [Mon16]). Recent progress towards implementing limited quantum devices has added urgency to the already important question of how a classical verifier can test a quantum device. This verifier could be an experimentalist running a new experimental setup; a consumer who has purchased a purported quantum device; a client who wishes to delegate some task to a quantum server. In all cases, the user would like to exert some form of control over the quantum device. For example, the experimentalist may think that she is testing that a particular experiment prepares a certain quantum state by performing a series of measurements, i.e. by state tomography, but this assumes some level of trust in the measurement apparatus being used. For a classical party to truly test a quantum system, that system should be modeled in a device-independent way, having classical inputs (e.g. measurement settings) and classical outputs (e.g. measurement results).

Tests of quantum mechanical properties of a system first appeared in the form of Bell tests [Bel64, CHSH69]. These tests make one crucial assumption on the system to be tested: that it consists of two spatially isolated components which are unable to communicate throughout the experiment. While the violation of a Bell inequality can be seen as a certificate of entanglement, the area of self-testing, first introduced in [MY04], allows for the certification of much stronger statements, including about which measurements are being performed, and on which state. Informally, a *robust rigidity theorem* is a statement about which kind of apparatus, quantum state and measurements, must be used by a pair of isolated devices in order to succeed in a given statistical test. Following a well-established tradition, we will refer to such tests as *games*, call the devices *players* (or *provers*), and the quantum state and measurements that they implement the *strategy* of the players. A rigidity theorem is a statement about the necessary structure of near-optimal strategies for a game.

In 2012, Reichardt, Unger and Vazirani proved a robust rigidity theorem for playing a sequence of  $n$  CHSH games [RUV13]. Aside from its intrinsic interest, this rigidity theorem had two important consequences. One was the first device-independent protocol for quantum key distribution. The second was a protocol whereby a completely classical verifier can test a universal quantum computer consisting of two space-like separated devices. This latter consequence of the rigidity theorem is made possible by the model of *quantum computation by teleportation*, whereby a quantum computation can be carried out through single-qubit gates applied to EPR halves and Bell basis measurements [GC99]. The resulting *two-prover classical-verifier* protocol for delegating quantum computations could have important practical consequences: for the foreseeable future, making use of a quantum computer will likely require delegating the computation to a potentially untrusted cloud service, such as that already announced by IBM [Cas17].

Unfortunately, the complexity overhead of the delegation protocol from [RUV13], in terms of both the number of EPR pairs needed for the provers and the overall time complexity of the provers as well as the (classical) verifier, while polynomial, is prohibitively large. Although the authors of [RUV13] do not provide an explicit value for the exponent, in [HPDF15] it is estimated that their protocol requires resources that scale like  $\Omega(g^{8192})$ , where  $g$  is the number of gates in the delegated circuit (notwithstanding the implicit constant, this already makes the approach thoroughly impractical for even a 2-gate circuit!). The large overhead is in part due to a very small (although still inverse polynomial) gap between the completeness and soundness parameters of the rigidity theorem; this requires the verifier to perform many more Bell tests than the actual number of EPR pairs needed to implement the computation, which would scale linearly with the circuit size. While the physical implementation of the protocol from [RUV13] for any non-trivial circuit remains out of reach, recently [HZM<sup>+</sup>17] provided an experimental demonstration of a two-prover delegation protocol based on [RUV13] for a 3-qubit quantum circuit based on Shor’s algorithm to factor the number 15. They are able to certify that the three pairs of devices shared by the provers share a state close to three required EPR pairs while making a reasonable number of Bell tests, around 6000, by making the strong assumption that the devices behave in an i.i.d. manner at each use; as argued above, without this assumption the protocol

	Provers	Rounds	Total Resources	Blind
RUV 2012 [RUV13]	2	$\text{poly}(n)$	$\geq g^{8192}$	yes
McKague 2013 [McK16]	$\text{poly}(n)$	$\text{poly}(n)$	$\geq 2^{153} g^{22}$	yes
GKW 2015 [GKW15]	2	$\text{poly}(n)$	$\geq g^{2048}$	yes
HDF 2015 [HPDF15]	$\text{poly}(n)$	$\text{poly}(n)$	$\Theta(g^4 \log g)$	yes
FH 2015 [FH15]	5	$\text{poly}(n)$	$> gn^3$	no
NV 2016 [NV16]	7	2	$> gn^3$	no
Verifier-on-a-Leash Protocol (Section 4)	2	$O(\text{depth})$	$\Theta(g \log g)$	yes
Dog-Walker Protocol (Section 5)	2	$O(1)$	$\Theta(g \log g)$	no

Table 1: Resource requirements of various delegation protocols in the multi-prover model. We use  $n$  to denote the number of qubits and  $g$  the number of gates in the delegated circuit. “depth” refers to the depth of the delegated circuit. “Total Resources” refers to the gate complexity of the provers, the number of EPR pairs of entanglement needed, and the number of bits of communication in the protocol. To ensure fair comparison, we require of each protocol that it produces the correct answer with probability 99%. For all protocols except NV 2016, and our two new protocols, this requires a polynomial number of sequential repetitions, which is taken into account when computing the total resources. (In [NV16] the gap amplification is performed by taking tensor powers and executing the verification procedure in parallel, so that the protocol achieves a constant gap with a single round of interaction; an additional round is needed to distribute the history state between the provers.)

would remain thoroughly impractical even for such a small circuit.

Subsequent work has presented significantly more efficient protocols for achieving the same functionality [McK16, GKW15, HPDF15, FH15, NV16]. We refer to Table 1 for a summary of our estimated lower bounds on the complexity of each of these results (not all papers provide explicit bounds, in which case our estimates, although generally conservative, should be taken with caution). Prior to this work, the most efficient multi-prover delegation protocols, in any model of computation, required resources that scale as at least  $\Omega(gn^3)$  for delegating a  $g$ -gate circuit on  $n$  qubits. Thus, while there exist a number of multi-prover classical-verifier delegation protocols, based on a wide variety of techniques, all require an amount of resources, both in terms of EPR pairs and time, that make them all but infeasible to implement in practice.

In contrast, more efficient schemes for verifiable delegation exist in the setting where the verifier has some limited quantum power, such as the ability to generate or measure single-qubit states. In that case, only a single prover is needed [ABE10, FK17, Mor14, Bro15, MF16] (see also [Fit16] for a recent survey), and the most efficient *single-prover quantum-verifier* protocols can evaluate a quantum circuit with  $g$  gates in time  $O(g)$ . The main reason these protocols are much more efficient than the classical-verifier multi-prover protocols is that they avoid the need for directly testing any of the qubits used by the prover, instead requiring the trusted verifier to directly either prepare or measure the qubits used for the computation. In addition, several of these protocols have an additional desirable property, called *blindness*, which states that the prover learns nothing about the circuit being implemented (some multi-prover protocols, such as [RUV13], are blind, but others, such as [FH15], are not).

**Our contributions** We overcome the efficiency limitations of multi-prover delegation protocols by introducing a new robust rigidity theorem. We use the theorem to adapt the efficient single-prover quantum-verifier delegation protocol introduced by Broadbent [Bro15] to obtain two new two-prover classical-verifier

protocols in which the complexity of verifiably delegating a  $g$ -gate quantum circuit scales as  $O(g \log g)$ <sup>1</sup>. Our protocols are the first protocols with a classical verifier to achieve this near-optimal complexity. We note that both our protocols require the verifier to communicate with one prover after at least one round of communication with the other has been completed. This means that the requirement that the two provers do not communicate throughout the protocol cannot be enforced through space-like separation, and should rather be seen as an additional assumption. This is also a drawback of previous protocols.

Before giving an overview of the delegation protocols, we introduce our robust rigidity theorem. The goal of the theorem is to provide the means to certify that a prover performs a desired measurement on  $m$  EPR pairs, where the measurement can be any  $m$ -fold tensor product of (possibly distinct) single-qubit Clifford observables.

**Theorem 1 (Informal).** *Let  $m \in \mathbb{Z}_{>0}$ . Let  $\mathcal{G}$  be a fixed, finite set of single-qubit Clifford observables. Then there exists an efficient two-prover test  $\text{RIGID}(\mathcal{G}, m)$  with  $O(m)$ -bit questions (a constant fraction of which are of the form  $W \in \mathcal{G}^m$ ) and answers such that the following properties hold:*

- (Completeness) *There is a strategy for the provers that uses  $m + 1$  EPR pairs and measurements of tensor products of single-qubit Clifford observables and succeeds with probability at least  $1 - e^{-\Omega(m)}$  in the test.*
- (Soundness) *For any  $\varepsilon > 0$ , any strategy for the provers that succeeds with probability  $1 - \varepsilon$  in the test must be  $\text{poly}(\varepsilon)$ -close, up to local isometries, to a strategy in which the provers begin with  $(m + 1)$  EPR pairs and is such that upon receipt of a question of the form  $W \in \mathcal{G}^m$  the prover measures the “correct” observable  $W$ .*

This theorem is proven in Section 3. The key point in the theorem is the “robustness”, which scales polynomially in  $\varepsilon$  but independently of  $m$ . (Although we do not strive to obtain the best dependence on  $\varepsilon$ , we believe it should be possible to obtain a scaling of the form  $C\sqrt{\varepsilon}$  for a reasonably controlled constant  $C$ .) The test is also very efficient. The analysis builds upon the “Pauli braiding test” of [NV16], which establishes a similar result for the more restricted case of Pauli  $\sigma_X$  and  $\sigma_Z$  observables. A first contribution of our result is to show how to also test for  $\sigma_Y$  observables. This is somewhat subtle due to an ambiguity in the complex phase that cannot be detected by a classical two-player test. In addition, we show how to test for more general Clifford observables by introducing a new “conjugation test” which tests how an observable applied by the provers acts on the Pauli group. We give a more detailed overview of the test in Section 3.

We now introduce our two-prover classical-verifier protocols. Both protocols use the single-prover quantum-verifier protocol introduced by Broadbent [Bro15] as a starting point and provide different methods to delegate the quantum computation performed by the quantum verifier to a second prover (call him PV for Prover  $V$ ). The rigidity test is used to verify that the second prover indeed performs the same actions as the honest verifier of [Bro15], which are sequences of single-qubit measurements of Clifford observables.

In the first protocol, one of the provers plays the role of Broadbent’s prover (call him PP for Prover  $P$ ), and the other plays the role of Broadbent’s verifier (PV). The protocol enforces correct behavior of PV using Theorem 1. We call it the *Verifier-on-a-Leash Protocol*, or “leash protocol” for short. The protocol has a constant soundness-completeness gap, in the sense that there is a constant gap between the acceptance probability of honest provers on a *yes* instance and the acceptance probability of any cheating provers on a *no* instance. It requires  $(2d + 1)$  rounds of interaction, where  $d$  is the depth of the circuit being delegated (see Section 2.3 for a precise definition of how this is computed). The first round of interaction requires communication linear in the size of the circuit; the subsequent  $2d$  rounds have communication that totals

---

<sup>1</sup>The  $\log g$  overhead is due to the complexity of sampling from the right distribution in rigidity tests. We leave the possibility of removing this by derandomization for future work. Another source of overhead is in achieving blindness, since in order to hide the circuit, we encode it as part of the input to a universal circuit, which introduces  $\log g$  overhead.

to an amount linear in the circuit size — in particular, each round corresponds to a *layer* of the circuit, and the communication in that round scales as the number of gates in that layer. The protocol requires  $O(n + g)$  EPR pairs to delegate a  $g$ -gate circuit on  $n$  qubits, and the overall time complexity of the protocol is  $O(g \log g)$ . The input to the circuit is hidden from the provers, meaning that the protocol can be made blind by encoding the circuit in the input, and delegating a universal circuit.

The completeness of the protocol follows directly from the completeness of [Bro15]. Once we ensure the correct behavior of PV using our rigidity test, soundness follows from [Bro15] as well, since the combined behavior of our verifier and an honest PV is nearly identical to that of Broadbent’s verifier.

The second protocol also starts from Broadbent’s protocol, but modifies it in a different way to achieve a protocol with a constant number of rounds of interaction: it is only required that the verifier exchange one round of classical messages with the first prover, followed by one round with the second prover. The proof of security is slightly more involved, but the key ideas are the same: we use a combination of our new self-testing results and the techniques of Broadbent’s protocol to control the two provers, one of which plays the role of Broadbent’s verifier, and the other the role of the prover. Because of the more complicated “leash” structure in this protocol, we call it the *Dog-Walker Protocol*. Like the leash protocol, the Dog-Walker Protocol has constant completeness-soundness gap, and overall time complexity  $O(g \log g)$ . One drawback of this protocol is that we do not currently know how to make it blind. In particular, while PV and PP would have to collude after the protocol is terminated to learn the input in the leash protocol, in the Dog-Walker protocol, PV simply receives the input in clear.

A constant-round protocol for verifiable, but not blind, delegated computation was previously given in [FH15]. There the idea is to ask the provers to share an encoding of the history state of the computation. The encoding can be verified with a single round of interaction by using a technique from [Ji16]. This “post-hoc” verification requires the provers to learn the verifier’s input even before they are separated, so that they can exchange the required history state. Alternatively, the state can be created by a single prover and teleported to the others with the help of the verifier; in this case the protocol requires two rounds of communication. In comparison with [FH15], our protocol has a lower overhead in terms of the computation required of the provers, and it only requires two provers, instead of five for [FH15].

Based on the Dog-Walker Protocol, it is possible to design a classical-verifier two-prover protocol for all languages in QMA. This is achieved along the same lines as the proof that  $\text{QMIP} = \text{MIP}^*$  from [RUV13]. The first prover, given the input, creates the QMA witness and teleports it to the second prover with the help of the verifier. The verifier then delegates the verification circuit to the second prover, as in the Dog-Walker Protocol; the first prover can be re-used to verify the operations of the second one.

**Open Questions** We have introduced a new rigidity theorem and shown how the theorem could be used to transform a specific quantum-verifier delegation protocol, due to Broadbent, into a classical-verifier protocol with an additional prover, while suffering very little overhead in terms of the efficiency of the protocol. We believe that a similar transformation could be performed starting from delegation protocols based on other models of computation, such as the protocol in the measurement-based model of [FK17] or the protocol based on computation by teleportation considered in [RUV13], and would lead to similar efficiency improvements.

A drawback of all these protocols, however, is that they require multiple rounds of adaptive communication. As a result, the no-communication requirement on the provers cannot be enforced by space-like separation, and needs to be taken as an a priori assumption. An important question left open by our work is whether there exists a multi-prover delegation protocol that consists of a single round of simultaneous communication with each prover, and is both blind and verifiable.

**Organization.** The remainder of this paper is organized as follows. In Section 2, we give the necessary preliminaries including outlining Broadbent’s EPR Protocol (Section 2.3). In Section 3, we present our new rigidity theorems. In Section 4, we present our first protocol, the leash protocol, and in Section 5, we present our second protocol, the Dog-Walker Protocol. In Section 6, we show that the constant soundness-completeness gap of our protocols can be amplified by sequential repetition.

**Acknowledgments.** We thank Anne Broadbent for useful discussions in the early stages of this work. All authors acknowledge the IQIM, an NSF Physics Frontiers Center at the California Institute of Technology, where this research was initiated.

## 2 Preliminaries

### 2.1 Notation

We often write  $\vec{x} = (x_1, \dots, x_n) \in \{0, 1\}^n$  for a string of bits, and  $W = W_1 \cdots W_m \in \Sigma^m$  for a string, where  $\Sigma$  is a finite alphabet. If  $S \subseteq \{1, \dots, m\}$  we write  $W_S$  for the sub-string of  $W$  indexed by  $S$ . For an event  $E$ , we use  $1_E$  to denote the indicator variable for that event, so  $1_E = 1$  if  $E$  is true, and otherwise  $1_E = 0$ . We write  $\text{poly}(\varepsilon)$  for  $O(\varepsilon^c)$ , where  $c$  is a universal constant that may change each time the notation is used.

$\mathcal{H}$  is a finite-dimensional Hilbert space. We denote by  $U(\mathcal{H})$  the set of unitary operators,  $\text{Obs}(\mathcal{H})$  the set of binary observables (we omit the term “binary” from here on; in this paper all observables are binary) and  $\text{Proj}(\mathcal{H})$  the set of projective measurements on  $\mathcal{H}$  respectively. We let  $|\text{EPR}\rangle$  denote an EPR pair:

$$|\text{EPR}\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle).$$

**Observables.** We use capital letters  $X, Z, W, \dots$  to denote observables. We use greek letters  $\sigma, \tau$  with a subscript  $\sigma_W, \tau_W$ , to emphasize that the observable  $W$  specified as subscript acts in a particular basis. For example,  $X$  is an arbitrary observable but  $\sigma_X$  is specifically the Pauli  $X$  matrix defined in (1).

For  $a \in \{0, 1\}^n$  and commuting observables  $\sigma_{W_1}, \dots, \sigma_{W_n}$ , we write  $\sigma_W(a) = \prod_{i=1}^n (\sigma_{W_i})^{a_i}$ . The associated projective measurements are  $\sigma_{W_i} = \sigma_{W_i}^0 - \sigma_{W_i}^1$  and  $\sigma_W^u = E_a(-1)^{u \cdot a} \sigma_W(a)$ . Often the  $\sigma_{W_i}$  will be single-qubit observables acting on distinct qubits, in which case each is implicitly tensored with identity outside of the qubit on which it acts.

**Pauli and Clifford groups.** Let

$$\sigma_I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{and} \quad \sigma_Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (1)$$

denote the standard Pauli matrices acting on a qubit. The single-qubit Weyl-Heisenberg group

$$\mathcal{H}^{(1)} = H(\mathbb{Z}_2) = \left\{ (-1)^c \sigma_X(a) \sigma_Z(b), a, b, c \in \{0, 1\} \right\}$$

is the matrix group generated by the Pauli  $\sigma_X$  and  $\sigma_Z$ . We let  $\mathcal{H}^{(n)} = H(\mathbb{Z}_2^n)$  be the direct product of  $n$  copies of  $\mathcal{H}^{(1)}$ . The  $n$ -qubit Clifford group is the normalizer of  $\mathcal{H}^{(n)}$  in the unitary group, up to phase:

$$G_C^{(n)} = \{ G \in U((\mathbb{C}^2)^{\otimes n}) : G \sigma G^\dagger \in \mathcal{H}^{(n)} \quad \forall \sigma \in \mathcal{H}^{(n)} \}.$$

Some Clifford observables we will use include

$$\sigma_H = \frac{\sigma_X + \sigma_Z}{\sqrt{2}}, \quad \sigma_{H'} = \frac{\sigma_X - \sigma_Z}{\sqrt{2}}, \quad \sigma_F = \frac{-\sigma_X + \sigma_Y}{\sqrt{2}}, \quad \sigma_G = \frac{\sigma_X + \sigma_Y}{\sqrt{2}}. \quad (2)$$

Note that  $\sigma_H$  and  $\sigma_{H'}$  are characterized by  $\sigma_X \sigma_H \sigma_X = \sigma_{H'}$  and  $\sigma_Z \sigma_H \sigma_Z = -\sigma_{H'}$ . Similarly,  $\sigma_F$  and  $\sigma_G$  are characterized by  $\sigma_X \sigma_F \sigma_X = -\sigma_G$  and  $\sigma_Y \sigma_F \sigma_Y = \sigma_G$ .

## 2.2 Quantum circuits

We use capital letters in sans-serif font to denote gates. We work with the universal quantum gate set  $\{\text{CNOT}, \text{H}, \text{T}\}$ , where the controlled-not gate is the two-qubit gate with the unitary action

$$\text{CNOT}|b_1, b_2\rangle = |b_1, b_1 \oplus b_2\rangle,$$

and the Hadamard and T gates are single-qubit gates with actions

$$\text{H}|b\rangle = \frac{1}{\sqrt{2}} (|0\rangle + (-1)^b |1\rangle) \quad \text{and} \quad \text{T}|b\rangle = e^{ib\pi/4} |b\rangle,$$

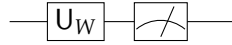
respectively. We will also use the following gates:

$$\text{X}|b\rangle = |b \oplus 1\rangle, \quad \text{Z}|b\rangle = (-1)^b |b\rangle, \quad \text{and} \quad \text{P}|b\rangle = i^b |b\rangle.$$

Measurements in the Z basis (or computational basis) will be denoted by the standard measurement symbol:



To measure another observable,  $W$ , we can perform a unitary change of basis  $U_W$ , so that the following circuit measures in the eigenbasis of  $W$ :



We assume that every circuit has a specified output wire, which is measured at the end of the computation to obtain the output bit. Without loss of generality, we can assume this is always the first wire. For an  $n$ -qubit system, we let  $\Pi_b$ , for  $b \in \{0, 1\}$ , denote the orthogonal projector onto states with  $|b\rangle$  in the output wire:  $|b\rangle\langle b| \otimes \text{Id}$ . For example, the probability that a circuit  $Q$  outputs 0 on input  $|\vec{x}\rangle$  is  $\|\Pi_0 Q|\vec{x}\rangle\|^2$ .

We can always decompose a quantum circuit into layers such that each layer contains at most one T gate applied to each wire. The minimum number of layers for which this is possible is called the *T depth* of the circuit. We note that throughout this work, we will assume circuits are compiled in a specific form that introduces extra T gates (see the paragraph on the H gadget in Section 2.3). The T depth of the resulting circuit is proportional to the depth of the original circuit.

## 2.3 Broadbent's EPR Protocol

In this section we summarize the main features of a delegation protocol introduced in [Bro15], highlighting the aspects that will be relevant to understanding our subsequent adaptation into two-prover protocols. The ‘‘EPR Protocol’’ from [Bro15] involves the interaction between a verifier  $V_{\text{EPR}}$  and a prover  $P$ . We write  $P_{\text{EPR}}$  for the ‘‘honest’’ behavior of the prover. The verifier  $V_{\text{EPR}}$  has limited quantum powers. Her goal is to delegate a BQP computation to the prover  $P$  in a verifiable way. Specifically, the verifier has as input a quantum circuit  $Q$  on  $n$  qubits and an input string  $\vec{x} \in \{0, 1\}^n$ , and the prover gets as input  $Q$ . The verifier and prover interact. At the end of the protocol, the verifier outputs either accept or reject. The protocol is such that there exist values  $p_{\text{sound}}$  and  $p_{\text{compl}}$  with  $p_{\text{sound}} < p_{\text{compl}}$  such that  $p_{\text{compl}} - p_{\text{sound}}$ , called the *soundness-completeness gap*, is a constant independent of input size, and moreover:



**Completeness:** If the prover is honest and  $\|\Pi_0 Q|\vec{x}\rangle\|^2 \geq 2/3$ , then the verifier outputs accept with probability at least  $p_{\text{compl}}$ ;

**Soundness:** If  $\|\Pi_0 Q|\vec{x}\rangle\|^2 \leq 1/3$ , then the probability the verifier outputs accept is at most  $p_{\text{sound}}$ .

In Section 6, we show that sequential repetition can be used to turn any such protocol, including our two-prover variants, into one in which the verifier outputs 0, 1, or abort, and we have the following: (1) If the prover(s) is/are honest, the verifier outputs abort with probability at most .01; and (2) If  $\|\Pi_b Q|\vec{x}\rangle\|^2 \geq 2/3$  for some  $b \in \{0, 1\}$  then the probability that the verifier outputs  $1 - b$  is at most .01.

In the EPR protocol,  $V_{\text{EPR}}$  and  $P_{\text{EPR}}$  are assumed to share  $(n + t)$  EPR pairs at the start of the protocol, where  $t$  is the number of T gates in  $Q$  and  $n$  the number of input bits. (In [Bro15] the EPR protocol is only considered in the analysis, and it is assumed that the EPR pairs are prepared by the verifier.) The first  $n$  EPR pairs correspond to the input to the computation; they are indexed by  $N = \{1, \dots, n\}$ . The remaining pairs are indexed by  $T = \{n + 1, \dots, n + t\}$ ; they will be used as ancilla qubits to implement each of the T gates in the delegated circuit.

The behavior of  $V_{\text{EPR}}$  depends on a *round type* randomly chosen by  $V_{\text{EPR}}$  after her interaction with  $P_{\text{EPR}}$ . There are three possible round types:

- Computation round ( $r = 0$ ): the verifier delegates the computation to  $P_{\text{EPR}}$ , and at the end of the round can recover its output if  $P_{\text{EPR}}$  behaves honestly;
- X-test round ( $r = 1$ ) and Z-test round ( $r = 2$ ): the verifier tests that  $P_{\text{EPR}}$  behaves honestly, and rejects if malicious behavior is detected.

For some constant  $p$ ,  $V$  chooses  $r = 0$  with probability  $p$ , and otherwise chooses  $r \in \{1, 2\}$  with equal probability. Since the choice of round type is made after interaction with  $P_{\text{EPR}}$ ,  $P_{\text{EPR}}$ 's behavior cannot depend on the round type. In particular, any deviating behavior in a computation round is reproduced in both types of test rounds. The analysis amounts to showing that any deviating behavior that affects the outcome of the computation will be detected in at least one of the test rounds.

In slightly more detail, the high-level structure of the protocol is the following.  $V_{\text{EPR}}$  measures her halves of the  $n$  qubits in  $N$  in order to prepare the input state on  $P_{\text{EPR}}$ 's system. As a result the input is quantum one-time padded with keys that depend on  $V_{\text{EPR}}$ 's measurement results. For example, in a computation round,  $V_{\text{EPR}}$  measures each input qubit in the Z basis, and gets some result  $\vec{d} \in \{0, 1\}^n$ , meaning the input on  $P_{\text{EPR}}$ 's side has been prepared as  $X^{\vec{d}}|0\rangle^{\otimes n}$ . In [Bro15], the input is always considered to be  $\vec{0}$ , but we can also prepare an arbitrary classical input  $\vec{x} \in \{0, 1\}^n$  by reinterpreting the one-time pad key as  $\vec{a} = \vec{d} \oplus \vec{x}$  so that the input state on  $P_{\text{EPR}}$ 's side is  $X^{\vec{a}}|\vec{x}\rangle$ . In a test round, on the other hand, the input is prepared as the one-time pad of either  $|0\rangle^{\otimes n}$  or  $|+\rangle^{\otimes n}$ . Note that as indicated in Figure 2 this choice of measurements will be made after the interaction with  $P_{\text{EPR}}$  has taken place.

The honest prover  $P_{\text{EPR}}$  applies the circuit  $Q$ , which we assume is compiled in the universal gate set  $\{H, T, \text{CNOT}\}$ , to his one-time padded input. We will shortly describe gadgets that  $P_{\text{EPR}}$  can apply in order to implement each of the three gate types. The gadgets are designed in a way that in a test round each gadget amounts to an application of an identity gate; this is what enables  $V_{\text{EPR}}$  to perform certain tests in those rounds that are meant to identify deviating behavior of a dishonest prover. After each gadget, the one-time padded keys can be updated by  $V_{\text{EPR}}$ , who is able to keep track of the keys at any point in the circuit using the *update rules* in Table 2. We now describe the three gadgets, before giving a complete description of the protocol.

**CNOT Gadget** To implement a CNOT gate on wires  $j$  and  $j'$ ,  $P_{\text{EPR}}$  simply performs the CNOT gate on those wires of his input qubits. The one-time pad keys are changed by the update rule in Table 2, because  $\text{CNOT} \cdot X^{a_j} Z^{b_j} \otimes X^{a_{j'}} Z^{b_{j'}} = X^{a_j} Z^{b_j + b_{j'}} \otimes X^{a_j + a_{j'}} Z^{b_{j'}} \cdot \text{CNOT}$ . Note that  $\text{CNOT}|0\rangle|0\rangle = |0\rangle|0\rangle$  and  $\text{CNOT}|+\rangle|+\rangle = |+\rangle|+\rangle$ , so in the test runs,  $P_{\text{EPR}}$  is applying the identity.

		Key Update Rule
T	Computation Round	$(a_j, b_j) \leftarrow (a_j + c_i, b_j + e_i + a_j + c_i + (a_j + c_i)z_i)$
	X-Test, even parity; or Z-test, odd parity	$(a_j, b_j) \leftarrow (e_i, 0)$
	Z-Test, even parity; or X-test, odd parity	$(a_j, b_j) \leftarrow (0, b_j + e_i + z_i)$
H		$(a_j, b_j) \leftarrow (b_j, a_j)$
CNOT		$(a_j, b_j, a_{j'}, b_{j'}) \leftarrow (a_j, b_j + b_{j'}, a_{j'} + b_{j'}, b_{j'})$

Table 2: Rules for updating the one-time-pad keys after applying each type of gate in the EPR Protocol, in particular: after applying the  $i$ -th T gate to the  $j$ -th wire; applying an H gate to the  $j$ -th wire; or applying a CNOT gate controlled on the  $j$ -th wire and targeting the  $j'$ -th wire.

**H Gadget** To implement an H gate on wire  $j$ ,  $P_{EPR}$  simply performs the H on wire  $j$ , and the one-time-pad keys are changed as in Table 2. Unlike CNOT, H does not act as the identity on  $|0\rangle$  and  $|+\rangle$ , so it is not the identity in a test round. To remedy this, assume that  $Q$  is compiled so that every H gate appears in a pattern  $H(TTH)^k$ , where the maximal such  $k$  is odd. This can be accomplished by replacing each H by HTTHTTHTTH, which implements the same unitary. In test rounds, the T gadget, described shortly, implements the identity, and since  $H(\text{Id } H)^k$  for odd  $k$  implements the identity,  $H(TTH)^k$  will also have no effect in test rounds.

**Parity of a T Gate** Within a pattern  $H(TTH)^k$ , the H has the effect of switching between an X-test round scenario (the state  $|0\rangle$ ) and a Z-test round scenario (the state  $|+\rangle$ ). In order to consistently talk about the type of a round while evaluating the circuit, we can associate a parity with each T gate in the circuit. The parity of the T gates that are not part of the pattern  $H(TTH)^k$  will be defined to be even. A H will always flip the parity, so that within such a pattern, the first two T gates will be odd, the next two will be even, etc., until the last two T gates will be odd again.

**T Gadget** The gadget for implementing the  $i$ -th T gate on the  $j$ -th wire is performed on  $P_{EPR}$ 's  $j$ -th input qubit, and his  $i$ -th auxiliary qubit (indexed by  $n + i$ ), which we can think of as being prepared in a particular auxiliary state by  $V_{EPR}$  measuring her half of the corresponding EPR pair, as shown in Figure 1. The gadget depends on a random bit  $z_i$  that is chosen by  $V_{EPR}$  and sent to the prover.

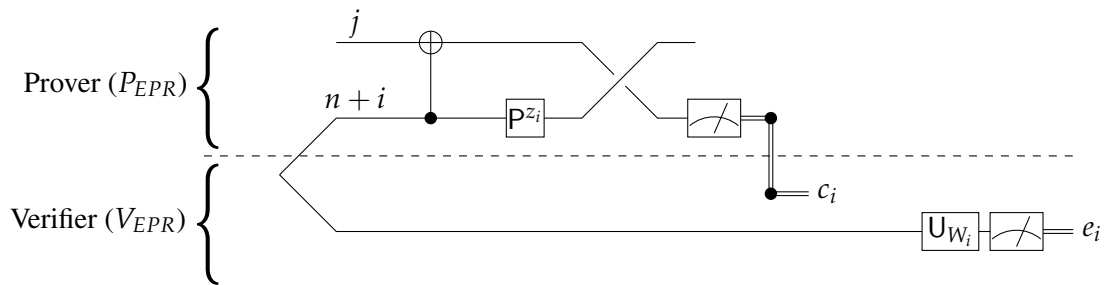


Figure 1: The gadget for implementing the  $i$ -th T gate on the  $j$ -th wire. The gate  $U_{W_i}$  implementing the change of basis associated with observable  $W_i$  is applied as part of the procedure  $V_{EPR}^r$  (see Figure 4) and is determined by the round type  $r$ , the parity of the  $i$ -th T gate,  $z_i$ ,  $c_i$ , and  $a'_i$  (the X-key going into the  $i$ -th T gate), as in Table 3.

		$U_{W_i}$ (observable $W_i$ )
Computation Round	$a'_i \oplus c_i \oplus z_i = 0$	HT (observable $G$ )
	$a'_i \oplus c_i \oplus z_i = 1$	HPT (observable $F$ )
X-Test Round	even T gate	Id (observable $Z$ )
	odd T gate	$z_i = 0$ H (observable $X$ )
		$z_i = 1$ HP (observable $Y$ )
Z-Test Round	odd T gate	Id (observable $Z$ )
	even T gate	$z_i = 0$ H (observable $X$ )
		$z_i = 1$ HP (observable $Y$ )

Table 3: The choice of  $U_{W_i}$  in the T gadget. We also indicate the observable  $W_i$  associated with the final measurement  $W_i = U_{W_i}^\dagger Z U_{W_i}$ .

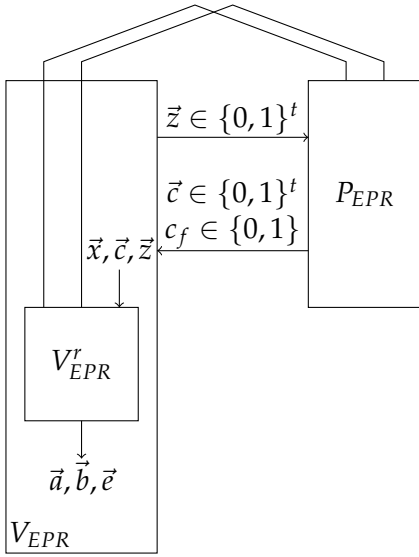


Figure 2: This figure describes how different pieces of the protocol fit together.  $V_{EPR}$  and  $P_{EPR}$  share  $n + t$  EPR pairs. The honest prover  $P_{EPR}$  can be seen as a procedure that acts on  $n + t$  qubits — the EPR pair halves — depending on a  $t$ -bit string  $\vec{z}$ . We have separated the quantum part of  $V_{EPR}$  into its own procedure, called  $V_{EPR}^r$ , where  $r \in \{0, 1, 2\}$  indicates the *round type*, which  $V_{EPR}$  runs on her  $n + t$  EPR halves, and the  $2t$  bits  $\vec{c}$  and  $\vec{z}$ . Aside from running  $V_{EPR}^r$ ,  $V_{EPR}$  is classical.

**The EPR Protocol.** We show how the gadgets just described are used in the complete protocol. We first describe the protocol for  $V_{EPR}$  below. For later convenience we have divided the action of  $V_{EPR}$  into classical actions and a single quantum subroutine  $V_{EPR}^r$  depending on the round type.

- 
1.  $V_{EPR}$  chooses  $\vec{z} \in_R \{0, 1\}^t$ , sends it to  $P_{EPR}$ ,<sup>a</sup> and receives back  $\vec{c} \in \{0, 1\}^t$  and  $c_f \in \{0, 1\}$ .
  2.  $V_{EPR}$  chooses a random round type  $r \in \{0, 1, 2\}$  and runs  $V_{EPR}^r$  (see Figure 4) on her EPR halves,  $\vec{x}$ ,  $\vec{c}$  and  $\vec{z}$ , to obtain bits  $\vec{a}, \vec{b} \in \{0, 1\}^n$  and  $\vec{e} \in \{0, 1\}^t$ .
  3.  $V_{EPR}$  applies the update rules from Table 2 on the initial keys  $(\vec{a}, \vec{b})$ , gate-by-gate, to obtain, for every  $i \in [t]$ , the X-key before the  $i$ -th T gate is applied,  $a'_i$ , and the final X key for the output wire,  $a'_f$ . If  $r = 1$  (X-test round) and there exists an  $i$  such that the  $i$ -th T gate is even and  $c_i \neq a'_i \oplus e_i$ , output reject. If  $r = 2$  (Z-test round) and there exists an  $i$  such that the  $i$ -th T gate is odd and  $c_i \neq a'_i \oplus e_i$ , output reject. If  $r \in \{0, 1\}$  (computation or X-test round) and  $c_f \oplus a'_f \neq 0$ , output reject. Otherwise, output accept.
- 

Figure 3: The EPR Protocol:  $V_{EPR}$ 's point of view.

<sup>a</sup>We note that in the original EPR Protocol of [Bro15], this is accomplished by  $P_{EPR}$  and  $V_{EPR}$  both measuring  $t$  EPR pairs in the Z basis to get a shared random string  $\vec{z}$ .

The procedure  $V_{EPR}^r$  measures each of the  $n + t$  EPR halves according to some observable that depends on  $r$ ,  $\vec{c}$ , and  $\vec{z}$ . In the case of a computation round,  $V_{EPR}^0$  measures the qubits in  $T$  adaptively.

---

Input: A circuit  $Q$  with  $t$  T gates,  $\vec{x} \in \{0, 1\}^n$ ,  $\vec{c}, \vec{z} \in \{0, 1\}^t$ , an  $n$ -qubit system indexed by  $N$ , and a  $t$ -qubit system indexed by  $T$ .

1. If  $r \in \{0, 1\}$ , measure each qubit in  $N$  in the Z basis, and otherwise measure in the X basis, to get results  $\vec{d} \in \{0, 1\}^n$ . If  $r = 0$ , set  $(\vec{a}, \vec{b}) = (\vec{d} \oplus \vec{x}, 0^n)$ ; if  $r = 1$ , set  $(\vec{a}, \vec{b}) = (\vec{d}, 0^n)$ ; and if  $r = 2$  set  $(\vec{a}, \vec{b}) = (0^n, \vec{d})$ .
  2. Going through  $Q$  gate-by-gate, use the update rules in Table 2 to update the one-time-pad keys. For every  $i \in [t]$ , when the  $i$ -th T gate is reached, let  $a'_i$  be the X key before the  $i$ -th T gate is applied. Choose an observable  $W_i$  according to Table 3 in which to measure the  $i$ -th qubit in  $T$ , corresponding to the  $i$ -th T gate, obtaining result  $e_i$ .
- 

Figure 4: The procedure  $V_{EPR}^r$ , employed by  $V_{EPR}$ .

Finally, we describe the behavior of an honest prover,  $P_{EPR}$ .

---

1. Receive  $\vec{z} \in \{0, 1\}^t$  from  $V_{EPR}$ .
  2. Evaluate  $Q$  gate-by-gate using the appropriate gadget for each gate. In particular, use  $z_i$  to implement the  $i$ -th T gadget, and obtain measurement result  $c_i$ .
  3. Measure the output qubit to obtain  $c_f$ , and return  $\vec{c}$  and  $c_f$  to  $V_{EPR}$ .
- 

Figure 5: The EPR Protocol: Honest prover strategy  $P_{EPR}$ .

**Completeness and Soundness.** We summarize the relevant part of the analysis of the EPR protocol from [Bro15]. First suppose  $P_{EPR}$  behaves honestly. If  $\|\Pi_0 Q|0^n\rangle\|^2 = p$ , then in a computation round,  $V_{EPR}$  outputs accept with probability  $p$ , whereas in a test round,  $V_{EPR}$  outputs accept with probability 1. This establishes completeness of the protocol:

**Theorem 2** (Completeness). *Suppose the verifier executes the EPR Protocol, choosing  $r = 0$  with probability  $p$ , on an input  $(Q, |\vec{x}\rangle)$  such that  $\|\Pi_0 Q|\vec{x}\rangle\|^2 \geq 1 - \delta$ . Then the probability that  $V_{EPR}$  accepts when interacting with the honest prover  $P_{EPR}$  is at least  $(1 - p) + p(1 - \delta)$ .*

The following theorem is implicit in [Bro15, Section 7.6], but we include a brief proof sketch:

**Theorem 3** (Soundness). *Suppose the verifier executes the EPR Protocol, choosing  $r = 0$  with probability  $p$ , on an input  $(Q, |\vec{x}\rangle)$  such that  $\|\Pi_0 Q|\vec{x}\rangle\|^2 \leq \delta$ . Let  $P_{EPR}^*$  be an arbitrary prover such that  $P_{EPR}^*$  is accepted by  $V_{EPR}$  with probability  $q_t$  conditioned on  $r \neq 0$ , and  $q_c$  conditioned on  $r = 0$ . Then the prover's overall acceptance probability is  $pq_c + (1 - p)q_t$ , and*

$$q_c \leq 2(q_t \delta + (1 - q_t)) - \delta.$$

*Proof sketch.* Using the notation of [Bro15], let  $A = \sum_k \sum_{Q \in B_{t,n}} |\alpha_{k,Q}|^2$ . For intuition,  $A$  should be thought of as the total weight on attacks that could change the outcome of the computation, called non-benign attacks in [Bro15]. By [Bro15], the probability of rejecting in a computation round is  $1 - q_c \geq (1 - \delta)(1 - A)$ , whereas the probability of rejecting in a test round is  $1 - q_t \geq \frac{1}{2}A$ . Combining these gives  $q_c \leq 2(q_t \delta + (1 - q_t)) - \delta$ .  $\square$

### 3 Rigidity

Each of our delegation protocols includes a *rigidity test* that is meant to verify that one of the provers measures his half of shared EPR pairs in a basis specified by the verifier, thereby preparing one of a specific family of post-measurement states on the other prover's space; the post-measurement states will form the basis for the delegated computation. This will be used to certify that one of the provers in our two-prover schemes essentially behaves as the quantum part of  $V_{EPR}$  would in the EPR protocol.

The main rigidity game is given in Section 3.4, while Sections 3.5 and 3.6 contain variants of it, which we later employ in the Leash and Dog-Walker protocols; here we give a brief overview of the structure of the test. The test is parametrized by the number  $m$  of EPR pairs to be used. The test  $\text{CLIFF}(\Sigma, m)$  is a single round of classical interaction between the verifier and the two provers. With constant probability the verifier sends one of the provers a string  $W$  chosen uniformly at random from  $\Sigma^m$  where the set  $\Sigma = \{X, Y, Z, F, G\}$  contains a label for each single-qubit observable to be tested. With the remaining probability other queries, requiring the measurement of observables not in  $\Sigma^m$  (such as the measurement of pairs of qubits in the Bell basis).

In general, an arbitrary strategy for the provers in the rigidity game consists of an arbitrary entangled state  $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  (which we take to be pure), and measurements (which we take to be projective) for each possible question.<sup>2</sup> This includes an  $m$ -bit outcome projective measurement  $\{W^u\}_{u \in \{0,1\}^m}$  for each of the queries  $W \in \Sigma^m$ . Our rigidity result states that any strategy that succeeds with probability  $1 - \varepsilon$  in the test is within  $\text{poly}(\varepsilon)$  of the honest strategy, up to local isometries (see Theorem 10 for a precise statement). This is almost true, but for an irreconcilable ambiguity in the definition of the complex phase  $\sqrt{-1}$ . The

<sup>2</sup>We make the assumption that the players employ a pure-state strategy for convenience, but it is easy to check that all proofs extend to the case of a mixed strategy. Moreover, it is always possible to consider (as we do) projective strategies only by applying Naimark's dilation theorem, and adding an auxiliary local system to each player as necessary, since no bound is assumed on the dimension of their systems.

fact that complex conjugation of observables leaves correlations invariant implies that no classical test can distinguish between the two nontrivial inequivalent irreducible representations of the Pauli group, which are given by the Pauli matrices  $\sigma_X, \sigma_Y, \sigma_Z$  and their complex conjugates  $\overline{\sigma_X} = \sigma_X, \overline{\sigma_Z} = \sigma_Z, \overline{\sigma_Y} = -\sigma_Y$  respectively. In particular, the provers may use a strategy that uses a combination of both representations; as long as they do so consistently no test will be able to detect this behavior.<sup>3</sup> The formulation of our result accommodates this irreducible degree of freedom by forcing the provers to use a single qubit, the  $(m+1)$ -st, to make their choice of representation (so honest provers require the use of  $(m+1)$  EPR pairs to test the operation of  $m$ -fold tensor products of observables from  $\Sigma$ , acting on  $m$  EPR pairs).

We introduce the language required to formulate our testing results in Section 3.1. We then start by giving a test for the conjugation of one observable to another by a unitary, the Conjugation Test, in Section 3.2. The test depends on a series of elementary tests, for commutation, anti-commutation, etc., that are presented in Appendix A. In Section 3.3, we apply the Conjugation Test to test the relations that dictate how an arbitrary  $m$ -qubit Clifford unitary acts by conjugation on the Pauli matrices. In Section 3.4 we specialize the test to the case of unitaries that can be expressed as the  $m$ -fold tensor product of Clifford observables taken from the set  $\Sigma$ . In Sections 3.5 and 3.6, we describe variants of the test from Section 3.4, which are later employed in the Leash and Dog-Walker protocols.

### 3.1 Testing

In this section we recall the standard formalisms from self-testing, including state-dependent distance measure, local isometries, etc. We also introduce a framework of “tests for relations” that will be convenient to formulate our results.

**Distance measures.** Ultimately our goal is to test that a player implements a certain tensor product of single-qubit or two-qubit measurements defined by observables such as  $\sigma_X, \sigma_Y$ , or  $\sigma_G$ . Since it is impossible to detect whether a player applies a certain operation  $X$  on state  $|\psi\rangle$ , or  $VXV^\dagger$  on state  $V|\psi\rangle$ , for any isometry  $V : L(\mathcal{H}) \rightarrow L(\mathcal{H}')$  such that  $V^\dagger V = \text{Id}$ , we will (as is standard in testing) focus on testing identity up to *local isometries*. Towards this, we introduce the following important piece of notation:

**Definition 4.** For finite-dimensional Hilbert spaces  $\mathcal{H}_A$  and  $\mathcal{H}_{A'}$ ,  $\delta > 0$ , and operators  $R \in L(\mathcal{H}_A)$  and  $S \in L(\mathcal{H}_{A'})$  we say that  $R$  and  $S$  are  $\delta$ -isometric with respect to  $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ , and write  $R \simeq_\delta S$ , if there exists an isometry  $V : \mathcal{H}_A \rightarrow \mathcal{H}_{A'}$  such that

$$\|(R - V^\dagger S V) \otimes \text{Id}_B |\psi\rangle\|^2 = O(\delta).$$

If  $V$  is the identity, then we further say that  $R$  and  $S$  are  $\delta$ -equivalent, and write  $R \approx_\delta S$  for  $\|(R - S) \otimes \text{Id}_B |\psi\rangle\|^2 = O(\delta)$ .

The notation  $R \simeq_\delta S$  carries some ambiguity, as it does not specify the state  $|\psi\rangle$ . The latter should always be clear from context: we will often simply write that  $R$  and  $S$  are  $\delta$ -isometric, without explicitly specifying  $|\psi\rangle$  or the isometry. The relation is transitive, but not reflexive: the operator on the right will always act on a space of dimension at least as large as that on which the operator on the left acts. The notion of  $\delta$ -equivalence is both transitive (its square root obeys the triangle inequality) and reflexive, and we will use it as our main notion of distance.

---

<sup>3</sup>See [RUV12, Appendix A] for an extended discussion of this issue, with a similar resolution to ours.

**Tests.** We formulate our tests as two-player games in which both players are treated symmetrically. We often use the same symbol, a capital letter  $X, Z, W, \dots$ , to denote a question in the game and the associated projective measurement  $\{W^a\}$  applied by the player upon receipt of that question. To a projective measurement with outcomes in  $\{0, 1\}^n$  we associate a family of observables  $W(u)$  parametrized by  $n$ -bit strings  $u \in \{0, 1\}^n$ , defined by  $W(u) = \sum_a (-1)^{u \cdot a} W^a$ . If  $n = 1$  we simply write  $W = W(1) = W^0 - W^1$ ; note that  $W(0) = \text{Id}$ .

With the exception of the Tomography Test TOM presented in Section 3.6, all the games, or tests, we consider implicitly include a “consistency test” which is meant to enforce that whenever both players are sent identical questions, they produce matching answers. More precisely, let  $T$  be any of the two-player tests described in the paper. Let  $\text{Pr}_T(W, W')$  be the distribution on questions  $(W, W')$  to the players that is specified by  $T$ . Since the players are always treated symmetrically,  $\text{Pr}_T(\cdot, \cdot)$  is permutation-invariant. Let  $\text{Pr}_T(\cdot)$  denote the marginal on either player. Then, instead of executing the test  $T$  as described, the verifier performs the following:

- (i) With probability  $1/2$ , execute  $T$ .
- (ii) With probability  $1/2$ , select a random question  $W$  according to  $\text{Pr}_T(W)$ . Send  $W$  to both players. Accept if and only if the players’ answers are equal.

Then, success with probability at least  $1 - \varepsilon$  in the modified test implies success with probability at least  $1 - 2\varepsilon$  in the original test, as well as in the consistency test. If  $\{W_A^a\}$  and  $\{W_B^b\}$  are the players’ corresponding projective measurements, the latter condition implies

$$\begin{aligned} \sum_a \|\langle \psi | W_A^a \otimes \text{Id} - \text{Id} \otimes W_B^a | \psi \rangle_{AB}\|^2 &= 2 - 2 \sum_a \langle \psi | W_A^a \otimes W_B^a | \psi \rangle \\ &\leq 4\varepsilon, \end{aligned} \tag{3}$$

so that  $W_A^a \otimes \text{Id} \approx_\varepsilon \text{Id} \otimes W_B^a$  (where the condition should be interpreted on average over the choice of a question  $W$  distributed as in the test). Similarly, if  $W_A, W_B$  are observables for the players that succeed in the consistency test with probability  $1 - 2\varepsilon$  we obtain  $W_A \otimes \text{Id} \approx_\varepsilon \text{Id} \otimes W_B$ . We will often use both relations to “switch” operators from one player’s space to the other’s; as a result we will also often omit an explicit specification of which player’s space an observable is applied to.

**Strategies.** Given a two-player game, or test, a strategy for the players consists of a bipartite entangled state  $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  together with families of projective measurements  $\{W_A^a\}$  for Alice and  $\{W_B^a\}$  for Bob, one for each question  $W$  that can be sent to either player in the test. As already mentioned, for convenience we restrict our attention to pure-state strategies employing projective measurements. We will loosely refer to a strategy for the players as  $(W, |\psi\rangle)$ , with the symbol  $W$  referring to the complete set of projective measurements used by the players in the game; taking advantage of symmetry we often omit the subscript A or B, as all statements involving observables for one player hold verbatim with the other player’s observables as well.

**Relations.** We use  $\mathcal{R}$  to denote a set of relations over variables  $X, Z, W, \dots$ , such as

$$\mathcal{R} = \{XZXZ = -\text{Id}, HX = ZH, X, Z, H \in \text{Obs}\}.$$

We only consider relations that can be brought in the form either  $f(W) = (-1)^a W_1 \cdots W_k = \text{Id}$ , where the  $W_i$  are (not necessarily distinct) unitary variables and  $a \in \mathbb{Z}_2$ , or  $f(W) = W_1 \cdot (\sum_a \omega_a W_2^a) = \text{Id}$ , where  $W_1$  is a unitary variable,  $\{W_2^a\}$  a projective measurement with  $s$  possible outcomes, and  $\omega_a$  are (arbitrary)  $s$ -th roots of unity.

**Definition 5** (Rigid self-test). We say that a set of relations  $\mathcal{R}$  is  $(c, \delta(\varepsilon))$ -testable, on average under the distribution  $\mathcal{D} : \mathcal{R} \rightarrow [0, 1]$ , if there exists a game (or test)  $G$  with question set  $\mathcal{Q}$  that includes (at least) a symbol for each variable in  $\mathcal{R}$  that is either an observable or a POVM and such that:

- (*Completeness*) There exists a set of operators which exactly satisfy all relations in  $\mathcal{R}$  and a strategy for the players which uses these operators (together possibly with others for the additional questions) that has success probability at least  $c$ ;
- (*Soundness*) For any  $\varepsilon > 0$  and any strategy  $(W, |\psi\rangle_{AB})$  that succeeds in the game with probability at least  $c - \varepsilon$ , the associated measurement operators satisfy the relations in  $\mathcal{R}$  up to  $\delta(\varepsilon)$ , in the state-dependent norm. More precisely, on average over the choice of a relation  $f(W) = \text{Id}$  from  $\mathcal{R}$  chosen according to  $\mathcal{D}$ , it holds that  $\|\text{Id} \otimes (f(W) - \text{Id})|\psi\rangle_{AB}\|^2 \leq \delta(\varepsilon)$ .

If both conditions hold, we also say that the game  $G$  is a robust  $(c, \delta(\varepsilon))$  self-test for the relations  $\mathcal{R}$ .

Most of the games we consider have perfect completeness,  $c = 1$ , in which case we omit explicitly mentioning the parameter. The distribution  $\mathcal{D}$  will often be implicit from context, and we do not always specify it explicitly (e.g. in case we only measure  $\delta(\varepsilon)$  up to multiplicative factors of order  $|\mathcal{R}|$  the exact distribution  $\mathcal{D}$  does not matter as long as it has complete support).

**Definition 6** (Stable relations). We say that a set of relations  $\mathcal{R}$  is  $\delta(\varepsilon)$ -stable, on average under the distribution  $\mathcal{D} : \mathcal{R} \rightarrow [0, 1]$ , if for any two families of operators  $W_A \in L(\mathcal{H}_A)$  and  $W_B \in L(\mathcal{H}_B)$  that are consistent on average, i.e.

$$\mathbb{E}_{f \sim \mathcal{D}} \mathbb{E}_{W \in_U f} \left\| (\text{Id} \otimes W_B - W_A \otimes \text{Id}) |\psi\rangle \right\|^2 \leq \varepsilon,$$

where  $W \in_U f$  is shorthand for  $W$  being a uniformly random operator among those appearing in the relation specified by  $f$ , and satisfy the relations on average, i.e.

$$\mathbb{E}_{\substack{f \sim \mathcal{D}: \\ f(W) = \text{Id} \in \mathcal{R}}} \left\| (f(W_A) - \text{Id}) \otimes \text{Id} |\psi\rangle \right\|^2 \leq \varepsilon,$$

there exists operators  $\hat{W}$  which satisfy the same relations exactly and are  $\delta(\varepsilon)$ -isometric to the  $W$  with respect to  $|\psi\rangle$ , on average over the choice of a random relation in  $\mathcal{R}$  and a uniformly random  $W$  appearing in the relation, i.e. there exists an isometry  $V_A$  such that

$$\mathbb{E}_{f \sim \mathcal{D}} \mathbb{E}_{W \in_U f} \left\| (\hat{W}_A - V_A^\dagger W_A V_A) \otimes \text{Id} |\psi\rangle \right\|^2 = O(\delta(\varepsilon)).$$

### 3.2 The conjugation test

We give a test which certifies that a unitary (not necessarily an observable) conjugates one observable to another. More precisely, let  $A, B$  be observables, and  $R$  a unitary, acting on the same space  $\mathcal{H}$ . The test  $\text{CONJ}(A, B, R)$ , given in Figure 6, certifies that the players implement observables of the form

$$X_R = \begin{pmatrix} 0 & R^\dagger \\ R & 0 \end{pmatrix} \quad \text{and} \quad C = C_{A,B} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \quad (4)$$

such that  $X_R$  and  $C$  commute. The fact that  $X_R$  is an observable implies that  $R$  is unitary,<sup>4</sup> while the commutation condition is equivalent to the relation  $RAR^\dagger = B$ . The test thus tests for the relations

$$\mathcal{C}\{R, C\} = \{X_R, C, X, Z \in \text{Obs}\} \cup \{XZ = -ZX\} \cup \{X_R C = C X_R, X_R Z = -Z X_R, CZ = ZC\}.$$

<sup>4</sup>Note that  $R$  will not be directly accessed in the test, since by itself it does not necessarily correspond to a measurement.



Here the anti-commuting observables  $X$  and  $Z$  are used to specify a basis in which  $X_R$  and  $C$  can be block-diagonalized. The anti-commutation and commutation relations with  $Z$  enforce that  $X_R$  and  $C$  respectively have the form described in (4). These relations are enforced using simple commutation and anti-commutation tests that are standard in the literature on self-testing. For convenience, we state those tests, COM and AC, in Appendix A. The conjugation test, which uses them as sub-tests, is given in Figure 6. Here, “Inputs” refers to a subset of designated questions in the test; “Relation” indicates a relation that the test aims to certify; “Test” describes the certification protocol. (Recall that all our protocols implicitly include a “consistency” test in which a question is chosen uniformly at random from the marginal distribution and sent to both players, whose answers are accepted if and only if they are equal.)

---

Test CONJ( $A, B, R$ )

- Inputs:  $A$  and  $B$  observables on the same space  $\mathcal{H}$ , and  $X$  and  $Z$  observables on  $\mathcal{H}'$ .  $X_R$  and  $C$  observables on  $\mathcal{H} \otimes \mathcal{H}'$ .
  - Relations:  $\mathcal{C}\{R, C\}$ , with  $R$  defined from  $X_R$ , and  $C$  related to  $A$  and  $B$ , as in (4).
  - Test: execute each of the following with equal probability
    - (a) With probability  $1/8$  each, execute tests  $\text{AC}(X, Z)$ ,  $\text{COM}(C, Z)$ ,  $\text{COM}(X_R, C)$ ,  $\text{AC}(X_R, Z)$  and  $\text{COM}(A, X)$ ,  $\text{COM}(B, X)$ ,  $\text{COM}(A, Z)$ ,  $\text{COM}(B, Z)$ .
    - (b) Ask one player to measure  $A$ ,  $B$ ,  $C$  or  $Z$  (with probability  $1/4$  each), and the other to jointly measure  $A$  or  $B$  (with probability  $1/2$  each) and  $Z$ . The first player returns one bit, and the second two bits. Reject if either:
      - The first player was asked  $C$ , the second player was asked  $(A, Z)$ , his second answer bit is 0, and his first answer bit does not match the first player’s;
      - The first player was asked  $C$ , the second player was asked  $(B, Z)$ , his second answer bit is 1, and his first answer bit does not match the first player’s.
      - The first player was asked  $A$ ,  $B$ , or  $Z$  and his answer bit does not match the corresponding answer from the second player.
- 

Figure 6: The conjugation test, CONJ( $A, B, R$ ).

**Lemma 7.** *The test CONJ( $A, B, R$ ) is a  $(1, \delta)$  self-test for the set of relations  $\mathcal{C}\{R, C\}$ , for some  $\delta = O(\sqrt{\epsilon})$ . Moreover, for any strategy that succeeds with probability at least  $1 - \epsilon$  in the test it holds that  $C \approx_\delta A(\text{Id} + Z)/2 + B(\text{Id} - Z)/2$ , where  $A, B, C$  and  $Z$  are the observables applied by the prover on receipt of a question with the same label.*

*Proof.* Completeness is clear, as players making measurements on a maximally entangled state on  $\mathcal{H}_A \otimes \mathcal{H}_B$ , tensored with an EPR pair on  $\mathbb{C}^2 \otimes \mathbb{C}^2$  for the  $X$  and  $Z$  observables, and using  $X_R$  and  $C$  defined in (4) (with the blocks specified by the space associated with each player’s half-EPR pair) succeed in each test with probability 1.

We now consider soundness. Success in  $\text{AC}(X, Z)$  in part (a) of the test implies the existence of local isometries  $V_A, V_B$  such that  $V_A : \mathcal{H}_A \rightarrow \mathcal{H}_{\hat{A}} \otimes \mathbb{C}_{A'}^2$ , with  $X \simeq_{\sqrt{\epsilon}} \text{Id}_{\hat{A}} \otimes \sigma_X$  and  $Z \simeq_{\sqrt{\epsilon}} \text{Id}_{\hat{A}} \otimes \sigma_Z$ . By Lemma 29, approximate commutation with both  $X$  and  $Z$  implies that under the same isometry,  $A \simeq_{\sqrt{\epsilon}} A_I \otimes \text{Id}$  and  $B \simeq_{\sqrt{\epsilon}} B_I \otimes \text{Id}$ , for observables  $A_I, B_I$  on  $\mathcal{H}_{\hat{A}}$ . Similarly, the parts of the test involving  $C$  and  $X_R$  imply that they each have the block decomposition specified in (4). In particular, anti-commutation of  $X_R$  with  $Z$  certifies that  $X_R$  has a decomposition of the form  $X_R \simeq R_X \otimes \sigma_X + R_Y \otimes \sigma_Y$ . Using that

$X_R$  is an observable, we deduce that there exists a unitary  $R$  on  $\mathcal{H}_{\hat{A}}$  such that  $R \approx R_X + iR_Y$ . Similarly, commutation of  $C$  with  $Z$  implies that  $C \simeq C_I \otimes I + C_Z \otimes \sigma_Z$ , for Hermitian  $C_I, C_Z$  such that  $C_I \pm C_Z$  are observables.

Next we analyze part (b) of the test. Let  $\{W_{AZ}^{a,z}\}$  be the projective measurement applied by the second player upon query  $(A, Z)$ . Success with probability  $1 - O(\varepsilon)$  in the first item ensures that

$$|\langle \psi | C \otimes (W_{AZ}^{00} - W_{AZ}^{10}) | \psi \rangle| = O(\varepsilon),$$

and a similar condition holds from the second item, with  $W_{BZ}$  instead of  $W_{AZ}$ . Success with probability  $1 - O(\varepsilon)$  in the third item ensures consistency of  $\{W_{AZ}^{a,z}\}$  (resp.  $\{W_{BZ}^{a,z}\}$ ) with the observable  $A$  (resp.  $B$ ) when marginalizing over the second outcome, and  $Z$  when marginalizing over the first outcome. Using the decompositions for  $A, B$  and  $C$  derived earlier, we obtain  $C_I \approx (A + B)/2$  and  $C_Z \approx (A - B)/2$ , giving the “Moreover” part of the lemma.

Finally, success in test  $\text{COM}(X_R, C)$  certifies the approximate commutation relation  $[X_R, C] \approx_{\sqrt{\varepsilon}} 0$ , which, given the decomposition of  $X_R$  and  $C$  obtained so far, implies  $RA \approx BR$ , as desired.  $\square$

### 3.3 Testing Clifford unitaries

Let  $m \geq 1$  be an integer, and  $R$  an  $m$ -qubit Clifford unitary.  $R$  is characterized, up to phase, by its action by conjugation on the  $m$ -qubit Weyl-Heisenberg group. This action is described by linear functions  $h_S : \{0, 1\}^m \times \{0, 1\}^m \rightarrow \mathbb{Z}_4$  and  $h_X, h_Z : \{0, 1\}^m \times \{0, 1\}^m \rightarrow \{0, 1\}^m$  such that

$$R\sigma_X(a)\sigma_Z(b)R^\dagger = (-1)^{h_S(a,b)}\sigma_X(h_X(a,b))\sigma_Z(h_Z(a,b)), \quad \forall a, b \in \{0, 1\}^m. \quad (5)$$

Using that  $(\sigma_X(a)\sigma_Z(b))^\dagger = (-1)^{a \cdot b}\sigma_X(a)\sigma_Z(b)$ , the same condition must hold of the right-hand side of (5), thus  $h_X(a, b) \cdot h_Z(a, b) = a \cdot b \pmod{2}$ . To any family of observables  $\{X(a), Z(b), a, b \in \{0, 1\}^m\}$  we associate, for  $a, b \in \{0, 1\}^m$ ,

$$A(a, b) = i^{a \cdot b} X(a)Z(b), \quad B(a, b) = i^{a \cdot b} X(h_X(a, b))Z(h_Z(a, b)), \quad (6)$$

where the phase  $i^{a \cdot b}$  is introduced to ensure that  $A(a, b)$  and  $B(a, b)$  are observables. Define  $C(a, b)$  in terms of  $A(a, b)$  and  $B(a, b)$  as in (4). The Clifford conjugation test aims to test for the conjugation relation  $RA(a, b)R^\dagger = B(a, b)$ , for all (in fact, on average over a randomly chosen)  $(a, b)$ . For this, we first need a test that ensures  $A(a, b)$  and  $B(a, b)$  themselves have the correct form, in terms of a tensor product of Pauli observables. Such a test was introduced in [NV16], where it is called “Pauli braiding test”. The test certifies the Pauli relations

$$\begin{aligned} \mathcal{P}^{(m)}\{X, Y, Z\} = & \left\{ W(a) \in \text{Obs}, W \in \{X, Y, Z\}^m, a \in \{0, 1\}^m \right\} \\ & \cup \left\{ W(a)W'(a') = (-1)^{|\{i: W_i \neq W'_i \wedge a_i a'_i = 1\}|} W'(a')W(a), \forall W, W' \in \{X, Y, Z\}^n, \forall a, a' \in \{0, 1\}^m \right\} \\ & \cup \left\{ W(a)W(a') = W(a + a'), \forall a, a' \in \{0, 1\}^m \right\}. \end{aligned}$$

The Pauli braiding test is recalled in Appendix B, and we refer to the test as  $\text{PBT}(X, Y, Z)$ . The original test from [NV16] only allows to test for tensor products of  $\sigma_X$  and  $\sigma_Z$  Pauli observables, and we extend the test to include Pauli  $\sigma_Y$ . This requires us to provide a means to accommodate the phase ambiguity discussed earlier. The result is described in the following lemma; we refer to Appendix B.2 for the proof. (In some cases a simpler variant of the test, which does not attempt to test for the  $Y$  observable, will suffice. This is essentially the original test from [NV16], which we call  $\text{PBT}(X, Z)$  and is introduced in Appendix B.1.)

**Lemma 8.** Suppose  $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  and  $W(a) \in \text{Obs}(\mathcal{H}_A)$ , for  $W \in \{X, Y, Z\}^m$  and  $a \in \{0, 1\}^m$ , specify a strategy for the players that has success probability at least  $1 - \varepsilon$  in the extended Pauli braiding test  $\text{PBT}(X, Y, Z)$  described in Figure 26. Then there exist a state  $|\text{AUX}\rangle_{\hat{A}\hat{B}}$  and isometries  $V_D : \mathcal{H}_D \rightarrow ((\mathbb{C}^2)^{\otimes m})_{D'} \otimes \hat{\mathcal{H}}_{\hat{D}}$  for  $D \in \{A, B\}$ , such that

$$\|(V_A \otimes V_B)|\psi\rangle_{AB} - |\text{EPR}\rangle_{A'B'}^{\otimes m} |\text{AUX}\rangle_{\hat{A}\hat{B}}\|^2 = O(\sqrt{\varepsilon}),$$

and on expectation over  $W \in \{X, Y, Z\}^m$ ,

$$\mathbb{E}_{a \in \{0,1\}^m} \|(W(a) - V_A^\dagger(\sigma_W(a) \otimes \Delta_W(a))V_A) \otimes \text{Id}_B |\psi\rangle\|^2 = O(\sqrt{\varepsilon}), \quad (7)$$

where  $\Delta_W(a) = \prod_i \Delta_{W_i}^{a_i} \in \text{Obs}(\mathcal{H}_{\hat{A}})$  are observables with  $\Delta_X = \Delta_Z = \text{Id}$  and  $\Delta_Y$  an arbitrary observable on  $\hat{\mathcal{H}}$  such that

$$\|\Delta_Y \otimes \Delta_Y |\text{AUX}\rangle - |\text{AUX}\rangle\|^2 = O(\sqrt{\varepsilon}).$$

Building on the Pauli braiding test and the conjugation test from the previous section, the Clifford conjugation test  $\text{CONJ-CLIFF}(R)$  described in Figure 7 provides a test for the set of relations

$$\begin{aligned} \mathcal{J}_{h_S, h_X, h_Z}\{R\} = & \mathcal{P}^{(m)}\{X, Y, Z\} \cup \{R \in \mathcal{U}\} \cup \{\Delta_Y \in \text{Obs}\} \\ & \cup \{RX(a)Z(b)R^\dagger = \Delta_Y^{h_S(a,b)} X(h_X(a, b))Z(h_Z(a, b)), \forall a, b \in \{0, 1\}^m\} \\ & \cup \{\Delta_Y X(a) = X(a)\Delta_Y, \Delta_Y Z(b) = Z(b)\Delta_Y, \forall a, b \in \{0, 1\}^m\}. \end{aligned} \quad (8)$$

Note the presence of the observable  $\Delta_Y$ , which arises from the conjugation ambiguity in the definition of  $Y$  (see Lemma 8).

---

Test  $\text{CONJ-CLIFF}(R)$ :

- Input:  $R$  an  $m$ -qubit Clifford unitary. Let  $h_S, h_X, h_Z$  be such that (5) holds, and  $A(a, b), B(a, b)$  the observables defined in (6).
  - Relations:  $\mathcal{J}_{h_S, h_X, h_Z}\{R\}$  defined in (8).
  - Test: execute each of the following with equal probability
    - (a) Execute test  $\text{PBT}(X, Y, Z)$  on  $(m + 1)$  qubits, where the last qubit is called the “control” qubit;
    - (b) Select  $a, b \in \{0, 1\}^m$  uniformly at random. Let  $C(a, b)$  be the observable defined from  $A(a, b)$  and  $B(a, b)$  in (4), with the block structure specified by the control qubit. Execute test  $\text{CONJ}\{A(a, b), B(a, b), R\}$ . In the test, to specify query  $A(a, b)$  or  $B(a, b)$ , represent each as a string in  $\{I, X, Y, Z\}^m$  and use the same label as for the same query when it is used in part (a).
- 

Figure 7: The Clifford conjugation test,  $\text{CONJ-CLIFF}(R)$ .

**Lemma 9.** Let  $R$  be an  $m$ -qubit Clifford unitary and  $h_S, h_X, h_Z$  such that (5) holds. Suppose a strategy for the players succeeds with probability at least  $1 - \varepsilon$  in test  $\text{CONJ-CLIFF}(R)$ . Let  $V_A : \mathcal{H}_A \rightarrow ((\mathbb{C}^2)^{\otimes (m+1)})_{A'} \otimes \mathcal{H}_{\hat{A}}$  be the isometry whose existence follows from part (a) of the test, and  $\Delta_Y$  the observable on  $\mathcal{H}_{\hat{A}}$  that represents the phase ambiguity (see Lemma 8). Then there exists a unitary  $\Lambda_R$  on  $\mathcal{H}_{\hat{A}}$ , commuting with  $\Delta_Y$ , such that

$$\|\Lambda_R \otimes \Lambda_R |\text{AUX}\rangle - |\text{AUX}\rangle\|^2 = O(\text{poly}(\varepsilon)). \quad (9)$$

Moreover, let  $\hat{\tau}_R$  be any  $m$ -qubit Clifford unitary, acting on the space  $(\mathbb{C}^2)^{\otimes m}$  into which the isometry  $V_A$  maps, which satisfies the relations specified in (8), where for any location  $i \in \{1, \dots, m\}$  such that  $a_i = b_i = 1$  we replace  $\sigma_X \sigma_Z$  by  $\tau_Y = \sigma_Y \otimes (i\Delta_Y)$ . Then, letting  $\tau_R = \hat{\tau}_R(\text{Id}_{A'} \otimes \Lambda_R)$  we have that under the same isometry,

$$R \simeq_{\text{poly}(\varepsilon)} \tau_R.$$

Note that  $\hat{\tau}_R$  is only defined up to phase in the lemma. Any representative will do, as the phase ambiguity can be absorbed in  $\Lambda_R$ . As an example, in this notation we have

$$\hat{\tau}_F = \frac{1}{\sqrt{2}}(-\sigma_X + \sigma_Y \otimes \Delta_Y), \quad \hat{\tau}_G = \frac{1}{\sqrt{2}}(\sigma_X + \sigma_Y \otimes \Delta_Y), \quad (10)$$

where the “honest” single-qubit Clifford observables  $\sigma_F$  and  $\sigma_G$  are defined in (2).

Completeness of the test is clear, as players making measurements on  $(m+1)$  shared EPR pairs using standard Pauli observables,  $R$ , and  $C(a, b)$  defined in (4) with  $A(a, b)$  and  $B(a, b)$  as in (6) will pass all tests with probability 1.

*Proof sketch.* For  $D \in \{A, B\}$  let  $V_D$  be the isometries that follow from part (a) of the test and Lemma 8. According to (6),  $A(a, b)$  and  $B(a, b)$  can each be expressed (up to phase) as a tensor product of  $X, Y, Z$  operators, where the number of occurrences of  $Y$  modulo 2 is  $a \cdot b$  for  $A(a, b)$  and  $h_X(a, b) \cdot h_Z(a, b) = a \cdot b \pmod{2}$  for  $B(a, b)$ . Thus the labels used to specify the observables in  $A(a, b)$  and  $B(a, b)$  in part (b), together with the analysis of part (a) and Lemma 8, imply that

$$A(a, b) \simeq_{\sqrt{\varepsilon}} \sigma_X(a) \sigma_Z(b) \otimes (i\Delta_Y)^{a \cdot b} \text{ and } B(a, b) \simeq_{\sqrt{\varepsilon}} \sigma_X(h_X(a, b)) \sigma_Z(h_Z(a, b)) \otimes (i\Delta_Y)^{a \cdot b + h_S(a, b)},$$

under the same isometry. Applying the analysis of the conjugation test given in Lemma 7 shows that  $X_R$  must have the form in (4), for some  $R$  that approximately conjugates  $A(a, b)$  to  $B(a, b)$ , on average over uniformly random  $a, b \in \{0, 1\}^m$ .

Let  $\hat{\tau}_R$  be as defined in the paragraph preceding the lemma. Note that  $\hat{\tau}_R$  acts on  $\mathcal{H}_{A'}$  and  $\mathcal{H}_{\hat{A}}$ . After application of the isometry,  $R$  has an expansion

$$R \simeq \hat{\tau}_R \cdot \left( \sum_{a, b} \sigma_X(a) \sigma_Z(b) \otimes \Lambda_R(a, b) \right), \quad (11)$$

for arbitrary  $\Lambda_R(a, b)$  on  $\mathcal{H}_{\hat{A}}$ ; since  $\hat{\tau}_R$  is invertible such an expansion exists for any operator. Using the approximate version of (5) certified by the conjugation test (Lemma 7),

$$RV_A^\dagger (\sigma_X(a) \sigma_Z(b) \otimes \Delta_Y^{a \cdot b}) V_A \approx V_A^\dagger (\sigma_X(h_X(a, b)) \sigma_Z(h_Z(a, b)) \otimes \Delta_Y^{a \cdot b + h_S(a, b)}) V_A R,$$

where the approximation holds on average over a uniformly random choice of  $(a, b)$  and up to error that is polynomial in  $\varepsilon$  but independent of  $m$ . Expanding out  $R$  and using the consistency relations between the two provers,

$$\begin{aligned} & \sum_{c, d} \hat{\tau}_R \left( \sigma_X(c) \sigma_Z(d) \otimes \Lambda_R(c, d) \right) \otimes \left( (-1)^{a \cdot b} \sigma_X(a) \sigma_Z(b) \otimes \Delta_Y^{a \cdot b} \right) \\ & \approx \sum_{c, d} \left( \sigma_X(h_X(a, b)) \sigma_Z(h_Z(a, b)) \otimes \Delta_Y^{a \cdot b + h_S(a, b)} \right) \hat{\tau}_R \left( \sigma_X(c) \sigma_Z(d) \otimes \Lambda_R(c, d) \right) \otimes \text{Id}, \end{aligned} \quad (12)$$

where the factor  $(-1)^{a \cdot b}$  comes from using

$$(\sigma_X(a) \sigma_Z(b) \otimes \text{Id}) |\text{EPR}\rangle^{\otimes m} = (\text{Id} \otimes (\sigma_X(a) \sigma_Z(b))^T) |\text{EPR}\rangle^{\otimes m}.$$

Using the conjugation relations satisfied, by definition, by  $\hat{\tau}_R$ , the right-hand side of (12) simplifies to

$$\sum_{c,d} \hat{\tau}_R \left( \sigma_X(a) \sigma_Z(b) \sigma_X(c) \sigma_Z(d) \otimes \Delta_Y^{a \cdot b} \Lambda_R(c, d) \right) \otimes \text{Id}. \quad (13)$$

Next using the fact that the state on which the approximations are measured is maximally entangled across registers **A** and **B**, together with the Pauli (anti-)commutation relations, to simplify the left-hand side of (12), together with (13) we arrive at the approximation

$$\begin{aligned} \sum_{c,d} \left( (-1)^{a \cdot d + b \cdot c} \sigma_X(a+c) \sigma_Z(b+d) \otimes \Lambda_R(c, d) \right) \otimes \left( \text{Id} \otimes \Delta_Y^{a \cdot b} \right) \\ \approx \sum_{c,d} \left( \sigma_X(a+c) \sigma_Z(b+d) \otimes \Delta_Y^{a \cdot b} \Lambda_R(c, d) \right) \otimes \text{Id}. \end{aligned}$$

If  $(c, d) \neq (0, 0)$  a fraction about half of all  $(a, b)$  such that  $a \cdot b = 0$  satisfy  $a \cdot d + b \cdot c = 1$ . Using that  $\{\sigma_X(a) \sigma_Z(b) \otimes \text{Id} | \text{EPR}\}$  are orthogonal for different  $(a, b)$ , the above then implies that  $\Lambda_R(c, d) \approx -\Lambda_R(c, d)$ , on average over  $(c, d) \neq (0, 0)$ . Hence  $\Lambda_R(c, d) \approx 0$ , on average over  $(c, d) \neq (0, 0)$ . Considering  $(a, b)$  such that  $a \cdot b = 1$  implies that  $\Lambda_R(0, 0)$  approximately commutes with  $\Delta_Y$ . Finally, the relation (9) follows from self-consistency of  $X_R$  implicitly enforced in the test.  $\square$

### 3.4 Tensor products of single-qubit Clifford observables

We turn to testing observables in the  $m$ -fold direct product of the Clifford group. Although the test can be formulated more generally, for our purposes it will be sufficient to specialize it to the case where each element in the direct product is an observable taken from the set  $\Sigma = \{X, Y, Z, F, G\}$  associated with the single-qubit Pauli observables defined in Section 2.1. Recall that the associated operators satisfy the conjugation relation  $\sigma_Y \sigma_F \sigma_Y = \sigma_G$ , which will be tested as part of our procedures (specifically, item (c) in Figure 8).

---

Test  $\text{CLIFF}(\Sigma, m)$ :

- Input: An integer  $m$  and a subset  $\Sigma = \{X, Y, Z, F, G\}$  of the single-qubit Clifford group.
  - Test: Select  $W \in \Sigma^m$  uniformly at random. Execute each of the following with equal probability:
    - (a) Execute the test  $\text{CONJ-CLIFF}(W)$ ;
    - (b) Send one player either the query  $W$ , or  $X_W$  and the other  $(W, X(e_{m+1}))$ , where  $e_{m+1}$  indicates the control qubit used for part (a). Receive one bit from the first player, and two from the second. If the query to the first player was  $W$ , check that the first player's answer is consistent with the second player's first answer bit. If the query to the first player was  $X_W$ , then: If the second player's second bit is 0, check that his first bit is consistent with the first player's; If the second player's second bit is 1, check that his first bit is different than the first player's.
    - (c) Let  $S$  and  $T$  be subsets of the positions in which  $W_i = F$  and  $W_i = G$  respectively, chosen uniformly at random. Let  $W'$  equal  $W$  except  $W'_i = G$  for  $i \in S$ , and  $W'_i = F$  for  $i \in T$ . Let  $R = Y(\sum_{i \in S \cup T} e_i)$ . Execute test  $\text{CONJ}(W, W', R)$ .
    - (d) Set  $W'_i = X$  (resp.  $Y$ ) whenever  $W_i = Y$  (resp.  $X$ ),  $W'_i = F$  (resp.  $G_i$ ) whenever  $W_i = G$  (resp.  $F$ ), and  $W'_i = X$  whenever  $W_i = Z$ . Execute test  $\text{PBT}(W, W')$  on  $m$  qubits.
    - (e) Let  $S$  and  $T$  be subsets of (non-overlapping) pairs of positions in which  $W_i = F$  and  $W_i = G$  respectively, chosen uniformly at random. Send one player the query  $W$ , with entries  $(i, j) \in S \cup T$  removed and replaced by  $\Phi_{i,j}$  (indicating a measurement in the Bell basis).
      - With probability  $1/2$ , send the other player the query  $W$ . Check consistency of outcomes associated with positions not in  $S \cup T$ . For outcomes in  $S \cup T$ , check the natural consistency as well: e.g. if the Bell measurement indicated the outcome  $\Phi_{00}$ , then the two outcomes reported by the other player at those locations should be identical.
      - With probability  $1/2$ , execute an independent copy of the Bell measurement test  $\text{BELL}$  (Figure 24) between the first and second players in each of the pair of qubits in  $S \cup T$ .
- 

Figure 8: The  $m$ -qubit Clifford test,  $\text{CLIFF}(\Sigma, m)$ .

The test is described in Figure 8. It is divided in five parts. Part (a) of the test executes  $\text{CONJ-CLIFF}(W)$  to verify that an observable  $W \in \Sigma^m$  satisfies the appropriate Pauli conjugation relations (5). Note that a priori test  $\text{CONJ-CLIFF}(W)$  only tests for the observable  $X_W$  obtained from  $W$  in blocks as  $X_R$  from  $R$  in (4) (indeed, in that test  $W$  need not be an observable). Thus part (b) of the test is introduced to verify that  $X_W \approx WX(e_{m+1})$ , where the  $(m+1)$ -st qubit is the one used to specify the block decomposition relating  $X_W$  to  $W$ . The result of parts (a) and (b) is that, under the same isometry as used to specify the Pauli  $X$  and  $Z$ ,  $W \simeq \hat{\tau}_W \cdot (\text{Id} \otimes \Lambda_W)$ , according to the same decomposition as shown in Lemma 9. The goal of the remaining three parts of the test is to verify that  $\Lambda_W = \Lambda_F^{|\{i: W_i \in \{F, G\}\}|}$ , for a single observable  $\Lambda_F$ . For this, part (c) of the test verifies that  $\Lambda_W$  only depends on the locations at which  $W_i \in \{F, G\}$ , but not on the specific observables at those locations. Part (d) verifies that  $\Lambda_W \approx \prod_{i: W_i \in \{F, G\}} \Lambda_i$  for commuting observables  $\Lambda_i$ . Finally, part (e) checks that  $\Lambda_i$  is (approximately) independent of  $i$ .

**Theorem 10.** *Suppose a strategy for the players succeeds in test  $\text{CLIFF}(\Sigma, m)$  (Figure 8) with probability at least  $1 - \varepsilon$ . Then for  $D \in \{A, B\}$  there exists an isometry*

$$V_D : \mathcal{H}_D \rightarrow (\mathbb{C}^2)_{D'}^{\otimes m} \otimes \mathcal{H}_D$$

such that

$$\|(V_A \otimes V_B)|\psi\rangle_{AB} - |\text{EPR}\rangle_{A'B'}^{\otimes m} |\text{AUX}\rangle_{\hat{A}\hat{B}}\|^2 = O(\sqrt{\varepsilon}), \quad (14)$$

and

$$\mathbb{E}_{W \in \Sigma^m, c \in \{0,1\}^m} \|\text{Id}_A \otimes (V_B W(c) - \tau_W(c) V_B) |\psi\rangle_{AB}\|^2 = O(\text{poly}(\varepsilon)). \quad (15)$$

Here  $\tau_W$  is defined from  $W$  as in Lemma 9, with  $\Lambda_{W_i} = \text{Id}$  if  $W_i \in \{X, Y, Z\}$  and  $\Lambda_{W_i} = \Lambda_F$  if  $W_i \in \{F, G\}$ , where  $\Lambda_F$  is an observable on  $\mathcal{H}_{\hat{B}}$  that commutes with  $\Delta_Y$ .

*Proof sketch.* The existence of the isometry, as well as (14) and (15) for  $W \in \{I, X, Y, Z\}^m$ , follows from the test  $\text{PBT}(X, Y, Z)$ , executed as part of the Clifford conjugation test from part (a), and Lemma 8. Using part (a) of the test and Lemma 9 it follows that every  $W \in \Sigma^m$  is mapped under the same isometry to

$$W \simeq_{\sqrt{\varepsilon}} \tau_W = \hat{\tau}_W(\text{Id} \otimes \Lambda_W), \quad (16)$$

where  $\hat{\tau}_W$  is as defined in the lemma and  $\Lambda_W$  is an observable on  $\mathcal{H}_{\hat{A}}$  which may depend on the whole string  $W$ ; here we also use the consistency check in part (b) to relate the observable  $X_W$  used in the Clifford conjugation test with the observable  $W$  used in part (c). Note that from the definition we can write  $\hat{\tau}_W = \otimes_i \hat{\tau}_{W_i}$ , where in particular  $\hat{\tau}_X = \sigma_X$ ,  $\hat{\tau}_Z = \sigma_Z$  and  $\hat{\tau}_Y = \sigma_Y \otimes \Delta_Y$ .

The analysis of the conjugation test given in Lemma 7 shows that success with probability  $1 - O(\varepsilon)$  in part (c) of the test implies the relations

$$\begin{aligned} \hat{\tau}_W \tau_R(\text{Id} \otimes \Lambda_W) &= \tau_R \hat{\tau}_W(\text{Id} \otimes \Lambda_W) \\ &\approx_{\sqrt{\varepsilon}} \hat{\tau}_{W'} \tau_R(\text{Id} \otimes \Lambda_{W'}), \end{aligned}$$

where the first equality is by definition of  $R$ , and uses that  $\tau_Y = \sigma_Y \otimes \Delta_Y$  and  $\Delta_Y$  commutes with  $\Lambda_W$ ; the approximation holds as a consequence of the conjugation test and should be understood on average over a uniformly random choice of  $W \in \Sigma^m$ . Thus  $\Lambda_W$  depends only on the locations at which  $W_i \in \{F, G\}$ , but not on the particular values of the observables at those locations.

Part (d) of the test and Lemma 8 imply that the observables  $W(a)$  satisfy approximate linearity conditions  $W(a)W(a') \approx W(a + a')$ , on average over a uniformly random choice of  $W \in \Sigma^n$  and  $a, a' \in \{0, 1\}^n$ . Using the form (16) for  $W$  and the fact that the  $\hat{\tau}_W(a)$  satisfy the linearity relations by definition, we deduce that  $\Lambda_{W(a)}\Lambda_{W(a')} \approx \Lambda_{W(a+a')}$  as well. Using the analysis of the Pauli braiding test (Lemma 8), this implies that for each  $i$  and  $W_i$  there is an observable  $\Lambda_{i, W_i}$  such that the  $\Lambda_{i, W_i}$  pairwise commute and  $\Lambda_W \approx \prod_i \Lambda_{i, W_i}$ . Using the preceding observations,  $\Lambda_{i, W_i} \approx \Lambda_i$  if  $W_i \in \{F, G\}$ , and  $\Lambda_{i, W_i} \approx \text{Id}$  if  $W_i \in \{X, Y, Z\}$ .

Success in part (e) of the test implies the condition  $\mathbb{E}_W \langle \psi | W \otimes W_\Phi | \psi \rangle \geq 1 - O(\varepsilon)$ , where  $W$  is distributed as in the test, and  $W_\Phi$  is the observable applied by the second player upon a query  $W$ , with some locations, indexed by pairs in  $S$  and  $T$ , have been replaced by the  $\Phi$  symbol (as described in the test). Let  $U$  be the set of  $i$  such that  $W_i \in \{F, G\}$ . Since  $\Delta_Y$  commutes with all observables in play, for clarity let us assume in the following that  $\Delta_Y = \text{Id}$ . From the decomposition of the observables  $W$  obtained so far and the analysis of the test BELL given in Lemma 30 it follows that

$$W \simeq \left( \otimes_i \hat{\tau}_{W_i} \right) \otimes \left( \prod_{i \in U} \Lambda_i \right), \quad \text{and} \quad W_\Phi \simeq \left( \otimes_{i \notin S \cup T} \hat{\tau}_{W_i} \right) \otimes \left( \otimes_{(i,j) \in S \cup T} \text{SW}_{i,j} \right) \otimes \left( \prod_{i \in U \setminus S \cup T} \Lambda_i \right),$$

where the ordering of tensor products does not respect the ordering of qubits, but it should be clear which registers each operator acts on. Using that for any operators  $A, B$  and  $\Delta$ ,

$$\langle \text{EPR} |^{\otimes 2} (A \otimes B \otimes |\Phi_{00}\rangle\langle\Phi_{00}|) | \text{EPR} \rangle^{\otimes 2} = \frac{1}{8} \text{Tr}(AB^T),$$

the above conditions imply

$$\mathbb{E}_{S=\{(s_i, s'_i)\}} \mathbb{E}_{T=\{(t_i, t'_i)\}} \Lambda_{s_i} \Lambda_{s'_i} \Lambda_{t_i} \Lambda_{t'_i} \approx \text{Id},$$

where the expectation is taken over sets  $S$  and  $T$  specified as in part (e), for a given  $W$ , and on average over the choice of  $W$ . Let  $\Lambda = \mathbb{E}_i \Lambda_i$ . By an averaging argument it follows that for  $U$  the set of locations such that  $W_i \in \{F, G\}$ ,  $\prod_{i \in U} \Lambda_i \approx \Lambda^{|S|}$ , again on average over the choice of  $W$ . To conclude we let  $\Lambda_F = \Lambda/|\Lambda|$ , which is an observable and satisfies the required conditions.  $\square$

### 3.5 Post-measurement states

We give a first corollary of Theorem 10 which expresses its conclusion (15) in terms of the post-measurement state of the first player. This corollary will be used in the analysis of the leash protocol from Section 4. To obtain a useful result we would like to “lift” the phase ambiguity  $\Lambda_W$  which remains in the statement of Theorem 10 (in contrast to the ambiguity  $\Delta_Y$ , which itself cannot be lifted solely by examining correlations). This ambiguity means that the provers have the liberty of choosing to report opposite outcomes whenever they apply an  $F$  or  $G$  observable, but they have to be consistent between themselves and across all of their qubits in doing so. To verify that the provers use the “right” labeling for their outcomes we incorporate a small tomography test in the test, described in Figure 9. Note that an inconvenience of the tomography is that the test no longer achieves perfect completeness (although completeness remains exponentially close to 1).

---

Test RIGID( $\Sigma, m$ ):

- Input: An integer  $m$  and a subset  $\Sigma = \{X, Y, Z, F, G\}$  of the single-qubit Clifford group.
  - Test: execute each of the following with equal probability:
    - (a) Execute the test CLIFF( $\Sigma, m$ );
    - (b) Send each player a uniformly random query  $W, W' \in \Sigma^m$ . Let  $T \subseteq \{1, \dots, m\}$  be the subset of positions  $i$  such that  $W_i \in \{X, Y\}$  and  $W'_i \in \{F, G\}$ . Reject if the fraction of answers  $(a_i, b_i)$ , for  $i \in T$ , from the provers that satisfy the CHSH correlations (i.e.  $a_i \neq b_i$  if and only if  $(W_i, W'_i) = (X, F)$ ) is not at least  $\cos^2 \frac{\pi}{8} - 0.1$ .
- 

Figure 9: The  $n$ -qubit rigidity test, RIGID( $\Sigma, m$ ).

For an observable  $W \in \Sigma$ , let  $\sigma_W = \sigma_W^{+1} - \sigma_W^{-1}$  be its eigendecomposition, where  $\sigma_W$  are the “honest” Pauli matrices defined in (1) and (2). For  $u \in \{\pm 1\}$  let  $\sigma_{W,+} = \sigma_W^u$  for  $W \in \Sigma$ , and

$$\sigma_{X,-}^u = \sigma_X^u, \quad \sigma_{Z,-}^u = \sigma_Z^u, \quad \sigma_{Y,-}^u = \sigma_Y^{-u}, \quad \sigma_{F,-}^u = \sigma_G^{-u}, \quad \sigma_{G,-}^u = \sigma_F^{-u}.$$

**Corollary 11.** *Let  $\varepsilon > 0$  and  $m$  an integer. Suppose a strategy for the players succeeds with probability  $1 - \varepsilon$  in test RIGID( $\Sigma, m$ ). Then for  $D \in \{A, B\}$  there exists an isometry*

$$V_D : \mathcal{H}_D \rightarrow (\mathbb{C}^2)_{D'}^{\otimes m} \otimes \mathcal{H}_{\hat{D}}$$

such that

$$\|(V_A \otimes V_B)|\psi\rangle_{AB} - |\text{EPR}\rangle^{\otimes m} \otimes |\text{AUX}\rangle_{\hat{A}\hat{B}}\|^2 = O(\sqrt{\varepsilon}), \quad (17)$$



and positive semidefinite matrices  $\tau_\lambda$  on  $\hat{\mathbf{A}}$  with orthogonal support, for  $\lambda \in \{+, -\}$ , such that  $\text{Tr}(\tau_+) + \text{Tr}(\tau_-) = 1$  and

$$\begin{aligned} \mathbb{E}_{W \in \Sigma^m} \sum_{u \in \{\pm 1\}^m} \left\| V_A \text{Tr}_B((\text{Id}_A \otimes W_B^u) |\psi\rangle\langle\psi|_{AB} (\text{Id}_A \otimes W_B^u)^\dagger) V_A^\dagger - \sum_{\lambda \in \{\pm\}} \left( \bigotimes_{i=1}^m \frac{\sigma_{W_i, \lambda}^{u_i}}{2} \right) \otimes \tau_\lambda \right\|_1 \\ = O(\text{poly}(\varepsilon)). \end{aligned} \quad (18)$$

Moreover, players employing the honest strategy succeed with probability  $1 - e^{-\Omega(m)}$  in the test.

*Proof.* From Theorem 10 we get isometries  $V_A, V_B$  and commuting observables  $\Delta_Y, \Lambda_F$  on  $\mathcal{H}_{\hat{\mathbf{A}}}$  such that the conclusions of the theorem hold. Write the eigendecomposition  $\Delta_Y = \Delta_Y^+ - \Delta_Y^-$  and  $\Lambda_F = \Lambda_F^+ - \Lambda_F^-$ . For  $\lambda \in \{+, -\}^2$  let

$$\tau_\lambda = \text{Tr}_B((\text{Id}_{\hat{\mathbf{A}}} \otimes \Delta_Y^{\lambda_1} \Lambda_F^{\lambda_2}) |AUX\rangle\langle AUX| (\text{Id}_{\hat{\mathbf{A}}} \otimes \Delta_Y^{\lambda_1} \Lambda_F^{\lambda_2})).$$

Using that  $\Delta_Y$  and  $\Lambda_F$  commute and satisfy

$$\Delta_Y \otimes \Delta_Y |AUX\rangle \approx \Lambda_F \otimes \Lambda_F |AUX\rangle \approx |AUX\rangle$$

it follows that the (sub-normalized) densities  $\tau_\lambda$  have (approximately) orthogonal support. In particular the provers' strategy in part (b) of the test is well-approximated by a mixture of four strategies, labeled by  $(\lambda_Y, \lambda_F) \in \{\pm 1\}^2$ , such that the strategy with label  $(\lambda_Y, \lambda_F)$  uses the observables

$$(X, Z, Y, F, G) = \left( \sigma_X, \sigma_Z, \lambda_Y \sigma_Y, \frac{1}{\sqrt{2}} \lambda_F (-\sigma_X + \lambda_Y \sigma_Y), \frac{1}{\sqrt{2}} \lambda_F (\sigma_X + \lambda_Y \sigma_Y) \right).$$

Among these four strategies, the two with  $\lambda_F = -1$  fail part (b) of the test with probability exponentially close to 1. Success in both parts of the test with probability at least  $1 - 2\varepsilon$  each thus implies

$$\text{Tr}(\tau_{+-}) + \text{Tr}(\tau_{--}) = \text{poly}(\varepsilon). \quad (19)$$

For  $W \in \Sigma^m$  and  $c \in \{0, 1\}^m$  the observable  $W(c) = \bigotimes_i W_i^{c_i}$  can be expanded in terms of a  $2^m$ -outcome projective measurement  $\{W^u\}$  as

$$W(c) = \sum_{u \in \{0, 1\}^m} (-1)^{u \cdot c} W^u.$$

Similarly, by definition the projective measurement associated with the commuting Pauli observables  $\tau_W(c) = \bigotimes_i \tau_{W_i}^{c_i}$ ,  $c \in \{0, 1\}^m$ , is

$$\tau_W^u = \bigotimes_i \left( \mathbb{E}_{c \in \{0, 1\}^m} (-1)^{u \cdot c} \tau_W(c) \right).$$

Thus,

$$\begin{aligned} \mathbb{E}_{c \in \{0, 1\}^m} \left\| \text{Id}_A \otimes (W(c) - V_B^\dagger \tau_W(c) V_B) |\psi\rangle_{AB} \right\|^2 \\ = \mathbb{E}_{c \in \{0, 1\}^m} \left\| \sum_u (-1)^{u \cdot c} \text{Id}_A \otimes (W^u - V_B^\dagger \tau_W^u V_B) |\psi\rangle_{AB} \right\|^2 \\ = \sum_{u \in \{0, 1\}^m} \left\| \text{Id}_A \otimes (W^u - V_B^\dagger \tau_W^u V_B) |\psi\rangle_{AB} \right\|^2, \end{aligned} \quad (20)$$

where the third line is obtained by expanding the square and using  $\mathbb{E}_{c \in \{0, 1\}^m} (-1)^{v \cdot c} = 1$  if  $v = 0^m$ , and 0 otherwise. Using (15), the expression in (20), when averaged over all  $W \in \Sigma^m$ , is bounded by  $O(\text{poly}(\varepsilon))$ .

Using the Fuchs-van de Graaf inequality and the fact that trace distance cannot increase under tracing out, we get that the following is  $O(\text{poly}(\varepsilon))$ :

$$\mathbb{E}_{W \in \Sigma^m} \sum_u \left\| V_A \text{Tr}_B((\text{Id}_A \otimes W^u) |\psi\rangle\langle\psi| (\text{Id}_A \otimes W^u)^\dagger) V_A^\dagger - \text{Tr}_B((\text{Id}_A \otimes \tau_W^u) |\psi\rangle\langle\psi| (\text{Id}_A \otimes \tau_W^u)^\dagger) \right\|_1. \quad (21)$$

Using that  $\tau_X = \sigma_X$ ,  $\tau_Z = \sigma_Z$ , and  $\tau_Y = \sigma_Y \Delta_Y$ , we deduce the post-measurement states for  $u \in \{\pm 1\}$

$$\tau_X^u = \sigma_X^u, \quad \tau_Z^u = \sigma_Z^u, \quad \tau_Y^u = \sigma_Y^u \otimes (\tau_{++} + \tau_{+-}) + \sigma_Y^{-u} \otimes (\tau_{-+} + \tau_{--}).$$

Similarly, from  $\tau_F = (-\tau_X + \tau_Y) \Lambda_F$  and  $\tau_G = (\tau_X + \tau_Y) \Lambda_F$  we get that e.g. the  $+1$  eigenspace of  $\tau_F$  is the combination of:

- The simultaneous  $+1$  eigenspace of  $\sigma_F = (-\sigma_X + \sigma_Y)/\sqrt{2}$ ,  $+1$  eigenspace of  $\Delta_Y$ , and  $+1$  eigenspace of  $\Lambda_F$ ;
- The simultaneous  $-1$  eigenspace of  $\sigma_F$ ,  $+1$  eigenspace of  $\Delta_Y$ , and  $-1$  eigenspace of  $\Lambda_F$ ;
- The simultaneous  $-1$  eigenspace of  $\sigma_G = -(-\sigma_X - \sigma_Y)/\sqrt{2}$ ,  $-1$  eigenspace of  $\Delta_Y$ , and  $+1$  eigenspace of  $\Lambda_F$ ;
- The simultaneous  $+1$  eigenspace of  $\sigma_G$ ,  $-1$  eigenspace of  $\Delta_Y$ , and  $-1$  eigenspace of  $\Lambda_F$ .

Proceeding similarly with  $\tau_G$ , we obtain

$$\begin{aligned} \tau_F^u &= \sigma_F^u \otimes \tau_{++} + \sigma_F^{-u} \otimes \tau_{+-} + \sigma_G^{-u} \otimes \tau_{-+} + \sigma_G^u \otimes \tau_{--}, \\ \tau_G^u &= \sigma_G^u \otimes \tau_{++} + \sigma_G^{-u} \otimes \tau_{+-} + \sigma_F^{-u} \otimes \tau_{-+} + \sigma_F^u \otimes \tau_{--}. \end{aligned}$$

Starting from (21) and using (14) we obtain

$$\begin{aligned} \mathbb{E}_{W \in \Sigma^m} \sum_u \left\| V_A \text{Tr}_B((\text{Id}_A \otimes W^u) |\psi\rangle\langle\psi| (\text{Id}_A \otimes W^u)^\dagger) V_A^\dagger \right. \\ \left. - \text{Tr}_B((\text{Id}_A \otimes \tau_W^u) |\text{EPR}\rangle\langle\text{EPR}|^{\otimes m} \otimes |\text{AUX}\rangle\langle\text{AUX}|_{\text{AB}} (\text{Id}_A \otimes \tau_W^u)^\dagger) \right\|_1 = O(\text{poly}(\varepsilon)). \end{aligned}$$

Since  $\text{Tr}_B(\text{Id} \otimes B |\text{EPR}\rangle\langle\text{EPR}|_{\text{AB}} \text{Id} \otimes B^\dagger) = (B^\dagger B)^T / 2$  for any single-qubit operator  $B$ , to conclude the bound claimed in the theorem it only remains to apply the calculations above and use (19) to eliminate the contribution of  $\tau_{+-}$  and  $\tau_{-+}$ ; the factor  $\frac{1}{2}$  comes from the reduced density matrix of an EPR pair.  $\square$

### 3.6 Tomography

Theorem 10 and Corollary 11 show that success in test  $\text{RIGID}(\Sigma, m)$  gives us control over the players' observables and post-measurement states in the test. This allows us to use one of the players to perform some kind of limited tomography (limited to post-measurement states obtained from measurements in  $\Sigma$ ), that will be useful for our analysis of the Dog-Walker Protocol from Section 5.

Let  $1 \leq m' \leq m$  and consider the test  $\text{TOM}(\Sigma, m', m)$  described in Figure 10. In this test, one player is sent a question  $W \in \Sigma^m$  chosen uniformly at random. Assuming the players are also successful in the test  $\text{RIGID}(\Sigma, m)$  (which can be checked independently, with some probability), using that the input distribution  $\mu$  in  $\text{RIGID}(\Sigma, m)$  assigns weight at least  $|\Sigma|^{-m}/2$  to any  $W' \in \Sigma^m$ , from Corollary 11 it follows that the second player's post-measurement state is close to a state consistent with the first player's reported outcomes. Now suppose the second player is sent a random subset  $S \subseteq [m]$  of size  $|S| = m'$ , and is allowed to report an arbitrary string  $W' \in \Sigma^{m'}$ , together with outcomes  $u$ . Suppose also that for each  $i \in S$ , we require that  $u_i = a_i$  whenever  $W'_i = W_i$ . Since the latter condition is satisfied by a constant fraction of  $i \in \{1, \dots, m'\}$ , irrespective of  $W'$ , with very high probability, it follows that the only possibility for the second player to satisfy the condition is to actually measure his qubits precisely in the basis that he indicates. This allows us

to check that a player performs the measurement that he claims, even if the player has the choice of which measurement to report.

---

Tomography Test  $\text{TOM}(\Sigma, m', m)$ :

- Input: Integer  $1 \leq m' \leq m$  and a subset  $\Sigma = \{X, Y, Z, F, G\}$  of the single-qubit Clifford group.
  - Test: Let  $S \subseteq [m]$  be chosen uniformly at random among all sets of size  $|S| = m'$ . Select  $W \in \Sigma^m$  uniformly at random. Send  $W$  to the first player, and the set  $S$  to the second. Receive  $a$  from the first player, and  $W' \in \Sigma^{m'}$  and  $u$  from the second. Accept only if  $a_i = u_i$  whenever  $i \in S$  and  $W_i = W'_i$ .
- 

Figure 10: The  $m$ -qubit tomography test  $\text{TOM}(\Sigma, m', m)$ .

**Corollary 12.** *Let  $\varepsilon > 0$  and  $1 \leq m' \leq m$  integer. Suppose a strategy for the players succeeds with probability  $1 - \varepsilon$  in both tests  $\text{RIGID}(\Sigma, m)$  (Figure 9) and  $\text{TOM}(\Sigma, m', m)$  (Figure 10). Let  $V_A, V_B$  be the isometries specified in Corollary 11. Let  $\{Q^{W',u}\}$  be the projective measurement applied by the second player in  $\text{TOM}(\Sigma, m', m)$ . Then there exists a distribution  $q$  on  $\Sigma^{m'} \times \{\pm\}$  such that*

$$\sum_{W' \in \Sigma^{m'}} \sum_{u \in \{\pm 1\}^{m'}} \left\| \text{Tr}_{AB}((\text{Id}_A \otimes V_B Q^{W',u}) |\psi\rangle\langle\psi|_{AB} (\text{Id}_A \otimes V_B Q^{W',u})^\dagger) \right. \\ \left. - \sum_{\lambda \in \{\pm\}} q(W', \lambda) \left( \bigotimes_{i=1}^{m'} \frac{1}{2} \sigma_{W'_i, \lambda}^{u_i} \right) \right\|_1 = O(\text{poly}(\varepsilon)),$$

where the notation is the same as in Corollary 11.

Moreover, players employing the honest strategy succeed with probability 1 in tomography part of the test.

*Proof.* Success in  $\text{RIGID}(\Sigma, m)$  allows us to apply Corollary 11. For any  $(W', u)$  let  $\rho_{A', \lambda}^{W', u}$  be the post-measurement state on the first player's space, conditioned on the second player's answer in test  $\text{TOM}(\Sigma, m', m)$  being  $(W', u)$ , after application of the isometry  $V_A$ , and conditioned on  $\mathcal{H}_A$  being in a state that lies in the support of  $\tau_\lambda$  (note this makes sense since  $\tau_+, \tau_-$  have orthogonal support). Using that for any  $i \in S$ ,  $W_i = W'_i$  with constant probability  $|\Sigma|^{-1}$ , it follows from (17) and (18) in Corollary 11 that success in  $\text{TOM}(\Sigma, m)$  implies the condition

$$\mathbb{E}_{\substack{S \subseteq \{1, \dots, m\} \\ |S| = m'}} \sum_{W', \lambda, u} \text{Tr}(\tau_\lambda) \text{Tr} \left( \left( \frac{|\Sigma| - 1}{|\Sigma|} \text{Id} + \frac{1}{|\Sigma|} \bigotimes_{i \in S} \sigma_{W'_i, \lambda}^{u_i} \right) \rho_{A', \lambda}^{W', u} \right) = 1 - O(\text{poly}(\varepsilon)). \quad (22)$$

Eq (22) concludes the proof, for some distribution  $q(W', \lambda) \approx \sum_u \text{Tr}(\rho_{A', \lambda}^{W', u}) \text{Tr}(\tau_\lambda)$  (the approximation is due to the fact that the latter expression only specifies a distribution up to error  $O(\text{poly}(\varepsilon))$ ).  $\square$

## 4 The Verifier-on-a-Leash Protocol

### 4.1 Protocol and statement of results

The Verifier-on-a-Leash Protocol (or “Leash Protocol” for short) involves a classical verifier and two quantum provers. The idea behind the Leash Protocol is to have a first prover, nicknamed PV for Prover V, carry out the quantum part of  $V_{\text{EPR}}$  from Broadbent's EPR Protocol by implementing the procedure  $V_{\text{EPR}}^r$ .

(See Section 2.3 for a summary of the protocol and a description of  $V_{EPR}$ . Throughout this section we assume that the circuit  $Q$  provided as input is compiled in the format described in Section 2.3.). A second prover, nicknamed PP for Prover  $P$ , will play the part of the prover  $P_{EPR}$ . Unlike in the EPR Protocol, the interaction with PV (i.e. running  $V_{EPR}^r$ ) will take place first, and PV will be asked to perform random measurements from the set  $\Sigma = \{X, Y, Z, F, G\}$ . The values  $\vec{z}$ , rather than being chosen at random, will be chosen based on the corresponding choice of observable. We let  $n$  be the number of input bits and  $t$  number of T gates in  $Q$ .

The protocol is divided into two sub-games; which game is played is chosen by the verifier by flipping a biased coin with probability  $(p_r, p_d = 1 - p_r)$ .

- The first game is a sequential version of the rigidity game  $\text{RIGID}(\Sigma, m)$  (from Section 3) described in Figure 16. This aims to enforce that PV performs precisely the right measurements;
- The second game is the delegation game, described in Figures 13, 14, and 15, and whose structure is summarized in Figure 11. Here the verifier guides PP through the computation in a similar way as in the EPR Protocol.

We call the resulting protocol the Leash Protocol with parameters  $(p_r, p_d)$ . In both sub-games the parameter  $m = \Theta(n + t)$  is chosen large enough so that with probability close to 1 each symbol in  $\Sigma$  appears in a random  $W \in \Sigma^m$  at least  $n + t$  times. It is important that PV is not able to tell which kind of game is being played. Notice also that in order to ensure blindness, we will require that the interaction with PV in the delegation game is sequential (more details on this are found in Section 4.4). In order for the two sub-games to be indistinguishable, we also require that the rigidity game  $\text{RIGID}(\Sigma, m)$  be played sequentially (i.e. certain subsets of questions and answers are exchanged sequentially, but the acceptance condition in the test is the same). Note, importantly, that the rigidity guarantees of  $\text{RIGID}(\Sigma, m)$  from Section 3 hold verbatim when the game is played sequentially, since this only reduces the number of ways that the provers can cheat. The following theorem states the guarantees of the Leash Protocol.

**Theorem 13.** *There are constants  $p_r, p_d = 1 - p_r$ , and  $\Delta > 0$  such that the following hold of the Verifier-on-a-Leash Protocol with parameters  $(p_r, p_d)$ , when executed on an input  $(Q, |\vec{x}\rangle)$ .*

- (Completeness:) *Suppose that  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \geq 2/3$ . Then there is a strategy for PV and PP that is accepted with probability at least  $p_{\text{compl}} = p_r(1 - e^{-\Omega(n+t)}) + 8p_d/9$ .*
- (Soundness:) *Suppose that  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \leq 1/3$ . Then any strategy for PV and PP is accepted with probability at most  $p_{\text{sound}} = p_{\text{compl}} - \Delta$ .*

*Further, the protocol leaks no information about  $\vec{x}$  to either prover individually, aside from an upper bound on the length of  $\vec{x}$ .*

The proof of the completeness property is given in Lemma 14. The soundness property is shown in Lemma 17. Blindness is established in Section 4.4. We first give a detailed description of the protocol. We start by describing the delegation game, specified in Figures 13, 14 and 15, which describe the protocol from the verifier's view, an honest PV's view, and an honest PP's view respectively. This will motivate the need for a sequential version of the game  $\text{RIGID}(\Sigma, m)$ , described in Figure 16. As we will show, the rigidity game forces PV to behave honestly. Thus, for the purpose of exposition, we assume for now that PV behaves honestly, which results in the joint behavior of PV and V being similar to that of the verifier  $V_{EPR}$  in the EPR Protocol.

From the rigidity game we may also assume that PV and PP share  $m$  EPR pairs, labeled  $\{1, \dots, m\}$ , for  $m = \Theta(n + t)$ . We will assume that the circuit  $Q$  is broken into  $d$  layers,  $Q = Q_1 \dots Q_d$ , such that in every  $Q_\ell$ , each wire has at most one T gate applied to it, after which no other gates are applied to that wire. We will suppose the T gates are indexed from 1 to  $t$ , in order of layer.

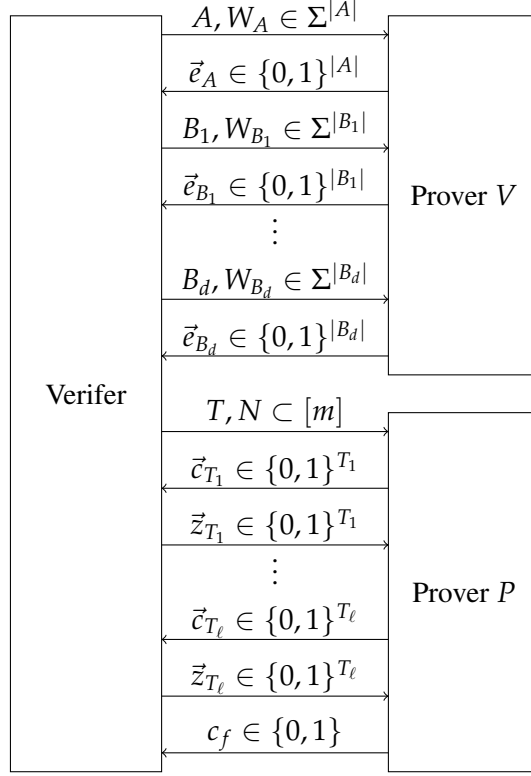


Figure 11: Structure of the delegation game.

The protocol begins with an interaction between the verifier and PV. The verifier selects a uniformly random partition  $A, B_1, \dots, B_d$  of  $\{1, \dots, m\}$ , with  $|A| = \Theta(n)$ , and for every  $\ell \in \{1, \dots, d\}$ ,  $|B_\ell| = \Theta(t_\ell)$ , where  $t_\ell$  is the number of T gates in  $Q_\ell$ . The verifier also selects a uniformly random  $W \in \Sigma^m$ , and partitions it into substrings  $W_A$  and  $W_{B_1}, \dots, W_{B_d}$ , meant to contain observables to initialize the computation qubits and auxiliary qubits for each layer of T gates respectively. The verifier instructs PV to measure his halves of the EPR pairs using the observables  $W_A$  first, and then  $W_{B_1}, \dots, W_{B_d}$ , sequentially. Upon being instructed to measure a set of observables, PV measures the corresponding half-EPR pairs and returns the results  $\vec{e}$  to the verifier. Breaking this interaction into multiple rounds is meant to enforce that, for example, the results output by PV upon receiving  $W_{B_\ell}$ , which we call  $\vec{e}_{B_\ell}$ , cannot depend on the choice of observables  $W_{B_{\ell+1}}$ . This is required for blindness.

Once the interaction with PV has been completed, as in the EPR Protocol, V selects one of three round types: computation ( $r = 0$ ), X-test ( $r = 1$ ), and Z-test ( $r = 2$ ). The verifier selects a subset  $N \subset A$  of size  $n$  of qubits to play the role of inputs to the computation. These are chosen from the subset of  $A$  corresponding to wires that PV has measured in the appropriate observable for the round type (see Table 4). For example, in an X-test round, PV's EPR halves corresponding to input wires should be measured in the Z basis so that PP is left with a one-time pad of the state  $|0\rangle^{\otimes n}$ , so in an X-test round, the computation wires are chosen from the set  $\{i \in A : W_i = Z\}$ . The input wires  $N$  are labeled by  $\mathcal{X}_1, \dots, \mathcal{X}_n$ .

The verifier also chooses subsets  $T_\ell = T_\ell^0 \cup T_\ell^1 \subset B_\ell$  of sizes  $t_{\ell,0}$  and  $t_{\ell,1} = t_\ell - t_{\ell,0}$  respectively, where  $t_{\ell,0}$  is the number of odd T gates in the  $\ell$ -th layer of  $Q$  (recall the definition of even and odd T gates from Section 2.3). The wires  $T_\ell^0$  and  $T_\ell^1$  will play the role of auxiliary states used to perform T gates from the  $\ell$ -th layer. They are chosen from those wires from  $B_\ell$  whose corresponding EPR halves have been measured in a correct basis, depending on the round type. For example, in an X-test round, the auxiliaries corresponding to odd T gates should be prepared by measuring the corresponding EPR half in either the X or Y basis (see

	Computation Round	X-test Round	Z-test Round
$N$ , input/computation qubits	$\{i \in A : W_i = Z\}$	$\{i \in A : W_i = Z\}$	$\{i \in A : W_i = X\}$
$T_\ell^0$ , even T gate auxiliaries	$\{i \in B_\ell : W_i \in \{G, F\}\}$	$\{i \in B_\ell : W_i = Z\}$	$\{i \in B_\ell : W_i \in \{X, Y\}\}$
$T_\ell^1$ , odd T gate auxiliaries	$\{i \in B_\ell : W_i \in \{G, F\}\}$	$\{i \in B_\ell : W_i \in \{X, Y\}\}$	$\{i \in B_\ell : W_i = Z\}$

Table 4: How the verifier chooses index sets  $T = T^0 \cup T^1$  and  $N$  for each type of round. These index sets determine which of the  $m$  systems are labeled by  $\{\mathcal{T}_i\}_{i=1}^t$  and  $\{\mathcal{X}_j\}_{j=1}^n$ , respectively.

Table 3), so in an X-test round,  $T_\ell^1$  is chosen from  $\{i \in B_\ell : W_i \in \{X, Y\}\}$  (see Table 4). We will let  $\mathcal{T}_1, \dots, \mathcal{T}_t$  label those EPR pairs that will be used as auxiliary states. In particular, the system  $\mathcal{T}_i$  will be used for the  $i$ -th T gate in the circuit, so if the  $i$ -th T gate is even,  $\mathcal{T}_i$  should be chosen from  $T^0 = \cup_\ell T_\ell^0$ , and otherwise it should be chosen from  $T_1 = \cup_\ell T_\ell^1$ . The verifier sends labels  $\mathcal{T}_1, \dots, \mathcal{T}_t$  and  $\mathcal{X}_1, \dots, \mathcal{X}_n$  to PP, who will act as  $P_{EPR}$  on the  $n + t$  qubits specified by these labels.

Just as in the EPR Protocol, the input on PP's system specified by  $\mathcal{X}_1, \dots, \mathcal{X}_n$  is a quantum one-time pad of either  $|\vec{x}\rangle$ ,  $|0\rangle^{\otimes n}$ , or  $|+\rangle^{\otimes n}$ , depending on the round type, with V holding the keys (determined by  $\vec{e}$ ). Throughout the interaction, PP always maintains a one-time pad of the current state of the computation, with the verifier in possession of the one-time-pad keys. The verifier updates her keys as the computation is carried out, using the rules in Table 2.

From PP's perspective, the protocol works just as the EPR Protocol, except that he does not receive the bit  $z_i$  needed to implement the T gadget until *during* the T gadget, after he has sent V his measurement result  $c_i$  (see Figure 12).

To perform the  $i$ -th T gate on the  $j$ -th wire, PP performs the circuit shown in Figure 12. As Figure 12 shows, PV has already applied the observable specified by V to his half of the EPR pair. The T gadget requires interaction with the verifier, to compute the bit  $z_i$ , which depends on the measured  $c_i$ , the value  $W_i$ , and one-time-pad key  $a_j$ , however, this interaction can be done in parallel for T gates in the same layer.

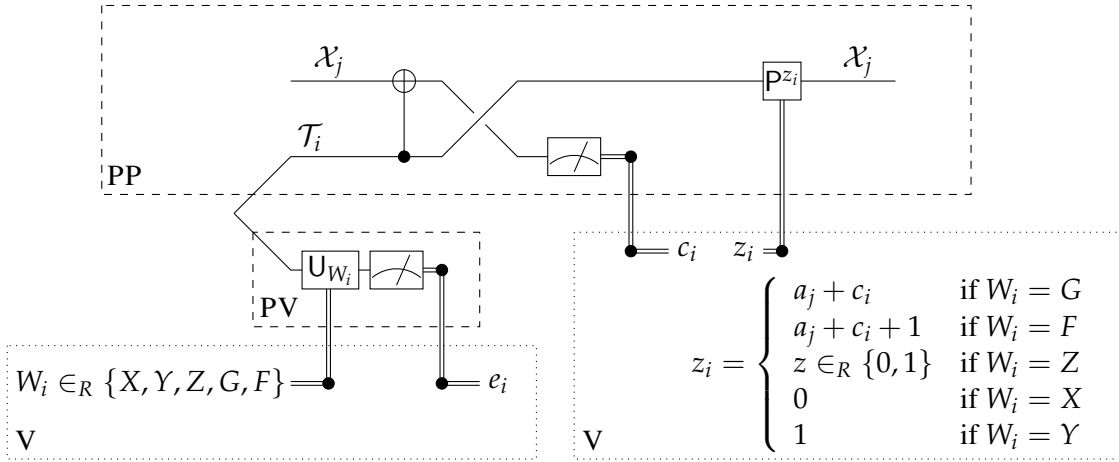


Figure 12: The gadget for implementing the  $i$ -th T gate, on the  $j$ -th wire.

It is simple to check that the T gadget in Figure 12 is the same as the T gadget for the EPR Protocol shown in Figure 1. In the case of the leash protocol,  $W$  is chosen at random, and then  $\vec{z}$  is chosen accordingly, whereas in the case of the EPR Protocol,  $\vec{z}$  is chosen at random and then  $W$  is chosen accordingly.

We now give the precise protocols for V (Figure 13) and honest provers PV (Figure 14) and PP (Figure 15). To see how these fit together, refer to Figure 11. For a string  $\vec{v}$  and an ordered set of indices  $I$ , we denote by  $\vec{v}_I$  the substring corresponding to indices in  $I$ .

---

Let  $(Q, \vec{x})$  be the input to the verifier, where  $Q$  is compiled in the form described in Section 2.3. Let  $n$  be the size of the input to  $Q$ . Let  $d$  be the T-depth, and for  $\ell \in \{1, \dots, d\}$  let  $t_\ell$  be the number of T gates in the  $\ell$ -th layer.

1. The verifier selects  $W \in_R \Sigma^m$ . She partitions  $\{1, \dots, m\}$  arbitrarily into non-overlapping subsets  $A$  and  $B_1, \dots, B_d$  such that  $W_A$  contains at least  $n$  copies of each symbol in  $\Sigma$ , and for each  $\ell \in \{1, \dots, d\}$ ,  $W_{B_\ell}$  contains at least  $t_\ell$  copies of each symbol in  $\Sigma$ .
2. The verifier sends  $A, W_A$  to PV, who returns  $\vec{e}_A \in \{0, 1\}^A$ . The verifier sequentially sends  $(B_1, W_{B_1}), \dots, (B_d, W_{B_d})$  to PV, each time receiving  $\vec{e}_{B_\ell} \in \{0, 1\}^{B_\ell}$  as answer.
3. The verifier selects a round type uniformly at random. She selects sets  $N \subseteq A$  and  $T_\ell \subseteq B_\ell$ , for  $\ell \in \{1, \dots, d\}$ , of sizes  $|N| = n$  and  $|T_\ell| = t_\ell$ , as follows:

**Computation Round:**  $N$  is chosen at random from  $\{i \in A : W_i = Z\}$ .  $T_\ell$  is chosen at random from  $\{i \in B_\ell : W_i \in \{G, F\}\}$ . She sets  $\vec{a} = \vec{e}_N + \vec{x}$  and  $\vec{b} = 0^n$ .

**X-test Round:**  $N$  is chosen at random from  $\{i \in A : W_i = Z\}$ .  $T_\ell = T_\ell^0 \cup T_\ell^1$ , where  $T_\ell^0$  is of size  $t_{\ell,0}$  chosen at random from  $\{i \in B_\ell : W_i = Z\}$  and  $T_\ell^1$  is of size  $t_{\ell,1}$  chosen at random from  $\{i \in B_\ell : W_i \in \{X, Y\}\}$ . She sets  $\vec{a} = \vec{e}_N$  and  $\vec{b} = 0^n$ .

**Z-test Round:**  $N$  is chosen at random from  $\{i \in A : W_i = X\}$ .  $T_\ell = T_\ell^0 \cup T_\ell^1$ , where  $T_\ell^0$  is of size  $t_{\ell,0}$  chosen at random from  $\{i \in B_\ell : W_i \in \{X, Y\}\}$  and  $T_\ell^1$  is of size  $t_{\ell,1}$  chosen at random from  $\{i \in B_\ell : W_i = Z\}$ . She sets  $\vec{a} = 0^n$  and  $\vec{b} = \vec{e}_N$ .

The verifier sends the sets  $N$  and  $T_\ell^0, T_\ell^1$ , for all  $\ell \in \{1, \dots, d\}$ , to PP.

4. For  $\ell = 1, \dots, d$ ,
    - (a) For each Clifford gate in the  $\ell$ -th layer, perform the appropriate key update from Table 2.
    - (b) The verifier receives  $\vec{c} = \{c_i\}_{i \in T_\ell}$  from PP. If it's an X-test round and  $i \in T_\ell^0$ , or it's a Z-test round and  $i \in T_\ell^1$ , reject if  $c_i \neq a_j + e_i$ , where  $j$  is the wire to which the  $i$ -th T gate is applied.
    - (c) For each  $i \in T_\ell$ , the verifier computes  $\vec{z} = \{z_i\}_{i \in T_\ell}$  as follows:
 

**Computation Round**  $z_i = a_j + 1_{W_i=F} + c_i$  ;

**X-test Round** if  $i \in T_\ell^0$ ,  $z_i \in_R \{0, 1\}$ ; else if  $i \in T_\ell^1$ ,  $z_i = 1_{W_i=Y}$ ;

**Z-test Round** if  $i \in T_\ell^0$ ,  $z_i = 1_{W_i=Y}$ ; else if  $i \in T_\ell^1$ ,  $z_i \in_R \{0, 1\}$ .
    - (d) The verifier sends  $\vec{z}$  to PP and updates keys  $(a_j, b_j)$  for each wire  $j$  that had a T gate applied, according to Table 2.
  5. The verifier receives a bit  $c_f$  from PP. She outputs reject if it's a computation or X-test round and  $c_f + a_f \neq 0$ , where  $a_f$  is the final X-key on the output wire; and accept otherwise.
- 

Figure 13: The Delegation Game: Verifier's point of view.

- 
1. For  $\ell = 0, 1, \dots, d$ ,
    - (a) PV receives a string  $W_S \in \Sigma^S$ , for some subset  $S$  of  $\{1, \dots, m\}$ , from V.
    - (b) For  $i \in S$ , PV measures his half of the  $i$ -th EPR pair using the observable indicated by  $W_i$ , obtaining an outcome  $e_i \in \{0, 1\}$ .
    - (c) PV returns  $\vec{e}_S$  to V.
- 

Figure 14: The Delegation Game: Honest strategy for PV.

---

1. PP receives subsets  $N$  and  $T_\ell^0, T_\ell^1$  of  $\{1, \dots, m\}$ , for  $\ell \in \{1, \dots, d\}$ , from the verifier.
  2. For  $\ell = 1, \dots, d$ ,
    - (a) PP does the Clifford computations in the  $\ell$ -th layer.
    - (b) For each  $i \in T_\ell = T_\ell^0 \cup T_\ell^1$ , PP applies a CNOT from  $\mathcal{T}_i$  into the input register corresponding to the wire on which this T gate should be performed,  $\mathcal{X}_j$ , and measures this wire to get a value  $c_i$ . The register  $\mathcal{T}_i$  is relabeled  $\mathcal{X}_j$ . He sends  $\vec{c}_{T_\ell} = \{c_i\}_{i \in T_\ell}$  to V. (See Figure 12).
    - (c) PP receives  $\vec{z}_{T_\ell} = \{z_i\}_{i \in T_\ell}$  from V. For each  $i \in T_\ell$ , he applies  $P^{z_i}$  to the corresponding  $\mathcal{X}_j$ .
  3. PP performs the final computations that occur after the  $d$ -th layer of T gates, measures the output qubit,  $\mathcal{X}_1$ , and sends the resulting bit,  $c_f$ , to V.
- 

Figure 15: The Delegation Game: Honest strategy for PP.

---

Finally, we describe the sequential version of the game  $\text{RIGID}(\Sigma, m)$  in Figure 16. It is no different than  $\text{RIGID}(\Sigma, m)$ , except for the fact that certain subsets of questions and answers are exchanged sequentially, but the acceptance condition is the same. As mentioned earlier, running the game sequentially only reduces the provers' ability to cheat. Hence, the guarantees from  $\text{RIGID}(\Sigma, m)$  in Section 3 hold verbatim for the sequential version.

---

Let  $m, n$ , and  $t_1, \dots, t_d$  be parameters provided as input, such that  $m = \Theta(n + t_1 + \dots + t_d)$ .

1. The verifier selects questions  $W, W' \in \Sigma^m$ , for the first and second player respectively, according to the distribution of questions in the game  $\text{RIGID}(\Sigma, m)$ . She partitions  $\{1, \dots, m\}$  at random into subsets  $A$  and  $B_\ell$ , for  $\ell \in \{1, \dots, d\}$ , of size  $|A| = \Theta(n)$  and  $|B_\ell| = \Theta(t_\ell)$ , exactly as in Step 1 of the Delegation Game.
  2. The verifier sends  $(A, W_A), (B_1, W_{B_1}), \dots, (B_d, W_{B_d})$  and  $(A, W'_A), (B_1, W'_{B_1}), \dots, (B_d, W'_{B_d})$  in sequence to the first and second prover respectively. They sequentially return respectively  $\vec{e}_A \in \{0, 1\}^{|A|}$ ,  $\vec{e}_{B_1} \in \{0, 1\}^{|B_1|}, \dots, \vec{e}_{B_d} \in \{0, 1\}^{|B_d|}$  and  $\vec{e}'_A \in \{0, 1\}^{|A|}$ ,  $\vec{e}'_{B_1} \in \{0, 1\}^{|B_1|}, \dots, \vec{e}'_{B_d} \in \{0, 1\}^{|B_d|}$ .
  3. The verifier accepts if and only if  $\vec{e}, \vec{e}'$  and  $W, W'$  satisfy the winning condition of  $\text{RIGID}(\Sigma, m)$ .
- 

Figure 16: Sequential version of  $\text{RIGID}(\Sigma, m)$ .



## 4.2 Completeness

**Lemma 14.** *Suppose the verifier executes the rigidity game with probability  $p_r$  and the delegation game with probability  $p_d = 1 - p_r$ , on an input  $(Q, |\vec{x}\rangle)$  such that  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \geq 2/3$ . Then there is a strategy for the provers which is accepted with probability at least  $p_{\text{compl}} = p_r(1 - e^{-\Omega(n+t)}) + \frac{8}{9}p_d$ .*

*Proof.* The provers PV and PP play the rigidity game according to the honest strategy, and the delegation game as described in Figures 14 and 15 respectively. Their success probability in the delegation game is the same as the honest strategy in the EPR Protocol, which is at least  $\frac{2}{3} + \frac{2}{3}\frac{1}{3} = \frac{8}{9}$ , by Theorem 2 and since in our protocol the verifier chooses each of the three types of rounds uniformly.  $\square$

## 4.3 Soundness

We divide the soundness analysis into three parts. First we analyze the case of an honest PV, and a cheating PP (Lemma 15). Then we show that if PV and PP pass the rigidity game with almost optimal probability, then one can construct new provers PV' and PP', with PV' honest, such that the probability that they are accepted in the delegation game is not changed by much (Lemma 16). In Lemma 17, we combine the previous to derive the desired constant soundness-completeness gap, where we exclude that the acceptance probability of the provers in the rigidity game is too low by picking a  $p_r$  large enough.

**Lemma 15** (Soundness against PP). *Suppose the verifier executes the delegation game on input  $(Q, |\vec{x}\rangle)$  such that  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \leq 1/3$  with provers  $(PV, PP^*)$  such that PV plays the honest strategy. Then the verifier accepts with probability at most  $7/9$ .*

*Proof.* Let  $PP^*$  be any prover. Assume that PV behaves honestly and applies the measurements specified by his query  $W$  on halves of EPR pairs shared with  $PP^*$ . As a result the corresponding half-EPR pair at  $PP^*$  is projected onto the post-measurement state associated with the outcome reported by PV to V.

From  $PP^*$ , we define another prover,  $P^*$ , such that if  $P^*$  interacts with  $V_{EPR}$ , the honest verifier for the EPR Protocol (Figure 3), then  $V_{EPR}$  rejects with the same probability that V would reject on interaction with  $PP^*$ . The main idea of the proof can be seen by looking at Figure 12, and noticing that: (1) the combined action of V and PV is unchanged if instead of choosing the  $W_i$ -values at random and then choosing  $z_i$  as a function of these, the  $z_i$  are chosen uniformly at random, and then the  $W_i$  are chosen as a function of these; and (2) with this transformation, the combined action of V and PV is now the same as the action of  $V_{EPR}$  in the EPR Protocol.

We now define  $P^*$ .  $P^*$  acts on a system that includes  $n + t$  qubits that, in an honest run of the EPR Protocol, are halves of EPR pairs shared with  $V_{EPR}$ .  $P^*$  receives  $\{z_i\}_{i=1}^t$  from  $V_{EPR}$ .  $P^*$  creates  $m - (n + t)$  half EPR pairs (i.e. single-qubit maximally mixed states) and randomly permutes these with his  $n + t$  unmeasured qubits,  $n$  of which correspond to computation qubits on systems  $\mathcal{X}_1, \dots, \mathcal{X}_n$  — he sets  $N$  to be the indices of these qubits — and  $t$  of which correspond to T-auxiliary states — he sets  $T^0$  and  $T^1$  to be the indices of these qubits.  $P^*$  simulates  $PP^*$  on these  $m$  qubits in the following way. First,  $P^*$  gives  $PP^*$  the index sets  $N$ ,  $T^0$ , and  $T^1$ . In the  $\ell$ -th iteration of the loop (Step 2. in Figure 15),  $PP^*$  returns some bits  $\{c_i\}_{i \in T_\ell}$ , and then expects inputs  $\{z_i\}_{i \in T_\ell}$ , which  $P^*$  provides, using the bits he received from  $V_{EPR}$ . Finally, at the end of the computation,  $PP^*$  returns a bit  $c_f$ , and  $P^*$  outputs  $\{c_i\}_{i \in T}$  and  $c_f$ .

This completes the description of  $P^*$ . To show the lemma we argue that for any input  $(Q, |\vec{x}\rangle)$  the probability that V outputs accept on interaction with PV and  $PP^*$  is the same as the probability that  $V_{EPR}$  outputs accept on interaction with  $P^*$ , which is at most  $\frac{2}{3}q_t + \frac{1}{3}q_c$  whenever  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \leq 1/3$ , by Theorem 3. Using  $\delta = \frac{1}{3}$ , Theorem 3 gives  $q_c \leq \frac{5}{3} - \frac{4}{3}q_t$ , which yields

$$\frac{2}{3}q_t + \frac{1}{3}q_c \leq \frac{5}{9} + \frac{2}{9}q_t \leq \frac{7}{9}.$$

There are two reasons that  $V_{EPR}$  might reject: (1) in a computation or  $X$ -test round, the output qubit decodes to 1; or (2) in an evaluation of the gadget in Figure 12 (either an  $X$ -test round for an even  $T$  gate, or a  $Z$ -test round for an odd  $T$  gate) the condition  $c_i = a_j \oplus e_i$  fails.

We first consider case (1). This occurs exactly when  $c_f \oplus a_f = 1$ , where  $a_f$  is the final  $X$  key of the output wire, held by  $V_{EPR}$ . We note that  $a_f$  is exactly the final  $X$  key that  $V$  would hold in the Verifier-on-a-Leash Protocol, which follows from the fact that the update rules in both the EPR Protocol and the leash protocol are the same. Thus, the probability that  $V_{EPR}$  finds  $v_f \oplus a_f = 1$  on interaction with  $P^*$  is exactly the probability that  $V$  finds  $c_f \oplus a_f = 1$  in Step 5 of Figure 13.

Next, consider case (2). The condition  $c_i \neq a_j \oplus e_i$  is exactly the condition in which a verifier interacting with  $P^*$  as in Figure 13 would reject (see Step 4.(b)).

Thus, the probability that  $V_{EPR}$  outputs reject upon interaction with  $P^*$  is exactly the probability that  $V$  outputs reject on interaction with  $PP^*$ , which, as discussed above, is at most  $7/9$ .  $\square$

The following lemma shows soundness against cheating  $PV^*$ .

**Lemma 16.** *Suppose the verifier executes the leash protocol on input  $(Q, |\vec{x}\rangle)$  such that  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \leq 1/3$  with provers  $(PV^*, PP^*)$ , such that the provers are accepted with probability  $1 - \varepsilon$ , for some  $\varepsilon > 0$ , in the rigidity game, and with probability at least  $q$  in the delegation game. Then there exist provers  $PP'$  and  $PV'$  such that  $PV'$  applies the honest strategy and  $PP'$  and  $PV'$  are accepted with probability at least  $q - \text{poly}(\varepsilon)$  in the delegation game.*

*Proof.* By assumption,  $PP^*$  and  $PV^*$  are accepted in the rigidity game with probability at least  $1 - \varepsilon$ . Let  $V_A, V_B$  be the local isometries guaranteed to exist by Corollary 11, and  $\{\tau_\lambda\}$  the sub-normalized densities associated with  $PP^*$ 's Hilbert space (recall that playing the rigidity game sequentially leaves the guarantees from Corollary 11 unchanged, since it only reduces the provers' ability to cheat).

First define provers  $PV''$  and  $PP''$  as follows.  $PP''$  and  $PV''$  initially share the state

$$|\psi'\rangle_{AB} = \otimes_{i=1}^m |\text{EPR}\rangle_{AB} \otimes \sum_{\lambda \in \{\pm\}} |\lambda\rangle_{A'} \langle \lambda|_{A'} \otimes |\lambda\rangle_{B'} \langle \lambda|_{B'} \otimes (\tau_\lambda)_{A''},$$

with registers  $AA'A''$  in the possession of  $PP''$  and  $BB'$  in the possession of  $PV''$ . Upon receiving a query  $W \in \Sigma^m$ ,  $PV''$  measures  $B'$  to obtain a  $\lambda \in \{\pm\}$ . If  $\lambda = +$  he proceeds honestly, measuring his half-EPR pairs exactly as instructed. If  $\lambda = -$  he proceeds honestly except that for every honest single-qubit observable specified by  $W$ , he instead measures the complex conjugate observable. Note that this strategy can be implemented irrespective of whether  $W$  is given at once, as in the game RIGID, or sequentially, as in the Delegation Game.  $PP''$  simply acts like  $PP^*$ , just with the isometry  $V_A$  applied.

First note that by Corollary 11, the distribution of answers of  $PV''$  to the verifier, as well as the subsequent interaction between the verifier and  $PP$ , generate (classical) transcripts that are within statistical distance  $\text{poly}(\varepsilon)$  from those generated by  $PV^*$  and  $PP^*$  with the same verifier.

Next we observe that taking the complex conjugate of both provers' actions does not change their acceptance probability in the delegation game, since the interaction with the verifier is completely classical. Define  $PP'$  as follows:  $PP'$  measures  $A'$  to obtain the same  $\lambda$  as  $PV''$ , and then executes  $PP''$  or its complex conjugate depending on the value of  $\lambda$ . Define  $PV'$  to execute the honest behavior (he measures to obtain  $\lambda$ , but then discards it and does not take any complex conjugates).

Then  $PV'$  applies the honest strategy, and  $(PV', PP')$  applies either the same strategy as  $(PV'', PP'')$  (if  $\lambda = +$ ) or its complex conjugate (if  $\lambda = -$ ). Therefore they are accepted in the delegation game with exactly the same probability.  $\square$

Combining Lemma 15 and Lemma 16 gives us the final soundness guarantee.

**Lemma 17.** (*Constant soundness-completeness gap*) *There exist constants  $p_r, p_d = 1 - p_r$  and  $\Delta > 0$  such that if the verifier executes the leash protocol with parameters  $(p_r, p_d)$  on input  $(Q, |\vec{x}\rangle)$  such that  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \leq 1/3$ , then any provers  $(PV^*, PP^*)$  are accepted with probability at most  $p_{\text{sound}} = p_{\text{compl}} - \Delta$ .*

*Proof.* Suppose provers  $PP^*$  and  $PV^*$  succeed in the delegation game with probability  $\frac{7}{9} + w$  for some  $w > 0$ , and the testing game with probability  $1 - \varepsilon_*(w)$ , where  $\varepsilon_*(w)$  will be specified below. By Lemma 16, this implies that there exist provers  $PP'$  and  $PV'$  such that  $PV'$  is honest and the provers succeed in the delegation game with probability at least  $\frac{7}{9} + w - g(\varepsilon_*(w))$ , where  $g(\varepsilon) = \text{poly}(\varepsilon)$  is the function from the guarantee of Lemma 16. Let  $\varepsilon_*(w)$  be such that  $g(\varepsilon_*(w)) \leq \frac{w}{2}$ . In particular,  $\frac{7}{9} + w - g(\varepsilon_*(w)) \geq \frac{7}{9} + \frac{w}{2} > \frac{7}{9}$ . This contradicts Lemma 15.

Thus if provers  $PP$  and  $PV$  succeed in the delegation game with probability  $\frac{7}{9} + w$  they must succeed in the rigidity game with probability less than  $1 - \varepsilon_*(w)$ . This implies that for any strategy of the provers, on any *no* instance, the probability that they are accepted is at most

$$\max \left\{ p_r + (1 - p_r) \left( \frac{7}{9} + \frac{1}{18} \right), p_r \left( 1 - \varepsilon_* \left( \frac{1}{18} \right) \right) + (1 - p_r) \cdot 1 \right\}. \quad (23)$$

Since  $\varepsilon_*(\frac{1}{18})$  is a positive constant, it is clear that one can pick  $p_r$  large enough so that

$$p_r \left( 1 - \varepsilon_* \left( \frac{1}{18} \right) \right) + (1 - p_r) \cdot 1 < p_r + (1 - p_r) \left( \frac{7}{9} + \frac{1}{18} \right). \quad (24)$$

Select the smallest such  $p_r$ . Then the probability that the two provers are accepted is at most

$$\begin{aligned} p_{\text{sound}} &:= p_r + (1 - p_r) \left( \frac{7}{9} + \frac{1}{18} \right) \\ &< p_r (1 - e^{-\Omega(n+t)}) + (1 - p_r) \frac{8}{9} \\ &= p_{\text{compl}}, \end{aligned}$$

which gives the desired constant completeness-soundness gap  $\Delta$ . □

#### 4.4 Blindness

In this subsection, we establish blindness of the Verifier-on-a-Leash Protocol. We will actually prove that the protocol presented in this section has the property that neither prover can learn anything about the input to the circuit,  $\vec{x}$ , aside from its length. Thus, our protocol can be turned into a blind protocol, where  $Q$  is also hidden, by modifying any input  $(Q, \vec{x})$  where  $Q$  has  $g$  gates and acts on  $n$  qubits, to an input  $(U_{g,n}, (Q, \vec{x}))$ , where  $U_{g,n}$  is a universal circuit that takes as input a description of a  $g$ -gate circuit  $Q$  on  $n$  qubits, and an input  $\vec{x}$ , and outputs  $Q|\vec{x}\rangle$ . The universal circuit  $U_{g,n}$  can be implemented in  $O(g \log n)$  gates. Running the Leash Protocol on  $(U_{g,n}, (Q, \vec{x}))$  reveals nothing about  $Q$  or  $\vec{x}$  aside from  $g$  and  $n$ .

In the form presented in Figure 13, the verifier  $V$  interacts first with  $PV$ , sending him random questions that are independent from the input  $\vec{x}$ , aside from the input length  $n$ . It is thus clear that the protocol is blind with respect to  $PV$ .

In contrast, the questions to  $PP$  depend on  $PV$ 's answers and on the input, so it may a priori seem like the questions can leak information to  $PP$ . To show that the protocol is also blind with respect to  $PP$ , we show that there is an alternative formulation, in which the verifier first interacts with  $PP$ , sending him random messages, and then only with  $PV$ , with whom the interaction is now adaptive. We argue that, for an arbitrary strategy of the provers, the reduced state of all registers available to either prover,  $PP$  or  $PV$ , is exactly the

same in both formulations of the protocol — the *original* and the *alternative* one. This establishes blindness for both provers. This technique for proving blindness is already used in [RUV13] to establish blindness of a two-prover protocol based on computation by teleportation.

**Lemma 18** (Blindness of the Leash Protocol). *For any strategy of  $PV^*$  and  $PP^*$ , the reduced state of  $PV^*$  (resp.  $PP^*$ ) at the end of the leash protocol is independent of the input  $\vec{x}$ , aside from its length.*

*Proof.* Let  $PV^*$  and  $PP^*$  denote two arbitrary strategies for the provers in the leash protocol. Each of these strategies can be modeled as a super-operator

$$\mathcal{T}_{PV} : L(\mathcal{H}_{T_{PV}} \otimes \mathcal{H}_{PV}) \rightarrow L(\mathcal{H}_{T'_{PV}} \otimes \mathcal{H}_{PV}), \quad \mathcal{T}_{PP,ad} : L(\mathcal{H}_{T_{PP}} \otimes \mathcal{H}_{PP}) \rightarrow L(\mathcal{H}_{T'_{PP}} \otimes \mathcal{H}_{PP})$$

respectively. Here  $\mathcal{H}_{T_{PV}}$  and  $\mathcal{H}_{T'_{PV}}$  (resp.  $\mathcal{H}_{T_{PP}}$  and  $\mathcal{H}_{T'_{PP}}$ ) are classical registers containing the inputs and outputs to and from  $PV^*$  (resp.  $PP^*$ ), and  $\mathcal{H}_{PV}$  (resp.  $\mathcal{H}_{PP}$ ) is the private space of  $PV^*$  (resp.  $PP^*$ ). Note that the interaction of each prover with the verifier is sequential, and we use  $\mathcal{T}_{PV}$  and  $\mathcal{T}_{PP,ad}$  to denote the combined action of the prover and the verifier across all rounds of interaction (formally these are sequences of superoperators).

Consider an alternative protocol, which proceeds as follows. The verifier first interacts with PP. From Figure 15 we see that the inputs required for PP are subsets  $N$  and  $T_1, \dots, T_d$ , and values  $\{z_i\}_{i \in T_\ell}$  for each  $\ell \in \{1, \dots, d\}$ . To select the former, the verifier proceeds as in the first step of the Delegation Game. She selects the latter uniformly at random. The verifier collects values  $\{c_i\}_{i \in T_\ell}$  from PP exactly as in the original Delegation Game.

Once the interaction with PP has been completed, the verifier interacts with PV. First, she selects a random string  $W_N \in \Sigma^N$ , conditioned on the event that  $W_N$  contains at least  $n$  copies of each symbol in  $\Sigma$ , and sends it to PV, collecting answers  $\vec{e}_N$ . The verifier then follows the same update rules as in the delegation game. We describe this explicitly for computation rounds. First, the verifier sets  $\vec{a} = \vec{e}_N$ . Depending on the values  $\{c_i\}_{i \in T_1}$  and  $\{z_i\}_{i \in T_1}$  obtained in the interaction with PP, using the equation  $z_i = a_j + 1_{W_i=F} + c_i$  she deduces a value for  $1_{W_i=F}$  for each  $i \in T_1 \subseteq B_1$ . She then selects a uniformly random  $W_{B_1} \in \Sigma^{B_1}$ , conditioned on the event that  $W_{B_1}$  contains at least  $t_1$  copies of each symbol from  $\Sigma$ , and for  $i \in T_1$  it holds that  $W_i = F$  if and only if  $z_i = a_j + 1 + c_i$ . The important observation is that, if  $T_1$  is a uniformly random, unknown subset, the marginal distribution on  $W_{B_1}$  induced by the distribution described above is independent of whether  $z_i = a_j + 1 + c_i$  or  $z_i = a_j + 0 + c_i$ : precisely, it is uniform conditioned on the event that  $W_{B_1}$  contains at least  $t_1$  copies of each symbol from  $\Sigma$ . The verifier receives outcomes  $\vec{e}_{B_1} \in \{0, 1\}^{B_1}$  from PV, and using these outcomes performs the appropriate key update rules; she then proceeds to the second layer of the circuit, until the end of the computation. Finally, the verifier accepts using the same rule as in the last step of the original delegation game.

We claim that both executions of the protocol, original and alternative, generate the same joint final state:

$$\mathcal{T}_{PP,ad} \circ \mathcal{T}_{PV}(\rho_{orig}) = \mathcal{T}_{PV,ad} \circ \mathcal{T}_{PP}(\rho_{alt}) \in \mathcal{H}_{PP} \otimes \mathcal{H}_{T'_{PP}} \otimes \mathcal{H}_V \otimes \mathcal{H}_{T'_{PV}} \otimes \mathcal{H}_{PV}, \quad (25)$$

where we use  $\rho_{orig}$  and  $\rho_{alt}$  to denote the joint initial state of the provers, as well as the verifier's initialization of her workspace, in the original and alternative protocols respectively, and  $\mathcal{T}_{PV,ad}$  and  $\mathcal{T}_{PP}$  are the equivalent of  $\mathcal{T}_{PV}$  and  $\mathcal{T}_{PP,ad}$  for the reversed protocol (in particular they correspond to the same strategies  $PV^*$  and  $PP^*$  used to define  $\mathcal{T}_{PV}$  and  $\mathcal{T}_{PP,ad}$ ). Notice that  $\mathcal{T}_{PV,ad}$  and  $\mathcal{T}_{PP}$  are well-defined since neither prover can distinguish between an execution of the original and the alternative protocol.<sup>5</sup> To see that the equality (25) holds, it is possible to re-write the final state of the protocol as the result of the following sequence of operations. First, the verifier initializes the message registers with  $PP^*$  and  $PV^*$  using half-EPR pairs,

<sup>5</sup>One must ensure that a prover does not realize if the alternative protocol is executed instead of the original; this is easily enforced by only interacting with any of the provers at specific, publicly decided times.

keeping the other halves in her private workspace. This simulates the generation of uniformly random messages to both provers. Then, the superoperator  $\mathcal{T}_{PV} \otimes \mathcal{T}_{PP}$  is executed. Finally, the verifier post-selects by applying a projection operator on  $\mathcal{H}_{T_{PV}} \otimes \mathcal{H}_{T'_{PV}} \otimes \mathcal{H}_{T_{PP}} \otimes \mathcal{H}_{T'_{PP}}$  which projects onto valid transcripts for the original protocol (i.e. transcripts in which the adaptive questions are chosen correctly). This projection can be implemented in two equivalent ways: either the verifier first measures  $\mathcal{H}_{T_{PV}} \otimes \mathcal{H}_{T'_{PV}}$ , and then  $\mathcal{H}_{T_{PP}} \otimes \mathcal{H}_{T'_{PP}}$ ; based on the outcomes she accepts a valid transcript for the original protocol or she rejects. Or, she first measures  $\mathcal{H}_{T_{PP}} \otimes \mathcal{H}_{T'_{PP}}$ , and then  $\mathcal{H}_{T_{PV}} \otimes \mathcal{H}_{T'_{PV}}$ ; based on the outcomes she accepts a valid transcript for the alternative protocol or she rejects. Using the commutation of the provers' actions, conditioned on the transcript being accepted, the first gives rise to the first final state in (25), and the second to the second final state. The two are equivalent because the acceptance condition for a valid transcript is identical in the two versions of the protocol.

Since in the first case the reduced state on  $\mathcal{H}_{T'_{PV}} \otimes \mathcal{H}_{PV}$  is independent of the input to the computation,  $\vec{x}$ , and in the second the reduced state on  $\mathcal{H}_{PP} \otimes \mathcal{H}_{T'_{PP}}$  is independent of  $\vec{x}$ , we deduce that the protocol is hides the input from each of  $PV^*$  and  $PP^*$ .  $\square$

## 5 The Dog-Walker Protocol

### 5.1 Protocol and statement of results

The Dog-Walker Protocol again involves a classical verifier and two provers PV and PP. As in the leash protocol presented in Section 4, PP and PV take the roles of  $P_{EPR}$  and  $V_{EPR}$  from [Bro15] respectively. The main difference is that the Dog-Walker Protocol gives up blindness in order to reduce the number of rounds to two (one round of interaction with each prover, played sequentially). After an initial round of communication with PP, who returns a sequence of measurement outcomes, the verifier V communicates all of PP's outcomes, except for the one corresponding to the output bit of the computation, as well as the input  $\vec{x}$  of the computation, to PV. Using these, PV can perform the required adaptive measurements without the need to interact with the verifier. It may seem quite risky to communicate bits sent by PP directly to PV — this seems to allow for communication between the two provers! Indeed, blindness is lost. However, if PP is honest, his outcomes  $\{c_i\}_i$  in the computation round are the result of measurements he performs on half-EPR pairs, and are uniform random bits. If he is dishonest, and does not return the outcomes obtained by performing the right measurements, he will be caught in the test rounds. It is only in computation rounds that V sends the measurement results  $\{c_i\}_i$  to PV.

To guarantee that PV behaves honestly, we combine the rigidity test  $\text{RIGID}(\Sigma, m)$  from Section 3.5 with the tomography test  $\text{TOM}(\Sigma, n + t, m)$  from Section 3.6. Part of the latter test requires PV to announce what measurements he has performed and the corresponding outcomes he obtained. He has to do so honestly in order to pass the test.

Throughout this section we let  $\Sigma = \{X, Y, Z, F, G\}$ , and let  $m = \Theta(n + t)$  be chosen large enough so that each symbol in  $\Sigma$  appears at least  $n + t$  times in a uniform random  $W \in \Sigma^m$ , with probability close to 1. Let  $\mu(W)$  denote the probability that a player receives input  $W$  while playing  $\text{RIGID}(\Sigma, m)$  (recall that both players have the same marginals in  $\text{RIGID}$ ). Let  $\mu(W'|W)$  denote the probability that one player receives  $W'$  given that the other player receives  $W$ .

The full protocols are presented in Figure 17 (verifier's point of view), Figure 19 (PV's point of view) and Figure 18 (PP's point of view). The protocol has two types of rounds: EPR and Rigidity. Within an EPR round are three types of sub-rounds: Computation sub-round, X-test sub-round, and Z-test sub-round. We will generally think of X- and Z-test sub-rounds as one sub-round type (Test sub-round). Within a Rigidity round are two types of sub-rounds: Tomography sub-round, which should be thought of as the Rigidity version of the EPR-Computation round; and Clifford sub-round, which should be thought of as the Rigidity version of the EPR-Test round. With some probability  $p_1$ , V runs a Rigidity round, Clifford sub-round; with

some probability  $p_2$ , V runs an EPR round, Test sub-round; with some probability  $p_3$ , V runs an EPR round, Computation sub-round; and with probability  $p_4 = 1 - p_1 - p_2 - p_3$ , V runs a Rigidity round, Tomography sub-round. We call this the Dog-Walker Protocol with parameters  $(p_1, p_2, p_3, p_4)$ .

The following theorem states the guarantees of the Dog-Walker Protocol.

**Theorem 19.** *There exist constants  $p_1, p_2, p_3, p_4 = 1 - p_1 - p_2 - p_3$ , and  $\Delta > 0$  such that the following hold of the Dog-Walker Protocol with parameters  $(p_1, p_2, p_3, p_4)$ , when executed on input  $(Q, |\vec{x}\rangle)$ .*

- (Completeness: ) *Suppose that  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \geq 2/3$ . Then there is a strategy for PV and PP that is accepted with probability at least  $p_{\text{compl}} = p_1(1 - e^{-\Omega(n+t)}) + p_2 + \frac{2}{3}p_3 + p_4$ .*
- (Soundness: ) *Suppose that  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \leq 1/3$ . Then any strategy for PV and PP is accepted with probability at most  $p_{\text{sound}} = p_{\text{compl}} - \Delta$ .*

The proof of the completeness property is given in Lemma 20, and the soundness property is given in Lemma 25.

---

1. Select a random round type **EPR** or **Rigidity**, and disjoint sets  $N, T^0, T^1 \subset \{1, \dots, m\}$  of sizes  $n, t_0$  and  $t - t_0$ .

**EPR** Choose  $\vec{z}$  uniformly at random from  $\{0, 1\}^t$  and send it, along with  $N, T^0$  and  $T^1$ , to PP. Receive measurement outcomes  $\vec{c} \in \{0, 1\}^t$  and  $c_f \in \{0, 1\}$  from PP.

**Rigidity** Choose  $W'$  according to  $\mu(\cdot)$  and send it to PP. Receive  $\vec{e}' \in \{0, 1\}^m$  from PP.

2. Select a sub-round type at random from **Computation**, **X Test** or **Z Test**.

**Computation** Based on whether it's an EPR or a Rigidity Round:

**EPR** (i) Send  $\vec{x}, \vec{z}, \vec{c}$  and sets  $N, T^0$  and  $T^1$  to PV, and receive measurement outcomes  $\vec{a}, \vec{b} \in \{0, 1\}^n$  and  $\vec{e} \in \{0, 1\}^t$ .

(ii) Apply the update rules from Table 2 gate-by-gate to obtain the final X key for the output wire  $a'_f$ . If  $c_f + a'_f \neq 0$ , reject.

**Rigidity (Tomography)** (i) Choose uniform random strings  $\vec{c}, \vec{z} \in \{0, 1\}^t, \vec{x} \in \{0, 1\}^n$  to send to PV, along with  $N$  and  $T$ , and receive measurement outcomes  $\vec{d} \in \{0, 1\}^n$  and  $\vec{e} \in \{0, 1\}^t$ .

(ii) From  $\vec{x}, \vec{c}, \vec{z}, \vec{d}$ , and  $\vec{e}$ , determine the adaptive measurements  $W \in \Sigma^{n+t}$  that  $V_{EPR}^0$  would have performed (based on Figure 4), and reject if the input-output pairs  $(W', \vec{e}')$  and  $(N \cup T, (W, \vec{e}))$  do not satisfy the winning criterion for  $\text{TOM}(\Sigma, n + t, m)$ .

**X-Test** Based on whether it's an EPR or a Rigidity Round:

**EPR** (i) Choose  $W \in \Sigma^m$  uniformly at random among all strings satisfying:  $W_i = Z$  for all  $i \in N$ ;  $W_i = Z$  for all  $i \in T^0$ ; and  $W_i \in \{X, Y\}$  for all  $i \in T^1$ . Send  $W$  to PV and receive measurement results  $\vec{e} \in \{0, 1\}^m$ . Let  $(\vec{a}, \vec{b}) = (\vec{e}_N, 0^n)$ .

(ii) Apply update rules from Table 2 gate-by-gate to obtain  $\forall i \in [t]$  the X key before the  $i$ -th T gate is applied,  $a'_i$ , and the final X key for the output wire,  $a'_f$ . If  $\exists i$  s.t. the  $i$ -th T gate is even and  $c_i \neq a'_i + e_i$ , reject. If  $c_f + a'_f \neq 0$ , reject.

**Rigidity (Clifford)** Choose  $W$  according to the marginal conditioned on  $W', \mu(\cdot|W')$ . Send  $W$  to PV and receive  $\vec{e} \in \{0, 1\}^m$ . Reject if  $(W', \vec{e}', W, \vec{e})$  doesn't win  $\text{RIGID}(\Sigma, m)$ .

**Z-Test** Based on whether it's an EPR or a Rigidity Round:

**EPR** (i) Choose  $W \in \Sigma^m$  uniformly at random among all strings satisfying:  $W_i = X$  for all  $i \in N$ ;  $W_i \in \{X, Y\}$  for all  $i \in T^0$ ; and  $W_i = Z$  for all  $i \in T^1$ . Send  $W$  to PV and receive measurement results  $\vec{e} \in \{0, 1\}^m$ . Let  $(\vec{a}, \vec{b}) = (0^n, \vec{e}_N)$ .

(ii) Apply update rules from Table 2 gate-by-gate to obtain  $\forall i \in [t]$ , the X key before the  $i$ -th T gate is applied,  $a'_i$ . If  $\exists i$  s.t. the  $i$ -th T gate is odd and  $c_i \neq a'_i + e_i$ , reject.

**Rigidity (Clifford)** Identical to X-Test case.

---

Figure 17: The Dog-Walker Protocol: Verifier's point of view.

- 
1. If PP receives a question  $W'$  from V (he is playing TOM or RIGID):

Measure the  $m$  qubits in the observable indicated by  $W'$  — for example, if  $W' \in \Sigma^m$ , for  $i \in \{1, \dots, m\}$ , measure the  $i$ -th qubit in the basis indicated by  $W'_i$  — and report the outcomes  $\vec{e}'$  to V.

2. If PP receives  $\vec{z}$ , and sets  $N$ ,  $T^0$  and  $T^1$  from V (he is playing the role of  $P_{EPR}$  from the EPR Protocol):

Run the prover  $P_{EPR}$  from Figure 5 on input  $\vec{z}$ , the  $n$  qubits in  $N$ , and the  $t$  qubits in  $T^0 \cup T^1$ . Report the outputs  $\vec{c} \in \{0, 1\}^t$  and  $c_f \in \{0, 1\}$  of  $P_{EPR}$  to V.

---

Figure 18: The Dog-Walker Protocol: Honest strategy for PP.

- 
1. If PV receives a question  $W$  from V (he is playing RIGID or an X- or Z-Test Round):

Measure the  $m$  qubits in the observable indicated by  $W$  — for example, if  $W \in \Sigma^m$ , for  $i \in \{1, \dots, m\}$ , measure the  $i$ -th qubit in the basis indicated by  $W_i$  — and report the outcomes  $\vec{e}$  to V.

2. If PV receives  $\vec{x}$ ,  $\vec{z}$ ,  $\vec{c}$  and sets  $N$ ,  $T^0$  and  $T^1$  from V (he is playing TOM or a Computation Round):

Run the procedure  $V_{EPR}^0$  from Figure 4 on input  $\vec{x}$ ,  $\vec{c}$ ,  $\vec{z}$ , the  $n$  qubits in  $N$ , and the  $t$  qubits in  $T^0 \cup T^1$ . Report the outputs  $\vec{d}$  and  $\vec{e}$  of  $V_{EPR}^0$  to V.

---

Figure 19: The Dog-Walker Protocol: Honest strategy for PV.

## 5.2 Completeness

**Lemma 20.** *Suppose V executes the Dog-Walker Protocol with parameters  $(p_1, p_2, p_3, p_4)$ . There is a strategy for the provers such that, on any input  $(Q, |\vec{x}\rangle)$  such that  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \geq \frac{2}{3}$ , V accepts with probability at least  $p_{\text{compl}} = p_1(1 - \delta_c) + p_2 + \frac{2}{3}p_3 + p_4$ , for some  $\delta_c = e^{-\Omega(n+t)}$ .*

*Proof.* The provers PV and PP play the strategy described in Figures 19 and 18 respectively. In the Rigidity-Tomography round, the verification performed by V amounts to playing  $\text{TOM}(\Sigma, n + t, m)$  with the provers (with an extra constraint on the output  $W$  of PV that is always satisfied by the honest strategy). This game has perfect completeness, which makes the V accept with probability 1 in the Rigidity-Tomography round. In the Rigidity-Clifford round, V plays  $\text{RIGID}(\Sigma, m)$  with the provers. The game has completeness at least  $1 - \delta_c$  for some  $\delta_c = e^{-\Omega(n+t)}$ , since  $m = \Omega(n + t)$ , therefore their success probability in this round is at least  $1 - \delta_c$ .

In the EPR round, the provers are exactly carrying out the EPR Protocol, with V using PV to run  $V_{EPR}^r$ , and PP playing the role of  $P_{EPR}$ . Thus, test rounds result in acceptance with probability 1, and the computation round results in acceptance with probability  $\|\Pi_0 Q |\vec{x}\rangle\|^2$ , by Theorem 2.  $\square$

## 5.3 Soundness

Figure 20 summarizes the high-level structure of the soundness analysis. Intuitively, our ultimate goal is to argue that both provers either apply the correct operations in EPR-Computation rounds, or are rejected



with constant probability. This will be achieved by employing a form of “hybrid argument” whereby it is argued that the provers, if they are not caught, must be using the honest strategies described in Figure 18 and Figure 19 in the different types of rounds considered in the protocol. Towards this, we divide the round types into the following four scenarios:

1. Ridity-Clifford: The round type is **Ridity** and the sub-round type is either **X-Test** or **Z-Test**. (When the provers are honest) PV behaves as in Item 1 of Figure 19, and PP behaves as in Item 1 of Figure 18.
2. EPR-Test: The round type is **EPR** and the sub-round type is either **X-Test** or **Z-Test**. PV behaves as in Item 1 of Figure 19, and PP behaves as in Item 2 of Figure 18.
3. EPR-Computation: The round type is **EPR** and the sub-round type is **Computation**. PV behaves as in Item 2 of Figure 19, and PP behaves as in Item 2 of Figure 18.
4. Ridity-Tomography: The round type is **Ridity** and the sub-round type is **Computation**. PV behaves as in Item 2 of Figure 19, and PP behaves as in Item 1 of Figure 18.

Examining Figure 17, we can see the following. In the Ridity-Clifford scenario, the verifier is precisely playing the game RIGID with the provers, as the provers receive questions  $W'$  and  $W$  distributed according to  $\mu(\cdot, \cdot)$ , the distribution of questions for  $\text{RIGID}(\Sigma, m)$ ; their answers are tested against the winning conditions of  $\text{RIGID}(\Sigma, m)$ . In the Ridity-Tomography scenario, the verifier plays a variant of the game TOM with the provers, in which PV’s choice of observable  $W$  is uniquely determined by his inputs  $\vec{x}$ ,  $\vec{c}$  and  $\vec{z}$ : it should match the observable implemented by  $V_{EPR}^0$  on these inputs. In EPR rounds, PV plays the part of  $V_{EPR}'$  from the EPR Protocol, and PP play the part of  $P_{EPR}$ . The EPR-Test scenario corresponds to X- and Z-tests from the EPR Protocol, whereas the EPR-Computation scenario corresponds to computation rounds from the EPR Protocol.

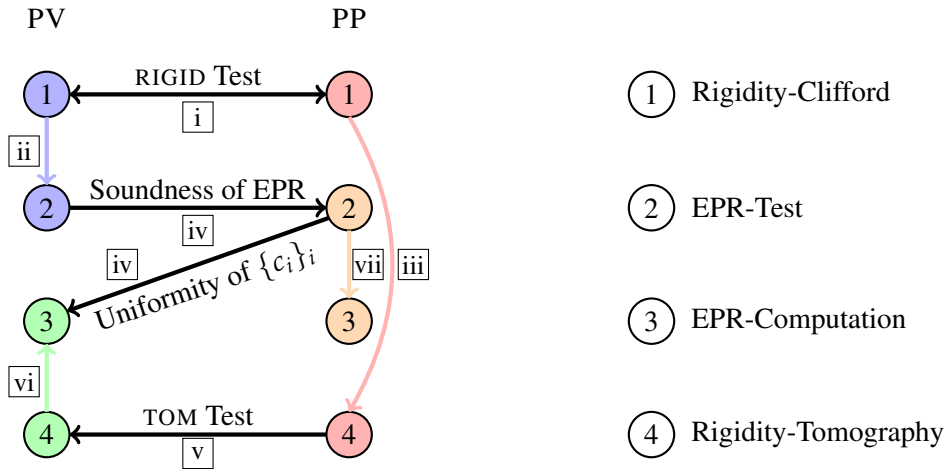


Figure 20: Overview of the soundness of the Dog-Walker Protocol

The structure of the proof is as follows (see also Figure 20):

- (i) By the game RIGID, in the Ridity-Clifford rounds, both PP and PV must be honest, or they would lose the game.
- (ii) Since PV can’t distinguish between Ridity-Clifford and EPR-Test (both are Figure 19 Item 1 from his perspective, and the input distributions, while not identical, are within constant total variation distance), PV must be honest in the EPR-Test rounds, by (i).

- (iii) Since PP can't distinguish between Rigidity-Clifford and Rigidity-Tomography (both are Figure 18 Item 1 from his perspective), PP must be honest in the Rigidity-Tomography rounds, by (i).
- (iv) Since PV is honest in EPR-Test rounds by (ii), PP must be honest in EPR-Test rounds or he will get caught, but in particular, he must output values  $\{c_i\}_{i \in [t]}$  that are uniform random and independent of  $\vec{z}$ . Since PP can't distinguish between EPR-Test and EPR-Computation rounds, this is also true in EPR-Computation rounds, when the verifier sends the values  $\{c_i\}_i$  to PV.
- (v) PV must be honest in Rigidity-Tomography rounds, or the provers would lose the game TOM.
- (vi) Since PV can't distinguish between Rigidity-Tomography rounds and EPR-Computation rounds (both are Figure 19 Item 2 from his perspective), PV must be honest in EPR-Computation rounds, by (v), and his input distribution to both rounds is within constant total variation distance, by (iv).
- (vii) Since PV is honest in EPR-Test rounds by (ii), and EPR-Computation rounds by (vi), the combined behavior of V and PV in the EPR rounds is that of  $V_{EPR}$  in the EPR Protocol, so by the soundness of the EPR Protocol, PP must be honest in EPR-Computation rounds, or get caught in the EPR-Test rounds with high probability.

The following lemma establishes (i), (ii) and (iii).

**Lemma 21.** *Suppose the verifier executes the Dog-Walker Protocol with provers  $(PV^*, PP^*)$  such that the provers are accepted with probability  $q_1 \geq 1 - \epsilon$  in the Rigidity-Clifford Round,  $q_2$  in the EPR-Test Round,  $q_3$  in the EPR-Computation Round, and  $q_4$  in the Rigidity-Tomography Round. Then there exist provers  $(PV', PP')$  such that:*

- *PV' and PP' both apply the honest strategy in the Rigidity-Clifford rounds, PV' applies the honest strategy in the EPR-Test rounds, and PP' applies the honest strategy in the Rigidity-Tomography rounds; in particular, the state shared by the provers at the beginning of the protocol is a tensor product of the honest state consisting of  $m$  shared EPR pairs and an arbitrary shared ancilla;*
- *The provers are accepted with probability  $q'_2 = q_2 - O(\text{poly}(\epsilon))$  in the EPR-Test Round,  $q'_3 = q_3$  in the EPR-Computation Round, and  $q'_4 = q_4 - O(\text{poly}(\epsilon))$  in the Rigidity-Tomography Round.*

*Proof.* Using a similar argument as in Lemma 16, the strategy of  $PV^*$  in Rigidity-Clifford rounds, which is also his strategy in EPR-Test rounds (Figure 19 Item 1); and the strategy of  $PP^*$  in Rigidity-Clifford rounds, which is also his strategy in Rigidity-Tomography rounds (Figure 18 Item 1); can both be replaced with the honest strategies. Since the distribution of inputs to  $PP^*$  in the Rigidity-Tomography rounds and Rigidity-Clifford rounds is the same, the success probability in the Rigidity-Tomography rounds is changed by at most  $O(\text{poly}(\epsilon))$  by using the honest strategy. On the other hand,  $PV^*$ 's input distribution in EPR-Test rounds is uniform on  $\Sigma^m$ , whereas his distribution in Rigidity-Clifford rounds is given by  $\mu$ . However, from the description of the test RIGID it is clear that for all  $W \in \Sigma^m$ ,  $\mu(W) \geq \frac{1}{c|\Sigma|^m}$  for some constant  $c > 1$ , thus the total variation distance between the two distributions is at most  $1 - \frac{1}{c}$ . Thus, replacing  $PV^*$  with the honest strategy in the EPR-Test rounds will change the success probability by at most  $O(\text{poly}(\epsilon))$ .

Finally, since the provers' strategy in the EPR-Computation round has not changed, the acceptance probability in it remains unchanged.  $\square$

Next, we will show that whenever  $PV^*$  is honest in the EPR-Test rounds this forces  $PP^*$  to output (close to) uniformly random  $\{c_i\}_{i \in [t]}$  that are independent of the round type, even given  $\vec{z}$ . This will allow us to verify that  $PP^*$  is unable to signal to  $PV^*$  whether the round is an EPR Round in the EPR-Computation round, when  $PV^*$  is sent  $\vec{z}$  and  $\vec{c}$ . This establishes (iv).

**Lemma 22.** *Suppose the verifier executes the Dog-Walker Protocol with provers  $(PV^*, PP^*)$  such that the initial shared state of the provers consists of  $m$  shared EPR pairs, together with an arbitrary shared auxiliary state;  $PV^*$  plays the honest strategy in the EPR-Test rounds; the provers are accepted with probability  $q_1$  in the Rigidity-Clifford Round,  $q_2 = 1 - \epsilon'$  in the EPR-Test Round,  $q_3$  in the EPR-Computation Round, and  $q_4$  in the Rigidity-Tomography Round. Then the input  $(\vec{c}, \vec{z})$  given by the verifier to  $PV^*$  in the EPR-Computation rounds has a distribution that is within  $O(\epsilon')$  total variation distance of uniform on  $\{0, 1\}^t \times \{0, 1\}^t$ .*

*Proof.* Let  $a'_i$  denote the X key of the wire to which the  $i$ -th T gate is applied, just before the  $i$ -th T gate is applied, and let  $D_i$  be a random variable defined as follows. If the  $i$ -th T gate is even, let  $D_i = e_i + a'_i$ , where we interpret  $e_i$  and  $a'_i$  as the random variables representing the measurement result and key V would get if she chooses to execute an X-Test round. If the  $i$ -th T gate is odd, let  $D_i = e_i + a'_i$ , where we interpret  $e_i$  and  $a'_i$  as the measurement result and key V would get if she chooses to execute an Z-Test round. Since  $PV^*$  is assumed to play honestly in EPR-Test rounds,  $\vec{D}$  is uniformly distributed in  $\{0, 1\}^t$ . In particular, we have, for any  $\vec{d}, \vec{z} \in \{0, 1\}^t$ ,

$$\Pr[\vec{D} = \vec{d}, \vec{Z} = \vec{z}] = \frac{1}{4^t}. \quad (26)$$

Let  $C_i$  be the random variable that corresponds to the measurement output of the  $i$ -th T gadget by  $PP^*$  in X-Test round if the  $i$ -th T gate is even, or the measurement output of the  $i$ -th T gadget by  $PP^*$  in Z-Test round if the  $i$ -th T gate is odd.

Let  $T^0 \subset [t]$  be the set of even T gates and  $T^1 \subset [t]$  the set of odd T gates. In an X-Test Round, the provers are rejected whenever  $i \in T^0$  and  $c_i \neq d_i$ , and in a Z-Test Round, they are rejected whenever  $i \in T^1$  and  $c_i \neq d_i$ . An EPR-Test Round consists of running one of these two rounds with equal probability, so:

$$\Pr[\vec{C} \neq \vec{D}] \leq 2\epsilon'. \quad (27)$$

We can express (27) as

$$\Pr[(\vec{C}, \vec{Z}) \neq (\vec{D}, \vec{Z})] \leq 2\epsilon'.$$

We conclude by using the easily verifiable fact that for any random variables  $X$  and  $Y$  such that  $\Pr[X = Y] \geq 1 - 2\epsilon'$ , the total variation distance between the marginal distributions on  $X$  and  $Y$  is at most  $2\epsilon'$ .  $\square$

Next, we can use the tomography test TOM to establish (v), and then the fact that by Lemma 22 the input to PV is not very different in EPR-Computation and Rigidity-Tomography rounds to establish (vi):

**Lemma 23.** *Suppose the verifier executes the Dog-Walker Protocol with provers  $(PV^*, PP^*)$  such that:  $PV^*$  applies the honest strategy in EPR-Test rounds;  $PP^*$  applies the honest strategy in the Rigidity-Tomography rounds; and the provers are accepted with probability  $q_1$  in the Rigidity-Clifford Round,  $q_2 = 1 - \epsilon'$  in the EPR-Test Round,  $q_3$  in the EPR-Computation Round, and  $q_4 = 1 - \epsilon$  in the Rigidity-Tomography Round. Then there exist provers  $(PV', PP')$  such that  $PV'$  applies the honest strategy in the Rigidity-Tomography rounds and EPR-Computation rounds,  $PP'$  applies the honest strategy in Rigidity-Tomography rounds, and the provers are accepted with probability  $q_1$  in the Rigidity-Clifford Round,  $q_2 = 1 - \epsilon'$  in the EPR-Test Round and  $q_3 - \text{poly}(\epsilon) - O(\epsilon')$  in the EPR-Computation round.*

*Proof.* The Rigidity-Tomography rounds can be seen as V playing the Tomography Game with the provers, except that whereas  $PV^*$  gets no non-trivial input in the Tomography Game, in the Rigidity-Tomography round, he gets random values  $\vec{c}$  and  $\vec{z}$  on which his strategy can depend. Fix  $\vec{x}$ , and let  $\{Q_{\vec{c}, \vec{z}}^u\}_u$  be the projective measurement that  $PV^*$  applies upon receiving  $\vec{c}, \vec{z}, \vec{x}$ , where  $u = (\vec{d}, \vec{e})$  is the string of outcomes obtained by PV on the  $n + t$  single-qubit measurements he is to perform according to Step 2 in Figure 19.

By Corollary 12, since the provers win the Rigidity-Tomography round with probability  $1 - \varepsilon$ , for every  $\vec{c}, \vec{z} \in \{0, 1\}^t$ , there exist distributions  $q_{\vec{c}, \vec{z}}$  on  $\Sigma^m \times \{\pm\}$  such that the following is  $O(\text{poly}(\varepsilon))$ :

$$\mathbb{E}_{\vec{c}, \vec{z}} \sum_{u \in \{0, 1\}^m} \left\| \text{Tr}_{\mathbf{A}, \mathbf{B}} \left( (\text{Id}_{\mathbf{A}} \otimes V_{\mathbf{B}} Q_{\vec{c}, \vec{z}}^u) |\psi\rangle\langle\psi|_{\mathbf{AB}} (\text{Id}_{\mathbf{A}} \otimes V_{\mathbf{B}} Q_{\vec{c}, \vec{z}}^u)^\dagger \right) - \sum_{\lambda \in \{\pm\}} q_{\vec{c}, \vec{z}}(W', \lambda) \left( \bigotimes_{i=1}^m \frac{\sigma_{W'_i, \lambda}^{u_i}}{2} \right) \right\|_1. \quad (28)$$

Here we use the notation from Corollary 11 and 12. The string  $W' = W(\vec{c}, \vec{z}, \vec{u}) \in \Sigma^m$  is uniquely determined by  $\vec{c}, \vec{z}$ , and the outcomes  $u$  reported by  $\text{PV}^*$ ; indeed it is using this string that  $\text{PV}^*$ 's answers are checked against the measurement outcomes obtained by  $\text{PP}^*$ , who by assumption applies the honest strategy. For any fixed  $(W', \lambda)$  the distribution on outcomes  $u$  obtained in the “honest” strategy represented by the right-hand side in (28) is uniform. Thus the outcomes  $u$  reported by  $\text{PV}^*$  are within  $\text{poly}(\varepsilon)$  of uniform. From this it follows that the joint distribution on transcripts  $(\vec{c}, \vec{z}, u, W' = W(\vec{c}, \vec{z}, u))$  that results from an interaction with  $\text{PV}^*$  is within statistical distance  $\text{poly}(\varepsilon)$  of the distribution generated by an interaction with the honest  $\text{PV}$ ; furthermore, by (28) the resulting post-measurement states on  $\text{PP}^*$  are also  $\text{poly}(\varepsilon)$  close to the honest ones, on average over this distribution.

We can now consider two provers  $\text{PV}'$  and  $\text{PP}'$  who, in Rigidity-Tomography rounds, first apply the isometries  $V_{\mathbf{A}}, V_{\mathbf{B}}$  from Corollary 12, then measure their auxiliary systems  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{B}}$  using  $\Delta_Y$ , obtaining a shared outcome  $\lambda \in \{\pm\}$ , and finally apply the honest strategy shown in Item 2 of Figure 19 ( $\lambda = +$ ) or its conjugate ( $\lambda = -$ ). Furthermore, conjugating the honest strategy produces exactly the same statistics as the honest strategy itself, so we may in fact assume that  $\text{PV}'$  and  $\text{PP}'$  both apply the honest strategy in Rigidity-Tomography rounds.

A consequence of  $\text{PV}'$  applying the honest strategy in Figure 19 Item 2 is that  $\text{PV}'$  also plays the honest strategy in EPR-Computation rounds. Since  $\text{PV}'$  is still honest in the EPR-Test round and  $q_2 = 1 - \varepsilon'$ , Lemma 22 implies that the distribution of the input to  $\text{PV}'$  in EPR-Computation rounds is within  $\text{poly}(\varepsilon) + O(\varepsilon')$  total variation distance of his input in Rigidity-Tomography rounds, therefore the provers' success probability in EPR-Computation rounds changes at most by  $\text{poly}(\varepsilon) + O(\varepsilon')$ .  $\square$

Finally, we show that if  $\text{PV}$  is honest,  $\text{PP}$  must be honest in EPR computation rounds, or the acceptance probability would be low, establishing (vii):

**Lemma 24.** *Suppose the verifier executes the Dog-Walker Protocol on an input  $(Q, |\vec{x}\rangle)$  such that  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \leq 1/3$  with provers  $(\text{PV}, \text{PP})$  such that  $\text{PV}$  plays the honest strategy. Let  $q_2$  be the provers' acceptance probability in EPR-Test rounds. Then the verifier accepts with probability at most  $p_1(1 - \delta_c) + p_2 q_2 + p_3(5/3 - 4q_2/3) + p_4$ .*

*Proof.* With probability  $p_2 + p_3$ ,  $V$  executes an EPR round, in which case, he executes EPR-Computation with probability  $\frac{p_3}{p_2 + p_3}$  and EPR-Test with probability  $\frac{p_2}{p_2 + p_3}$ . In the former case, since  $\text{PV}$  is honest, he is executing  $V_{\text{EPR}}^0$ . In fact, the behavior of an honest  $\text{PV}$  in the EPR-Test rounds is also that of  $V_{\text{EPR}}^r$ . Thus, the combined behavior of  $V$  and  $\text{PV}$  is that of  $V_{\text{EPR}}$ . Then the result follows from Theorem 3.  $\square$

We can now combine Lemmas 21, 23, and 24 to get the main result of this section, the “soundness” part of Theorem 19.

**Lemma 25** (Constant soundness-completeness gap). *There exist constants  $p_1, p_2, p_3, p_4 = 1 - p_1 - p_2 - p_3$  and  $\Delta > 0$  such that if the verifier executes the Dog-Walker Protocol with parameters  $(p_1, p_2, p_3, p_4)$  on input  $(Q, |\vec{x}\rangle)$  such that  $\|\Pi_0 Q |\vec{x}\rangle\|^2 \leq 1/3$ , then any provers  $(\text{PV}^*, \text{PP}^*)$  are accepted with probability at most  $p_{\text{sound}} = p_{\text{compl}} - \Delta$ .*

*Proof.* Suppose the provers  $\text{PV}^*$  and  $\text{PP}^*$  are such that the lowest acceptance probability in either the Rigidity-Clifford round or the Rigidity-Tomography round is  $1 - \varepsilon$ , and they are accepted with probability

$1 - \varepsilon'$  in the EPR-Test round, and with probability  $1/3 + w$  in the Computation Round. Applying Lemma 21 and Lemma 23 in sequence, we deduce the existence of provers  $(PV', PP')$  for which

$$\begin{aligned} q'_1 &= 1 - O(\delta_c), \\ q'_2 &= 1 - \varepsilon' - \text{poly}(\varepsilon), \\ q'_3 &= \frac{1}{3} + w - \text{poly}(\varepsilon) - O(\varepsilon'), \\ q'_4 &= 1, \end{aligned}$$

where  $q'_1, q'_2, q'_3$  and  $q'_4$  are their success probabilities in the four types of rounds, and  $1 - \delta_c$  is the completeness of the RIGID test; from Corollary 11 we have  $\delta_c = 2^{-\Omega(n+t)}$ . Moreover  $PV'$  applies the honest strategy in all rounds, while  $PP'$  applies the honest strategy in the Rigidity-Clifford and Rigidity-Tomography rounds. Applying Lemma 24, it follows that

$$w \leq O(\varepsilon') + \text{poly}(\varepsilon) + p_1 \cdot O(\delta_c).$$

Therefore the prover's overall success probability is at most

$$\begin{aligned} &\min(p_1, p_4)(1 - \varepsilon) + \max(p_1, p_4) + p_2(1 - \varepsilon') + p_3 \left( \frac{1}{3} + w \right) \\ &\leq p_{\text{compl}} - \left( \frac{p_3}{3} + \varepsilon' p_2 + \varepsilon \min(p_1, p_4) \right) + p_3 (O(\varepsilon') + \text{poly}(\varepsilon)) + (p_1 + p_3 p_1) \cdot O(\delta_c), \end{aligned}$$

where recall from Lemma 20 that  $p_{\text{compl}} = p_1(1 - \delta_c) + p_2 + p_4 + \frac{2}{3}p_3$ . Fixing  $p_2$  to be a large enough multiple of  $p_1$  and of  $p_3$  we can ensure that the net contribution of the terms involving  $\varepsilon'$  and  $\delta_c$  on the right-hand side is always non-positive. Choosing  $p_1 = p_4$  and  $p_3$  so that the ratio  $p_3/p_1$  is small enough we can ensure that the right-hand side is less than  $p_{\text{compl}} - \Delta$ , for some universal constant  $\Delta > 0$  and all  $\varepsilon, \varepsilon' \geq 0$ . □

## 6 Running our protocols in sequence

In order to make a fair comparison between previous delegated computation protocols and ours (see Figure 1) we analyzed their resource requirements under the condition that they produce the correct outcome of the computation with 99% probability. For most protocols, this is achieved by sequentially repeating the original version, in order to amplify the completeness-soundness gap.

In this section, we describe a sequential procedure that, starting from our protocols in Sections 4 and 5, ensures that either the verifier aborts, or she obtains the correct outcome of the computation with probability 99%. Moreover, for honest provers, the probability that the procedure aborts is exponentially small in the number of sequential repetitions. Our sequential procedure has a number of rounds which depends on the desired soundness. As long as one only requires amplification of an arbitrarily small, but constant, soundness, to a fixed constant, the number of sequential repetitions remains constant.

To emphasize the importance of having such a sequential procedure, we note that, firstly, the current completeness-soundness gap between acceptance probability on *yes* and *no* instances, for both the leash and the Dog-Walker protocol, is a very small constant. Secondly, if a classical client wishes to employ our protocols to delegate a computation, we need to specify what the client interprets, at the end of the protocol, as the outcome of the delegated computation. The natural approach is to have the verifier interpret accept as a *yes* outcome and reject as a *no* outcome. However, this is not enough, as our security model based on the constant gap between acceptance probability for *yes* and *no* instances means that, while the provers have a

low probability of making the verifier accept a *no* instance as a *yes*, they can always make the verifier accept a *yes* instance as a *no*, simply by behaving so that they are rejected.

The first point is addressed by running copies of the original protocol in sequence to amplify the completeness-soundness gap. The second point is addressed by having the verifier run the protocol twice: once for the circuit  $Q$ , and once for the circuit  $Q'$  defined by appending an  $X$  gate to the output wire of  $Q$ . If  $f : X \rightarrow \{0, 1\}$  for some  $X \subseteq \{0, 1\}^n$  is defined by  $f(x) = 1$  if  $\|\Pi_0 Q|x\rangle\|^2 \geq 2/3$ , and  $f(x) = 0$  if  $\|\Pi_0 Q|x\rangle\|^2 \leq 1/3$ , i.e.  $Q$  decides  $f$  with bounded error  $1/3$ , then it is easy to see that  $Q'$  decides  $1 - f$  with bounded error  $1/3$ . Thus, the verifier will accept  $x$  as a *yes* instance of  $f$  if the protocol outputs accept when running  $Q$  on  $x$  and outputs reject when running  $Q'$  on  $x$ . The verifier accepts  $x$  as a *no* instance of  $f$  if the protocol outputs reject when running  $Q$  on  $x$  and outputs accept when running  $Q'$  on  $x$ . The verifier aborts if she sees accept-accept or reject-reject.

## 6.1 Sequential version of our protocols

Let  $P$  denote either the Verifier-on-a-leash or the Dog-Walker protocol from Sections 4 and 5 respectively, and let  $c$  and  $\Delta$  denote the completeness and completeness-soundness gap. Let  $\kappa$  be a security parameter.

---

Protocol  $\text{Seq}(P, c, \Delta, \kappa)$ : Let  $(Q, x)$  be the verifier's input.

1. The verifier runs  $\kappa$  copies of protocol  $P$  in sequence on input  $(Q, x)$  with PP and PV. Then she runs  $\kappa$  copies in sequence on input  $(Q', x)$ .
  2. Let  $\vec{o}, \vec{\bar{o}} \in \{0, 1\}^\kappa$  be such that  $o_i = 1$  iff the  $i$ -th copy on input  $(Q, x)$  accepts, and  $\bar{o}_i = 1$  iff the  $i$ -th copy on input  $(Q', x)$  accepts. Let  $wt(\vec{o})$  and  $wt(\vec{\bar{o}})$  be their Hamming weights. Then, the verifier accepts 1 as the outcome of the delegated computation if  $wt(\vec{o}) \geq (c - \frac{\Delta}{2}) \cdot \kappa$  and  $wt(\vec{\bar{o}}) < (c - \frac{\Delta}{2}) \cdot \kappa$ , and she accepts 0 as the outcome of the computation if  $wt(\vec{o}) < (c - \frac{\Delta}{2}) \cdot \kappa$  and  $wt(\vec{\bar{o}}) \geq (c - \frac{\Delta}{2}) \cdot \kappa$ . Otherwise the verifier aborts.
- 

Figure 21: Sequential version of our protocols

We state and prove completeness and soundness for the sequential protocol.

**Theorem 26.** *Let  $c$  and  $\Delta$  be respectively the completeness and completeness-soundness gap of protocol  $P$ . On input  $(Q, x)$ :*

- *If the provers are honest,*

$$\Pr(\text{Seq}(P, c, \Delta, \kappa) \text{ outputs } f(x)) \geq 1 - 2 \exp\left(-\frac{\Delta^2 \kappa}{2}\right).$$

- *For any cheating provers,*

$$\Pr(\text{Seq}(P, c, \Delta, \kappa) \text{ outputs } 1 - f(x)) \leq \exp\left(-\frac{\Delta^2 \kappa}{8}\right).$$

*Proof.* We first show completeness. Let  $s = c - \Delta$  be the soundness of protocol  $P$ . Suppose  $f(x) = 1$  (the

case  $f(x) = 0$  is analogous). If the provers are honest, then the probability that the verifier outputs 1 is:

$$\begin{aligned} \Pr(\text{Verifier outputs 1}) &= \Pr\left(wt(\vec{d}) \geq \left(c - \frac{\Delta}{2}\right) \cdot \kappa \wedge wt(\vec{d}) < \left(c - \frac{\Delta}{2}\right) \cdot \kappa\right) \\ &\geq 1 - \Pr\left(wt(\vec{d}) < \left(c - \frac{\Delta}{2}\right) \cdot \kappa\right) - \Pr\left(wt(\vec{d}) \geq \left(c - \frac{\Delta}{2}\right) \cdot \kappa\right) \\ &\geq 1 - 2 \exp\left(-\frac{\Delta^2 \kappa}{2}\right) \end{aligned}$$

by Hoeffding's inequality.

Next we show soundness. Again suppose  $f(x) = 1$  (the case  $f(x) = 0$  is analogous). Let  $W_j$  be an indicator random variable for the event  $\tilde{d}_j = 1$ , and let  $F_j = W_j - s$ . Define  $X_l = \sum_{j=1}^l F_j$ , for  $l = 1, \dots, \kappa$ . The  $F_j$  define a submartingale with  $|F_j| \leq 1 \forall j$ . Hence, by Azuma's inequality, for any  $\kappa \geq 1$ ,  $\Pr(X_\kappa \geq t) \leq \exp(-\frac{t^2}{2\kappa})$ . This implies that

$$\Pr\left(\sum_{j=1}^{\kappa} W_j - \kappa \cdot s \geq t\right) = \Pr\left(\sum_{j=1}^{\kappa} F_j \geq t\right) = \Pr(X_\kappa \geq t) \leq \exp\left(-\frac{t^2}{2\kappa}\right).$$

Then, for any provers PP and PV,

$$\begin{aligned} \Pr(\text{Verifier outputs 0}) &\leq \Pr\left(wt(\vec{d}) \geq \left(c - \frac{\Delta}{2}\right) \cdot \kappa\right) \\ &= \Pr\left(\sum_{j=1}^{\kappa} W_j \geq \left(c - \frac{\Delta}{2}\right) \cdot \kappa\right) \\ &= \Pr\left(\sum_{j=1}^{\kappa} W_j - \kappa \cdot s \geq \kappa \cdot \frac{\Delta}{2}\right) \\ &\leq \exp\left(-\frac{\Delta^2 \kappa}{8}\right). \end{aligned} \quad \square$$

## A Some simple tests

In this appendix we collect simple tests that will be used as building blocks. In Section A.1 and Section A.2 we review elementary tests whose analysis is either immediate or can be found in the literature. In Section A.3 we formulate a simple test for measurements in the Bell basis and the associated two-qubit SWAP observable.

### A.1 The Magic Square game

We use the Magic Square game [Mer90] as a building block, noting that it provides a robust self-test test for the two-qubit Weyl-Heisenberg group (see Section 2.1 for the definition). Questions in this game are specified by a triple of labels corresponding to the same row or column from the square pictured in Figure 22 (so a typical question could be  $(IZ, XI, XZ)$ ; there are 6 questions in total, each a triple). An answer is composed of three values in  $\{\pm 1\}$ , one for each of the labels making up the question. Answers from the prover should be entrywise consistent, and such that the product of the answers associated to any row or column except the last should be  $+1$ ; for the last column it should be  $-1$ . The labels indicate the “honest” strategy for the game, which consists of each prover measuring two half-EPR pairs using the commuting Pauli observables indicated by the labels of his question.

$IZ$	$ZI$	$ZZ$
$XI$	$IX$	$XX$
$XZ$	$ZX$	$YY$

Figure 22: Questions, and a strategy, for the Magic Square game

The following lemma states some properties of the Magic Square game, interpreted as a self-test (see e.g. [WBMS16]).

**Lemma 27.** *Suppose a strategy for the provers, using state  $|\psi\rangle$  and observables  $W$ , succeeds with probability at least  $1 - \varepsilon$  in the Magic Square game. Then there exist isometries  $V_D : \mathcal{H}_D \rightarrow (\mathbb{C}^2 \otimes \mathbb{C}^2)_{D'} \otimes \mathcal{H}_{\hat{D}}$ , for  $D \in \{A, B\}$  and a state  $|\text{AUX}\rangle_{\hat{A}\hat{B}} \in \mathcal{H}_{\hat{A}} \otimes \mathcal{H}_{\hat{B}}$  such that*

$$\|(V_A \otimes V_B)|\psi\rangle_{AB} - |\text{EPR}\rangle_{A'B'}^{\otimes 2} |\text{AUX}\rangle_{\hat{A}\hat{B}}\|^2 = O(\sqrt{\varepsilon}),$$

and for  $W \in \{I, X, Z\}^2 \cup \{YY\}$ ,

$$\|(W - V_A^\dagger \sigma_W V_A) \otimes \text{Id}_B |\psi\rangle\|^2 = O(\sqrt{\varepsilon}).$$

## A.2 Elementary tests

Figure 23 summarizes some elementary tests. For each test, “Inputs” refers to a subset of designated questions in the test; “Relation” indicates a relation that the test aims to certify (in the sense of Section 3.1); “Test” describes the certification protocol. (Recall that all our protocols implicitly include a “consistency” test in which a question is chosen uniformly at random from the marginal distribution and sent to both provers, whose answers are accepted if and only if they are equal.)



---

Test ID( $A, B$ ):

- Inputs:  $A, B$  two observables on the same space  $\mathcal{H}$ .
- Relation:  $A = B$ .
- Test: Send  $W \in \{A, B\}$  and  $W' \in \{A, B\}$ , chosen uniformly at random, to the first and second prover respectively. Receive an answer in  $\{\pm 1\}$  from each prover. Accept if and only if the answers are equal whenever the questions are identical.

Test AC( $X, Z$ ):

- Inputs:  $X, Z$  two observables on the same space  $\mathcal{H}$ .
- Relation:  $XZ = -ZX$ .
- Test: Execute the Magic Square game, using the label “X” for the “XI” query, and “Z” for the “ZI” query.

Test COM( $A, B$ ):

- Inputs:  $A, B$  two observables on the same space  $\mathcal{H}$ .
- Relation:  $AB = BA$ .
- Test: Send  $W \in \{A, B\}$  chosen uniformly at random to the first prover. Send  $(A, B)$  to the second prover. Receive a bit  $c \in \{\pm 1\}$  from the first prover, and two bits  $(a', b') \in \{\pm 1\}^2$  from the second. Accept if and only if  $c = a'$  if  $W = A$ , and  $c = b'$  if  $W = B$ .

Test PROD( $A, B, C$ ):

- Inputs:  $A, B$  and  $C$  three observables on the same space  $\mathcal{H}$ .
- Relations:  $AB = BA = C$ .
- Test: Similar to the commutation game, but use  $C$  to label the question  $(A, B)$ .

---

Figure 23: Some elementary tests.

**Lemma 28.** *Each of the tests described in Figure 23 is a robust  $(1, \delta)$  self-test for the indicated relation(s), for some  $\delta = O(\varepsilon^{1/2})$ .*

*Proof.* The proof for each test is similar. As an example we give it for the commutation test COM( $A, B$ ).

First we verify completeness. Let  $A, B$  be two commuting observables on  $\mathcal{H}_A = \mathcal{H}_B = \mathcal{H}$ , and  $|\text{EPR}\rangle_{AB}$  the maximally entangled state in  $\mathcal{H}_A \otimes \mathcal{H}_B$ . Upon receiving question  $A$  or  $B$ , the prover measures the corresponding observable. If the question is  $(A, B)$ , he jointly measures  $A$  and  $B$ . This strategy succeeds with probability 1 in the test.

Next we establish soundness. Let  $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  be a state shared by the provers,  $A, B$  their observables on questions  $A, B$ , and  $\{C^{a,b}\}$  the four-outcome PVM applied on question  $(A, B)$ . Assume the strategy succeeds with probability at least  $1 - \varepsilon$ . Recall that this includes both the test described in Figure 23, and the automatic consistency test. Let  $C_A = \sum_{a,b} (-1)^a C^{a,b}$  and  $C_B = \sum_{a,b} (-1)^b C^{a,b}$ . Then  $C_A$

and  $C_B$  commute. Thus

$$\begin{aligned}
A_A B_A \otimes \text{Id}_B &\approx_{\sqrt{\varepsilon}} A_A \otimes (C_B)_B \\
&\approx_{\sqrt{\varepsilon}} \text{Id}_A \otimes (C_B)_B (C_A)_B \\
&= \text{Id}_A \otimes (C_A)_B (C_B)_B \\
&\approx_{\sqrt{\varepsilon}} B_A \otimes (C_A)_B \\
&\approx_{\sqrt{\varepsilon}} B_A A_A \otimes \text{Id}_B.
\end{aligned}$$

Here each approximation uses the consistency condition provided by the test, as explained in (3). Thus  $[A, B] = (AB - BA) \approx_{\sqrt{\varepsilon}} 0$ , as desired.  $\square$

We will often make use of the following simple lemma, which expresses an application of the above tests.

**Lemma 29.** *Let  $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  and  $A, X$  observables on  $\mathcal{H}_A$  such that there exists an isometry  $\mathcal{H}_A \simeq \mathbb{C}^2 \otimes \mathcal{H}_{\hat{A}}$  under which the following conditions hold, for some  $\delta_1, \delta_2, \delta_3$ .<sup>6</sup>*

- (i) *There exists an observable  $A'$  on  $\mathcal{H}_B$  such that  $A \otimes \text{Id} \approx_{\delta_1} \text{Id} \otimes A'$ ;*
- (ii)  *$|\psi\rangle \simeq_{\delta_1} |\text{EPR}\rangle_{\text{AUX}}$  and  $X \simeq_{\delta_1} \sigma_X \otimes \text{Id}$ ;*
- (iii)  *$[A, X] \approx_{\delta_2} 0$ ;*
- (iv)  *$\{A, X\} \approx_{\delta_3} 0$ .*

*Then there exist Hermitian  $A_I, A_X, A_Y, A_Z$  on  $\mathcal{H}_{\hat{A}}$  such that  $A \simeq_{\delta_1+\delta_2} \text{Id} \otimes A_I + \sigma_X \otimes A_X$  and  $A \simeq_{\delta_1+\delta_3} \sigma_Y \otimes A_Y + \sigma_Z \otimes A_Z$ . (A similar claim holds with  $X$  replaced by  $Z$ .)*

*Proof.* After application of the isometry, an arbitrary observable  $\tilde{A}$  on  $\mathbb{C}^2 \otimes \mathcal{H}_{\hat{A}}$  has a decomposition  $\tilde{A} = \sum_{P \in \{I, X, Y, Z\}} \sigma_P \otimes A_P$ , for Hermitian operators  $A_P$  on  $\mathcal{H}_{\hat{A}}$ . We can compute

$$[\tilde{A}, \sigma_X \otimes \text{Id}] = -2i \sigma_Z \otimes A_Y + 2i \sigma_Y \otimes A_Z, \quad (29)$$

$$\{\tilde{A}, \sigma_X \otimes \text{Id}\} = 2 \sigma_X \otimes A_I + 2 \sigma_I \otimes A_X. \quad (30)$$

Assumptions (i) and (ii) imply  $[A, X] \simeq_{\delta_1} [\tilde{A}, \sigma_X \otimes \text{Id}]$ , so by (iii) and (29) we get  $\|A_Y|_{\text{AUX}}\|^2 + \|A_Z|_{\text{AUX}}\|^2 = O(\delta_1 + \delta_2)$ . Similarly, (iv) and (30) give  $\|A_I|_{\text{AUX}}\|^2 + \|A_X|_{\text{AUX}}\|^2 = O(\delta_1 + \delta_3)$ .  $\square$

### A.3 The Bell basis

Given two commuting pairs of anti-commuting observables  $\{X_1, Z_1\}$  and  $\{X_2, Z_2\}$  we provide a test for a four-outcome projective measurement in the Bell basis specified by these observables, i.e. the joint eigenbasis of  $X_1 X_2$  and  $Z_1 Z_2$ . The same test can be extended to test the “SW” observable,

$$\text{SW} = \frac{1}{2} (\text{Id} + X_1 X_2 + Z_1 Z_2 - (X_1 Z_1)(X_2 Z_2)), \quad (31)$$

which exchanges the qubits specified by each pair of observables. The Bell measurement test described in Figure 24 tests for both.

---

<sup>6</sup>Note that we allow either  $\delta_i$  to equal 1, leading to a vacuous condition.

---

Test BELL( $X_1, X_2, Z_1, Z_2$ ):

- Inputs: For  $i \in \{1, 2\}$ ,  $\{X_i, Z_i\}$  observables,  $\{\Phi^{ab}\}_{a,b \in \{0,1\}}$  a four-outcome projective measurement, and SW an observable, all acting on the same space  $\mathcal{H}$ .
  - Relations: for all  $a, b \in \{0, 1\}$ ,  $\Phi^{ab} = \frac{1}{4}(\text{Id} + (-1)^a Z_1 Z_2)(\text{Id} + (-1)^b X_1 X_2)$ , and  $\text{SW} = \Phi^{00} + \Phi^{01} + \Phi^{10} - \Phi^{11}$ .
  - Test: execute each of the following with equal probability:
    - (a) Execute the Magic Square game, labeling each entry of the square from Figure 22 (except entry  $(3, 3)$ , labeled as  $Y_1 Y_2$ ) using the observables  $X_1, Z_1$  and  $X_2, Z_2$ .
    - (b) Send  $\Phi$  to one prover and the labels  $(X_1 X_2, Z_1 Z_2, Y_1 Y_2)$  associated with the third column of the Magic Square to the other. The first prover replies with  $a, b \in \{0, 1\}$ , and the second with  $c, d, e \in \{\pm 1\}$ . The referee checks the provers' answers for the obvious consistency conditions. For example, if the first prover reports the outcome  $(0, 0)$ , then the referee rejects if  $(c, d) \neq (+1, +1)$ .
    - (c) Send  $\Phi$  to one prover and SW to the other. The first prover replies with  $a, b \in \{0, 1\}$ , and the second with  $c \in \{\pm 1\}$ . Accept if and only  $c = (-1)^{ab}$ .
- 

Figure 24: The Bell measurement test.

**Lemma 30.** *The test BELL( $X_1, X_2, Z_1, Z_2$ ) is a robust  $(1, \delta)$  self-test for*

$$\mathcal{R} = \left\{ \left\{ \Phi^{ab} \right\}_{a,b \in \{0,1\}} \in \text{Proj}, \text{SW} \in \text{Obs} \right\} \cup \left\{ \Phi^{ab} = \frac{1}{4}(1 + (-1)^a Z_1 Z_2)(1 + (-1)^b X_1 X_2) \right\} \\ \cup \left\{ \text{SW} = \Phi^{00} + \Phi^{01} + \Phi^{10} - \Phi^{11} \right\},$$

for some  $\delta(\varepsilon) = O(\sqrt{\varepsilon})$ .

*Proof.* Completeness is clear: the provers can play the honest strategy for the Magic Square game, use a measurement in the Bell basis on their two qubits for  $\Phi$ , and measure the observable in (31) for SW.

For soundness, let  $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ ,  $\{W_1 W'_2 : W, W' \in \{I, X, Z\}\}$ ,  $\{\Phi^{ab}\}$  and SW denote a state and operators for a strategy that succeeds with probability at least  $1 - \varepsilon$  in the test. From the analysis of the Magic Square game (Lemma 27) it follows that the provers' observables  $X_1 X_2$  and  $Z_1 Z_2$  associated to questions with those labels approximately commute, and are each the product of two commuting observables  $X_1 I$ ,  $I X_2$  and  $Z_1 I$ ,  $I Z_2$  respectively, such that  $X_1 I$  and  $Z_1 I$ , and  $I X_2$  and  $I Z_2$ , anti-commute; all approximate identities hold up to error  $O(\sqrt{\varepsilon})$ .

Since  $X_1 X_2$  and  $Z_1 Z_2$  appear together in the same question (the last column of the Magic Square, Figure 22), each prover has a four-outcome projective measurement  $\{W^{c,d}\}_{c,d \in \{0,1\}}$  such that  $\sum_d (-1)^c W^{c,d} = X_1 X_2$  and  $\sum_c (-1)^d W^{c,d} = Z_1 Z_2$ , from which it follows that  $W^{c,d} = (1/4)(1 + (-1)^c Z_1 Z_2)(1 + (-1)^d X_1 X_2)$ .

The prover's success probability in part (b) of the test is then

$$\sum_{a,b} \langle \psi | \Phi^{ab} \otimes W^{a,b} | \psi \rangle = \sum_{a,b} \langle \psi | \Phi^{ab} \otimes \frac{1}{4}(1 + (-1)^a Z_1 Z_2)(1 + (-1)^b X_1 X_2) | \psi \rangle.$$

Using that, by assumption,  $\{\Phi^{ab}\}$  is a projective measurement, the condition that this expression be at least  $1 - O(\varepsilon)$  implies

$$\Phi^{ab} \otimes \text{Id} \approx_{\sqrt{\varepsilon}} \text{Id} \otimes \frac{1}{4}(1 + (-1)^a Z_1 Z_2)(1 + (-1)^b X_1 X_2).$$

Combining this with the implicit consistency test yields the first relation. The last is guaranteed by part (c) of the test, which checks for the correct relationship between  $\text{SW}$  and  $\Phi$ ; the analysis is similar.  $\square$

## B The $m$ -qubit Pauli group

In this section we formulate a robust self-test for the  $m$ -qubit Pauli group. The result is a slight extension of the results from [NV16] to allow testing of  $\sigma_Y$  observables.

### B.1 The $m$ -qubit Weyl-Heisenberg group

We start by giving a self-test for tensor products of  $\sigma_X$  and  $\sigma_Z$  observables acting on  $m$  qubits, i.e. the  $m$ -qubit Weyl-Heisenberg group  $\mathcal{H}^{(m)}$  (see Section 2.1). Let  $\mathcal{P}^{(m)}$  denote the relations

$$\begin{aligned} \mathcal{P}^{(m)}\{X, Z\} = & \left\{ W(a) \in \text{Obs}, W \in \prod_{i=1}^m \{X_i, Z_i\}, a \in \{0, 1\}^m \right\} \\ & \cup \left\{ W(a)W'(a') = (-1)^{|\{i: W_i \neq W'_i \wedge a_i a'_i = 1\}|} W'(a')W(a), \forall a, a' \in \{0, 1\}^m \right\} \\ & \cup \left\{ W(a)W(a') = W(a + a'), \forall a, a' \in \{0, 1\}^m \right\}. \end{aligned}$$

Recall the notation  $W(a)$  for the string that is  $W_i$  when  $a_i = 1$  and  $I$  otherwise. The first set of relations expresses the canonical anti-commutation relations. The second set of relations expresses the obvious relations  $\sigma_W \text{Id} = \text{Id} \sigma_W$  and  $\sigma_W^2 = \text{Id}$ , for  $W \in \{X, Z\}$ , coordinate-wise. It is easy to verify that  $\mathcal{P}^{(m)}$  forms a defining set of relations for  $\mathcal{H}^{(m)}$ . Our choice of relations is suggested by the Pauli braiding test introduced in [NV16], which shows that the relations are testable with a robustness parameter  $\delta(\varepsilon)$  that is independent of  $m$ . The underlying test is called the Pauli braiding test, and denoted  $\text{PBT}(X, Z)$ . For convenience here we use a slight variant of the test, which includes more questions; the test is summarized in Figure 25.

---

Test  $\text{PBT}(X, Z)$ :

- Inputs:  $(W, a)$ , for  $W \in \prod_{i=1}^m \{X_i, Z_i\}$  and  $a \in \{0, 1\}^m$ .
  - Relations:  $\mathcal{P}^{(m)}\{X, Z\}$ .
  - Test: Perform the following with probability  $1/2$  each:
    - (a) Select  $W, W' \in \prod_i \{X_i, Z_i\}$ , and  $a, a' \in \{0, 1\}^m$ , uniformly at random. If  $\{i : W_i \neq W'_i \wedge a_i = a'_i = 1\}$  has even cardinality then execute test  $\text{COM}(W(a), W'(a'))$ . Otherwise, execute test  $\text{AC}(W(a), W'(a'))$ .
    - (b) Select  $(a, a') \in \{0, 1\}^m$  and  $W \in \prod_{i=1}^m \{X_i, Z_i\}$  uniformly at random. Execute test  $\text{PROD}(W(a), W(a'), W(a + a'))$ .
- 

Figure 25: The Pauli braiding test,  $\text{PBT}(X, Z)$ .

The following lemma follows immediately from the definition of the relations  $\mathcal{P}^{(m)}\{X, Z\}$  and the analysis of the tests  $\text{COM}$ ,  $\text{PROD}$  and  $\text{AC}$  given in Section A.2.

**Lemma 31** (Theorem 13 [NV16]). *The test  $\text{PBT}(X, Z)$  is a robust  $(1, \delta)$  self-test for  $\mathcal{P}^{(m)}\{X, Z\}$ , for some  $\delta(\varepsilon) = O(\varepsilon^{1/2})$ .*

In addition we need the following lemma, which states that observables approximately satisfying the relations  $\mathcal{P}^{(m)}\{X, Z\}$  are close to operators which, up to a local isometry, behave exactly as a tensor product of Pauli  $\sigma_X$  and  $\sigma_Z$  observables.

**Lemma 32** (Theorem 14 [NV16]). *The set of relations  $\mathcal{P}^{(n)}$  is  $\delta$ -stable, with  $\delta(\varepsilon) = O(\varepsilon)$ .*

Lemma 32 is proved in [NV16] with a polynomial dependence of  $\delta$  on  $\varepsilon$ . The linear dependence can be established by adapting the results of [GH15] to the present setting; we omit the details (see [Vid17]).

The following lemma is an extension of Lemma 29 to the case of multi-qubit Pauli observables; the lemma avoids any dependence of the error on the number of qubits, as would follow from a sequential application of Lemma 29.

**Lemma 33.** *Let  $n$  be an integer,  $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  and  $A$  and  $X(a)$ , for  $a \in \{0, 1\}^m$ , observables on  $\mathcal{H}_A$  such that there exists an isometry  $\mathcal{H}_A \simeq (\mathbb{C}^2)^{\otimes m} \otimes \mathcal{H}_{\hat{A}}$  under which the following conditions hold, for some  $\delta_1, \delta_2, \delta_3$ :*

- (i) *There exists an observable  $A'$  on  $\mathcal{H}_B$  such that  $A \otimes \text{Id} \simeq_{\delta_1} \text{Id} \otimes A'$ ;*
- (ii)  *$|\psi\rangle \simeq_{\delta_1} |\text{EPR}\rangle^{\otimes m} |\text{AUX}\rangle$ , and  $X(a) \simeq_{\delta_1} \sigma_X(a) \otimes \text{Id}$ ;*
- (iii)  *$[A, X(a)] \simeq_{\delta_2} 0$ ;*
- (iv) *For some  $c \in \{0, 1\}^m$  and  $a \cdot c = 1$ ,  $\{A, X(a)\} \simeq_{\delta_3} 0$ ;*

where the first two conditions are meant on average over a uniformly random  $a \in \{0, 1\}^m$ , and the last over a uniformly random  $a$  such that  $a \cdot c = 1$ . For  $P \in \{I, X, Y, Z\}^m$  let  $x_P \in \{0, 1\}^m$  be such that  $(x_P)_i = 1$  if and only if  $P_i \in \{Y, Z\}$ . Then there exists Hermitian  $A_P$ , for  $P \in \{I, X, Y, Z\}^m$ , on  $\mathcal{H}_{\hat{A}}$  such that

$$A \simeq_{\delta_1 + \delta_2} \sum_{P \in \{I, X\}^m} \sigma_P \otimes A_P, \quad \text{and} \quad A \simeq_{\delta_1 + \delta_3} \sum_{\substack{P \in \{I, X, Y, Z\}^m: \\ c_i=1 \implies P_i \in \{Y, Z\} \\ c_i=0 \implies P_i \in \{I, X\}}} \sigma_P \otimes A_P.$$

(A similar claim holds with  $X$  replaced by  $Z$ .)

*Proof.* After application of the isometry, an arbitrary observable  $\tilde{A}$  on  $(\mathbb{C}^2)^{\otimes m} \otimes \mathcal{H}_{\hat{A}}$  has a decomposition  $\tilde{A} = \sum_{P \in \{I, X, Y, Z\}^m} \sigma_P \otimes A_P$ , for Hermitian operators  $A_P$  on  $\mathcal{H}_{\hat{A}}$ . Then the analogue of (29) is

$$[\tilde{A}, \sigma_X(a) \otimes \text{Id}] = 2 \sum_{P: a \cdot x_P = 1} \sigma_P \sigma_X(a) \otimes A_P.$$

Using that any string  $x_P$  which is not the  $0^m$  string satisfies  $a \cdot x_P = 1$  with probability almost  $1/2$  for a uniform choice of  $a$ , orthogonality of the  $\sigma_P \sigma_X(a)$  for distinct  $P$  lets us conclude the proof of the first relation as in Lemma 29. Similarly, the analogue of (30) gives

$$\{\tilde{A}, \sigma_X(a) \otimes \text{Id}\} = 2 \sum_{P: a \cdot x_P = 0} \sigma_P \sigma_X(a) \otimes A_P.$$

Using that any string  $x_P$  which is not  $c$  satisfies  $a \cdot x_P = 0$  with probability almost  $1/2$  for a uniform choice of  $a$  such that  $a \cdot c = 1$ , orthogonality of the  $\sigma_P \sigma_X(a)$  for distinct  $P$  lets us conclude the proof of the second relation.  $\square$

## B.2 The $m$ -qubit Pauli group

We will use an extended version of the Pauli braiding test introduced in Section B.1 which allows to test for a third observable,  $Y_i$ , on each system. Ideally we would like to enforce the relation  $Y_i = \sqrt{-1}X_iZ_i$ . Unfortunately, the complex phase cannot be tested from classical correlations alone: complex conjugation leaves correlations invariant, but does not correspond to a unitary change of basis (see [RUV12, Appendix A] for a discussion of this issue).

We represent the “choice” of complex phase,  $\sqrt{-1}$  or its conjugate  $-\sqrt{-1}$ , by an observable  $\Delta$  that the prover measures on a system that is in a tensor product with all other systems on which the prover acts. Informally, the outcome obtained when measuring  $\Delta$  tells the prover to use  $Y = iXZ$  or  $Y = -iXZ$ .

We first introduce  $Y$  and test that the triple  $\{X, Y, Z\}$  pairwise anticommute at each site. This corresponds to the following set of relations:

$$\begin{aligned} \mathcal{P}^{(m)}\{X, Y, Z\} = & \left\{ W(a) \in \text{Obs}, W \in \{X, Y, Z\}^n, a \in \{0, 1\}^n \right\} \\ & \cup \left\{ W(a)W'(a') = (-1)^{|\{i: W_i \neq W'_i \wedge a_i a'_i = 1\}|} W'(a')W(a), \forall a, a' \in \{0, 1\}^n \right\} \\ & \cup \left\{ W(a)W(a') = W(a + a'), \forall a, a' \in \{0, 1\}^n \right\}. \end{aligned}$$

---

Test PBT( $X, Y, Z$ ):

- Inputs:  $W \in \prod_{i=1}^m \{X, Y, Z\}$
  - Relations:  $\mathcal{P}^{(m)}\{X, Y, Z\}$ .
  - Test: Perform the following with equal probability:
    - (a) Execute test PBT( $X^m, Z^m$ ).
    - (b) Execute test PBT( $Y^m, X^m$ ) or test PBT( $Y^m, Z^m$ ), chosen with probability 1/2 each.
    - (c) Select a random permutation  $\sigma \in \mathfrak{S}_{m/2}$ , and  $W \in \{I, Y\}^m$  uniformly at random. Write  $W = W_1 W_2$ , where  $W_1, W_2 \in \{I, Y\}^{m/2}$ . Let  $W_1^\sigma$  be the string  $W_1$  with its entries permuted according to  $\sigma$ . Do the following with equal probability:
      - (i) Send one prover  $W_1 W_1^\sigma$  and the other either  $W_1 W_2$  or  $W_2 W_1^\sigma$  (chosen with probability 1/2), and check consistency of the first or second half of the provers' answer bits.
      - (ii) Send one prover  $W_1 W_1^\sigma$ , and the other  $\prod_i \Phi_{i, \sigma(i)}$ , where each  $\Phi_{i, \sigma(i)}$  designates a measurement in the Bell basis for the  $(i, m/2 + \sigma(i))$  pair of qubits. The first prover replies with  $a \in \{\pm 1\}^m$ , and the second with  $b \in \{00, 01, 10, 11\}^{m/2}$ . For each  $i \in \{1, \dots, m/2\}$  such that  $b_i = 00$ , check that  $a_i = a_{m/2 + \sigma(i)}$ .
      - (iii) Execute  $m/2$  copies of test BELL (in parallel), for qubit pairs  $(i, m/2 + \sigma(i))$ , for  $i \in \{1, \dots, m/2\}$ .
- 

Figure 26: The extended Pauli braiding test, PBT( $X, Y, Z$ ).

The test is described in Figure 26. It has three components. Part (a) of the test executes test PBT( $X^m, Z^m$ ), which gives us multi-qubit Pauli  $X$  and  $Z$  observables. Part (b) of the test introduces observables labeled  $Y(c)$ , and uses tests PBT( $Y^m, X^m$ ) and PBT( $Y^m, Z^m$ ) to enforce appropriate anti-commutation relations with the Pauli  $X$  and  $Z$  observables obtained in part (a). Using Lemma 33, this part of the test will establish that the  $Y(c)$  observables approximately respect the same  $n$ -qubit tensor product structure as  $X(a)$  and  $Z(b)$ .

Part (c) of the test is meant to control the “phase” ambiguity in the definition of  $Y(c)$  that remains after the analysis of part (b). Indeed, from that part it will follow that  $Y(c) \simeq \sigma_Y(c) \otimes \Delta(c)$ , where  $\Delta(c)$  is an arbitrary observable acting on the ancilla system produced by the isometry obtained in part (a). We would like to impose  $\Delta(c) \approx \Delta_Y^{[c]}$  for a fixed observable  $\Delta_Y$  which represents the irreducible phase degree of freedom in the definition of  $Y$ , as discussed above. To obtain this, part (c) of the test performs a form of SWAP test between different  $Y(c)$  observables, enforcing that e.g.  $Y(1,0,1)$  is consistent with  $Y(0,1,1)$  after an appropriate Bell measurement has “connected” registers 1 and 2. The swapping is defined using Pauli  $\sigma_X$  and  $\sigma_Z$ , which leave the ancilla register invariant; consistency will then imply  $\Delta(1,0,1) \approx \Delta(0,1,1)$ .

**Lemma 8 (restated).** Suppose  $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  and  $W(a) \in \text{Obs}(\mathcal{H}_A)$ , for  $W \in \{X, Y, Z\}^m$  and  $a \in \{0,1\}^m$ , specify a strategy for the players that has success probability at least  $1 - \varepsilon$  in the extended Pauli braiding test  $\text{PBT}(X, Y, Z)$  described in Figure 26. Then there exist isometries  $V_D : \mathcal{H}_D \rightarrow ((\mathbb{C}^2)^{\otimes m})_{D'} \otimes \hat{\mathcal{H}}_{\hat{D}}$ , for  $D \in \{A, B\}$ , such that

$$\|(V_A \otimes V_B)|\psi\rangle_{AB} - |\text{EPR}\rangle_{A'B'}^{\otimes n} |\text{AUX}\rangle_{\hat{A}\hat{B}}\|^2 = O(\sqrt{\varepsilon}),$$

and on expectation over  $W \in \{X, Y, Z\}^m$ ,

$$\mathbb{E}_{a \in \{0,1\}^m} \|(W(a) - V_A^\dagger(\sigma_W(a) \otimes \Lambda_W(a))V_A) \otimes \text{Id}_B |\psi\rangle\|^2 = O(\sqrt{\varepsilon}), \quad (32)$$

where  $\Lambda_W(a) = \prod_i \Lambda_{W_i}^{a_i} \in \text{Obs}(\mathcal{H}_{\hat{A}})$  are observables with  $\Delta_X = \Delta_Z = \text{Id}$  and  $\Delta_Y$  an arbitrary observable on  $\hat{\mathcal{H}}$  such that

$$\|\Delta_Y \otimes \Delta_Y |\text{AUX}\rangle - |\text{AUX}\rangle\|^2 = O(\sqrt{\varepsilon}).$$

*Proof sketch.* The existence of the isometries  $V_A$  and  $V_B$  follows from part (a) of the test and the combination of Lemma 31 and Lemma 32; see e.g. [NV16] for an explicit construction. Under this isometry we have  $X(a) \simeq_{\sqrt{\varepsilon}} \sigma_X(a)$  and  $Z(b) \simeq_{\sqrt{\varepsilon}} \sigma_Z(b)$ , on average over  $a, b \in \{0,1\}^m$ . Applying the second part of Lemma 33, the anti-commutation relations between  $Y(c)$  and  $X(a)$  and  $Z(b)$  verified in part (b) of the test imply that under the same isometry,

$$Y(c) \simeq \sigma_Y(c) \otimes \Delta(c),$$

for some observable  $\Delta(c)$  on  $\mathcal{H}_{\hat{A}}$ . Using the linearity relations that are verified in the PBT test, we may in addition express  $\Delta(c) = \prod_i \Delta_i^{c_i}$  for (perfectly) commuting observables  $\Delta_i$ . Using Claim 34 below, success at least  $1 - O(\varepsilon)$  in part (c) of the test then implies that on average over a random permutation  $\sigma \in \mathcal{S}_{n/2}$ ,

$$\mathbb{E}_{\sigma} \mathbb{E}_{c \in \{0,1\}^{m/2}} 2^{-m} \text{Tr}(\sigma_Y(c, c^\sigma)) \langle \text{AUX} | \left( \prod_{i=1}^{m/2} (\Delta_i \Delta_{m/2+\sigma(i)})^{c_i} \right) | \text{AUX} \rangle = 1 - O(\sqrt{\varepsilon}), \quad (33)$$

where we wrote  $(c, c^\sigma)$  for the  $m$ -bit string  $(c_1, \dots, c_{m/2}, c_{\sigma(1)}, \dots, c_{\sigma(m/2)})$ . Defining

$$\Delta_Y = \mathbb{E}_{i \in \{\frac{m}{2}+1, \dots, m\}} \frac{\Delta_i}{|\mathbb{E}_i \Delta_i|}, \quad (34)$$

Eq. (33) readily implies that  $\Delta(c) \approx_{\sqrt{\varepsilon}} \Delta_Y^{[c]}$ . In slightly more detail, we first observe that

$$\mathbb{E}_{c \in \{0,1\}^{m/2}} \left\| \left( \Delta(c) - \left( \mathbb{E}_{i \in \{\frac{m}{2}+1, \dots, m\}} \Delta_i \right)^{[c]} \right) | \text{AUX} \rangle \right\|^2 \leq \mathbb{E}_c \mathbb{E}_{g: \{1, \dots, \frac{m}{2}\} \rightarrow \{\frac{m}{2}+1, \dots, m\}} \left\| \left( \Delta(c) - \prod_i \Delta_{g(i)}^{c_i} \right) | \text{AUX} \rangle \right\|^2. \quad (35)$$

where the first inequality is by convexity, with the expectation taken over a random function  $g$ . We would like to relate this last term to the expectation over a random permutation  $\sigma \in \mathcal{S}_{m/2}$ . One way to do this is to observe that with probability  $1 - O(1/m)$  over the choice of a uniformly random  $g$  it is possible to write

$$\prod_i \Delta_{g(i)}^{c_i} = \left( \prod_i \Delta_{m/2+\tau'(i)}^{c'_i} \right) \left( \prod_i \Delta_{m/2+\tau''(i)}^{c''_i} \right),$$

where  $c'_i + c''_i = c_i$  for all  $i$ ,  $\tau', \tau''$  are permutations such that  $m/2 + \tau'(i) = g(i)$  if  $c'_i = c_i$ , and  $m/2 + \tau''(i) = g(i)$  if  $c''_i = c_i$ ; this is possible because  $g$  might have two-element collisions, but is unlikely to have any three-element collisions. Moreover, for uniformly random  $c$  and  $g$  we can ensure that the marginal distribution on  $(c', \tau')$  and  $(c', \tau'')$  is uniform. This allows us to use (33) twice to bound the right-hand side of (35) by  $O(\sqrt{\varepsilon})$  (after having expanded the square). As a consequence,  $E_i \Delta_i$  is close to an observable, and it is then routine to show that  $\Delta_Y$  defined in (34) satisfies  $\Delta(c) \approx_{\sqrt{\varepsilon}} \Delta_Y^{|c|}$ , on average over a uniformly random  $c$ .

The last condition in the lemma follows from the consistency relations, which imply that  $X(a) \otimes X(a)$ ,  $Z(b) \otimes Z(b)$  and  $Y(c) \otimes Y(c)$  all approximately stabilize  $|\psi\rangle$ ; then  $\Delta_Y^{|a|} \otimes \Delta_Y^{|a|} \approx X(a)Z(a)Y(a) \otimes X(a)Z(a)Y(a)$  also does.  $\square$

**Claim 34.** *Let  $A \in \text{Obs}(\mathbb{C}_{A_1}^2 \otimes \cdots \otimes \mathbb{C}_{A_k}^2 \otimes \mathcal{H})$  and  $B \in \text{Obs}(\mathbb{C}_{B_1}^2 \otimes \cdots \otimes \mathbb{C}_{B_k}^2 \otimes \mathcal{H})$  be  $k$ -qubit observables acting on distinct registers  $A_j, B_j$ , as well as a common space  $\mathcal{H}$ , and  $\Phi_{A'B'} = \prod_{j=1}^k |\text{EPR}\rangle\langle\text{EPR}|_{A'_j, B'_j}$  the the projector on  $k$  EPR pairs across registers  $A'_j$  and  $B'_j$ . Then*

$$\begin{aligned} & \left( \bigotimes_j \langle \text{EPR} |_{A_j A'_j} \langle \text{EPR} |_{B_j B'_j} \otimes \text{Id}_{\mathcal{H}} \right) \left( (A_{A\mathcal{H}} \otimes \text{Id}_B) (\text{Id}_A \otimes B_{B\mathcal{H}}) \otimes \Phi_{A'B'} \right) \left( \bigotimes_j |\text{EPR}\rangle_{A_j A'_j} |\text{EPR}\rangle_{B_j B'_j} \otimes \text{Id}_{\mathcal{H}} \right) \\ &= \frac{1}{2^{2k}} \sum_i \text{Tr}(A_i B_i) A'_i B'_i, \end{aligned} \quad (36)$$

where we write  $A = \sum_i A_i \otimes A'_i$  and  $B = \sum_i B_i \otimes B'_i$ , for  $A_i$  on  $\mathcal{H}_A$ ,  $B_i$  on  $\mathcal{H}_B$ , and  $A'_i, B'_i$  on  $\mathcal{H}$ .

*Proof.* We do the proof for  $k = 1$ , as the general case is similar. Using that for any operators  $X_{AB}$  and  $Y_{A'B'}$ ,

$$\langle \text{EPR} |_{AA'} \langle \text{EPR} |_{BB'} (X_{AB} \otimes Y_{A'B'}) |\text{EPR}\rangle_{AA'} |\text{EPR}\rangle_{BB'} = \frac{1}{4} \text{Tr}(XY^T),$$

the left-hand side of (36) evaluates to

$$4^{-1} \text{Tr}_{AB} \left( (A_{A\mathcal{H}} \otimes \text{Id}_B) (\text{Id}_A \otimes B_{B\mathcal{H}}) (\Phi_{A'B'}^T \otimes \text{Id}_{\mathcal{H}}) \right),$$

which using the same identity again gives the right-hand side of (36).  $\square$

## References

- [AAV13] Dorit Aharonov, Itai Arad, and Thomas Vidick. Guest column: the quantum PCP conjecture. *SIGACT News*, (2):47–79, 2013.
- [ABE10] Dorit Aharonov, Michael Ben-Or, and Elad Eban. Interactive proofs for quantum computations. In *Proceedings of the first Symposium on Innovations in Computer Science (ICS 2010)*, pages 453–469, 2010.



- [ALM<sup>+</sup>98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.
- [AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.
- [Bel64] John S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.
- [BGKW88] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Multi-prover interactive proofs: How to remove intractability assumptions. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing (STOC 1988)*, pages 113–131, New York, NY, USA, 1988. ACM.
- [Bro15] Anne Broadbent. How to verify a quantum computation, 2015. arXiv preprint [arXiv:1509.09180](https://arxiv.org/abs/1509.09180).
- [Cas17] Davide Castelvecchi. IBM’s quantum cloud computer goes commercial. *Nature News*, 543(7644), 6 March 2017.
- [CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880–884, 1969.
- [FH15] Joseph F. Fitzsimons and Michal Hajdušek. Post hoc verification of quantum computation, 2015. arXiv preprint [arXiv:1512.04375](https://arxiv.org/abs/1512.04375).
- [Fit16] Joseph F. Fitzsimons. Private quantum computation: An introduction to blind quantum computing and related protocols, 2016. arXiv preprint [arXiv:1611.10107](https://arxiv.org/abs/1611.10107).
- [FK17] Joseph F. Fitzsimons and Elham Kashefi. Unconditionally verifiable blind quantum computation. *Physical Review A*, 96(012303), 2017. arXiv preprint [arXiv:1203.5217](https://arxiv.org/abs/1203.5217).
- [FV15] Joseph Fitzsimons and Thomas Vidick. A multiprover interactive proof system for the local Hamiltonian problem. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS 2015)*, pages 103–112, 2015.
- [GC99] Daniel Gottesman and Isaac L. Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402:390–393, 1999.
- [GH15] W. T. Gowers and O. Hatami. Inverse and stability theorems for approximate representations of finite groups, 2015. arXiv preprint [arXiv:1510.04085](https://arxiv.org/abs/1510.04085).
- [GKW15] Alexandru Gheorghiu, Elham Kashefi, and Petros Wallden. Robustness and device independence of verifiable blind quantum computing. *New Journal of Physics*, 17, 2015.
- [HPDF15] Michal Hajdušek, Carlos A. Pérez-Delgado, and Joseph F. Fitzsimons. Device-independent verifiable blind quantum computation, 2015. arXiv preprint [arXiv:1502.02563](https://arxiv.org/abs/1502.02563).
- [HZM<sup>+</sup>17] He-Liang Huang, Qi Zhao, Xiongfeng Ma, Chang Liu, Zu-En Su, Xi-Lin Wang, Li Li, Nai-Le Liu, Barry C. Sanders, Chao-Yang Lu, and Jian-Wei Pan. Experimental Blind Quantum Computing for a Classical Client. *Physical Review Letters*, 119:050503, Aug 2017.

- [Ji16] Zhengfeng Ji. Classical verification of quantum proofs. In *Proceedings of the Forty-eighth Annual ACM SIGACT Symposium on Theory of Computing (STOC 2016)*, pages 885–898, 2016.
- [McK16] Matthew McKague. Interactive proofs for BQP via self-tested graph states. *Theory of Computing*, 12(3):1–42, 2016. arXiv preprint [arXiv:1309.5675](https://arxiv.org/abs/1309.5675).
- [Mer90] N. David Mermin. Simple unified form for the major no-hidden-variables theorems. *Physical Review Letters*, 65:3373–3376, 1990.
- [MF16] Tomoyuki Morimae and Joseph F. Fitzsimons. Post hoc verification with a single prover, 2016. arXiv preprint [arXiv:1603.06046](https://arxiv.org/abs/1603.06046).
- [Mon16] Ashely Montanaro. Quantum algorithms: an overview. *npj Quantum Information*, 2(15023), 2016.
- [Mor14] Tomoyuki Morimae. Verification for measurement-only blind quantum computing. *Physical Review A*, 89, 2014.
- [MY04] Dominic Mayers and Andrew Yao. Self testing quantum apparatus. *Quantum Information & Computation*, 4:273–286, 2004.
- [NV16] Anand Natarajan and Thomas Vidick. Robust self-testing of many-qubit states, 2016. arXiv preprint [arXiv:1610.03574](https://arxiv.org/abs/1610.03574).
- [Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.
- [RUV12] Ben W. Reichardt, Falk Unger, and Umesh Vazirani. A classical leash for a quantum system: Command of quantum systems via rigidity of CHSH games, 2012. arXiv preprint [arXiv:1209.0448](https://arxiv.org/abs/1209.0448).
- [RUV13] Ben W. Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496:456–460, 2013. Full version [arXiv:1209.0448](https://arxiv.org/abs/1209.0448).
- [Vid17] Thomas Vidick. The Pauli braiding test. Available at <https://mycqstate.wordpress.com/2017/06/28/pauli-braiding/>, 2017.
- [WBMS16] Xingyao Wu, Jean-Daniel Bancal, Matthew McKague, and Valerio Scarani. Device-independent parallel self-testing of two singlets. *Physical Review A*, 93:062121, 2016.