

# Self-testing of a single quantum device under computational assumptions

Tony Metger<sup>1</sup> and Thomas Vidick<sup>2</sup>

<sup>1</sup>Institute for Theoretical Physics, ETH Zürich, 8093 Zürich, Switzerland

<sup>2</sup>Department of Computing and Mathematical Sciences, California Institute of Technology, CA 91125, United States

## Abstract

Self-testing is a method to characterise an arbitrary quantum system based only on its classical input-output correlations, and plays an important role in device-independent quantum information processing as well as quantum complexity theory. Prior works on self-testing require the assumption that the system's state is shared among multiple parties that only perform local measurements and cannot communicate. Here, we replace the setting of *multiple non-communicating* parties, which is difficult to enforce in practice, by a *single computationally bounded* party. Specifically, we construct a protocol that allows a classical verifier to robustly certify that a single computationally bounded quantum device must have prepared a Bell pair and performed single-qubit measurements on it, up to a change of basis applied to both the device's state and measurements. This means that under computational assumptions, the verifier is able to certify the presence of entanglement, a property usually closely associated with two separated subsystems, inside a single quantum device. To achieve this, we build on techniques first introduced by Brakerski et al. (2018) and Mahadev (2018) which allow a classical verifier to constrain the actions of a quantum device assuming the device does not break post-quantum cryptography.

Tony Metger: [tmetger@ethz.ch](mailto:tmetger@ethz.ch)

Thomas Vidick: [vidick@caltech.edu](mailto:vidick@caltech.edu)

Note: a short version of this work has appeared in the [Proceedings of the 12th Innovations in Theoretical Computer Science Conference \(ITCS 2021\)](#).

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Self-testing in the multi- and single-prover settings . . . . .	5
1.2	Cryptographic primitives . . . . .	7
1.3	Our self-testing protocol . . . . .	7
1.4	Soundness proof . . . . .	8
1.5	Discussion . . . . .	9
<b>2</b>	<b>Preliminaries</b>	<b>11</b>
2.1	Notation . . . . .	11
2.2	Extended trapdoor claw-free functions . . . . .	11
2.3	Efficiency and computational indistinguishability . . . . .	12
2.4	Distance measures . . . . .	14
2.5	Lifting state-dependent operator relations using computational indistinguishability . . . . .	20
<b>3</b>	<b>Self-testing protocol</b>	<b>23</b>
3.1	Completeness of self-testing protocol . . . . .	23
<b>4</b>	<b>Soundness of self-testing protocol</b>	<b>25</b>
4.1	Devices . . . . .	26
4.2	Success probabilities of a device . . . . .	29
4.3	Lifting relations from one basis choice to another . . . . .	32
4.4	Uniform normalisation and answers . . . . .	32
4.5	Anti-commutation relations . . . . .	33
4.6	Commutation relations . . . . .	39
4.7	Approximate equality of non-tilde observables and Pauli observables . . . . .	40
4.8	Approximate equality of tilde observables and Pauli observables . . . . .	48
4.9	Products of observables . . . . .	49
4.10	Certifying Bell states . . . . .	52

# 1 Introduction

The *device-independent* approach to quantum information processing treats quantum devices as black boxes which we can interact with classically to observe their input-output correlations. Based solely on these correlations and the assumption that quantum mechanics is correct, the goal is to prove statements about the devices, e.g., to show that they can be used for secure quantum key distribution (see e.g. [MY04]) or delegated quantum computation (see e.g. [RUV13]). At a fundamental level, this provides a theory-of-computation approach to the study of classical signatures of quantum mechanics and their use as a “leash” to control and characterize quantum devices.

*Self-testing* is arguably the most effective method in device-independent quantum information processing. The goal in self-testing is to characterise the quantum state and measurements of multiple black-box quantum devices using only their classical input-output correlations. In analogy to the setting of interactive proof systems, the classical party observing the input-output correlations is sometimes called the *verifier*, and the black-box quantum devices are called *provers*. More specifically, the verifier can interact with multiple quantum provers by sending (classical) questions as inputs and receiving (classical) answers as outputs. The provers can share any (finite-dimensional) entangled quantum state at the start of the interaction and are computationally unbounded; however, it is assumed that after having received the verifier’s questions, the provers can no longer communicate. Based on the question-answer correlations, the verifier would like to deduce that the provers must have shared a certain initial state and performed certain measurements on it, up to a local change of basis on each prover’s Hilbert space. We will describe this scenario in more detail in Section 1.1. We emphasize that self-testing is a uniquely quantum phenomenon: for classical devices, there is simply a function that is implemented by the device, and it is not meaningful to ask *how* the function is implemented “on the inside”. In contrast, for quantum devices, in certain cases knowledge of the function (the observed input-output behaviour) implies an essentially unique realization in terms of a quantum state and measurements on it.

The term *self-testing* was introduced by Mayers and Yao in [MY04] in the context of proofs of security for quantum key distribution, but the notion was already present in earlier works [SW87, PR92]. For a review covering a large number of different self-testing protocols, as well as applications such as randomness expansion and delegated quantum computation, see [SB19]. In addition to more practical applications, self-testing has also proved to be a powerful tool in quantum complexity theory for the study of multi-prover interactive proof systems in the quantum setting and is at the heart of the recent characterisation of the complexity class  $\text{MIP}^*$  [JNV+20].

The starting point for our work is the observation that, while the model of non-communicating quantum provers used in existing self-testing results is appealing in theory, it is difficult to enforce this non-communication assumption in practice. Motivated by the many applications of self-testing in quantum cryptography (e.g. device-independent quantum key distribution) and complexity theory, we are compelled to search for protocols that allow for a self-testing-like certification of a *single* untrusted quantum device.

Self-testing protocols in the multi-prover setting are typically based on the violation of Bell inequalities [Bel64], for which the non-communication assumption is necessary. Hence, different techniques or additional assumptions are necessary when considering the single-device scenario. What could a “computational Bell inequality” look like?

In this paper we give an answer to this question by constructing a self-testing protocol for a *single computationally bounded* quantum device. Specifically, the only assumptions required are the correctness of quantum mechanics and that the prover does not have the ability to break the Learning with Errors (LWE) assumption [Reg09], a common assumption in post-quantum cryptography, *during* the protocol execution (whereas breaking the LWE assumption *after* the end of the protocol is allowed). Our protocol is a three-round interaction between a classical verifier and a quantum prover, at the end of which the verifier decides to either “accept” or “reject” the prover. Informally, the guarantee provided by the protocol is the following:

**Theorem** (Informal). *A prover’s strategy in the protocol is described by a quantum state and the measurements that the prover makes on the state to obtain the (classical) answers received by the verifier. If a*

computationally bounded prover is accepted by the verifier with probability  $1 - \varepsilon$ , then there exists an isometry  $V$  such that for a universal constant  $c > 0$  and under the isometry  $V$ :

- (i) the prover’s state is  $O(\varepsilon^c)$ -close (in trace distance) to a Bell pair,
- (ii) (a subset of) the prover’s measurements are  $O(\varepsilon^c)$ -close to single-qubit measurements in the computational or Hadamard basis, where the measurement bases are chosen by the verifier. Here, “closeness” is measured in a distance measure suitable for measurements acting on a state.

We emphasize that the theorem not only guarantees the preparation of an entangled state by the prover, but also the implementation of specific measurements on it. As such, it provides a complete analogue of foundational self-testing results for the CHSH inequality [SW87, MYS12].

The proof of our main result builds on techniques introduced in recent works [Mah18, BCM+18, GV19] to allow a classical verifier to leverage post-quantum cryptography to control a computationally bounded quantum prover. Because they are relevant for understanding the proof of our results, we now give a brief overview of these works and explain their relation to self-testing.

In [Mah18], Mahadev gives the first protocol to classically verify a delegated quantum computation with a single untrusted quantum prover. The central ingredient in Mahadev’s verification protocol is a “measurement protocol” that allows the verifier to force the prover to report classical outcomes obtained by performing certain measurements on a quantum state that the prover has “committed to” using classical information. The main guarantee of the measurement protocol is this: if the prover is accepted in the protocol, *there exists* a quantum state such that the distribution over the prover’s answers could have been produced by performing the requested measurements on this state. In other words, all of the prover’s answers must be self-consistent in the sense that they could have originated from performing different measurements on (copies of) the same quantum state.

To verify a quantum computation, the statement that the prover’s answers are *consistent* with measurements on a quantum state is sufficient, as the *existence* of a quantum state with the right properties can certify the outcome of the quantum computation (this is due to Kitaev’s “circuit-to-Hamiltonian” construction, which we do not explain here). However, in this work we seek to make a stronger statement: we want to certify that the prover *actually constructed* the desired quantum state *and performed the desired measurements* on it (up to an isometry). While the honest prover in Mahadev’s protocol does indeed construct the desired quantum state, the protocol does not guarantee that an arbitrary prover must do, too. Hence, our self-testing protocol is stronger in the sense that it allows for a more stringent characterisation of the prover’s actions, namely its actual states and measurements. To emphasize the difference, we note that the guarantee of Mahadev’s protocol *does not* directly imply that a successful prover must have performed any quantum computation; the guarantee is only that, if the correct state preparation and measurements were to be performed, the outcome would be as claimed by the prover.

Another closely related work is that of Brakerski et al. [BCM+18], who give a protocol between a classical verifier and a quantum prover that allows the verifier to generate certified information-theoretic randomness, again assuming that the prover does not break the LWE assumption; in other words, their protocol generates information-theoretic randomness from a computational assumption. For this, the authors show that two of the prover’s measurements must be maximally incompatible, as defined by a quantity that they call the “overlap”. Informally, one can think of two maximally incompatible measurements as being close to a computational and Hadamard basis measurement, up to some global change of basis. Hence, this result already resembles self-testing in the sense that the verifier can make a statement about the actual measurements used by the prover. In particular, it does serve as a “test of quantumness” for the prover.

Building on [BCM+18] and using techniques from [Mah18], Gheorghiu and Vidick construct a protocol for a task that they call verifiable remote state preparation (RSP) [GV19]. They consider a set of single-qubit pure states  $\{|\psi_1\rangle, \dots, |\psi_n\rangle\}$ . Under the same LWE assumption as before, the protocol enables the verifier to certify that the prover has prepared one of these states, up to a global change of basis (i.e., some isometry  $V$  that is applied to all  $|\psi_i\rangle$ ). More precisely, the verifier cannot decide beforehand on a particular  $|\psi_i\rangle$ , but after executing the protocol, the verifier knows which  $|\psi_i\rangle$  the prover has prepared, and

the distribution over  $i$  can be made uniform. The prover, on the other hand, does not know which  $|\psi_i\rangle$  he has prepared.

This result resembles a self-testing statement even more than that of [BCM<sup>+</sup>18] because it explicitly characterises a family of single-qubit quantum states, one of which is certified to be present in the prover’s space. However, it differs from a standard self-testing statement in that it is defined for a family of states, not an individual state: because the prover’s isometry  $V$  is arbitrary, any individual state  $|\psi_i\rangle$  can be mapped to another arbitrary state. Hence, what is certified in RSP is not any individual state, but the relationships (e.g., orthogonality) between different states in some family. Alternatively, one can also take the view that RSP characterises the relationships between the prover’s states and measurements. We return to this issue in more detail in Section 1.1. The idea of certifying a family of states has also been considered by Cojocaru et al. [CCKW19], who call this notion “blind self-testing”. They analyze a different protocol under a restricted adversarial model and conjecture that their protocol yields similar guarantees as [GV19] for single-qubit states and tensor products of single-qubit states.

This lengthy overview of previous works makes explicit a progression towards the task that we tackle here, that of genuine self-testing of a single quantum device. We note that this presentation clearly benefits from hindsight, and that none of the cited works mentions any relation to self-testing; indeed, the results are too weak to be used in this setting. In particular, none of the previous works provides a sufficiently strong guarantee on the measurements performed by the quantum device and goes beyond the setting of a single qubit, which is arguably the main technical challenge. Indeed, moving from a single-qubit state to an entangled two-qubit state means that the verifier has to enforce a tensor product structure on the prover’s space, which is one of the main difficulties in our soundness proof (Section 4). On a technical level, it requires the certification of compatibility relations between different measurements meant to act on different qubits. Additionally, having two qubits instead of one prevents us from using Jordan’s lemma, a standard tool in self-testing also used in [GV19], to characterise the prover’s measurements; in Section 4.7, we show how to characterise the prover’s measurements using a different method starting with a *partial* characterisation of the prover’s measurements, using that to partially characterise the prover’s states, which in turn is used for a stronger partial characterisation of the measurements, etc., until we reach the full statement that shows that the prover makes single-qubit measurements on a Bell pair.

## 1.1 Self-testing in the multi- and single-prover settings

In this section, we give a brief overview of the standard multi-prover self-testing scenario, and explain how it can be extended to a single prover. For more details on the multi-prover scenario, see [ŠB19] or [Sca19, Chapter 7]. For simplicity, let us consider the case of two provers  $A$  and  $B$ , with Hilbert spaces  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , respectively. Hence, the total Hilbert space is  $\mathcal{H}_A \otimes \mathcal{H}_B$ . The verifier interacts with  $A$  and  $B$  by sending questions and receiving answers. The question-answer correlations can be described by a family of probability distributions  $\{p(a, b|x, y)\}_{x, y}$ , where for each choice of questions  $x$  and  $y$  sent to  $A$  and  $B$ , respectively,  $p(a, b|x, y)$  is a probability distribution over their answers  $a$  and  $b$ . We say that a quantum state  $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$  is *compatible* with the correlations  $p(a, b|x, y)$  if there are local measurements  $\{P_x^{(a)}\}_a$  on  $\mathcal{H}_A$  for every input  $x$ , and  $\{Q_y^{(b)}\}_b$  on  $\mathcal{H}_B$  for every input  $y$ , that realise the correlations  $p(a, b|x, y)$ , i.e.,  $p(a, b|x, y) = \langle \psi | P_x^{(a)} \otimes Q_y^{(b)} | \psi \rangle_{AB}$  for all  $x, y, a, b$ .

**Definition 1.1** (Self-testing of states, informal). The correlations  $p(a, b|x, y)$  self-test a state  $|\phi\rangle_{AB}$  if for any state  $|\psi\rangle_{AB}$  compatible with these correlations, there exists a *local* isometry  $V = V_A \otimes V_B$  (with  $V_A$  only acting on  $\mathcal{H}_A$ , and  $V_B$  only acting on  $\mathcal{H}_B$ ) such that  $V|\psi\rangle_{AB} = |\phi\rangle_{AB}|AUX\rangle$  for some ancillary state  $|AUX\rangle$ .

A more operational view of this statement is that it must be possible to “extract” the state  $|\phi\rangle_{AB}$  from  $|\psi\rangle_{AB}$  only by performing local operations. The condition that the isometry must be local is crucial: if we would allow a global isometry, we could map any state  $|\psi\rangle_{AB}$  to the desired state  $|\phi\rangle_{AB}$ . In the two-prover case, the notion of a *local* isometry is natural, since the separation between the two provers induces a tensor product structure  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$  on the global Hilbert space  $\mathcal{H}$ . In contrast, for a single prover no such tensor product structure exists and we cannot define *local* isometries in a meaningful way.

In Definition 1.1, we only dealt with the provers' state, not his measurements. A stronger notion of self-testing is to characterise both the provers' state and measurements. This is the version of self-testing originally considered by Mayers and Yao [MY04], and we will see that it can be meaningfully extended to the single-prover setting.

**Definition 1.2** (Self-testing of states and measurements, informal). The correlations  $p(a, b|x, y)$  self-test a state  $|\phi\rangle_{AB}$  and measurements  $\{M_x^{(a)}\}, \{N_y^{(b)}\}$  if for any state  $|\psi\rangle_{AB}$  and measurements  $\{P_x^{(a)}\}, \{Q_y^{(b)}\}$  that realise the correlations  $p(a, b|x, y)$ , there exists a local isometry  $V = V_A \otimes V_B$  such that

- (i)  $V|\psi\rangle_{AB} = |\phi\rangle_{AB}|AUX\rangle$ ,
- (ii)  $V(P_x^{(a)} \otimes Q_y^{(b)})|\psi\rangle_{AB} = \left( (M_x^{(a)} \otimes N_y^{(b)})|\phi\rangle_{AB} \right) |AUX\rangle$ , for some ancillary state  $|AUX\rangle$ .

The first condition is the same as in Definition 1.1. The second condition roughly says that the “physical” measurements  $\{P_x^{(a)}\}$  and  $\{Q_y^{(b)}\}$  used by  $A$  and  $B$ , respectively, act on the state  $|\psi\rangle_{AB}$  in the same way that the desired measurements  $\{M_x^{(a)}\}$  and  $\{N_y^{(b)}\}$  act on the desired state  $|\phi\rangle_{AB}$ .

Self-testing of states and measurements still has meaning in the single-prover setting. In this setting, one can imagine that the verifier sends both questions  $x$  and  $y$  to the same prover, and the prover replies with two answers  $a$  and  $b$ . To compute his answers, the prover prepares a quantum state  $|\psi\rangle$  and, on inputs  $x, y$ , performs a measurement  $\{P_{x,y}^{(a,b)}\}_{a,b}$  to obtain answers  $a, b$ .

**Definition 1.3** (Self-testing for a single prover, informal). The correlations  $p(a, b|x, y)$  self-test a state  $|\phi\rangle$  and measurements  $\{K_{x,y}^{(a,b)}\}_{a,b}$  if for any state  $|\psi\rangle$  and measurements  $\{P_{x,y}^{(a,b)}\}_{a,b}$  that realise the correlations  $p(a, b|x, y)$ , there exists an isometry  $V$  such that

- (i)  $V|\psi\rangle = |\phi\rangle|AUX\rangle$ ,
- (ii)  $V P_{x,y}^{(a,b)} |\psi\rangle = \left( K_{x,y}^{(a,b)} |\phi\rangle \right) |AUX\rangle$ , for some ancillary state  $|AUX\rangle \in \mathcal{H}'$ .

This definition is rather informal because whenever the number of possible questions and answers is fixed and independent of the security parameter (as is the case in this paper), single-round question-answer correlations  $p(a, b|x, y)$  alone cannot be sufficient: a prover can always succeed in the protocol simply by answering the verifier's questions according to a look-up table; such a prover is classical and does not actually perform any computation. Therefore, our protocol will have multiple rounds of interaction between the verifier and the prover: the questions and answers in the initial “setup rounds” will involve a public key that scales with the security parameter; then, in the last round, the verifier observes question-answer correlations  $p(a, b|x, y)$  similar to standard self-testing, i.e., with a fixed question and answer length. Instead of using multi-round interaction, one could also try to build a single-round protocol with questions that depend on the security parameter (e.g., the question would include a public key). A number of recent works have shown that under the (quantum) random oracle assumption, the protocol for certifying the quantumness of a prover from [BCM<sup>+</sup>18] and the verification protocol from [Mah18] can be adapted to this single-round setting [ACGH19, CCY19, BKVV20]. We leave it for future work to investigate whether the interaction in our protocol can also be removed with the random oracle assumption.

To obtain a statement that is more similar to the two-prover scenario, we consider the stronger constraint that the desired measurements have a tensor product form  $K_{x,y}^{(a,b)} = M_x^{(a)} \otimes N_y^{(b)}$ . In particular, this means that answer  $a$  only depends on question  $x$  and  $b$  only depends on  $y$ , and it enforces a natural tensor product structure on the prover's space.

Specifically, we define Hilbert spaces  $\mathcal{H}_A, \mathcal{H}_B$  and  $\mathcal{H}'$  and deduce the existence of an isometry  $V$  from the prover's physical space  $\mathcal{H}$  to  $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}'$  such that under the isometry, the measurements operators  $P_{x,y}^{(a,b)}$  act on  $|\psi\rangle$  in the same way that tensor product measurement operators of the form  $M_x^{(a)} \otimes N_y^{(b)}$  act on  $|\phi\rangle_{AB}$ , where  $M_x^{(a)}$  acts only on  $\mathcal{H}_A$ ,  $N_y^{(b)}$  acts only on  $\mathcal{H}_B$ , and  $|\phi\rangle_{AB}$  is the state that we are self-testing for (e.g., a Bell state).

**Definition 1.4** (Self-testing of tensor product strategies for a single prover, informal). The correlations  $p(a, b|x, y)$  self-test a state  $|\phi\rangle_{AB}$  and measurements  $\{M_x^{(a)}\}$  on system  $A$  and  $\{N_y^{(b)}\}$  on system  $B$  if for any state  $|\psi\rangle \in \mathcal{H}$  and measurements  $\{P_{x,y}^{(a,b)}\}_{a,b}$  on  $\mathcal{H}$  that realise the correlations  $p(a, b|x, y)$ , there exists an isometry  $V : \mathcal{H} \rightarrow \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}'$  such that

- (i)  $V|\psi\rangle = |\phi\rangle_{AB}|AUX\rangle$ ,
- (ii)  $V P_{x,y}^{(a,b)} |\psi\rangle = \left( (M_x^{(a)} \otimes N_y^{(b)}) |\phi\rangle_{AB} \right) |AUX\rangle$ , for some ancillary state  $|AUX\rangle \in \mathcal{H}'$ .

Again, this definition is informal for the same reason as for Definition 1.3. A formal statement of such a single-prover self-testing result with a tensor product structure is given in Theorem 4.42, the main result of this paper.

## 1.2 Cryptographic primitives

The main cryptographic primitive underlying our self-testing protocol is a so-called extended noisy trapdoor claw-free function family (ENTCF family). ENTCF families were introduced by Mahadev in [Mah18], building on the construction of noisy trapdoor claw-free function families by Brakerski et al. in [BCM<sup>+</sup>18]. Here, we only give a brief informal description of the main properties of an ENTCF family (see Section 2.2 for references and details).

An ENTCF family consists of two families  $\mathcal{F}$  and  $\mathcal{G}$  of function pairs. A function pair  $(f_{k,0}, f_{k,1}) \in \mathcal{F}$  is called a *claw-free pair* and is indexed by a public key  $k$ . Similarly, an *injective pair* is a pair of functions  $(f_{k,0}, f_{k,1}) \in \mathcal{G}$ , also indexed by a public key  $k$ . Informally, the most important properties are the following:

- (i) For fixed  $k \in \mathcal{K}_{\mathcal{F}}$ ,  $f_{k,0}$  and  $f_{k,1}$  are bijections with the same image, i.e., for every  $y$  in their image there exists a unique pair  $(x_0, x_1)$ , called a *claw*, such that  $f_{k,0}(x_0) = f_{k,1}(x_1) = y$ .
- (ii) Given a key  $k \in \mathcal{K}_{\mathcal{F}}$  for a claw-free pair, it is quantum-computationally intractable (without access to trapdoor information) to compute both a preimage  $x_i$  and a single generalised bit of  $x_0 \oplus x_1$  (i.e.,  $d \cdot (x_0 \oplus x_1)$  for any non-trivial bit string  $d$ ), where  $(x_0, x_1)$  forms a valid claw. This is called the *adaptive hardcore bit property*.
- (iii) For fixed  $k \in \mathcal{K}_{\mathcal{G}}$ ,  $f_{k,0}$  and  $f_{k,1}$  are injective functions with disjoint images.
- (iv) Given a key  $k \in \mathcal{K}_{\mathcal{F}} \cup \mathcal{K}_{\mathcal{G}}$ , it is quantum-computationally hard (without access to trapdoor information) to determine the “function type”, i.e., to decide whether  $k$  is a key for a claw-free or an injective pair. This is called *injective invariance*.
- (v) For every key  $k \in \mathcal{K}_{\mathcal{F}} \cup \mathcal{K}_{\mathcal{G}}$ , there exists a trapdoor  $t_k$ , which can be sampled together with  $k$  and with which (ii) and (iv) are computationally easy.

## 1.3 Our self-testing protocol

We now give an informal description of our self-testing protocol with the honest prover behaviour and provide some intuition for its soundness. A full description of the protocol is given in Figure 1, and for a more detailed overview of the soundness proof, see the introduction to Section 4.

On a very high level, one can view the protocol as first executing the RSP protocol from [GV19] twice in parallel to prepare two qubits in the provers space. Then, the prover is asked to perform an entangling operation on these two qubits. Because the prover does not know which states the qubits are in, and the entangling operation acts differently on different states, to pass the checks in the protocol the prover has to apply the entangling operation honestly.

In more detail, the protocol begins with the verifier sampling two uniformly random bits  $\theta_1, \theta_2$ , each bit denoting a *basis choice* (either the computational or the Hadamard basis). The case where both bits denote the Hadamard basis will be the one where the prover prepares a Bell pair, whereas the other basis choices

serve as tests that prevent the prover from cheating. Depending on these basis choices, the verifier then samples two key-trapdoor pairs  $(k_1, t_{k_1})$  and  $(k_2, t_{k_2})$  from the ENTCF family: for the computational basis, it samples an injective pair, and for the Hadamard basis a claw-free pair. The verifier sends the keys to the prover and keeps the trapdoors private.

The honest prover treats the two keys separately. For each key  $k_i$ , he prepares the state

$$|\psi_i\rangle = \frac{1}{\sqrt{2|\mathcal{X}|}} \sum_{x \in \mathcal{X}, b \in \{0,1\}} |b\rangle |x\rangle |f_{k_i,b}(x)\rangle. \quad (1.1)$$

Here,  $\mathcal{X}$  is the domain of the ENTCF family. Note that even though the prover does not know which kind of function (claw-free or injective) he is dealing with, the definition of ENTCF families still allows him to construct this state. The prover now measures both image registers (i.e., the registers storing “ $f_{k_i,b}(x)$ ”), obtains images  $y_1, y_2$ , and sends these to the verifier. (In the terminology of [Mah18], this is called a “commitment”.) Depending on the choice of function family by the verifier, the prover’s post-measurement state has one of two forms: if the verifier sampled the key  $k_i$  from the injective family, the post-measurement state is a computational basis state:

$$|\psi'_i\rangle = |b\rangle |x_b\rangle, \quad (1.2)$$

where  $x_b$  is the unique preimage of  $y_i$ . If the key  $k_i$  belongs to a claw-free family, the post-measurement state is a superposition over a claw:

$$|\psi'_i\rangle = \frac{1}{\sqrt{2}}(|0\rangle |x_0\rangle + |1\rangle |x_1\rangle), \quad (1.3)$$

where  $(x_0, x_1)$  form a claw, i.e.,  $f_{k_i,0}(x_0) = f_{k_i,1}(x_1) = y$ .

At this point, the verifier selects a round type, either a “preimage round” or a “Hadamard round”, uniformly at random and sends the round type to the prover. For a preimage round, the honest prover measures his entire state in the computational basis and returns the result; the verifier checks that the prover has indeed returned correct preimages for the submitted  $y_1, y_2$ . The preimage round is an additional test that is required for us to leverage the adaptive hardcore bit property, but we do not discuss this further in this overview.

For a Hadamard round, the honest prover measures both of his preimage registers (i.e., the registers containing “ $x_b$ ”) in the Hadamard basis, obtains two bit strings  $d_1, d_2$ , and sends these to the verifier. This results in the following states (using the notation from above):

$$|\psi''_i\rangle = \begin{cases} |b\rangle & \text{if } k_i \text{ belongs to an injective family,} \\ \frac{1}{\sqrt{2}}(|0\rangle + (-1)^{d_i \cdot (x_0 \oplus x_1)} |1\rangle) & \text{if } k_i \text{ belongs to a claw-free family.} \end{cases} \quad (1.4)$$

Note that the phase in the second case is exactly the adaptive hardcore bit from the definition of ENTCF families. At this point, the verifier selects two additional bases  $q_1, q_2$  uniformly at random (again from either the computational or Hadamard basis), and sends these to the prover. In analogy to self-testing, we call these bases “questions”. The honest prover now applies a  $CZ$  gate (an entangling two-qubit gate that applies a  $\sigma_Z$  operation to the second qubit if the first qubit is in state  $|1\rangle$ ) to its state  $|\psi''_1\rangle |\psi''_2\rangle$ . In the case where both  $\theta_1$  and  $\theta_2$  specify the Hadamard basis, this results in a Bell state (rotated by a single-qubit Hadamard gate). The prover measures the individual qubits of the resulting state in the bases specified by  $q_1, q_2$ . The outcomes  $v_1, v_2$  are returned to the verifier.

The verifier can use the prover’s answers  $y_1, y_2, d_1, d_2$  and her trapdoor information  $t_{k_1}, t_{k_2}$  to determine which state  $CZ|\psi''_1\rangle |\psi''_2\rangle$  the prover should have prepared. The verifier accepts the prover if his answers  $v_1, v_2$  are consistent with making the measurements specified by  $q_1, q_2$  on the honest prover’s state  $CZ|\psi''_1\rangle |\psi''_2\rangle$ .

## 1.4 Soundness proof

We now give a brief intuition for the soundness of the protocol; the actual soundness proof is given in Section 4. Let us first consider a version of the protocol where the prover is not supposed to perform a  $CZ$  operation.



As noted before, this would be (a simplified version of) the RSP protocol [GV19], executed twice in parallel. For the purposes of this overview, let us assume that the only way for the prover to pass these two parallel executions of the RSP protocol is to treat each execution separately, i.e., use a tensor product Hilbert space  $\mathcal{H}_1 \otimes \mathcal{H}_2$  and execute each instance of the RSP protocol on a different part of the space (enforcing such a tensor product structure is reminiscent of the classic question of parallel repetition [Raz98] and is actually one of the main difficulties in our soundness proof, but we leave the details of this for Section 4). It now follows from the security of the RSP protocol that the prover must have prepared one of  $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$  in each part of his space (up to a “local” change of basis for each space), but he does not know which one.

Now consider how a  $CZ$  operation acts on these different states: if both states are Hadamard basis states (e.g.,  $|+\rangle|-\rangle$ ), the  $CZ$  operation will entangle them and produce a Bell state (rotated by a single-qubit Hadamard gate); in contrast, if at least one of the states is a computational basis state (e.g.,  $|1\rangle|-\rangle$ ), the resulting state will still be a product state of computational and Hadamard basis states (albeit a different one). This means that in the latter case, the  $CZ$  operation essentially only relabels the states. Therefore, if the verifier adapts her checks to account for the relabelling, in the latter case the guarantees from the RSP protocol still hold. Because the prover does not know which bases the verifier has selected, we can extend these guarantees to the case of two Hadamard basis states, too.

We stress that this only provides a rough intuition, and that the actual proof proceeds quite differently from this because we cannot just assume the existence of a tensor product structure on the prover’s Hilbert space. Deducing this tensor product structure poses technical difficulties. In two-prover self-testing proofs, the first step is to show that the measurement operators used by each prover approximately satisfy certain relations, e.g. anti-commutation. Because the measurement operators of different provers act on different Hilbert spaces, they exactly commute. Combining the approximate relations from the first step with the exact commutation relations, one can show that the prover’s measurement operators must be close to some desired operators, e.g. the Pauli operators. This last “rounding step” typically uses Jordan’s lemma or a stability theorem for approximate group representations [GH17]. In our case, we cannot show exact commutation relations between operators — commutation can only be enforced via the protocol, which tolerates a small failure probability. Hence, we are only able to show approximate commutation relations, which prevents us from applying Jordan’s lemma or the result of [GH17]. We therefore develop an alternative approach to “rounding” the prover’s operators that only requires approximate commutation and leverages the cryptographic assumptions. This method might also be useful for other applications that require a very tight “cryptographic leash” on a quantum prover.

## 1.5 Discussion

Self-testing has developed into a versatile tool for quantum information processing and quantum complexity theory and presents one of the strongest possible black-box certification techniques of quantum devices. The standard self-testing setting involves multiple non-communicating quantum provers, which is difficult to enforce in practice. The main contribution of this paper is the construction of a self-testing protocol that allows a classical verifier to certify that a single computationally bounded quantum prover has prepared a Bell state and measured the individual qubits of the state in the computational or Hadamard basis, up to a global change of basis applied to both the state and measurements. This means that we are able to certify the existence of entanglement in a single quantum device.

Due to the interactive nature of our protocol, this certification remains valid even if it turned out that any quantum computation is classically simulable, i.e.,  $BQP = BPP$ . It therefore constitutes a “test of quantumness” in the sense of [BCM<sup>+</sup>18] and differs from proposals for testing quantum supremacy such as [BFNV19], which only certify the “quantumness” of a device under the assumption that  $BQP \neq BPP$ .

Existing multi-prover self-testing protocols are typically based on non-local games, e.g., the CHSH game [MYS12]. Our self-testing protocol follows a more “custom” approach guided by the available cryptographic primitives. While this enables us to construct a single-prover self-test for single-qubit measurements on a Bell state, arguably the most important quantum state for many applications, it does not allow us to extend the result to other states for which multi-prover self-tests are known [CGS17]. To better make use of the extensive existing self-testing literature, it would be desirable to construct a procedure that allows for the

“translation” of multi-prover non-local games to single-prover games with computational assumptions. In classical cryptography, similar attempts have been made to construct single-prover argument systems from multi-prover proof systems using fully homomorphic encryption [ABOR00, KRR14, DHRW16].

Another approach to constructing single-prover self-tests for a larger class of states might be to strengthen Mahadev’s measurement protocol [Mah18] from guaranteeing the existence of a state compatible with the measurement results to certifying that the prover actually has prepared this state. As a step in this direction, the second author and Zhang recently showed that Mahadev’s protocol is a classical proof of quantum knowledge [VZ20]. The concept of a proof of quantum knowledge, first introduced in [BG20, CVZ20] for the setting of a quantum verifier and extended to the setting of a classical verifier in [VZ20], is still less stringent than a self-test and in particular lacks the strong characterisation of the prover’s measurements that we obtain in self-testing.

Beyond the conceptual appeal of gaining more fine-grained control over untrusted quantum devices, our self-testing protocol presents a first step towards translating multi-prover protocols for applications such as delegated computation [RUV13, CGJV19], randomness expansion [Col06, VV12, MS17], or secure multi-party quantum computation [CGS02, BCG<sup>+</sup>06] to a single-prover setting. There are already computationally secure single-prover protocols for delegated quantum computation [Mah18] and randomness expansion [BCM<sup>+</sup>18]; however, establishing a more general link between self-testing-based multi-prover protocols and computationally secure single-prover protocols is still desirable: it might lead to conceptually simpler single-prover protocols and will be useful for constructing single-prover protocols for other applications without resorting to a low-level cryptographic analysis.

For example, using our self-testing theorem in a black-box way, the first author and others have recently constructed a protocol for device-independent quantum key distribution (DIQKD) [MDCAF20]. In contrast to previous DIQKD protocols, which rely on a non-communication similar to the one in standard self-testing, this new DIQKD protocol requires no non-communication assumption and more closely models how DIQKD protocols are expected to be implemented experimentally. Crucially, the security analysis of this DIQKD protocol can be reduced to our self-testing theorem without any intricate cryptographic analysis involving computational hardness assumptions.

We believe that, in a similar vein, our protocol will also serve as a useful building block for other future protocols for computationally bounded quantum devices, in the same way that self-testing for EPR pairs in the multi-prover scenario has proved to be a versatile tool in physics, cryptography, and complexity theory.

**Organisation.** The paper is organised as follows. In Section 2, we give preliminary definitions and technical lemmas, most importantly involving the state-dependent distance between operators. In Section 3, we describe our self-testing protocol and show that it has completeness negligibly close to 1, i.e., that there exists an honest prover that is accepted with all but negligible probability. In Section 4, we show that our protocol is sound, meaning that any prover that is accepted with high probability must use states and measurements close to the desired ones. The main result that formalises this statement is Theorem 4.42.

**Acknowledgements.** We thank Andrea Coladangelo, Andru Gheorghiu, Anand Natarajan, and Tina Zhang for helpful discussions; Andrea Coladangelo, Andru Gheorghiu, Urmila Mahadev, Akihiro Mizutani, and the Quantum referees for valuable comments on the manuscript; and Lídia del Río for pointing out the reference [BRV<sup>+</sup>19]. Tony Metger acknowledges support from ETH Zürich and the ETH Foundation through the *Excellence Scholarship & Opportunity Programme*, from the IQIM, an NSF Physics Frontiers Center (NSF Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028), and from the National Centres of Competence in Research (NCCRs) QSIT and SwissMAP. Thomas Vidick is supported by NSF CAREER Grant CCF-1553477, AFOSR YIP award number FA9550-16-1-0495, a CIFAR Azrieli Global Scholar award, MURI Grant FA9550-18-1-0161, and the IQIM, an NSF Physics Frontiers Center (NSF Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028). Most of this work was carried out while Tony Metger was a visiting student researcher at the Department of Computing and Mathematical Sciences at Caltech.

## 2 Preliminaries

This section establishes a number of definitions and technical lemmas that we will use in the soundness proof in Section 4. We assume basic familiarity with quantum mechanics and start with a description of the notation in this paper. On a first reading of this paper, we recommend skipping most of the preliminary section and only referring back to the relevant results when they are referenced in the soundness proof. The most relevant parts for a first reading are Section 2.2, Definition 2.10, Definition 2.14, Lemma 2.21, and Lemma 2.25.

### 2.1 Notation

For a bit  $b \in \{0, 1\}$ ,  $\bar{b}$  denotes the flipped bit, i.e.,  $\bar{b} = b \oplus 1$ .

A function  $n : \mathbb{N} \rightarrow \mathbb{R}_+$  is called *negligible* if  $\lim_{\lambda \rightarrow \infty} n(\lambda)p(\lambda) = 0$  for any polynomial  $p$ . We use  $\text{negl}(\lambda)$  to denote an arbitrary negligible function.

We use  $\mathcal{H}$  to denote an arbitrary finite-dimensional Hilbert space, and use indices to differentiate between distinct spaces. The set of linear operators on a Hilbert space  $\mathcal{H}$  is denoted  $\mathcal{L}(\mathcal{H})$ , and the set of unitary operators is  $\mathcal{U}(\mathcal{H})$ . The map  $\text{Tr} : \mathcal{L}(\mathcal{H}) \rightarrow \mathbb{C}$  denotes the trace, and  $\text{Tr}_B : \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \rightarrow \mathcal{L}(\mathcal{H}_A)$  is the partial trace over subsystem  $B$ .  $\text{Pos}(\mathcal{H})$  denotes the set of positive semidefinite operators on  $\mathcal{H}$ , and  $\mathcal{D}(\mathcal{H}) = \{A \in \text{Pos}(\mathcal{H}) \mid \text{Tr}[A] = 1\}$  is the set of density matrices on  $\mathcal{H}$ .

For  $A \in \mathcal{L}(\mathcal{H})$  and  $p \in \mathbb{N}$ , the Schatten  $p$ -norm is  $\|A\|_p = \text{Tr}[|A|^p]^{1/p}$  with  $|A| = \sqrt{A^\dagger A}$ , and  $\|A\|_\infty$  is the operator norm (largest singular value). For  $A, B \in \mathcal{L}(\mathcal{H})$ , we use the commutator  $[A, B] = AB - BA$  and the anti-commutator  $\{A, B\} = AB + BA$ . The single qubit Pauli operators are  $\sigma_X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$  and  $\sigma_Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ . The Hadamard basis states are written as  $|(-)^b\rangle = \frac{1}{\sqrt{2}}(|0\rangle + (-1)^b|1\rangle)$ .

An *observable* on  $\mathcal{H}$  is a Hermitian linear operator on  $\mathcal{H}$ . A *binary observable* is an observable that only has eigenvalues  $\in \{-1, 1\}$ . For a binary observable  $O$  and  $b \in \{0, 1\}$ , we denote by  $O^{(b)}$  the projector onto the  $(-1)^b$ -eigenspace of  $O$ . For any procedure which takes a quantum state as input and produces a bit (or more generally an integer) as output, e.g., by measuring the input state, we denote the probability distribution over outputs  $b$  on input state  $\psi$  by  $\text{Pr}[b \mid \psi]$ .

The self-testing protocol that we will introduce in Section 3 has a *security parameter*  $\lambda$ . The quantities in the rest of the paper are typically families indexed by this security parameters, but we leave this implicit most of the time.

### 2.2 Extended trapdoor claw-free functions

As mentioned in Section 1.2, we rely on a cryptographic primitive called extended noisy trapdoor claw-free function families (ENTCF families) [BCM<sup>+</sup>18, Mah18], a brief description of which was also given in Section 1.2. We refer the reader to [Mah18, section 4] for the formal definition of ENTCF families, and we will use the notation therein throughout the rest of this paper. For the construction of ENTCF families from the Learning with Errors problem [Reg09], see [BCM<sup>+</sup>18, section 4] and [Mah18, Section 9]. These works also give the conditions which parameters in the construction need to satisfy. Security of our protocol holds under the same conditions. Since the conditions are quite involved, we do not reproduce them here. For convenience, we define the following maps that “decode” the output of an ENTCF.

**Definition 2.1** (Decoding maps).

- (i) For a key  $k \in \mathcal{K}_G \cup \mathcal{K}_F$ , an image  $y \in \mathcal{Y}$ , a bit  $b \in \{0, 1\}$ , and a preimage  $x \in \mathcal{X}$ , we define  $\text{CHK}(k, y, b, x)$  to return 1 if  $y \in \text{SUPP}(f_{k,b}(x))$ , and 0 otherwise. (This definition is as in [Mah18, Definition 4.1 and 4.2]).
- (ii) For a key  $k \in \mathcal{K}_G$  and a  $y \in \mathcal{Y}$ , we define  $\hat{b}(k, y)$  by the condition  $y \in \cup_x \text{SUPP}(f_{k, \hat{b}(k,y)}(x))$ . (This is well-defined because  $f_{k,1}$  and  $f_{k,2}$  form an injective pair.)

- (iii) For a key  $k \in \mathcal{K}_{\mathcal{G}} \cup \mathcal{K}_{\mathcal{F}}$  and a  $y \in \mathcal{Y}$ , we define  $\hat{x}_b(k, y)$  by the condition  $y \in \text{SUPP}(f_{k,b}(\hat{x}_b(k, y)))$ , and  $\hat{x}_b(k, y) = \perp$  if  $y \notin \cup_x \text{SUPP}(f_{k,b}(x))$ . For  $k \in \mathcal{K}_{\mathcal{G}}$ , we also use the shorthand  $\hat{x}(k, y) = \hat{x}_{\hat{b}(k,y)}(k, y)$ .
- (iv) For a key  $k \in \mathcal{K}_{\mathcal{F}}$ , a  $y \in \mathcal{Y}$ , and a  $d \in \{0, 1\}^w$ , we define  $\hat{u}(k, y, d)$  by the condition  $d \cdot (\hat{x}_0(k, y) \oplus \hat{x}_1(k, y)) = \hat{u}(k, y, d)$ .

### 2.3 Efficiency and computational indistinguishability

In this section, we define what it means for actions performed by a quantum device, e.g., unitaries or measurements, to be efficient. We also define the notion of computational indistinguishability for quantum states. In these definitions, we make the dependence on the security parameter  $\lambda$  explicit for the sake of clarity.

**Definition 2.2** (Efficient unitaries, isometries, measurements, and observables).

- (i) We call a family of unitaries  $\{U_\lambda \in \mathcal{U}(\mathcal{H}_\lambda)\}_{\lambda \in \mathbb{N}}$  efficient if there exists a (classical) polynomial-time Turing machine  $M$  that, on input  $1^\lambda$ , outputs a description of a circuit (with a fixed gate set) that implements the unitary.
- (ii) We call a family of isometries  $\{V_\lambda : \mathcal{H}_{A_\lambda} \rightarrow \mathcal{H}_{B_\lambda}\}_{\lambda \in \mathbb{N}}$  efficient if there exists an efficient family of unitaries  $\{U_\lambda \in \mathcal{U}(\mathcal{H}_{B_\lambda})\}_{\lambda \in \mathbb{N}}$  such that  $V_\lambda = U_\lambda(\mathbb{1}_{A_\lambda} \otimes |0_{k(\lambda)}\rangle)$ , where  $k(\lambda) = \dim(\mathcal{H}_{B_\lambda}) - \dim(\mathcal{H}_{A_\lambda})$ .
- (iii) We call a family of binary observables  $\{Z_\lambda \in \text{Herm}(\mathcal{H}_{A_\lambda})\}_{\lambda \in \mathbb{N}}$  efficient if there exists a family of Hilbert spaces  $\mathcal{H}_{B_\lambda}$  with  $\dim \mathcal{H}_{B_\lambda} = \text{poly}(\lambda)$ , and a family of efficient unitaries  $\{U_\lambda \in \mathcal{U}(\mathcal{H}_{A_\lambda} \otimes \mathcal{H}_{B_\lambda})\}_{\lambda \in \mathbb{N}}$  such that for any  $|\psi\rangle_A \in \mathcal{H}_A$ :

$$U_\lambda^\dagger(\sigma_Z \otimes \mathbb{1})U_\lambda(|\psi\rangle_A |0\rangle_B) = (Z_\lambda |\psi\rangle_A) \otimes |0\rangle_B. \quad (2.1)$$

- (iv) We call a family of measurements  $\{M_\lambda = \{M_\lambda^{(i)} \in \mathcal{L}(\mathcal{H}^{A_\lambda})\}_{i \in \mathcal{A}}\}_{\lambda \in \mathbb{N}}$  efficient if the isometry

$$|\psi\rangle \mapsto \sum_{i \in \mathcal{A}} |i\rangle \otimes M^{(i)} |\psi\rangle \quad (2.2)$$

is efficient.

**Lemma 2.3.** *For any efficient two-outcome measurement  $\{M^{(0)}, M^{(1)}\}$ ,  $O = M^{(0)} - M^{(1)}$  is an efficient binary observable.*

*Proof.* We have  $O^\dagger = O$  and  $O^2 = M^{(0)} + M^{(1)} = \mathbb{1}$ , so  $O$  is a binary observable. To see that  $O$  is efficient, take  $U$  to be the unitary extension of the isometry  $|0\rangle \otimes M^{(0)} + |1\rangle \otimes M^{(1)}$ . Because  $\{M^{(0)}, M^{(1)}\}$  is an efficient measurement,  $U$  is an efficient unitary. A direct calculation shows that

$$U^\dagger(\sigma_Z \otimes \mathbb{1})U |0\rangle |\psi\rangle = |0\rangle \otimes O |\psi\rangle \quad (2.3)$$

for any  $|\psi\rangle$ . □

**Lemma 2.4.** *Let  $A$  be an efficient binary observable. Then, the isometry*

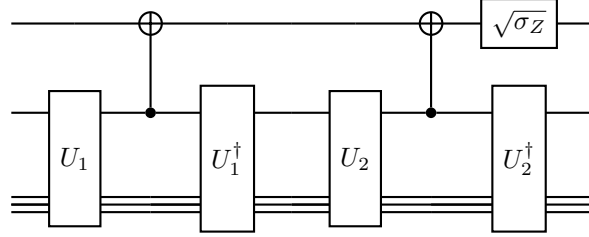
$$|\psi\rangle \mapsto |0\rangle \otimes A^{(0)} |\psi\rangle + |1\rangle \otimes A^{(1)} |\psi\rangle \quad (2.4)$$

*is efficient.*

*Proof.* Let  $U$  be the unitary associated with  $A$ . We can construct the desired isometry as follows: first, we apply  $U$  to  $|\psi\rangle$ . Then, we apply a CNOT gate with the first qubit of  $U|\psi\rangle$  being the control, and an ancillary qubit in state  $|0\rangle$  being the target. Finally, we apply  $U^\dagger$ . To see that this indeed implements the correct isometry, note that  $A^{(b)} = U^\dagger(|b\rangle\langle b| \otimes \mathbb{1})U$  for  $b \in \{0, 1\}$  and that the CNOT gate can be written as  $\mathbb{1}_2 \otimes |0\rangle\langle 0| + \sigma_X \otimes |1\rangle\langle 1|$  □

**Lemma 2.5.** *Let  $A_1$  and  $A_2$  be efficient commuting binary observables. Then  $A_1A_2$  is also an efficient binary observable.*

*Proof.* Let  $U_1, U_2$  be the efficient unitaries such that  $A_i = U_i^\dagger(\sigma_Z \otimes \mathbb{1})U_i$ . We define the unitary  $U$  by the following circuit:



We claim that  $U^\dagger(\sigma_Z \otimes \mathbb{1})U = \mathbb{1}_2 \otimes A_1A_2$ . Using  $U_i^\dagger(|b\rangle\langle b| \otimes \mathbb{1})U_i = A_i^{(b)}$  for  $b \in \{0, 1\}$ , we can write  $U$  as

$$U = (\mathbb{1}_2 \otimes A_1^{(0)} + \sigma_X \otimes A_1^{(1)})(\mathbb{1}_2 \otimes A_2^{(0)} + \sigma_X \otimes A_2^{(1)})(\sqrt{\sigma_Z} \otimes \mathbb{1}). \quad (2.5)$$

Since we can write the binary observables  $A_i$  as  $A_i = (-1)^b(2A_i^{(b)} - \mathbb{1})$ , it follows from  $[A_1, A_2] = 0$  that  $[A_1^{(b_1)}, A_2^{(b_2)}] = 0$  for any  $b_1, b_2 \in \{0, 1\}$ . Using this, the orthogonality of the projectors, and the anti-commutation of  $\sigma_Z$  and  $\sigma_X$ , the lemma follows by a direct calculation.  $\square$

**Lemma 2.6.** *Let  $U_1, U_2$  be efficient unitaries on  $\mathcal{H}$ . Then,  $(U_1 + U_2)^\dagger(U_1 + U_2)$  and  $(U_1 - U_2)^\dagger(U_1 - U_2)$  are observables and there exists an efficient procedure that, given a state  $\psi \in \mathcal{D}(\mathcal{H})$ , outputs a bit  $b$  with*

$$\Pr[b = 0|\psi] = \frac{1}{4}\text{Tr}[(U_1 + U_2)^\dagger(U_1 + U_2)\psi], \quad \Pr[b = 1|\psi] = \frac{1}{4}\text{Tr}[(U_1 - U_2)^\dagger(U_1 - U_2)\psi]. \quad (2.6)$$

(Note that  $(U_1 + U_2)^\dagger(U_1 + U_2) + (U_1 - U_2)^\dagger(U_1 - U_2) = 4 \cdot \mathbb{1}$ .)

*Proof.* The fact that both operators are observables, i.e., Hermitian, is immediate. We construct the following efficient procedure: given  $\psi$ , we (efficiently) prepare  $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H}'$ , a purification of  $\psi$  (this is only to simplify the calculation). Because  $U_1$  and  $U_2$  are efficient unitaries, so are controlled versions of  $U_1$  and  $U_2$ . Therefore, using an ancilla in the state  $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$  as the control qubit, we can efficiently prepare the state

$$\frac{1}{\sqrt{2}}(U_1 \otimes \mathbb{1}_{\mathcal{H}'}|\psi\rangle)|0\rangle + \frac{1}{\sqrt{2}}(U_2 \otimes \mathbb{1}_{\mathcal{H}'}|\psi\rangle)|1\rangle \quad (2.7)$$

$$= \frac{1}{2}((U_1 + U_2) \otimes \mathbb{1}_{\mathcal{H}'}|\psi\rangle)|+\rangle + \frac{1}{2}((U_1 - U_2) \otimes \mathbb{1}_{\mathcal{H}'}|\psi\rangle)|-\rangle. \quad (2.8)$$

Measuring the last qubit in the Hadamard basis produces the desired distribution:

$$\Pr[+] = \frac{1}{4}\langle\psi|(U_1 + U_2)^\dagger(U_1 + U_2) \otimes \mathbb{1}_{\mathcal{H}'}|\psi\rangle = \frac{1}{4}\text{Tr}[(U_1 + U_2)^\dagger(U_1 + U_2)\psi], \quad (2.9)$$

$$\Pr[-] = \frac{1}{4}\langle\psi|(U_1 - U_2)^\dagger(U_1 - U_2) \otimes \mathbb{1}_{\mathcal{H}'}|\psi\rangle = \frac{1}{4}\text{Tr}[(U_1 - U_2)^\dagger(U_1 - U_2)\psi]. \quad (2.10)$$

$\square$

**Corollary 2.7.** *Let  $C, D$  be efficient binary observables on  $\mathcal{H}$  and  $\psi$  a state on  $\mathcal{H}$ . Then,  $\{C, D\}^\dagger\{C, D\}$  and  $[C, D]^\dagger[C, D]$  are observables and there exists an efficient procedure that, given a state  $\psi \in \mathcal{D}(\mathcal{H})$ , outputs a bit  $b$  with*

$$\Pr[b = 1|\psi] = \frac{1}{4}\text{Tr}[\{C, D\}^\dagger\{C, D\}\psi], \quad \Pr[b = 0|\psi] = \frac{1}{4}\text{Tr}[[C, D]^\dagger[C, D]\psi]. \quad (2.11)$$

*Proof.* Since  $C$  and  $D$  are efficient binary observables, they are also efficient unitaries by definition. Hence, the result follows from Lemma 2.6 with  $U_1 = CD$  and  $U_2 = DC$ .  $\square$

**Definition 2.8.** We call two (families of) states  $\psi, \psi' \in \mathcal{D}(\mathcal{H})$  *computationally indistinguishable up to  $O(\delta)$*  if for any efficient procedure (called a *distinguisher*) that takes as input  $\psi$  or  $\psi'$  and produces an output bit  $b$ , we have

$$\Pr[b = 0|\psi] \approx_\delta \Pr[b = 0|\psi'] . \quad (2.12)$$

We use the notation

$$\psi \stackrel{c}{\approx}_\delta \psi' . \quad (2.13)$$

The following lemma states the simple fact that for an efficient measurement, the post-measurement states of two indistinguishable states must also be indistinguishable.

**Lemma 2.9.** *Let  $\psi, \psi' \in \mathcal{D}(\mathcal{H})$  such that  $\psi \stackrel{c}{\approx}_\delta \psi'$  for some  $\delta$ . If  $\{M^{(a)}\}_{a \in \mathcal{A}}$  is an efficient measurement on  $\mathcal{H}$ , then*

$$\sum_{a \in \mathcal{A}} M^{(a)} \psi M^{(a)} \stackrel{c}{\approx}_\delta \sum_{a \in \mathcal{A}} M^{(a)} \psi' M^{(a)} . \quad (2.14)$$

*Proof.* The proof is a simple reduction: given an efficient distinguisher  $D$  that distinguishes  $\sum_{a \in \mathcal{A}} M^{(a)} \psi M^{(a)}$  and  $\sum_{a \in \mathcal{A}} M^{(a)} \psi' M^{(a)}$  with advantage  $\delta$ , the following distinguisher  $D'$  is efficient and distinguishes  $\psi$  and  $\psi'$  with advantage  $\delta$ : given  $\psi$  or  $\psi'$ ,  $D'$  applies the isometry associated with the measurement  $\{M^{(a)}\}_{a \in \mathcal{A}}$ , traces out the pointer register to create  $\sum_{a \in \mathcal{A}} M^{(a)} \psi M^{(a)}$  or  $\sum_{a \in \mathcal{A}} M^{(a)} \psi' M^{(a)}$ , and runs the distinguisher  $D$  on this state.  $\square$

## 2.4 Distance measures

In self-testing, the verifier wants to make statements about the states and measurements used by quantum provers. The verifier can never make an “absolute” statement about any of the prover’s measurements (i.e., one that only depends on the prover’s measurement operators, not the state), since the only information available to the verifier is the prover’s classical output, which he generates by applying his measurement operators to his state. Therefore, to make statements about the prover’s operators, it is helpful to define a *state-dependent* distance between operators. Informally, if the state-dependent distance between two operators is small, this means that the two operators act on the state in the same way. A more detailed motivation of the state-dependent distance can be found in [Vid11, section 4.1], and a useful collection of many of its properties is given in [NW19, section 4.5].

**Definition 2.10** (State-dependent inner product and norm). Let  $\mathcal{H}$  be a finite-dimensional Hilbert space and  $A, B \in L(\mathcal{H})$  be linear operators on  $\mathcal{H}$ . Let  $\psi \in \text{Pos}(\mathcal{H})$ . We define the state-dependent (semi) inner product of  $A$  and  $B$  w.r.t  $\psi$  as

$$\langle A, B \rangle_\psi = \text{Tr}[A^\dagger B \psi] . \quad (2.15)$$

This induces the state-dependent (semi) norm

$$\|A\|_\psi^2 = \langle A, A \rangle_\psi = \text{Tr}[A^\dagger A \psi] . \quad (2.16)$$

*Remark 2.11.* The state dependent (semi) norm can also be expressed as a Schatten 2-norm (commonly called the Hilbert-Schmidt norm):

$$\|A\|_\psi = \left\| A \psi^{1/2} \right\|_2 . \quad (2.17)$$

**Lemma 2.12.** *The state-dependent semi inner product satisfies the properties of a semi inner product.*

*Proof.* We check the required properties.

(i) Symmetry:

$$\langle A, B \rangle_\psi = \text{Tr}[A^\dagger B \psi] = \text{Tr}[(A^\dagger B \psi)^\dagger]^* = \text{Tr}[\psi^\dagger B^\dagger A]^* = \text{Tr}[B^\dagger A \psi]^* = \langle B, A \rangle_\psi^*. \quad (2.18)$$

(ii) Linearity in the second argument: follows directly from the linearity of the trace.

(iii) Positive semi-definite: Because  $\psi$  is positive, so is  $A\psi A^\dagger$ . Therefore,

$$\langle A, A \rangle_\psi = \text{Tr}[A\psi A^\dagger] \geq 0. \quad (2.19)$$

□

*Remark 2.13.* The Cauchy-Schwarz inequality holds for semi inner products, so we have

$$|\langle X, Y \rangle_\psi| \leq \|X\|_\psi \cdot \|Y\|_\psi. \quad (2.20)$$

We will frequently make statements about two quantities (e.g., two linear operators) being approximately equal. The following definition introduces a short-hand notation for making such statements more compactly.

**Definition 2.14** (Approximate equality). We overload the symbol “ $\approx$ ” in the following ways (leaving the dependence on the security parameter implicit in the quantities on the left):

(i) **Complex numbers:** For  $a, b \in \mathbb{C}$  we define:

$$a \approx_\varepsilon b \iff |a - b| = O(\varepsilon) + \text{negl}(\lambda). \quad (2.21)$$

(ii) **Operators:** For  $A, B \in \mathcal{L}(\mathcal{H})$ , we define:

$$A \approx_\varepsilon B \iff \|A - B\|_1^2 = O(\varepsilon) + \text{negl}(\lambda). \quad (2.22)$$

(We will most frequently use this for (possibly subnormalised) quantum states  $A, B \in \text{Pos}(\mathcal{H})$ .)

(iii) **Operators on a state:** For  $A, B \in \mathcal{L}(\mathcal{H})$  and  $\psi \in \text{Pos}(\mathcal{H})$ , we define:

$$A \approx_{\varepsilon, \psi} B \iff \|A - B\|_\psi^2 = O(\varepsilon) + \text{negl}(\lambda). \quad (2.23)$$

If we write  $\approx_0$ , we mean that the quantities are negligibly close. All asymptotic statements are understood to be in the limits  $\varepsilon \rightarrow 0$  and  $\lambda \rightarrow \infty$ .

*Remark 2.15.* Note that we use a mixed convention, where the difference for states and operators is squared, but the difference for complex numbers is not. This is so that we have

$$A \approx_{\varepsilon, \psi} B \iff \text{Tr}[(A - B)^\dagger (A - B) \psi] \approx_\varepsilon 0 \quad (2.24)$$

with the same index on both sides.

#### 2.4.1 Properties of the state-dependent distance

The following lemma will be useful for showing that two operators are close in the state-dependent distance, up to an isometry.

**Lemma 2.16.** *Let  $\mathcal{H}_1, \mathcal{H}_2$  be Hilbert spaces with  $\dim(\mathcal{H}_1) \leq \dim(\mathcal{H}_2)$  and  $V : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  an isometry. Let  $A$  and  $B$  be binary observables on  $\mathcal{H}_1$  and  $\mathcal{H}_2$ , respectively,  $\psi_1 \in \text{Pos}(\mathcal{H}_1)$ ,  $\psi_2 \in \text{Pos}(\mathcal{H}_2)$ , and  $\varepsilon \geq 0$ . Then:*

$$\text{Tr}[V^\dagger B V A \psi_1] \approx_\varepsilon \text{Tr}[A \psi_1] \implies V^\dagger B V \approx_{\varepsilon, \psi_1} A, \quad (2.25)$$

$$\text{Tr}[V A V^\dagger B \psi_2] \approx_\varepsilon \text{Tr}[B \psi_2] \implies V A V^\dagger \approx_{\varepsilon, \psi_2} B. \quad (2.26)$$

*Proof.* We first show the first relation. By the definition of the state-dependent distance, we need to show that

$$\mathrm{Tr}\left[(V^\dagger BV - A)^\dagger (V^\dagger BV - A) \psi_1\right] \approx_\varepsilon 0. \quad (2.27)$$

Expanding the left hand side yields:

$$\mathrm{Tr}[V^\dagger BVV^\dagger BV\psi_1] + \mathrm{Tr}[A^2\psi_1] - \mathrm{Tr}[V^\dagger BV A\psi_1] - \mathrm{Tr}[AV^\dagger BV\psi_1]. \quad (2.28)$$

For the first term, note that  $(VV^\dagger)^2 = VV^\dagger$ , so  $VV^\dagger$  is a projector and in particular less than or equal to  $\mathbb{1}$ . Therefore, we have

$$\mathrm{Tr}[V^\dagger BVV^\dagger BV\psi_1] = \mathrm{Tr}\left[\left(\psi_1^{1/2}V^\dagger B\right)VV^\dagger\left(\psi_1^{1/2}V^\dagger B\right)^\dagger\right] \quad (2.29)$$

$$\leq \mathrm{Tr}\left[\left(\psi_1^{1/2}V^\dagger B\right)\left(\psi_1^{1/2}V^\dagger B\right)^\dagger\right] \quad (2.30)$$

$$= \mathrm{Tr}[\psi_1], \quad (2.31)$$

where we used  $B^2 = \mathbb{1}$  and  $V^\dagger V = \mathbb{1}$  in the last line. Note that since

$$\mathrm{Tr}\left[(V^\dagger BV - A)^\dagger (V^\dagger BV - A) \psi_1\right] \geq 0, \quad (2.32)$$

this also upper-bounds the absolute value.

The second term equals  $\mathrm{Tr}[\psi_1]$  because  $A^2 = \mathbb{1}$ . For the third and fourth terms, we can rewrite

$$\mathrm{Tr}[AV^\dagger BV\psi_1] = \mathrm{Tr}\left[(AV^\dagger BV\psi_1)^\dagger\right]^* \quad (2.33)$$

$$= \mathrm{Tr}[V^\dagger BV A\psi_1]^*. \quad (2.34)$$

Therefore, we can combine the third and fourth term and have

$$\mathrm{Tr}\left[(V^\dagger BV - A)^\dagger (V^\dagger BV - A) \psi_1\right] \leq 2 \mathrm{Tr}[\psi_1] - 2 \Re \mathrm{Tr}[V^\dagger BV A\psi_1]. \quad (2.35)$$

Since

$$|\mathrm{Tr}[V^\dagger BV A\psi_1] - \mathrm{Tr}[\psi_1]|^2 \geq (\Re \mathrm{Tr}[V^\dagger BV A\psi_1] - \mathrm{Tr}[\psi_1])^2, \quad (2.36)$$

$\mathrm{Tr}[V^\dagger BV A\psi_1] \approx_\varepsilon \mathrm{Tr}[\psi_1]$  implies  $\Re \mathrm{Tr}[V^\dagger BV A\psi_1] \approx_\varepsilon \mathrm{Tr}[\psi_1]$ , which completes the proof of the first relation.

The second relation follows analogously from the expansion

$$\mathrm{Tr}[(VAV^\dagger - B)^\dagger (VAV^\dagger - B)\psi_2] \quad (2.37)$$

$$= \mathrm{Tr}[VV^\dagger\psi_2] + \mathrm{Tr}[\psi_2] - 2 \Re \mathrm{Tr}[VAV^\dagger B\psi_2] \quad (2.38)$$

$$\leq 2 \mathrm{Tr}[\psi_2] - 2 \Re \mathrm{Tr}[VAV^\dagger B\psi_2]. \quad (2.39)$$

□

**Lemma 2.17** (Relation between state-dependent and operator norms). *Let  $\psi \in \mathrm{Pos}(\mathcal{H})$  with  $\mathrm{Tr}[\psi] \leq 1$  and  $C \in \mathcal{L}(\mathcal{H})$  a linear operator. Then we have:*

$$\|C\|_\psi \leq \sqrt{\|C^\dagger C\|_\infty} \leq \|C\|_\infty. \quad (2.40)$$

*Proof.* Consider the spectral decomposition  $\psi = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$ . Then

$$\|C\|_\psi^2 = \mathrm{Tr}[C^\dagger C\psi] = \sum_i \lambda_i \langle\psi_i|C^\dagger C|\psi_i\rangle. \quad (2.41)$$



It is a standard result from linear algebra that for any *Hermitian* linear operator  $A$ :

$$\|A\|_\infty = \max_{|\phi\rangle \in \mathcal{H}, \langle \phi|\phi\rangle=1} \langle \phi|A|\phi\rangle. \quad (2.42)$$

Since  $C^\dagger C$  is Hermitian and  $|\psi_i\rangle$  normalised, this implies  $\langle \psi_i|C^\dagger C|\psi_i\rangle \leq \|C^\dagger C\|_\infty$ . Therefore, using  $\sum_i \lambda_i = \text{Tr}[\psi] \leq 1$ , we have

$$\|C\|_\psi^2 = \sum_i \lambda_i \langle \psi_i|C^\dagger C|\psi_i\rangle \leq \sum_i \lambda_i \|C^\dagger C\|_\infty \leq \|C^\dagger C\|_\infty. \quad (2.43)$$

The second inequality,  $\sqrt{\|C^\dagger C\|_\infty} \leq \|C\|_\infty$ , follows immediately from the standard properties  $\|AB\|_\infty \leq \|A\|_\infty \|B\|_\infty$  and  $\|A^\dagger\|_\infty = \|A\|_\infty$  for any linear operators  $A, B \in \mathcal{L}(\mathcal{H})$ .  $\square$

We will require two further miscellaneous properties of the state-dependent distance.

**Lemma 2.18.**

(i) Let  $\psi \in \text{Pos}(\mathcal{H})$ , and  $A, B \in \mathcal{L}(\mathcal{H})$ . For  $C \in \mathcal{L}(\mathcal{H})$  such that  $C^\dagger C \leq \mathbb{1}$  we have

$$A \approx_{\varepsilon, \psi} B \implies CA \approx_{\varepsilon, \psi} CB. \quad (2.44)$$

(ii) Let  $\psi_i \in \text{Pos}(\mathcal{H})$  for  $i \in \{1, \dots, n\}$  with constant  $n$ , and  $A, B \in \mathcal{L}(\mathcal{H})$ . Define  $\psi = \sum_i \psi_i$ . Then:

$$\forall i \in \{1, \dots, n\} : A \approx_{\varepsilon, \psi_i} B \iff A \approx_{\varepsilon, \psi} B \quad (2.45)$$

*Proof.*

(i) Since  $\psi$  is positive, we have  $\psi = \psi^{1/2} \psi^{1/2}$ . Therefore, we can use  $C^\dagger C \leq \mathbb{1}$  in the following bound:

$$\text{Tr}[(CA - CB)^\dagger (CA - CB)\psi] = \text{Tr}[\psi^{1/2} (A - B)^\dagger C^\dagger C (A - B) \psi^{1/2}] \quad (2.46)$$

$$\leq \text{Tr}[(A - B)^\dagger (A - B)\psi]. \quad (2.47)$$

(ii) Inserting the definition of  $\psi$ :

$$\sum_i \text{Tr}[(A - B)^\dagger (A - B)\psi_i] = \text{Tr}[(A - B)^\dagger (A - B)\psi]. \quad (2.48)$$

The implication from left to right in the lemma follows because each term in the sum is  $O(\varepsilon)$  by assumption and there are constantly many terms. The implication from right to left follows because each  $\text{Tr}[(A - B)^\dagger (A - B)\psi_i]$  is positive.  $\square$

The following two lemmas state that if the outcome of measuring a binary observable on a state is almost certain, then the observable is close to identity on the state. Informally, this can be viewed as a variant of the gentle-measurement lemma (see e.g. [Will11, lemma 9.4.1]) expressed in the formalism of the state-dependent distance.

**Lemma 2.19.** Let  $\psi \in \text{Pos}(\mathcal{H})$ ,  $\{M^{(a)}\}_{a \in \mathcal{A}}$  a projective measurement with index set  $\mathcal{A}$ , and  $O$  a binary observable

$$O = \sum_a (-1)^{s_a} M^{(a)}, \quad (2.49)$$

where  $s_a \in \{0, 1\}$ . Suppose there exists an  $a' \in \mathcal{A}$  for which

$$\text{Tr}[M^{(a')}\psi] \approx_\varepsilon \text{Tr}[\psi]. \quad (2.50)$$

Then

$$O \approx_{\varepsilon, \psi} (-1)^{s_{a'}} \mathbb{1} \quad (2.51)$$

*Proof.* Using the fact that  $O$  is a binary observable, we can expand

$$\mathrm{Tr}\left[(O - (-1)^{s_{a'}}\mathbf{1})^\dagger (O - (-1)^{s_{a'}}\mathbf{1})\psi\right] = 2\mathrm{Tr}[\psi] - 2(-1)^{s_{a'}}\mathrm{Tr}[O\psi]. \quad (2.52)$$

Inserting the definition of  $O$  and using  $\mathrm{Tr}[M^{(a)}\psi] \geq 0$  for all  $a$  for the inequality, as well  $\sum_a M^{(a)} = \mathbf{1}$  for the last equality, we get

$$(-1)^{s_{a'}}\mathrm{Tr}[O\psi] = \sum_a (-1)^{s_{a'}+s_a}\mathrm{Tr}[M^{(a)}\psi] \quad (2.53)$$

$$= \mathrm{Tr}[M^{(a')}\psi] + \sum_{a \neq a'} (-1)^{s_{a'}+s_a}\mathrm{Tr}[M^{(a)}\psi] \quad (2.54)$$

$$\geq \mathrm{Tr}[M^{(a')}\psi] - \sum_{a \neq a'} \mathrm{Tr}[M^{(a)}\psi] \quad (2.55)$$

$$= 2\mathrm{Tr}[M^{(a')}\psi] - \mathrm{Tr}[\psi]. \quad (2.56)$$

Inserting this and using the assumption  $\mathrm{Tr}[M^{(a')}\psi] \approx_\varepsilon \mathrm{Tr}[\psi]$ :

$$\mathrm{Tr}\left[(O - (-1)^{s_{a'}}\mathbf{1})^\dagger (O - (-1)^{s_{a'}}\mathbf{1})\psi\right] \leq 4\mathrm{Tr}[\psi] - 4\mathrm{Tr}[M^{(a')}\psi] \approx_\varepsilon 0. \quad (2.57)$$

□

We will often use the previous lemma together with the following simple statement:

**Lemma 2.20.** *Let  $O$  be a binary observable on  $\mathcal{H}$  and  $\psi \in \mathrm{Pos}(\mathcal{H})$ . Then:*

$$O \approx_{\varepsilon,\psi} (-1)^b \mathbf{1} \implies O^{(b)} \approx_{\varepsilon,\psi} \mathbf{1} \text{ and } O^{(\bar{b})} \approx_{\varepsilon,\psi} 0. \quad (2.58)$$

*Proof.* This follows immediately from the fact that since  $O$  is a binary observable, we have

$$O = (-1)^b (2O^{(b)} - \mathbf{1}).$$

□

The main feature of the state-dependent distance is that if two operators are close in the state-dependent distance, we can replace one operator by the other *acting on either side of the state*. The following two lemmas formalise this replacement step. In addition to replacing operators with one another, we will also need to replace states, as shown in Lemma 2.21(ii).

**Lemma 2.21** (Replacement lemma).

(i) *Let  $\psi \in \mathrm{Pos}(\mathcal{H})$ , and  $A, B, C \in \mathcal{L}(\mathcal{H})$ . If  $A \approx_{\varepsilon,\psi} B$  and  $\|C\|_\infty = O(1)$ , then*

$$\mathrm{Tr}[CA\psi] \approx_{\varepsilon^{1/2}} \mathrm{Tr}[CB\psi], \quad (2.59)$$

$$\mathrm{Tr}[AC\psi] \approx_{\varepsilon^{1/2}} \mathrm{Tr}[BC\psi]. \quad (2.60)$$

(ii) *Let  $\psi, \psi' \in \mathrm{Pos}(\mathcal{H})$ , and  $A \in \mathcal{L}(\mathcal{H})$ . If  $\psi \approx_\varepsilon \psi'$  and  $\|A\|_\infty = O(1)$ , then*

$$\mathrm{Tr}[A\psi] \approx_{\varepsilon^{1/2}} \mathrm{Tr}[A\psi']. \quad (2.61)$$

*Proof.*

(i) We show the first relation, the second one is analogous. We rewrite the expression as an inner product and apply the Cauchy-Schwarz inequality (Remark 2.13):

$$|\mathrm{Tr}[C(A - B)\psi]| = |\langle C^\dagger, A - B \rangle_\psi| \quad (2.62)$$

$$\leq \|C^\dagger\|_\psi \cdot \|A - B\|_\psi \quad (2.63)$$

$$= O(\varepsilon^{1/2}). \quad (2.64)$$

In the last line, we used  $\|C^\dagger\|_\psi \leq \|C^\dagger\|_\infty = \|C\|_\infty$  from Lemma 2.17 and  $\|C\|_\infty = O(1)$  by assumption.

(ii) By Hölder's inequality:

$$|\mathrm{Tr}[A(\psi - \psi')]| \leq \|A\|_\infty \cdot \|\psi - \psi'\|_1. \quad (2.65)$$

The result follows since  $\|A\|_\infty = O(1)$  and  $\|\psi - \psi'\|_1 = O(\sqrt{\varepsilon})$  by assumption. □

**Lemma 2.22.** *Let  $A, B \in \mathcal{L}(\mathcal{H})$  be linear operators,  $C \in \mathcal{L}(\mathcal{H})$  a linear operator with constant operator norm, and  $\psi \in \mathrm{Pos}(\mathcal{H})$  with  $\mathrm{Tr}[\psi] \leq 1$ . Then, the following holds:*

$$A \approx_{\varepsilon, \psi} B \implies A\psi C \approx_\varepsilon B\psi C \quad \text{and} \quad C\psi A^\dagger \approx_\varepsilon C\psi B^\dagger. \quad (2.66)$$

*Proof.* We show the first relation, the second one is analogous. Suppose  $A \approx_{\varepsilon, \psi} B$ . By Hölder's inequality, we have

$$\|(A - B)\psi C\|_1^2 \leq \|(A - B)\psi^{1/2}\|_2^2 \cdot \|\psi^{1/2}C\|_2^2 \quad (2.67)$$

$$= \|A - B\|_\psi^2 \cdot \|C\|_\psi^2 \quad \text{by Remark 2.11} \quad (2.68)$$

$$\leq \|A - B\|_\psi^2 \cdot \|C\|_\infty^2 \quad \text{by Lemma 2.17} \quad (2.69)$$

$$= O(\varepsilon) \quad (2.70)$$

□

A self-testing statement always involves showing the existence of an isometry  $V$  from the prover's Hilbert space into some larger Hilbert space (see Section 1.1). The main technical difficulty that arises from this is that the application of  $V^\dagger$ , i.e., the mapping from the larger space to the smaller space, cannot be inverted in general:  $VV^\dagger \neq \mathbb{1}$ . The following two lemmas deal with how the state-dependent distance behaves under the application of an isometry.

**Lemma 2.23.** *Let  $\mathcal{H}_1, \mathcal{H}_2$  be Hilbert spaces with  $\dim(\mathcal{H}_1) \leq \dim(\mathcal{H}_2)$ ,  $V : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  an isometry, and  $A$  and  $B$  binary observables on  $\mathcal{H}_1$  and  $\mathcal{H}_2$ , respectively. Then, the following holds for any  $\psi \in \mathrm{Pos}(\mathcal{H}_1)$ :*

$$VAV^\dagger \approx_{\varepsilon, V\psi V^\dagger} B \implies A \approx_{\varepsilon, \psi} V^\dagger B V. \quad (2.71)$$

$$A \approx_{\varepsilon, \psi} V^\dagger B V \implies VAV^\dagger \approx_{\varepsilon^{1/2}, V\psi V^\dagger} B. \quad (2.72)$$

*Proof.* We prove each relation in turn.

**Proof of the first relation.**

Using  $V^\dagger V = \mathbb{1}$ :

$$\mathrm{Tr}\left[(A - V^\dagger B V)^\dagger (A - V^\dagger B V) \psi\right] = \mathrm{Tr}\left[V^\dagger (VAV^\dagger - B) VV^\dagger (VAV^\dagger - B) V\psi\right] \quad (2.73)$$

Since  $\psi$  is positive, we have  $\psi = \psi^{1/2}\psi^{1/2}$ :

$$= \text{Tr} \left[ \psi^{1/2} V^\dagger (VAV^\dagger - B) VV^\dagger (VAV^\dagger - B) V\psi^{1/2} \right] \quad (2.74)$$

We have  $(VV^\dagger)^2 = VV^\dagger$ , so  $VV^\dagger$  is a projector and in particular less than or equal to  $\mathbf{1}$ . Since the expression has the form  $\text{Tr}[MVV^\dagger M^\dagger]$ , we can bound it as:

$$\leq \text{Tr} \left[ \psi^{1/2} V^\dagger (VAV^\dagger - B) (VAV^\dagger - B) V\psi^{1/2} \right] \quad (2.75)$$

$$= \text{Tr} \left[ (VAV^\dagger - B) (VAV^\dagger - B) V\psi V^\dagger \right] \quad (2.76)$$

Since we are assuming  $VAV^\dagger \approx_{\varepsilon, V\psi V^\dagger} B$ :

$$\approx_\varepsilon 0. \quad (2.77)$$

### Proof of the second relation.

By Lemma 2.16, we only need to show

$$\text{Tr} [VAV^\dagger BV\psi V^\dagger] = \text{Tr} [AV^\dagger BV\psi] \approx_{\varepsilon^{1/2}} \mathbf{1}. \quad (2.78)$$

This follows immediately from the replacement (Lemma 2.21(i)), the assumption  $V^\dagger BV \approx_{\varepsilon, \psi} A$ , and the fact that  $A^2 = \mathbf{1}$ .  $\square$

**Lemma 2.24.** *Let  $\mathcal{H}_1, \mathcal{H}_2$  be Hilbert spaces with  $\dim(\mathcal{H}_1) \leq \dim(\mathcal{H}_2)$  and  $V : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  an isometry. Let  $A$  and  $B$  be binary observables on  $\mathcal{H}_1$  and  $\mathcal{H}_2$ , respectively,  $\psi \in \text{Pos}(\mathcal{H}_1)$ , and  $\varepsilon \geq 0$ . Then for any  $b \in \{0, 1\}$ :*

$$V^\dagger BV \approx_{\varepsilon, \psi} A \implies V^\dagger B^{(b)} V \approx_{\varepsilon, \psi} A^{(b)}, \quad (2.79)$$

$$B \approx_{\varepsilon, V\psi V^\dagger} VAV^\dagger \implies B^{(b)} \approx_{\varepsilon, V\psi V^\dagger} VA^{(b)}V^\dagger. \quad (2.80)$$

*Proof.*

**Proof of the first relation.** For any binary observable  $O$ , we can write  $O = (-1)^b (2O^{(b)} - \mathbf{1})$ . Therefore,

$$V^\dagger BV - A = (-1)^b V^\dagger (2B^{(b)} - \mathbf{1}) V - (-1)^b (2A^{(b)} - \mathbf{1}) \quad (2.81)$$

$$= (-1)^b \cdot 2 \cdot (V^\dagger B^{(b)} V - A^{(b)}). \quad (2.82)$$

This means that

$$\text{Tr} \left[ (V^\dagger B^{(b)} V - A^{(b)})^\dagger (V^\dagger B^{(b)} V - A^{(b)}) \psi \right] = \frac{1}{4} \text{Tr} \left[ (V^\dagger BV - A)^\dagger (V^\dagger BV - A) \psi \right] \approx_\varepsilon 0. \quad (2.83)$$

**Proof of the second relation.** Similarly to the first case, we have

$$B - VAV^\dagger = (-1)^b \cdot 2 \cdot (B^{(b)} - VA^{(b)}V^\dagger) - (-1)^b (\mathbf{1} - VV^\dagger). \quad (2.84)$$

The result then follows from  $(\mathbf{1} - VV^\dagger) V\psi V^\dagger = 0$ .  $\square$

## 2.5 Lifting state-dependent operator relations using computational indistinguishability

The following lemma collects a number of statements that allow us to replace computationally indistinguishable states with one another in the state-dependent distance. This means that if two states are computationally indistinguishable and a state-dependent operator relation holds for one of the states, we can “lift” this relation to the other state, provided the operators are efficient. We will make use of this many times throughout the rest of the paper.

**Lemma 2.25** (Lifting lemma). *Let  $\psi, \psi' \in \mathcal{D}(\mathcal{H})$  such that  $\psi \stackrel{c}{\approx}_\delta \psi'$ .*

(i) *Let  $A$  be an efficient binary observable on  $\mathcal{H}$ . Then:*

$$\mathrm{Tr}[A\psi] \approx_\delta \mathrm{Tr}[A\psi'] . \quad (2.85)$$

(ii) *Let  $A, B$  be efficient binary observables on  $\mathcal{H}$ . Then:*

$$A \approx_{\varepsilon, \psi} B \implies A \approx_{\delta+\varepsilon, \psi'} B . \quad (2.86)$$

(iii) *Let  $A, B$  be efficient binary observables on  $\mathcal{H}$ . Then:*

$$[A, B] \approx_{\varepsilon, \psi} 0 \implies [A, B] \approx_{\delta+\varepsilon, \psi'} 0 . \quad (2.87)$$

(iv) *Let  $A, B$  be efficient binary observables on  $\mathcal{H}$ . Then:*

$$\{A, B\} \approx_{\varepsilon, \psi} 0 \implies \{A, B\} \approx_{\delta+\varepsilon, \psi'} 0 . \quad (2.88)$$

(v) *Let  $\mathcal{H}'$  be another Hilbert space with  $\dim(\mathcal{H}') \geq \dim(\mathcal{H})$ ,  $A$  an efficient binary observable on  $\mathcal{H}$ ,  $B$  an efficient binary observable on  $\mathcal{H}'$ , and  $V : \mathcal{H} \rightarrow \mathcal{H}'$  an efficient isometry. Then:*

$$A \approx_{\varepsilon, \psi} V^\dagger B V \implies A \approx_{\varepsilon^{1/2}+\delta, \psi'} V^\dagger B V . \quad (2.89)$$

(vi) *Let  $\mathcal{H}'$  be another Hilbert space with  $\dim(\mathcal{H}') \geq \dim(\mathcal{H})$ . For this case, let  $\psi, \psi' \in \mathcal{D}(\mathcal{H}')$  such that  $\psi \stackrel{c}{\approx}_\delta \psi'$ . Let  $A$  be an efficient binary observable on  $\mathcal{H}$ ,  $B$  an efficient binary observable on  $\mathcal{H}'$ , and  $V : \mathcal{H} \rightarrow \mathcal{H}'$  an efficient isometry. Then:*

$$V A V^\dagger \approx_{\varepsilon, \psi} B \implies V A V^\dagger \approx_{\varepsilon^{1/4}+\delta, \psi'} B . \quad (2.90)$$

*Proof.*

(i) Since  $A$  is efficient, the procedure that makes the measurement  $\{A^{(0)}, A^{(1)}\}$  and outputs the result is efficient. The probability of outputting 0 given state  $\rho$  is  $\mathrm{Tr}[A^{(0)}\rho]$ . Therefore, by the definition of computational indistinguishability (Definition 2.8), we have

$$\mathrm{Tr}[A\psi] = 2 \mathrm{Tr}\left[A^{(0)}\psi\right] - 1 \approx_\delta 2 \mathrm{Tr}\left[A^{(0)}\psi'\right] - 1 = \mathrm{Tr}[A\psi'] . \quad (2.91)$$

(ii) By assumption,

$$\mathrm{Tr}[(A - B)^\dagger (A - B)\psi] \approx_\varepsilon 0 . \quad (2.92)$$

By Lemma 2.6 and Definition 2.8, we also have

$$\mathrm{Tr}[(A - B)^\dagger (A - B)\psi] \approx_\delta \mathrm{Tr}[(A - B)^\dagger (A - B)\psi'] . \quad (2.93)$$

The result follows by the triangle inequality.

(iii) As above, using Corollary 2.7.

(iv) As above, using Corollary 2.7.

- (v) We first remark that the result of (ii) does not directly apply here, since  $V^\dagger BV$  is in general not a binary observable. Let  $U \in \mathcal{U}(\mathcal{H}')$  be an efficient unitary such that  $V = U(\mathbb{1} \otimes |0_k\rangle)$ , where  $k = \dim(\mathcal{H}')/\dim(\mathcal{H})$  (assuming that  $\dim(\mathcal{H})$  divides  $\dim(\mathcal{H}')$ , which is without loss of generality since we can add extra dimensions to  $\mathcal{H}'$  if necessary). Then by Lemma 2.6 and because we can efficiently prepare the state  $|0_k\rangle\langle 0_k|$ , there exists an efficient procedure that outputs a bit  $b$  with

$$4 \cdot \Pr[b = 0 | \psi'] = \text{Tr}[(A \otimes \mathbb{1}_k - U^\dagger B U)^2 (\psi' \otimes |0_k\rangle\langle 0_k|)] . \quad (2.94)$$

By the assumption  $\psi \stackrel{c}{\approx}_\delta \psi'$ , we therefore have

$$\text{Tr}[(A \otimes \mathbb{1}_k - U^\dagger B U)^2 (\psi' \otimes |0_k\rangle\langle 0_k|)] \approx_\delta \text{Tr}[(A \otimes \mathbb{1}_k - U^\dagger B U)^2 (\psi \otimes |0_k\rangle\langle 0_k|)] . \quad (2.95)$$

Since we are assuming  $A \approx_{\varepsilon, \psi} V^\dagger B V$ , it now suffices to show

$$\text{Tr}[(A \otimes \mathbb{1}_k - U^\dagger B U)^2 (\psi \otimes |0_k\rangle\langle 0_k|)] \approx_{\varepsilon^{1/2}} \text{Tr}[(A - V^\dagger B V)^2 \psi] . \quad (2.96)$$

Multiplying out the expression on the left hand side and using that we can move  $|0_k\rangle$  and  $\langle 0_k|$  past  $A \otimes \mathbb{1}_k$ , we get

$$\text{Tr}[(A \otimes \mathbb{1}_k - U^\dagger B U)^2 (\psi \otimes |0_k\rangle\langle 0_k|)] = 1 + 1 - \text{Tr}[A V^\dagger B V \psi] - \text{Tr}[V^\dagger B V A \psi] . \quad (2.97)$$

Expanding the right hand side of Equation (2.96), one gets the same terms, except for  $\text{Tr}[V^\dagger B V \psi V^\dagger B V]$  instead of 1. However, using the assumption  $A \approx_{\varepsilon, \psi} V^\dagger B V$  and the replacement lemma (Lemma 2.21) twice:

$$\text{Tr}[V^\dagger B V \psi V^\dagger B V] \approx_{\varepsilon^{1/2}} \text{Tr}[A \psi A] = 1 . \quad (2.98)$$

The last equality is true because  $A$  squares to identity and  $\psi$  is normalised.

- (vi) Let  $U$  be as in (v), again assuming without loss of generality that  $\dim(\mathcal{H})$  divides  $\dim(\mathcal{H}')$ . By the same reasoning as in (v), we have

$$\text{Tr}[(U(A \otimes \mathbb{1}_k)U^\dagger - B)^2 \psi'] \approx_\delta \text{Tr}[(U(A \otimes \mathbb{1}_k)U^\dagger - B)^2 \psi] , \quad (2.99)$$

so it suffices to show

$$\text{Tr}[(U(A \otimes \mathbb{1}_k)U^\dagger - B)^2 \psi] \approx_{\varepsilon^{1/4}} \text{Tr}[(V A V^\dagger - B)^2 \psi] . \quad (2.100)$$

As a first step, we show  $V V^\dagger \approx_{\varepsilon^{1/2}, \psi} \mathbb{1}$ . For this, observe that since  $V^\dagger V = \mathbb{1}$ :

$$\text{Tr}[(V V^\dagger - \mathbb{1})^2 \psi] = \text{Tr}[(V V^\dagger V V^\dagger + \mathbb{1} - 2 V V^\dagger) \psi] \quad (2.101)$$

$$= 1 - \text{Tr}[V V^\dagger \psi] \quad (2.102)$$

Using that  $B^2 = \mathbb{1}$ ,  $V^\dagger V = \mathbb{1}$ , and  $\psi$  is normalised:

$$= 1 - (\text{Tr}[(V A V^\dagger - B)^2 \psi] - 1 + \text{Tr}[V A V^\dagger B \psi] + \text{Tr}[B V A V^\dagger \psi]) \quad (2.103)$$

Since by assumption  $V A V^\dagger \approx_{\varepsilon, \psi} B$ , the first trace is  $O(\varepsilon)$ . For the other two traces, we can use the replacement lemma (Lemma 2.21) to replace  $V A V^\dagger$  with  $B$ :

$$\approx_{\varepsilon^{1/2}} 2 - 2 \text{Tr}[B^2 \psi] \quad (2.104)$$

$$= 0 . \quad (2.105)$$

This allows us to show Equation (2.100) as follows. On the one hand we have from expanding as in Lemma 2.16:

$$\text{Tr}[(U(A \otimes \mathbb{1}_k)U^\dagger - B)^2 \psi] = 2 - 2 \Re \text{Tr}[B U (A \otimes \mathbb{1}_k) U^\dagger \psi] \quad (2.106)$$

Using  $VV^\dagger \approx_{\varepsilon^{1/2}} \mathbb{1}$  and the replacement lemma (Lemma 2.21):

$$\approx_{\varepsilon^{1/4}} 2 - 2 \Re \operatorname{Tr} [BU(A \otimes \mathbb{1}_k)U^\dagger VV^\dagger \psi] \quad (2.107)$$

Rewriting  $V = U(\mathbb{1} \otimes |0_k\rangle)$ , using  $U^\dagger U = \mathbb{1}$ , and evaluating  $(A \otimes \mathbb{1}_k)(\mathbb{1} \otimes |0_k\rangle) = (\mathbb{1} \otimes |0_k\rangle)A$ :

$$= 2 - 2 \Re \operatorname{Tr} [BU(\mathbb{1} \otimes |0_k\rangle)AV^\dagger \psi] \quad (2.108)$$

$$= 2 - 2 \Re \operatorname{Tr} [BVA V^\dagger \psi] . \quad (2.109)$$

On the other hand, expanding in the same manner as in Equation (2.106):

$$\operatorname{Tr} [(VA V^\dagger - B)^2 \psi] = 1 + \operatorname{Tr} [VV^\dagger \psi] - 2 \Re \operatorname{Tr} [BVA V^\dagger \psi] \quad (2.110)$$

Using  $VV^\dagger \approx_{\varepsilon^{1/2}, \psi} \mathbb{1}$ :

$$\approx_{\varepsilon^{1/2}} 2 - 2 \Re \operatorname{Tr} [BVA V^\dagger \psi] . \quad (2.111)$$

The result follows from the triangle inequality. □

### 3 Self-testing protocol

In this section, we introduce the self-testing protocol and describe the behaviour of an honest prover that succeeds with probability negligibly close to 1. The protocol is described in detail in Figure 1, an informal description was already given in the introduction (Section 1.3).

#### 3.1 Completeness of self-testing protocol

**Proposition 3.1.** *There is an efficient quantum prover that is accepted in the self-testing protocol with probability negligibly close to 1 (as a function of the security parameter).*

*Proof.* We describe the honest strategy. Given keys  $k_1, k_2$ , the prover initially treats each key separately (i.e., it prepares a product state). For each  $k_i$ , the prover prepares the state

$$\frac{1}{\sqrt{2 \cdot |\mathcal{X}|}} \sum_{b \in \{0,1\}} \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \sqrt{f_{k_i, b}(x)(y) |b\rangle |x\rangle |y\rangle} . \quad (3.1)$$

Preparing this state can be efficiently done (up to negligible error) using the SAMP procedure from the definition of ENTFCF families ([BCM<sup>+</sup>18, definition 3.1] and [Mah18, definition 4.2]). The prover then measures the two “image registers” (i.e., the ones where  $y$  is stored) to obtain images  $y_1, y_2 \in \mathcal{Y}$  and sends these back to the verifier. The post-measurement for each  $i \in \{1, 2\}$  is

$$\begin{cases} |\hat{b}(k_i, y_i)\rangle |\hat{x}(k_i, y_i)\rangle & \text{if } k_i \in \mathcal{K}_{\mathcal{G}} , \\ \frac{1}{\sqrt{2}} (|0\rangle |\hat{x}_0(k_i, y_i)\rangle + |1\rangle |\hat{x}_1(k_i, y_i)\rangle) & \text{if } k_i \in \mathcal{K}_{\mathcal{F}} . \end{cases} \quad (3.2)$$

If the verifier selects a preimage round, the prover measures both registers in the computational basis and returns the result. From the states in Equation (3.2) it is clear that the prover succeeds with probability negligibly close to 1 in the preimage round.

If the verifier selects a Hadamard round, the prover measures both “ $x$ -registers” in the Hadamard basis to obtain strings  $d_1, d_2$  and returns these to the verifier. We introduce the shorthand  $b_i = \hat{b}(k_i, y_i)$  and  $x_{b,i} = \hat{x}_b(k_i, y_i)$ . At this point, the prover’s state for each  $i \in \{1, 2\}$  is (up to a global phase)

$$\begin{cases} |b_i\rangle & \text{if } k_i \in \mathcal{K}_{\mathcal{G}} , \\ |(-)^{d_i \cdot (x_0, i \oplus x_1, i)}\rangle & \text{if } k_i \in \mathcal{K}_{\mathcal{F}} . \end{cases} \quad (3.3)$$

---

Let  $\lambda$  be a security parameter and  $(\mathcal{F}, \mathcal{G})$  an ENTCF family.

1. The verifier selects bases  $\theta_1, \theta_2 \in_R \{0, 1\}$ , where 0 corresponds to the computational and 1 to the Hadamard basis. We call the basis choices  $(0, 1)$  and  $(1, 0)$  the *test case*, and the basis choice  $(1, 1)$  the *Bell case*.
2. The verifier samples keys and trapdoors  $(k_1, t_{k_1}; k_2, t_{k_2})$  using

$$\begin{cases} (k_i, t_{k_i}) \leftarrow \text{GEN}_{\mathcal{K}_{\mathcal{G}}}(1^\lambda) & \text{if } \theta_i = 0, \\ (k_i, t_{k_i}) \leftarrow \text{GEN}_{\mathcal{K}_{\mathcal{F}}}(1^\lambda) & \text{if } \theta_i = 1. \end{cases}$$

The verifier sends  $(k_1, k_2)$  to the prover (but keeps the trapdoors  $t_{k_i}$  private).

3. The verifier receives  $y_1, y_2 \in \mathcal{Y}$  from the prover.
4. The verifier selects a round type  $\in \{\text{preimage round, Hadamard round}\}$  uniformly at random and sends the round type to the prover.
  - (i) For a *preimage round*: The verifier receives  $(b_1, x_1; b_2, x_2)$  from the prover, with  $b_i \in \{0, 1\}$  and  $x_i \in \mathcal{X}$ . The verifier sets  $\text{flag} \leftarrow \text{fail}_{\text{pre}}$  if  $\text{CHK}(k_i, y_i, b_i, x_i) = 0$  for  $i = 1$  or  $i = 2$ .
  - (ii) For a *Hadamard round*: The verifier receives  $d_1, d_2 \in \{0, 1\}^w$  from the prover (for some  $w$  depending on the security parameter). The verifier selects questions  $q_1, q_2 \in_R \{0, 1\}$ , sends these to the prover, and receives answers  $v_1, v_2 \in \{0, 1\}$ . Depending on the basis choice, the verifier performs the following checks:

Basis choice $(\theta_1, \theta_2)$	Verifier's check
$(0, 0)$	None
$(0, 1)$	Set $\text{flag} \leftarrow \text{fail}_{\text{Test}}$ if one of the following is true: $q_1 = 0$ and $\hat{b}(k_1, y_1) \neq v_1$ . $q_2 = 1$ and $\hat{u}(k_2, y_2, d_2) \neq v_2 \oplus \hat{b}(k_1, y_1)$ .
$(1, 0)$	Set $\text{flag} \leftarrow \text{fail}_{\text{Test}}$ if one of the following is true: $q_1 = 1$ and $\hat{u}(k_1, y_1, d_1) \neq v_1 \oplus \hat{b}(k_2, y_2)$ . $q_2 = 0$ and $\hat{b}(k_2, y_2) \neq v_2$ .
$(1, 1)$	Set $\text{flag} \leftarrow \text{fail}_{\text{Bell}}$ if one of the following is true: $(q_1, q_2) = (0, 1)$ and $\hat{u}(k_2, y_2, d_2) \neq v_1 \oplus v_2$ $(q_1, q_2) = (1, 0)$ and $\hat{u}(k_1, y_1, d_1) \neq v_1 \oplus v_2$

---

Figure 1: The self-testing protocol. Some of the verifier's checks, such as that for  $(\theta_1, \theta_2) = (1, 0)$ ,  $\hat{u}(k_1, y_1, d_1)$  must equal  $v_1 \oplus \hat{b}(k_2, y_2)$ , not  $v_1$ , might look counter-intuitive. They are defined this way because the verifier must effectively "decode" the CZ gate that the honest prover applies. For the honest prover behaviour, see the proof of Proposition 3.1.



Now the prover applies a controlled- $Z$  gate ( $CZ$ ) between the two qubits (with  $i = 1$  being the control and  $i = 2$  being the target qubit). This results in the state (again up to global phases)

$$\begin{cases} |b_1\rangle|b_2\rangle & \text{if } k_1, k_2 \in \mathcal{K}_{\mathcal{G}}; \\ |b_1\rangle|(-)^{b_1 \oplus d_2 \cdot (x_{0,2} \oplus x_{1,2})}\rangle & \text{if } k_1 \in \mathcal{K}_{\mathcal{G}}, k_2 \in \mathcal{K}_{\mathcal{F}}; \\ |(-)^{b_2 \oplus d_1 \cdot (x_{0,1} \oplus x_{1,1})}\rangle|b_2\rangle & \text{if } k_1 \in \mathcal{K}_{\mathcal{F}}, k_2 \in \mathcal{K}_{\mathcal{G}}; \\ |\phi_H^{(d_2 \cdot (x_{0,2} \oplus x_{1,2}), d_1 \cdot (x_{0,1} \oplus x_{1,1}))}\rangle & \text{if } k_1, k_2 \in \mathcal{K}_{\mathcal{F}}, \end{cases} \quad (3.4)$$

with the (Hadamard-rotated) Bell states

$$|\phi_H^{(a,b)}\rangle = (\sigma_X^a \otimes \sigma_X^b)(|00\rangle + |01\rangle + |10\rangle - |11\rangle). \quad (3.5)$$

When the prover receives questions  $q_1, q_2 \in \{0, 1\}$  from the verifier, he measures each qubit individually in the computational (if  $q_i = 0$ ) or Hadamard (if  $q_i = 1$ ) basis and returns the outcomes  $v_1, v_2$ . For the first three cases in Equation (3.4), it is easy to see that the prover will be accepted. For the last case, this follows from

$$(\sigma_Z \otimes \sigma_X)|\phi_H^{(a,b)}\rangle = (-1)^a|\phi_H^{(a,b)}\rangle, \quad (3.6)$$

$$(\sigma_X \otimes \sigma_Z)|\phi_H^{(a,b)}\rangle = (-1)^b|\phi_H^{(a,b)}\rangle. \quad (3.7)$$

□

## 4 Soundness of self-testing protocol

The goal of this section is to prove the soundness of the self-testing protocol in Figure 1, i.e., to show that any computationally bounded prover that succeeds in the protocol must have prepared a Bell state and measured the individual qubits in the computational or Hadamard basis, up to a global change of basis. This statement is made formal in Theorem 4.42. All statements in this section are under the assumption that no efficient quantum device can break the LWE assumption [Reg09] (for the same parameters as those used in [BCM<sup>+</sup>18] and [Mah18]). Informally, the main steps of the soundness proof are the following:

1. We first formalise the actions of a quantum prover as a device (Section 4.1). A device is essentially a collection of states and measurements used by the prover to compute his answers to the verifier. These states and measurements are the ones that the verifier can characterise with the self-testing protocol. We then express the success probability of a device in terms of its states and measurements in section 4.2.
2. In the self-testing protocol, the verifier chooses which type of function (claw-free or injective) to use. We show that because it is computationally hard to determine the function type given only the key, different states prepared by the prover for different key choices are computationally indistinguishable. This will allow us to use different key choices to characterise different aspects of the prover's behaviour, and “lift” these characterisations to another key choice using the lifting lemma (Lemma 2.25).
3. We show that different observables used by the prover either anti-commute (Section 4.5) or commute (Section 4.6) on the prover's state.
4. In the self-testing protocol, the prover gets two questions indicating the measurement bases for the first and second “qubit” (though at this point in the proof, we do not yet have a characterisation of the prover's states in terms of qubits). Depending on whether the bases for both “qubits” are the same or different, the prover's measurements are described by different observables, which we call “non-tilde observables” if the questions are the same, and “tilde observables” if they are different. To fully characterise the prover's measurements, we need to characterise both tilde and non-tilde observables. In

particular, to analyse the “Bell case” in the protocol, the tilde observables are required. For technical reasons, characterising non-tilde observables is easier. Hence, the next step is to characterise the non-tilde observables as follows:

- (i) We define an isometry  $V_S$  (Definition 4.27), which is a single-prover version of the two-prover “swap isometry” in [MYS12]. The isometry is defined in terms of the prover’s non-tilde observables. This is the isometry for which we will show that it maps the prover’s states and observables to the desired Bell states and two-qubit Pauli observables.
  - (ii) We show that under this isometry, the prover’s observables on the first “qubit” are close to Pauli observables (Equation (4.88) and Lemma 4.30).
  - (iii) We use this characterisation of the observables to obtain a characterisation of the prover’s first qubit in the “test case” of the self-testing protocol (Lemmas 4.31 and 4.32).
  - (iv) We use the characterisation of the prover’s first qubit to show that the prover’s observables on the second qubit are also approximately equal to Pauli observables (Lemma 4.33).
5. We show that non-tilde observables and tilde observables are approximately equal on the state (Lemma 4.35), and hence tilde observables are also close to Pauli observables under the isometry  $V_S$  (Corollary 4.36).
  6. We use the characterisation of the prover’s observables to also characterise the prover’s second “qubit” in the “test case” (Lemma 4.39). The computational indistinguishability of the verifier’s basis choices allows us to extend this characterisation to the Bell case (Corollary 4.40).
  7. The characterisation of both of the prover’s “qubits” allows us to show that products of the prover’s observables are close to tensor products of Pauli observables (Lemma 4.41).
  8. Using this characterisation of products of observables, we can show that in the Bell case, the prover must have produced a Bell pair and measured its individual qubits in the computational or Hadamard basis (Theorem 4.42).

## 4.1 Devices

We model the actions of a general prover by a “device”. This formalises all possible actions that can be taken by the prover to compute his answers  $y_1, y_2, d_1, d_2$ , and  $v_1, v_2$  to the verifier. By Naimark’s theorem, up to adding dimensions to the prover’s Hilbert space, we can assume without loss of generality that the prover only performs projective measurements (instead of more general POVMs).

**Definition 4.1** (Devices). A device  $D = (S, \Pi, M, P)$  is specified by the following:

- (i) A set  $S = \{\psi^{(\theta_1, \theta_2)}\}_{\theta_1, \theta_2 \in \{0,1\}}$  of states  $\psi^{(\theta_1, \theta_2)} \in \mathcal{D}(\mathcal{H}_D \otimes \mathcal{H}_Y)$ , where  $\dim(\mathcal{H}_Y) = |\mathcal{Y}|^2$  and the states are classical on  $\mathcal{H}_Y$ :

$$\psi^{(\theta_1, \theta_2)} = \sum_{y_1, y_2 \in \mathcal{Y}} \psi_{y_1, y_2}^{(\theta_1, \theta_2)} \otimes |y_1, y_2\rangle\langle y_1, y_2|_Y. \quad (4.1)$$

In the context of the self-testing protocol,  $\psi^{(\theta_1, \theta_2)}$  is the prover’s state after returning  $y_1, y_2$  for the case where the verifier makes basis choices  $\theta_1, \theta_2$ . Each  $\psi^{(\theta_1, \theta_2)}$  also implicitly depends on the specific keys chosen by the verifier (not just the key type); all the statements we make hold on average over key choices.

- (ii) A projective measurement  $\Pi$  on  $\mathcal{H}_D \otimes \mathcal{H}_Y$ :

$$\Pi = \left\{ \Pi^{(b_1, x_1; b_2, x_2)} = \sum_{y_1, y_2} \Pi_{y_1, y_2}^{(b_1, x_1; b_2, x_2)} \otimes |y_1, y_2\rangle\langle y_1, y_2|_Y \right\}_{b_1, b_2 \in \{0,1\}; x_1, x_2 \in \mathcal{X}}. \quad (4.2)$$

This is the measurement used by the prover to compute his answer  $(b_1, x_1; b_2, x_2)$  in the preimage challenge.

(iii) A projective measurement  $M$  on  $\mathcal{H}_D \otimes \mathcal{H}_Y$

$$M = \left\{ M^{(d_1, d_2)} = \sum_{y_1, y_2} M_{y_1, y_2}^{(d_1, d_2)} \otimes |y_1, y_2\rangle\langle y_1, y_2|_Y \right\}_{d_1, d_2 \in \{0, 1\}^w}. \quad (4.3)$$

This is the measurement used by the prover to compute his answer  $(d_1, d_2)$  in the Hadamard challenge. We use an additional Hilbert spaces  $\mathcal{H}_R$  to record the outcomes of measuring  $M$  and write the post-measurement state after applying  $M$  to  $\psi^{(\theta_1, \theta_2)}$  as

$$\sigma^{(\theta_1, \theta_2)} := \sum_{y_1, y_2, d_1, d_2} M_{y_1, y_2}^{(d_1, d_2)} \psi_{y_1, y_2}^{(\theta_1, \theta_2)} M_{y_1, y_2}^{(d_1, d_2)} \otimes |y_1, y_2; d_1, d_2\rangle\langle y_1, y_2; d_1, d_2|_{YR}. \quad (4.4)$$

(iv) A set  $P = \{P_{0,0}, P_{0,1}, P_{1,0}, P_{1,1}\}$ , where for each  $q_1, q_2 \in \{0, 1\}$ ,  $P_{q_1, q_2}$  is a projective measurement on  $\mathcal{H}_D \otimes \mathcal{H}_Y \otimes \mathcal{H}_R$ :

$$P_{q_1, q_2} = \left\{ P_{q_1, q_2}^{(v_1, v_2)} = \sum_{y_1, y_2, d_1, d_2} P_{q_1, q_2; y_1, y_2; d_1, d_2}^{(v_1, v_2)} \otimes |y_1, y_2; d_1, d_2\rangle\langle y_1, y_2; d_1, d_2|_{YR} \right\}_{v_1, v_2 \in \{0, 1\}}. \quad (4.5)$$

In the context of the self-testing protocol, given questions  $q_1, q_2$ , the prover will measure  $\{P_{q_1, q_2}\}$  and return the outcomes  $v_1, v_2$  as his answer.

*Remark 4.2.* Mirroring the above notation, for an operator  $A$  on  $\mathcal{H}_D \otimes \mathcal{H}_Y \otimes \mathcal{H}_R$  that is classical on  $\mathcal{H}_Y$  and  $\mathcal{H}_R$ , we will write  $A_{y_1, y_2, d_1, d_2}$  for the operators defined by

$$A = \sum_{y_1, y_2, d_1, d_2} A_{y_1, y_2, d_1, d_2} \otimes |y_1, y_2; d_1, d_2\rangle\langle y_1, y_2; d_1, d_2|, \quad (4.6)$$

and similarly for operators that are just defined on  $\mathcal{H}_D \otimes \mathcal{H}_Y$  or  $\mathcal{H}_D \otimes \mathcal{H}_R$ .

**Definition 4.3** (Efficient devices). A device is called *efficient* if the states  $\psi^{(\theta_1, \theta_2)}$  can be prepared efficiently and the measurements  $\Pi$ ,  $M$ , and  $P_{q_1, q_2}$  can be performed efficiently (as defined in Definition 2.2)

#### 4.1.1 Marginal measurements

In the standard self-testing scenario for a single Bell pair (as in e.g. [MYS12]), each prover returns a single bit. Therefore, for a fixed question, the measurement performed by each prover can be described by a binary observable. In contrast, in the single-prover setting, the prover sees both questions at once. Hence, for fixed questions  $q_1, q_2$ , its measurements are described by the 4-outcome measurements  $\{P_{q_1, q_2}^{(v_1, v_2)}\}_{v_1, v_2}$ . We relate the single-prover scenario to the two-prover scenario by defining marginal observables. These intuitively correspond to the observables used by each prover in the two-prover setting. However, in the single-prover setting, the observable used to obtain the first answer bit  $v_1$  can also depend on the second question bit  $q_2$ . Therefore, there are two different sets of marginal observables: the “non-tilde observables”, which result from marginalising over projectors with  $q_1 = q_2$ ; and the “tilde observables”, which result from marginalising over projectors with  $q_1 \neq q_2$ . A formal definition follows.

**Definition 4.4** (Marginal observables). For a device  $D = (S, \Pi, M, P)$ , we define the following binary observables:

$$\begin{aligned} Z_1 &:= \sum_{i,j} (-1)^i P_{0,0}^{(i,j)}, & Z_2 &:= \sum_{i,j} (-1)^j P_{0,0}^{(i,j)}, & X_1 &:= \sum_{i,j} (-1)^i P_{1,1}^{(i,j)}, & X_2 &:= \sum_{i,j} (-1)^j P_{1,1}^{(i,j)}, \\ \tilde{Z}_1 &:= \sum_{i,j} (-1)^i P_{0,1}^{(i,j)}, & \tilde{Z}_2 &:= \sum_{i,j} (-1)^j P_{1,0}^{(i,j)}, & \tilde{X}_1 &:= \sum_{i,j} (-1)^i P_{1,0}^{(i,j)}, & \tilde{X}_2 &:= \sum_{i,j} (-1)^j P_{0,1}^{(i,j)}. \end{aligned}$$

*Remark 4.5.* By the same reasoning as in Lemma 2.3, all of the above are efficient binary observables.

#### 4.1.2 Partial post-measurement states

In Equation (4.4), we defined the prover's post-measurement state as

$$\sigma^{(\theta_1, \theta_2)} = \sum_{y_1, y_2, d_1, d_2} \underbrace{M_{y_1, y_2}^{(d_1, d_2)} q_{y_1, y_2}^{(\theta_1, \theta_2)} M_{y_1, y_2}^{(d_1, d_2)}}_{:= \sigma_{y_1, y_2; d_1, d_2}^{(\theta_1, \theta_2)}} \otimes |y_1, y_2; d_1, d_2\rangle\langle y_1, y_2; d_1, d_2|_R. \quad (4.7)$$

Depending on the values of  $y_1, y_2, d_1, d_2$ , the prover has to give different answers  $v_1, v_2$  to the verifier in order to be accepted in the self-testing protocol. For the analysis, it will be useful to split the state  $\sigma^{(\theta_1, \theta_2)}$  according to these correct answers. A formal definition follows.

**Definition 4.6** (Notation for partial post-measurement states). We define the following states:

$$\sigma^{(0, v_1; 0, v_2)} = \sum_{d_1, d_2} \sum_{\substack{y_1: \hat{b}(k_1, y_1) = v_1 \\ y_2: \hat{b}(k_2, y_2) = v_2}} \sigma_{y_1, y_2; d_1, d_2}^{(0, 0)} \otimes |y_1, y_2, d_1, d_2\rangle\langle y_1, y_2, d_1, d_2| \quad (4.8)$$

$$\sigma^{(0, v_1; 1, v_2)} = \sum_{d_1, y_2} \sum_{\substack{y_1: \hat{b}(k_1, y_1) = v_1 \\ d_2: \hat{u}(k_2, y_2, d_2) = v_2 \oplus \hat{b}(k_1, y_1)}} \sigma_{y_1, y_2; d_1, d_2}^{(0, 1)} \otimes |y_1, y_2, d_1, d_2\rangle\langle y_1, y_2, d_1, d_2| \quad (4.9)$$

$$\sigma^{(1, v_1; 0, v_2)} = \sum_{y_1, d_2} \sum_{\substack{d_1: \hat{u}(k_1, y_1, d_1) = v_1 \oplus \hat{b}(k_2, y_2) \\ y_2: \hat{b}(k_2, y_2) = v_2}} \sigma_{y_1, y_2; d_1, d_2}^{(1, 0)} \otimes |y_1, y_2, d_1, d_2\rangle\langle y_1, y_2, d_1, d_2| \quad (4.10)$$

$$\sigma^{(1, s_1; 1, s_2)} = \sum_{y_1, y_2} \sum_{\substack{d_1: \hat{u}(k_1, y_1, d_1) = s_2 \\ d_2: \hat{u}(k_2, y_2, d_2) = s_1}} \sigma_{y_1, y_2; d_1, d_2}^{(1, 1)} \otimes |y_1, y_2, d_1, d_2\rangle\langle y_1, y_2, d_1, d_2| \quad (4.11)$$

Note that for any  $\theta_1, \theta_2 \in \{0, 1\}$ :

$$\sigma^{(\theta_1, \theta_2)} = \sum_{v_1, v_2} \sigma^{(\theta_1, v_1; \theta_2, v_2)}. \quad (4.12)$$

*Remark 4.7.* This indexing scheme has slightly different interpretations for the test case and the Bell case. In the test case, i.e., for  $(\theta_1, \theta_2) = (0, 1)$  or  $(\theta_1, \theta_2) = (1, 0)$ ,  $v_1$  and  $v_2$  are the bits that the prover has to return to pass both of the verifier's checks (provided the questions ask for a measurement in the basis chosen at the start, i.e.,  $q_i = \theta_i$ ). In the Bell case, only the sum  $v_1 \oplus v_2$  of the bits returned by the prover is checked. Here,  $\sigma^{(1, s_1; 1, s_2)}$  is that part of the state for which  $s_1$  is the accepted sum  $v_1 \oplus v_2$  on question  $(0, 1)$ , and  $s_2$  is the accepted sum on question  $(1, 0)$ .

In particular, for the *honest* prover, we have (after tracing out the classical registers  $Y$  and  $R$ ):

$$\begin{aligned} \text{Tr}_{YR} [\sigma^{(0, v_1; 0, v_2)}] &= \frac{1}{4} |v_1, v_2\rangle\langle v_1, v_2|, \\ \text{Tr}_{YR} [\sigma^{(0, v_1; 1, v_2)}] &= \frac{1}{4} |v_1, (-)^{v_2}\rangle\langle v_1, (-)^{v_2}|, \\ \text{Tr}_{YR} [\sigma^{(1, v_1; 0, v_2)}] &= \frac{1}{4} |(-)^{v_1}, v_2\rangle\langle (-)^{v_1}, v_2|, \\ \text{Tr}_{YR} [\sigma^{(1, s_1; 1, s_2)}] &= \frac{1}{4} |\phi_H^{(s_1, s_2)}\rangle\langle \phi_H^{(s_1, s_2)}|, \quad \phi_H^{(s_1, s_2)} = (\sigma_X^{s_1} \otimes \sigma_X^{s_2})(|00\rangle + |01\rangle + |10\rangle - |11\rangle). \end{aligned}$$

We now prove a simple technical lemma which we will frequently use in the soundness proof.

**Lemma 4.8.** *Let  $D = (S, \Pi, M, P)$  be an efficient device. For any binary observable  $O$ , bits  $\theta_1, \theta_2 \in \{0, 1\}$ ,  $i \in \{1, 2\}$ , and  $\varepsilon \geq 0$ :*

$$\sum_{v_1, v_2} \text{Tr} [O^{(v_i)} \sigma^{(\theta_1, v_1; \theta_2, v_2)}] \approx_\varepsilon 1 \implies \forall v_1, v_2 \in \{0, 1\} : \text{Tr} [O^{(v_i)} \sigma^{(\theta_1, v_1; \theta_2, v_2)}] \approx_\varepsilon \text{Tr} [\sigma^{(\theta_1, v_1; \theta_2, v_2)}]. \quad (4.13)$$

*Proof.* Fix  $\theta_1, \theta_2$ , and  $i$ . For any  $v_1, v_2 \in \{0, 1\}$ ,  $O^{(v_i)}$  is a projector and therefore less than or equal to  $\mathbb{1}$ , and  $\sigma^{(\theta_1, v_1; \theta_2, v_2)}$  is positive for any  $v_i$ . Hence, there exist  $\delta_{v_1, v_2} \geq 0$  such that

$$\mathrm{Tr}\left[O^{(v_i)}\sigma^{(\theta_1, v_1; \theta_2, v_2)}\right] = \mathrm{Tr}\left[\sigma^{(\theta_1, v_1; \theta_2, v_2)}\right] - \delta_{v_1, v_2}. \quad (4.14)$$

Summing over  $v_1, v_2$  on both sides and using that  $\sigma^{(\theta_1, \theta_2)} = \sum_{v_1, v_2} \sigma^{(\theta_1, v_1; \theta_2, v_2)}$  is normalised, we get

$$\sum_{v_1, v_2} \mathrm{Tr}\left[O^{(v_i)}\sigma^{(\theta_1, v_1; \theta_2, v_2)}\right] = 1 - \sum_{v_1, v_2} \delta_{v_1, v_2}. \quad (4.15)$$

By assumption, the left hand side is  $\approx_\varepsilon 1$ , so  $\sum_{v_1, v_2} \delta_{v_1, v_2} \approx_\varepsilon 0$ . Since all  $\delta_{v_1, v_2}$  are positive, this implies  $\delta_{v_1, v_2} \approx_\varepsilon 0$  for all  $v_1, v_2$ , which completes the proof.  $\square$

**Corollary 4.9.** *Let  $D = (S, \Pi, M, P)$  be an efficient device. For any binary observable  $O$ , bits  $\theta_1, \theta_2 \in \{0, 1\}$ ,  $i \in \{1, 2\}$ , and  $\varepsilon \geq 0$ :*

$$\sum_{v_1, v_2} \mathrm{Tr}\left[O^{(v_i)}\sigma^{(\theta_1, v_1; \theta_2, v_2)}\right] \approx_\varepsilon 1 \implies \forall v_1, v_2 \in \{0, 1\}: O \approx_{\varepsilon, \sigma^{(\theta_1, v_1; \theta_2, v_2)}} (-1)^{v_i} \mathbb{1}. \quad (4.16)$$

*Proof.* This follows immediately by combining Lemmas 2.19 and 4.8.  $\square$

## 4.2 Success probabilities of a device

During the self-testing protocol, the verifier applies certain checks to the answers given by the prover. If the prover fails these checks, the verifier sets **flag** to **fail<sub>pre</sub>**, **fail<sub>test</sub>**, or **fail<sub>bell</sub>**. Here, we relate the probabilities that the prover passes these checks to the states and measurements used in the definition of devices (Definition 4.1).

**Lemma 4.10** (Success probability). *Let  $D = (S, \Pi, M, P)$  be a device.*

(i) **Preimage check:** *We define*

$$1 - \gamma_P(D) = \min P, \quad (4.17)$$

*with*

$$P = \left( \sum_{y_1, y_2, b_2, x_2} \mathrm{Tr}\left[\Pi_{y_1, y_2}^{(\hat{b}(k_1, y_1), \hat{x}(k_1, y_1); b_2, x_2)} \psi_{y_1, y_2}^{(0, \theta)}\right], \sum_{y_1, y_2, b_1, x_1} \mathrm{Tr}\left[\Pi_{y_1, y_2}^{(b_1, x_1; \hat{b}(k_2, y_2), \hat{x}(k_2, y_2))} \psi_{y_1, y_2}^{(\theta, 0)}\right], \right. \\ \left. \sum_{y_1, y_2, b_1, x_1, b} \mathrm{Tr}\left[\Pi_{y_1, y_2}^{(b_1, x_1; b, \hat{x}_b(k_2, y_2))} \psi_{y_1, y_2}^{(\theta, 1)}\right], \sum_{y_1, y_2, b_2, x_2, b} \mathrm{Tr}\left[\Pi_{y_1, y_2}^{(b, \hat{x}_b(k_1, y_1); b_2, x_2)} \psi_{y_1, y_2}^{(1, \theta)}\right] \right)_{\theta \in \{0, 1\}}.$$

*Then,  $\gamma_P(D)$  is bounded by the prover's success probability as follows:*

$$\gamma_P(D) \leq 8 \cdot \Pr[\mathbf{flag} = \mathbf{fail}_{\text{pre}}]. \quad (4.18)$$

*Here,  $k_1$  and  $k_2$  are the keys chosen by the verifier. (As always, we take an implicit expectation value over the keys chosen by the verifier.)*

(ii) **Test case:** *We define*

$$1 - \gamma_T(D) = \min T, \quad (4.19)$$

with

$$\begin{aligned}
T = & \left( \text{Tr} \left[ \sum_{v_1, v_2} Z_1^{(v_1)} \sigma^{(0, v_1; 1, v_2)} \right], \text{Tr} \left[ \sum_{v_1, v_2} \tilde{Z}_1^{(v_1)} \sigma^{(0, v_1; 1, v_2)} \right], \right. \\
& \text{Tr} \left[ \sum_{v_1, v_2} X_1^{(v_1)} \sigma^{(1, v_1; 0, v_2)} \right], \text{Tr} \left[ \sum_{v_1, v_2} \tilde{X}_1^{(v_1)} \sigma^{(1, v_1; 0, v_2)} \right], \\
& \text{Tr} \left[ \sum_{v_1, v_2} Z_2^{(v_2)} \sigma^{(1, v_1; 0, v_2)} \right], \text{Tr} \left[ \sum_{v_1, v_2} \tilde{Z}_2^{(v_2)} \sigma^{(1, v_1; 0, v_2)} \right], \\
& \left. \text{Tr} \left[ \sum_{v_1, v_2} X_2^{(v_2)} \sigma^{(0, v_1; 1, v_2)} \right], \text{Tr} \left[ \sum_{v_1, v_2} \tilde{X}_2^{(v_2)} \sigma^{(0, v_1; 1, v_2)} \right] \right) \quad (4.20)
\end{aligned}$$

Then,  $\gamma_T(D)$  is bounded by the prover's success probability as follows:

$$\gamma_T(D) \leq 8 \cdot \Pr[\text{flag} = \text{fail}_{\text{Test}}]. \quad (4.21)$$

(iii) **Bell case:** We define

$$1 - \gamma_B(D) = \min \left( \sum_{s_1, s_2} \text{Tr} \left[ (\tilde{Z}_1 \tilde{X}_2)^{(s_1)} \sigma^{(1, s_1; 1, s_2)} \right], \sum_{s_1, s_2} \text{Tr} \left[ (\tilde{X}_1 \tilde{Z}_2)^{(s_2)} \sigma^{(1, s_1; 1, s_2)} \right] \right). \quad (4.22)$$

Then,  $\gamma_B(D)$  is bounded by the prover's success probability as follows:

$$\gamma_B(D) \leq 2 \cdot \Pr[\text{flag} = \text{fail}_{\text{Bell}}]. \quad (4.23)$$

*Proof.* For a tuple  $P$ , we denote its elements by  $P_j$ .

(i) From the self-testing protocol, the definition of the CHK-procedure (Definition 2.1), and the fact that the verifier chooses bases  $\theta_1, \theta_2$  and questions  $q_1, q_2$  uniformly at random, it is clear that

$$\Pr[\text{flag} \geq \text{fail}_{\text{pre}}] = 1 - \frac{1}{8} \sum_{j=1}^8 P_j. \quad (4.24)$$

Because each element of  $P$  is upper-bounded by 1, this implies

$$\Pr[\text{flag} = \text{fail}_{\text{pre}}] \geq 1 - \frac{7}{8} - \frac{1}{8} \min P = \frac{\gamma_P(D)}{8}. \quad (4.25)$$

(ii) From the self-testing protocol and the definition of  $\sigma^{(\theta_1, v_1; \theta_2, v_2)}$ , it is clear that because the verifier chooses bases  $\theta_1, \theta_2$  and questions  $q_1, q_2$  uniformly at random, and  $\text{flag} = \text{fail}_{\text{Test}}$  occurs if the verifier's check fails on at least one answer  $v_i$ , we have

$$\Pr[\text{flag} = \text{fail}_{\text{Test}}] \geq 1 - \frac{1}{8} \sum_{j=1}^8 T_j. \quad (4.26)$$

The result now follows as in (i).

(iii) In the Bell case, the verifier checks one of

$$\tilde{Z}_1 \tilde{X}_2 = P_{0,1}^{(0,0)} - P_{0,1}^{(0,1)} - P_{0,1}^{(1,0)} + P_{0,1}^{(1,1)}, \quad \tilde{X}_1 \tilde{Z}_2 = P_{1,0}^{(0,0)} - P_{1,0}^{(0,1)} - P_{1,0}^{(1,0)} + P_{1,0}^{(1,1)} \quad (4.27)$$

uniformly at random. Considering the self-testing protocol and the definition of  $\sigma^{(1, s_1; 1, s_2)}$ , the result now follows as in (i).

□

*Remark 4.11.* Throughout the entire soundness proof, we will implicitly assume that for the device  $D$  under consideration, the quantities  $\gamma_P(D)$ ,  $\gamma_T(D)$ , and  $\gamma_B(D)$  are bounded away from 1 by a non-negligible amount. Indeed, in the case where one of these quantities is  $1 - \text{negl}(\lambda)$ , our soundness result (Theorem 4.42) trivially holds. To see this, note that in this case, for any constant  $c \geq 0$ ,

$$\gamma_P(D)^c + \gamma_T(D)^c + \gamma_B(D)^c \geq 1 - \text{negl}(\lambda).$$

Therefore, recalling that the definition of approximate equality (Definition 2.14) always includes a term  $\text{negl}(\lambda)$ , we see that

$$\rho \approx_{\gamma_P(D)^c + \gamma_T(D)^c + \gamma_B(D)^c} \sigma$$

holds for any two (potentially subnormalised) quantum states  $\rho$  and  $\sigma$ , since the trace distance  $\|\rho - \sigma\|_1$  is always at most  $2 = O(1)$ . Hence, the statements of Theorem 4.42 are trivially satisfied.

#### 4.2.1 Reduction to perfect device

The purpose of this section is to show that for the rest of the soundness proof, we can restrict ourselves to devices that pass the preimage round of the protocol with probability  $1 - \text{negl}(\lambda)$ . This is primarily a technical convenience that simplifies the arguments in later parts of the proof.

**Definition 4.12** (Perfect device). We call a device  $D$  *perfect* if  $\gamma_P(D) = \text{negl}(\lambda)$ .

The following lemma says that for any *efficient* device  $D$ , there exists another *efficient perfect* device  $D'$ , which uses the same measurements as  $D$ , and whose initial state is close to the initial state of  $D$ . Since we are ultimately interested in characterising the states and measurements of a device, this will allow us to first replace the arbitrary device by a perfect one, characterise the states and measurements of the perfect device, and finally argue that this characterisation also applies (up to some error) to the arbitrary device.

**Lemma 4.13.** *Let  $D = (S, \Pi, M, P)$  be an efficient device with  $\gamma_P(D) < 1$ , where  $S = \{\psi^{(\theta_1, \theta_2)}\}$ . Then there exists an efficient perfect device  $D' = (S', \Pi, M, P)$ , which uses the same measurements  $\Pi, M, P$  and whose states  $S' = \{\psi'^{(\theta_1, \theta_2)}\}$  satisfy for any  $\theta_1, \theta_2 \in \{0, 1\}$ :*

$$\left\| \psi'^{(\theta_1, \theta_2)} - \psi^{(\theta_1, \theta_2)} \right\|_1 \approx_{\gamma_P(D)^{1/2}} 0. \quad (4.28)$$

*Proof.* The idea of the proof is the same as in [Mah18, claim 7.2] and [GV19, Lemma 3.9]. We give a construction of  $D'$  as follows:  $D'$  first prepares the states  $\psi^{(\theta_1, \theta_2)}$  as  $D$  does.  $D'$  then applies the efficient unitary  $U_\Pi$  associated with the measurement  $\Pi$ :

$$|0_{2(1+|\mathcal{X}|)}\rangle\langle 0_{2(1+|\mathcal{X}|)}|_R \otimes \psi^{(\theta_1, \theta_2)} \xrightarrow{U_\Pi} |b_1, x_1; b_2, x_2\rangle\langle b_1, x_1; b_2, x_2|_R \otimes \Pi^{(b_1, x_1; b_2, x_2)} \psi^{(\theta_1, \theta_2)} \Pi^{(b_1, x_1; b_2, x_2)}.$$

Now  $D'$  coherently evaluates the (efficient) CHK-function on the  $Y$ -register of  $\Pi^{(b_1, x_1; b_2, x_2)} \psi^{(\theta_1, \theta_2)} \Pi^{(b_1, x_1; b_2, x_2)}$  and the new register containing  $b_i, x_i$ . If CHK succeeds,  $D'$  applies  $U^\dagger$  to the state, traces out the ancillary register  $R$ , and uses this as  $\psi'^{(\theta_1, \theta_2)}$ . Otherwise,  $D'$  repeats the process up to polynomially (in the security parameter) many times, and aborts if the CHK procedure never succeeds. Since  $\gamma_P(D)$  is defined as the maximum failure probability of the preimage check on one of the two qubits, and the CHK procedure fails if the preimage check fails on either qubit, the probability of the CHK procedure failing is at most  $2\gamma_P(D)$  by a union bound. Hence, recalling Remark 4.11, the probability that CHK fails polynomially many times is negligible. (We remark that the prover described by  $D'$  has to run this checking procedure before actually returning the images  $y, y'$  to the verifier.)

It is clear that  $D'$  is efficient and perfect. Fix  $\theta_1, \theta_2$ . We need to show  $\left\| \psi'^{(\theta_1, \theta_2)} - \psi^{(\theta_1, \theta_2)} \right\|_1 \approx_{\gamma_P(D)^{1/2}} 0$ . Since the probability of the CHK to succeed is at least  $1 - 2\gamma_P(D)$ , by the gentle measurement lemma (see e.g. [Wil11, lemma 9.4.1]), the post-measurement state after CHK has succeeded is  $O(\gamma_P(D)^{1/2})$ -close in trace distance to  $U(|0_{2(1+|\mathcal{X}|)}\rangle\langle 0_{2(1+|\mathcal{X}|)}|_R \otimes \psi^{(\theta_1, \theta_2)})U^\dagger$ . Because the trace distance is unitarily invariant, this implies that the state  $\psi'^{(\theta_1, \theta_2)}$  is also  $O(\gamma_P(D)^{1/2})$ -close in trace distance to  $\psi^{(\theta_1, \theta_2)}$ . □

### 4.3 Lifting relations from one basis choice to another

A lot of the leverage that the verifier has over the prover stems from the fact that the prover does not know the verifier's basis choices  $\theta_1, \theta_2$ . In particular, the prover does not know whether he is in the “Bell case”  $\theta_1 = \theta_2 = 1$ , where the honest prover prepares an entangled state, or in the “test case”  $\theta_1 \neq \theta_2$ , where the honest prover prepares a product state. The test case is useful as a testing procedure because the two bits that the prover has to return as an answer in the Hadamard round are determined by the  $y_1, y_2$  and  $d_1, d_2$  that the prover returned in the previous rounds. Using the trapdoor, the verifier can check each answer *individually*. In contrast, in the Bell case, only the sum of both answers is checked by the verifier.

In the soundness proof, we often want to “lift” approximate-equality relations that we can certify for one of the test cases, e.g.  $\theta_1 = 0, \theta_2 = 1$ , to any other choices of  $\theta_1, \theta_2$ . Intuitively, this is possible because the prover does not know which case it is in, so it cannot adapt its behaviour accordingly. We have already shown in Lemma 2.25 that we can lift relations from one state to another if the states are computationally indistinguishable and the relation only involves efficient quantities. Therefore, we only need to show that the different  $\sigma^{(\theta_1, \theta_2)}$  are computationally indistinguishable, which we do in the following simple lemma.

**Lemma 4.14.** *Let  $D = (S, \Pi, M, P)$  be an efficient device. Given a state  $\sigma^{(\theta_1, \theta_2)}$  with uniformly sampled  $(\theta_1, \theta_2)$ , no efficient procedure can correctly guess  $(\theta_1, \theta_2)$  with probability non-negligibly larger than  $1/4$ . The same holds for the states  $\psi^{(\theta_1, \theta_2)}$ .*

*Proof.* The values for  $\theta_i$  directly indicate whether  $k_i \in \mathcal{K}_G$  or  $k_i \in \mathcal{K}_F$ . Therefore, a procedure that correctly guesses  $\theta_1, \theta_2$  with probability non-negligibly larger than  $1/4$  would violate the injective invariance property of the ENTFCF family [Mah18, definition 4.3]. The same reasoning applies for  $\psi^{(\theta_1, \theta_2)}$ .  $\square$

### 4.4 Uniform normalisation and answers

In Definition 4.6, we defined the partial post-measurement states  $\sigma^{(\theta_1, v_1; \theta_2, v_2)}$ . In this section we show that in the test case (i.e.,  $\theta_1 \neq \theta_2$ ), the normalisation of a certain marginalisation of these states is uniform, i.e., the same for both  $v = 0$  and  $v = 1$ . This is a weaker statement than showing that the normalisation of  $\sigma^{(\theta_1, v_1; \theta_2, v_2)}$  itself (without marginalising) is uniform, but it will be sufficient. The proof reduces the uniform normalisation to the adaptive hardcore bit property of the ENTFCF family (informally, item (ii) in the list in Section 1.2).

**Lemma 4.15.** *Let  $D = (S, \Pi, M, P)$  be an efficient perfect device. Then:*

$$\sum_{v_2} \text{Tr} \left[ \sigma^{(1,0; 0, v_2)} \right] \approx_0 \sum_{v_2} \text{Tr} \left[ \sigma^{(1,1; 0, v_2)} \right], \quad \sum_{v_1} \text{Tr} \left[ \sigma^{(0, v_1; 1, 0)} \right] \approx_0 \sum_{v_1} \text{Tr} \left[ \sigma^{(0, v_1; 1, 1)} \right]. \quad (4.29)$$

*Proof.* We show the first relation, the second one is analogous. Assume for the sake of contradiction that

$$\sum_{v_2} \text{Tr} \left[ \sigma^{(1,0; 0, v_2)} \right] - \sum_{v_2} \text{Tr} \left[ \sigma^{(1,1; 0, v_2)} \right] \geq \mu(\lambda), \quad (4.30)$$

for some non-negligible positive  $\mu(\lambda)$ . Here, we assumed for concreteness that the left hand side is positive, but the proof is easily seen to also hold for the case where the left hand side is *smaller* than a non-negligible *negative* function by flipping the final bit in the output of the procedure  $\mathcal{A}$  below.

We want to show that this contradicts the adaptive hardcore bit property. To this end, we define the following efficient procedure  $\mathcal{A}$ :  $\mathcal{A}$  is given a key  $k_1 \in \mathcal{K}_F$  and samples another key and a trapdoor  $(k_2, t_{k_2}) \leftarrow \text{GEN}_G(1^\lambda)$ .  $\mathcal{A}$  first prepares the state  $\psi^{(\theta_1, \theta_2)}$  by performing the same operations as the device  $D$ , obtaining  $y_1, y_2$  in the process; this is efficient because  $D$  is efficient. Then,  $\mathcal{A}$  performs the preimage measurement  $\Pi$ , obtaining outcomes  $(b_1, x_1)$  and  $(b_2, x_2)$ . Finally, it measures  $M$ , obtaining outcomes  $d_1, d_2$ . Again, these measurements are efficient because  $D$  is efficient.  $\mathcal{A}$  outputs the tuple  $(b_1, x_1, d_1, \hat{b}(k_2, y_2))$ . Note that since  $\mathcal{A}$  has access to  $t_{k_2}$ , computing  $\hat{b}(k_2, y_2)$  is efficient.



We now argue that this indeed breaks the adaptive hardcore bit property. Because the device  $D$  is perfect, the preimage measurement yields a correct preimage with probability negligibly close to 1. Then, by the collapsing property [GV19, Lemma A.7], the states before and after the preimage measurement are computationally indistinguishable. Since  $M$  is an efficient measurement, this means that the outcome distributions obtained by measuring  $M$  directly on  $\psi^{(\theta_1, \theta_2)}$  and measuring  $M$  on the post-measurement state after having measured  $\Pi$  must be negligibly close. Hence,  $y_1, y_2, d_1, d_2$  obtained by  $\mathcal{A}$  have the same distribution (up to negligible difference) as the images and equation strings obtained by the device  $D$ . Using the definition of  $\sigma^{(1, v_1; 0, v_2)}$ , this means that on average over  $\mathcal{A}$ 's distribution over  $k_i, y_i, d_i$ :

$$\Pr\left[\hat{u}(k_1, y_1, d_1) = v_1 \oplus \hat{b}(k_2, y_2)\right] = \text{Tr}\left[\sum_{v_2} \sigma^{(1, v_1; 0, v_2)}\right]. \quad (4.31)$$

Combining this with the assumption in Equation (4.30), we see that  $\mathcal{A}$ 's output  $(b_1, x_1, d_1, \hat{b}(k_2, y_2))$  is “correct” (i.e., in the set  $H_{k_1}$  in [BCM<sup>+</sup>18, definition 3.1(iv)]) with non-negligible advantage.  $\square$

As a corollary to the above lemma, we can also show that for a device that succeeds with high probability in the test case, the answers returned on question  $q = 1$  (i.e., a Hadamard basis measurement) must be close to uniform.

**Corollary 4.16.** *Let  $D = (S, \Pi, M, P)$  be an efficient perfect device. Then:*

$$\text{Tr}\left[X_1 \sigma^{(\theta_1, \theta_2)}\right] \approx_{\gamma_T(D)} 0, \quad \text{Tr}\left[X_2 \sigma^{(\theta_1, \theta_2)}\right] \approx_{\gamma_T(D)} 0. \quad (4.32)$$

*Proof.* We show the first relation, the second is analogous. Since  $X_1$  is efficient, by the lifting lemma (Lemma 2.25(i)) and the indistinguishability of  $\sigma^{(\theta_1, \theta_2)}$  (Lemma 4.14), it suffices to show  $\text{Tr}\left[X_1 \sigma^{(1, 0)}\right] \approx_{\gamma_T(D)} 0$ . By the definition of  $\gamma_T(D)$  (Equation (4.19)) and Lemma 4.8, we have (using that  $X_1 = (-1)^{v_1} (2X^{(v_1)} - \mathbb{1})$  since  $X_1$  is a binary observable):

$$\text{Tr}\left[X_1 \sigma^{(1, 0)}\right] = \sum_{v_1, v_2} \text{Tr}\left[(-1)^{v_1} (2X^{(v_1)} - \mathbb{1}) \sigma^{(1, v_1; 0, v_2)}\right] \quad (4.33)$$

$$\approx_{\gamma_T(D)} \sum_{v_1, v_2} (-1)^{v_1} \text{Tr}\left[\sigma^{(1, v_1; 0, v_2)}\right]. \quad (4.34)$$

The result now follows from Lemma 4.15.  $\square$

## 4.5 Anti-commutation relations

The goal of this section is to prove the following proposition.

**Proposition 4.17.** *For any efficient perfect device  $D = (S, \Pi, M, P)$ , the following approximate anti-commutation relations hold for any  $\theta_1, \theta_2 \in \{0, 1\}$  and  $i \in \{1, 2\}$ :*

$$\{Z_i, X_i\} \approx_{\gamma_T(D)^{1/2}, \sigma^{(\theta_1, \theta_2)}} 0. \quad (4.35)$$

The proof is given at the end of this section. We first show a number of auxiliary lemmas.

**Lemma 4.18.** *For any efficient device  $D = (S, \Pi, M, P)$ , the following holds for any  $\theta_1, \theta_2 \in \{0, 1\}$  and  $i \in \{1, 2\}$ :*

$$\sum_b \text{Tr}\left[X_i Z_i^{(b)} \sigma^{(\theta_1, \theta_2)} Z_i^{(b)}\right] \approx_{\gamma_T(D)^{1/2}} 0. \quad (4.36)$$

*Proof.* To simplify the notation, we show this for  $i = 1$ ; the proof for  $i = 2$  is analogous. First note that because  $Z_1$  is efficient, by Lemmas 2.4, 2.9, and 4.14, the states  $\sum_b Z_1^{(b)} \sigma^{(\theta_1, \theta_2)} Z_1^{(b)}$  are computationally indistinguishable for different  $\theta_1, \theta_2$ . Because  $X_1$  is an efficient binary observable, by the lifting lemma (Lemma 2.25(i)) and the indistinguishability of  $\sigma^{(\theta_1, \theta_2)}$  (Lemma 4.14), it suffices to show the lemma for a particular choice of  $\theta_1, \theta_2$ .

We choose  $\theta_1 = 0, \theta_2 = 1$ . By the definition of  $\gamma_T$  (Equation (4.19)), and Corollary 4.9 combined with Lemma 2.20, we have that for all  $v_1, v_2 \in \{0, 1\}$ :

$$Z_1^{(v_1)} \approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} \mathbb{1}, \quad Z_1^{(\bar{v}_1)} \approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} 0. \quad (4.37)$$

We can use this and Lemma 2.22 to obtain

$$\sum_b Z_1^{(b)} \sigma^{(0, v_1; 1, v_2)} Z_1^{(b)} \approx_{\gamma_T(D)} \sigma^{(0, v_1; 1, v_2)}. \quad (4.38)$$

Using this and the replacement lemma (Lemma 2.21(ii)), we get:

$$\sum_b \text{Tr} \left[ X_1 Z_1^{(b)} \sigma^{(0, 1)} Z_1^{(b)} \right] = \sum_{b, v_1, v_2} \text{Tr} \left[ X_1 Z_1^{(b)} \sigma^{(0, v_1; 1, v_2)} Z_1^{(b)} \right] \quad (4.39)$$

$$\approx_{\gamma_T(D)^{1/2}} \sum_{b, v_1, v_2} \text{Tr} \left[ X_1 \sigma^{(0, v_1; 1, v_2)} \right] \quad (4.40)$$

$$= \text{Tr} \left[ X_1 \sigma^{(0, 1)} \right] \quad (4.41)$$

$$\approx_{\gamma_T(D)} 0. \quad (4.42)$$

In the last line, we used Corollary 4.16.  $\square$

For the proof of Proposition 4.17, Lemma 4.18 will not be sufficient. We will need the stronger statement that e.g.  $\sum_{b, v_2} \text{Tr} \left[ X_i Z_i^{(b)} \sigma^{(1, v_1; 0, v_2)} Z_i^{(b)} \right]$  is small *for every*  $v_1$ , not just their sum. This is shown in Lemma 4.21. For the proof, we have to make use of the preimage test, which will enable us to relate the statement of Lemma 4.21 to the adaptive hardcore bit property. This is achieved with the following lemma.

**Lemma 4.19.** *We define the following projectors, which project onto the correct preimage answer (for given keys  $k_1, k_2$ ):*

$$\tilde{\Pi}_{y_1, y_2}^{(b_1, b_2)} = \Pi_{y_1, y_2}^{(b_1, \hat{x}_{b_1}(k_1, y_1); b_2, \hat{x}_{b_2}(k_2, y_2))}, \quad (4.43)$$

with  $\tilde{\Pi}_{y_1, y_2}^{(b_1, b_2)} := 0$  if  $\perp \in \{\hat{x}_{b_1}(k_1, y_1), \hat{x}_{b_2}(k_2, y_2)\}$ . We denote their marginals by

$$\tilde{\Pi}_{1, y_1, y_2}^{(b)} = \tilde{\Pi}_{y_1, y_2}^{(b, 0)} + \tilde{\Pi}_{y_1, y_2}^{(b, 1)}, \quad \tilde{\Pi}_{2, y_1, y_2}^{(b)} = \tilde{\Pi}_{y_1, y_2}^{(0, b)} + \tilde{\Pi}_{y_1, y_2}^{(1, b)}. \quad (4.44)$$

For any efficient perfect device  $D = (S, \Pi, M, P)$ , the following holds for any  $\theta_1, \theta_2 \in \{0, 1\}$  and  $i \in \{1, 2\}$ :

$$\sum_{b, y_1, y_2, d_1, d_2} \left\| M_{y_1, y_2}^{(d_1, d_2)} \tilde{\Pi}_{i, y_1, y_2}^{(b)} - Z_{i, y_1, y_2, d_1, d_2}^{(b)} M_{y_1, y_2}^{(d_1, d_2)} \right\|_{\psi_{y_1, y_2}^{(\theta_1, \theta_2)}}^2 \approx_{\gamma_T(D)} 0. \quad (4.45)$$

Note that each term in the sum is positive, so the statement also holds for any sums over subsets of  $b, y_i, d_i$ .

*Proof.* We show this for  $i = 1$ , the proof for  $i = 2$  is analogous. Inserting the definition of the state-dependent norm and multiplying out the terms, we find that the left hand side of Equation (4.45) equals

$$\begin{aligned} & \sum_{b, d_1, d_2} \text{Tr} \left[ M^{(d_1, d_2)} \tilde{\Pi}_1^{(b)} \psi^{(\theta_1, \theta_2)} \tilde{\Pi}_1^{(b)} M^{(d_1, d_2)} \right] + \text{Tr} \left[ Z_{1, d_1, d_2}^{(b)} M^{(d_1, d_2)} \psi^{(\theta_1, \theta_2)} M^{(d_1, d_2)} Z_{1, d_1, d_2}^{(b)} \right] \\ & - \text{Tr} \left[ Z_{1, d_1, d_2}^{(b)} M^{(d_1, d_2)} \left( \tilde{\Pi}_1^{(b)} \psi^{(\theta_1, \theta_2)} + \psi^{(\theta_1, \theta_2)} \tilde{\Pi}_1^{(b)} \right) M^{(d_1, d_2)} \right], \end{aligned} \quad (4.46)$$

where we defined

$$\tilde{\Pi}_i^{(b)} = \sum_{y_1, y_2} \tilde{\Pi}_{i, y_1, y_2}^{(b)} \otimes |y_1, y_2\rangle\langle y_1, y_2|. \quad (4.47)$$

We treat each term in turn:

**1st term.** Since  $\{M^{(d_1, d_2)}\}_{d_1, d_2}$  forms a projective measurement, this term equals

$$\sum_b \text{Tr} \left[ \tilde{\Pi}_1^{(b)} \psi^{(\theta_1, \theta_2)} \right]. \quad (4.48)$$

With the definition of  $\tilde{\Pi}_{1, y_1, y_2}^{(b)}$  (Equation (4.44)) and  $\gamma_P$  (Equation (4.17)), one can see that this is negligibly close to 1 for a perfect device.

**2nd term.** This term equals 1, since  $\{M^{(d_1, d_2)}\}_{d_1, d_2}$  and  $\{Z_{1, d_1, d_2}^{(b)}\}_b$  form projective measurements.

**3rd term.** To bound this term, first note that since  $\tilde{\Pi}_1^{(0)} + \tilde{\Pi}_1^{(1)}$  is a projector, we have

$$\text{Tr} \left[ \left( \tilde{\Pi}_1^{(0)} + \tilde{\Pi}_1^{(1)} - \mathbf{1} \right)^2 \psi^{(\theta_1, \theta_2)} \right] = 1 - \text{Tr} \left[ \left( \tilde{\Pi}_1^{(0)} + \tilde{\Pi}_1^{(1)} \right) \psi^{(\theta_1, \theta_2)} \right]. \quad (4.49)$$

Therefore, for a perfect device:

$$\tilde{\Pi}_1^{(0)} + \tilde{\Pi}_1^{(1)} \approx_{0, \psi^{(\theta_1, \theta_2)}} \mathbf{1}. \quad (4.50)$$

We can therefore bound the third term as follows:

$$\begin{aligned} & \sum_{b, d_1, d_2} \text{Tr} \left[ Z_{1, d_1, d_2}^{(b)} M^{(d_1, d_2)} \left( \tilde{\Pi}_1^{(b)} \psi^{(\theta_1, \theta_2)} + \psi^{(\theta_1, \theta_2)} \tilde{\Pi}_1^{(b)} \right) M^{(d_1, d_2)} \right] \\ &= \sum_b \text{Tr} \left[ \left( \sum_{d_1, d_2} M^{(d_1, d_2)} Z_{1, d_1, d_2}^{(b)} M^{(d_1, d_2)} \right) \left( \tilde{\Pi}_1^{(b)} \psi^{(\theta_1, \theta_2)} + \psi^{(\theta_1, \theta_2)} \tilde{\Pi}_1^{(b)} \right) \right] \end{aligned}$$

Note that  $\|\sum_{d_1, d_2} M^{(d_1, d_2)} Z_{1, d_1, d_2}^{(b)} M^{(d_1, d_2)}\|_\infty \leq 1$ , since  $Z_{1, d_1, d_2}^{(b)}$  is a projector (and therefore  $Z_{1, d_1, d_2}^{(b)} \leq \mathbf{1}$ ), and the  $M^{(d_1, d_2)}$  are orthogonal projectors that sum to  $\mathbf{1}$ . Therefore, we can apply the replacement lemma (Lemma 2.21(i)) and use Equation (4.50) to find:

$$\begin{aligned} & \approx_0 \sum_{b, d_1, d_2} \text{Tr} \left[ Z_{1, d_1, d_2}^{(b)} M^{(d_1, d_2)} \left( \tilde{\Pi}_1^{(b)} \psi^{(\theta_1, \theta_2)} \left( \tilde{\Pi}_1^{(b)} + \tilde{\Pi}_1^{(\bar{b})} \right) + \left( \tilde{\Pi}_1^{(b)} + \tilde{\Pi}_1^{(\bar{b})} \right) \psi^{(\theta_1, \theta_2)} \tilde{\Pi}_1^{(b)} \right) M^{(d_1, d_2)} \right] \\ &= 2 \sum_{b, d_1, d_2} \text{Tr} \left[ Z_{1, d_1, d_2}^{(b)} M^{(d_1, d_2)} \left( \tilde{\Pi}_1^{(b)} \psi^{(\theta_1, \theta_2)} \tilde{\Pi}_1^{(b)} \right) M^{(d_1, d_2)} \right] \\ & \quad + \text{Tr} \left[ Z_{1, d_1, d_2}^{(b)} M^{(d_1, d_2)} \left( \tilde{\Pi}_1^{(b)} \psi^{(\theta_1, \theta_2)} \tilde{\Pi}_1^{(\bar{b})} + \tilde{\Pi}_1^{(\bar{b})} \psi^{(\theta_1, \theta_2)} \tilde{\Pi}_1^{(b)} \right) M^{(d_1, d_2)} \right] \\ &= 2 \sum_{b, d_1, d_2} \text{Tr} \left[ Z_{1, d_1, d_2}^{(b)} M^{(d_1, d_2)} \left( \tilde{\Pi}_1^{(b)} \psi^{(\theta_1, \theta_2)} \tilde{\Pi}_1^{(b)} \right) M^{(d_1, d_2)} \right]. \quad (4.51) \end{aligned}$$

For the last line, we used that  $\tilde{\Pi}_1^{(b)} \psi^{(\theta_1, \theta_2)} \tilde{\Pi}_1^{(\bar{b})} + \tilde{\Pi}_1^{(\bar{b})} \psi^{(\theta_1, \theta_2)} \tilde{\Pi}_1^{(b)}$  is actually independent of  $b$ , so we can perform the sums  $\sum_b Z_{1, d_1, d_2}^{(b)} = \mathbf{1}$  for any  $d_1, d_2$ , and  $\sum_{d_1, d_2} M^{(d_1, d_2)} = \mathbf{1}$ , and finally use the cyclicity of the trace and  $\tilde{\Pi}^{(0)} \tilde{\Pi}^{(1)} = 0$ .

Let us consider  $(\theta_1, \theta_2) = (0, 1)$ . Since  $D$  is a perfect device (i.e., fails the preimage test with negligible probability) and  $\tilde{\Pi}_1^{(b)}$  projects onto the correct preimage answer with the first bit being  $b$ , it follows from Definition 4.6 that

$$\sum_{d_1, d_2} M^{(d_1, d_2)} \tilde{\Pi}_1^{(b)} \psi^{(0, 1)} \tilde{\Pi}_1^{(b)} M^{(d_1, d_2)} \otimes |d_1, d_2\rangle\langle d_1, d_2| \approx_0 \sum_{v_2} \sigma^{(0, b; 1, v_2)}.$$

Therefore, from the definition of  $\gamma_T$  (Equation (4.19)) we have that

$$\sum_{b, d_1, d_2} \text{Tr} \left[ Z_{1, d_1, d_2}^{(b)} M^{(d_1, d_2)} \left( \tilde{\Pi}_1^{(b)} \psi^{(0, 1)} \tilde{\Pi}_1^{(b)} \right) M^{(d_1, d_2)} \right] \approx_{\gamma_T(D)} 1. \quad (4.52)$$

To conclude, we need to extend the statement to any choice of  $\theta_1, \theta_2$ . For this, we use the same reasoning as in the proof of the lifting lemma (Lemma 2.25). Specifically, we observe that there exists an efficient procedure  $\mathcal{A}$  that, given a state  $\psi$ , estimates the l.h.s. of Equation (4.52): the procedure  $\mathcal{A}$  first (efficiently) measures  $\{\sum_{b_2, x_2} \Pi^{(b, x_1; b_2, x_2)}\}_{b, x_1}$ , records the bit  $b$ , and discards  $x_1$ . Since we are dealing with a perfect device, this measurement returns a correct preimage  $x_1 = \hat{x}_b(k_1, y_1)$  with overwhelming probability, and hence produces a post-measurement state negligibly close to  $\tilde{\Pi}_1^{(b)} \psi^{(0,1)} \tilde{\Pi}_1^{(b)}$ . The procedure  $\mathcal{A}$  now measures  $\{M^{(d_1, d_2)}\}$  and discards the result, leaving the system in a state negligibly close to  $\sum_{d_1, d_2} M^{(d_1, d_2)} \tilde{\Pi}_1^{(b)} \psi^{(0,1)} \tilde{\Pi}_1^{(b)} M^{(d_1, d_2)}$ . Finally, the procedure  $\mathcal{A}$  measures  $Z$  to obtain a bit  $b'$  and outputs 1 if  $b = b'$ , and 0 otherwise. By construction, it is clear that this estimates the l.h.s. of Equation (4.52). Since the procedure  $\mathcal{A}$  is efficient, its probability of returning 1 must be negligibly close for the computationally indistinguishable input states  $\psi^{(\theta_1, \theta_2)}$ , as otherwise  $\mathcal{A}$  could be used to distinguish these states with non-negligible advantage. It follows that the l.h.s. of Equation (4.52) must be negligibly close for different  $\psi^{(\theta_1, \theta_2)}$ , so Equation (4.52) holds for all  $\theta_1, \theta_2$ .  $\square$

The following is a purely technical lemma that will be required for the proof of Lemma 4.21.

**Lemma 4.20.** *For any efficient perfect device  $D = (S, \Pi, M, P)$ , the following holds for any  $\theta_1, \theta_2, b_1, b_2 \in \{0, 1\}$ :*

$$\tilde{\Pi}^{(b_1, b_2)} \psi^{(\theta_1, 0)} \tilde{\Pi}^{(b_1, \bar{b}_2)} \approx_0 0, \quad \tilde{\Pi}^{(b_1, b_2)} \psi^{(0, \theta_2)} \tilde{\Pi}^{(\bar{b}_1, b_2)} \approx_0 0, \quad (4.53)$$

with  $\tilde{\Pi}^{(b_1, b_2)} := \sum_{y_1, y_2} \tilde{\Pi}_{y_1, y_2}^{(b_1, b_2)} \otimes |y_1, y_2\rangle\langle y_1, y_2|$ .

*Proof.* We show the first relation, the second one is analogous. Let  $k_1, k_2$  be the keys used by the verifier (which we usually leave implicit), and note that  $k_2 \in \mathcal{K}_{\mathcal{G}}$  because we are considering the state  $\psi^{(\theta_1, 0)}$ .

Define  $\mathcal{Y}_b = \{y_2 \in \mathcal{Y} \mid \hat{b}(k_2, y_2) = b\}$ . We claim that

$$\tilde{\Pi}^{(b_1, b_2)} \approx_{0, \psi^{(\theta_1, 0)}} \sum_{\substack{y_1 \in \mathcal{Y}, \\ y_2 \in \mathcal{Y}_{b_2}}} \tilde{\Pi}_{y_1, y_2}^{(b_1, b_2)} \otimes |y_1, y_2\rangle\langle y_1, y_2|. \quad (4.54)$$

This implies the lemma by the following argument. Making use of Lemma 2.22 and expanding  $\psi^{(\theta_1, 0)} = \sum_{y_1, y_2} \psi_{y_1, y_2}^{(\theta_1, 0)} \otimes |y_1, y_2\rangle\langle y_1, y_2|$ , all inner products between different  $y_2$  in  $\tilde{\Pi}^{(b_1, b_2)} \psi^{(\theta_1, 0)} \tilde{\Pi}^{(b_1, \bar{b}_2)}$  equal 0, since  $\mathcal{Y}_{b_2}$  and  $\mathcal{Y}_{\bar{b}_2}$  are disjoint for  $k_2 \in \mathcal{K}_{\mathcal{G}}$ .

Equation (4.54) can be shown as follows. We have

$$\mathrm{Tr} \left[ \left( \tilde{\Pi}^{(b_1, b_2)} - \sum_{y_1 \in \mathcal{Y}, y_2 \in \mathcal{Y}_{b_2}} \tilde{\Pi}_{y_1, y_2}^{(b_1, b_2)} \otimes |y_1, y_2\rangle\langle y_1, y_2| \right)^2 \psi^{(\theta_1, 0)} \right] = \sum_{y_1 \in \mathcal{Y}, y_2 \in \mathcal{Y} \setminus \mathcal{Y}_{b_2}} \mathrm{Tr} \left[ \tilde{\Pi}_{y_1, y_2}^{(b_1, b_2)} \psi_{y_1, y_2}^{(\theta_1, 0)} \right].$$

Comparing the definition of  $\gamma_P$  (Equation (4.17)) and the definition of  $\mathcal{Y}_{b_2}$ , the right hand side has to be negligibly close to 0 for a perfect device.  $\square$

We are now in a position to show that the statement of Lemma 4.18 also holds if we do not sum over both  $v_1$  and  $v_2$ , but only over the  $v_i$  associated with a computational basis measurement.

**Lemma 4.21.** *For any efficient perfect device  $D = (S, \Pi, M, P)$ , the following holds for any  $v_1, v_2 \in \{0, 1\}$ :*

$$\sum_{b, v'_2} \mathrm{Tr} \left[ X_1 Z_1^{(b)} \sigma^{(1, v_1; 0, v'_2)} Z_1^{(b)} \right] \approx_{\gamma_T(D)^{1/2}} 0, \quad (4.55)$$

$$\sum_{b, v'_1} \mathrm{Tr} \left[ X_2 Z_2^{(b)} \sigma^{(0, v'_1; 1, v_2)} Z_2^{(b)} \right] \approx_{\gamma_T(D)^{1/2}} 0. \quad (4.56)$$

*Proof.* We show the first relation, the proof of the second is analogous. Define the shorthand

$$\chi^{(v_1)} := \sum_{b, v_2'} \text{Tr} \left[ X_1 Z_1^{(b)} \sigma^{(1, v_1; 0, v_2')} Z_1^{(b)} \right] \quad (4.57)$$

First note that by Lemma 4.18, we have  $\chi^{(0)} + \chi^{(1)} \approx_{\gamma_T(D)^{1/2}} 0$ . Therefore, to show this lemma, it suffices to show

$$\chi^{(0)} \approx_{\gamma_T(D)^{1/2}} \chi^{(1)}. \quad (4.58)$$

Inserting the definition of  $\sigma^{(1, v_1; 0, v_2)}$ , we have

$$\chi^{(v_1)} = \sum_{b, y_1, y_2, d_2} \sum_{d_1: \hat{u}(k_1, y_1, d_1) = v_1 \oplus \hat{b}(k_2, y_2)} \text{Tr} \left[ X_{1, y_1, y_2, d_1, d_2} Z_{1, y_1, y_2, d_1, d_2}^{(b)} M_{y_1, y_2}^{(d_1, d_2)} \psi_{y_1, y_2}^{(1, 0)} M_{y_1, y_2}^{(d_1, d_2)} Z_{1, y_1, y_2, d_1, d_2}^{(b)} \right].$$

We would know like to use Lemma 4.19 to replace terms of the form  $ZM$  by terms of the form  $M\Pi$ . More specifically, we need to show:

$$\chi^{(v_1)} \approx_{\gamma_T(D)^{1/2}} \xi^{(v_1)}, \quad (4.59)$$

where

$$\xi^{(v_1)} := \sum_{b, y_1, y_2, d_2} \sum_{d_1: \hat{u}(k_1, y_1, d_1) = v_1 \oplus \hat{b}(k_2, y_2)} \text{Tr} \left[ X_{1, y_1, y_2, d_1, d_2} M_{y_1, y_2}^{(d_1, d_2)} \tilde{\Pi}_{1, y_1, y_2}^{(b)} \psi_{y_1, y_2}^{(1, 0)} \tilde{\Pi}_{1, y_1, y_2}^{(b)} M_{y_1, y_2}^{(d_1, d_2)} \right]. \quad (4.60)$$

Due to the sums in Lemma 4.19, we cannot apply the replacement lemma (Lemma 2.21) directly to show Equation 4.60. However, we can use the Cauchy-Schwarz inequality in a very similar manner as in the proof of Lemma 2.21. We give the full details in Lemma 4.22 afterwards.

Expanding  $\tilde{\Pi}_{1, y_1, y_2}^{(b)} = \sum_{b_2} \tilde{\Pi}_{y_1, y_2}^{(b, b_2)}$  and using Lemma 4.20 with the replacement lemma (Lemma 2.21) to discard the cross-terms,

$$\xi^{(v_1)} \approx_0 \sum_{b, b_2, y_1, y_2, d_2} \sum_{d_1: \hat{u}(k_1, y_1, d_1) = v_1 \oplus \hat{b}(k_2, y_2)} \text{Tr} \left[ X_{1, y_1, y_2, d_1, d_2} M_{y_1, y_2}^{(d_1, d_2)} \tilde{\Pi}_{y_1, y_2}^{(b, b_2)} \psi_{y_1, y_2}^{(1, 0)} \tilde{\Pi}_{y_1, y_2}^{(b, b_2)} M_{y_1, y_2}^{(d_1, d_2)} \right]. \quad (4.61)$$

It now suffices to show  $\xi^{(0)} \approx_0 \xi^{(1)}$ . For the sake of contradiction, assume  $\xi^{(0)} - \xi^{(1)} \geq \mu$  for a non-negligible positive  $\mu$ . As in the proof of Lemma 4.15, we assume that the left hand side is positive for concreteness; if  $\xi^{(0)} - \xi^{(1)} \leq -\mu$ , the proof also holds, but we have to flip the final bit in the output tuple of the procedure  $\mathcal{A}$  below. We show that this enables us to construct an efficient procedure  $\mathcal{A}$  that breaks the adaptive hardcore bit property [BCM<sup>+</sup>18, definition 3.1(4)].

The procedure  $\mathcal{A}$  takes as input a key  $k \in \mathcal{K}_{\mathcal{F}}$ . It sets  $k_1 := k$  and samples  $(k_2, t_{k_2}) \leftarrow \text{GEN}_{\mathcal{K}_{\mathcal{G}}}(1^\lambda)$ . It then uses  $k_1, k_2$  to construct the state  $\psi^{(1, 0)}$  in the same way as the device  $D$ , which is efficient by assumption. In the process, it obtains images  $y_1$  and  $y_2$ . It now performs the projective measurement  $\{\Pi^{(b_1, x_1; b_2, x_2)}\}$ , obtaining outcome  $(b_1, x_1; b_2, x_2)$ . Next,  $\mathcal{A}$  performs the measurement  $M$ , obtaining outcomes  $d_1, d_2$ . Finally,  $\mathcal{A}$  measures  $X_1$  to get an outcome  $u$ . Since the device  $D$  is efficient, all measurements performed by  $\mathcal{A}$  are efficient. Because  $\mathcal{A}$  sampled  $k_2$  itself, it has access to the trapdoor, so it can efficiently compute  $\hat{b}(k_2, y_2)$ . The output of  $\mathcal{A}$  is the tuple  $(b_1, x_1, d_1, u \oplus \hat{b}(k_2, y_2))$ .

Since  $D$  is a perfect device, with probability negligibly close to 1,  $\mathcal{A}$ 's measurement outcomes  $(b_1, x_1), (b_2, x_2)$  are valid preimages for  $y_1, y_2$ , respectively. Hence, after measuring  $M$ ,  $\mathcal{A}$  has prepared a state that is negligibly close to one of

$$\left\{ \sum_{\substack{b_1, b_2, \\ y_1, y_2, d_2}} \sum_{d_1: \hat{u}(k_1, y_1, d_1) = v_1 \oplus \hat{b}(k_2, y_2)} M_{y_1, y_2}^{(d_1, d_2)} \tilde{\Pi}_{y_1, y_2}^{(b_1, b_2)} \psi_{y_1, y_2}^{(1, 0)} \tilde{\Pi}_{y_1, y_2}^{(b_1, b_2)} M_{y_1, y_2}^{(d_1, d_2)} \right\}_{v_1}. \quad (4.62)$$

Since we assumed  $\xi^{(0)} - \xi^{(1)} \geq \mu$ , the bit  $u$  obtained by  $\mathcal{A}$  from the  $X$ -measurement is equal to  $v_1$  with non-negligible advantage over  $1/2$ . Since  $\mathcal{A}$  can compute  $\hat{b}(k_2, y_2)$  correctly,  $\mathcal{A}$ 's output  $u \oplus \hat{b}(k_2, y_2)$  equals  $\hat{u}(k_1, y_1, d_1)$  with non-negligible advantage over  $1/2$ . This contradicts the adaptive hardcore bit property [BCM<sup>+</sup>18, definition 3.1(4.)] and completes the proof.  $\square$

We prove the remaining step from the previous lemma.

**Lemma 4.22.** *With the notation from the previous lemma,*

$$\chi^{(v_1)} \approx_{\gamma_T(D)^{1/2}} \xi^{(v_1)}.$$

*Proof.* To simplify the notation, we use the shorthand  $y = (y_1, y_2)$ ,  $d = (d_1, d_2)$ , and  $S_y = \{d_1 \mid \hat{u}(k_1, y_1, d_1) = v_1 \oplus \hat{b}(k_2, y_2)\}$ . Then, we have

$$\begin{aligned} \chi^{(v_1)} - \xi^{(v_1)} &= \sum_{b, y_1, y_2, d_2} \sum_{d_1 \in S_y} \left\langle X_{1,y,d} Z_{1,y,d}^{(b)} M_y^{(d)}, Z_{1,y,d}^{(b)} M_y^{(d)} - M_y^{(d)} \tilde{\Pi}_{1,y}^{(b)} \right\rangle_{\psi_y^{(1,0)}} \\ &\quad + \left\langle Z_{1,y,d}^{(b)} M_y^{(d)} - M_y^{(d)} \tilde{\Pi}_{1,y}^{(b)}, X_{1,y,d} M_y^{(d)} \tilde{\Pi}_{1,y}^{(b)} \right\rangle_{\psi_y^{(1,0)}}. \end{aligned} \quad (4.63)$$

(To see this, one can simply insert Definition 2.10 to write the r.h.s. in terms of traces, and cancel crossterms.) Applying the triangle inequality and the Cauchy-Schwarz inequality for the state-dependent inner product, we find

$$\begin{aligned} \left| \chi^{(v_1)} - \xi^{(v_1)} \right| &\leq \sum_{b, y_1, y_2, d_2} \sum_{d_1 \in S_y} \left\| X_{1,y,d} Z_{1,y,d}^{(b)} M_y^{(d)} \right\|_{\psi_y^{(1,0)}} \cdot \left\| Z_{1,y,d}^{(b)} M_y^{(d)} - M_y^{(d)} \tilde{\Pi}_{1,y}^{(b)} \right\|_{\psi_y^{(1,0)}} \\ &\quad + \left\| Z_{1,y,d}^{(b)} M_y^{(d)} - M_y^{(d)} \tilde{\Pi}_{1,y}^{(b)} \right\|_{\psi_y^{(1,0)}} \cdot \left\| X_{1,y,d} M_y^{(d)} \tilde{\Pi}_{1,y}^{(b)} \right\|_{\psi_y^{(1,0)}}. \end{aligned} \quad (4.64)$$

Since each term in the sum is non-negative, we can extend the sum from  $d_1 \in S_y$  to all  $d_1$ . Applying the standard Cauchy-Schwarz inequality, we obtain:

$$\begin{aligned} \left| \chi^{(v_1)} - \xi^{(v_1)} \right| &\leq \left( \sum_{b, y_1, y_2, d_1, d_2} \left\| X_{1,y,d} Z_{1,y,d}^{(b)} M_y^{(d)} \right\|_{\psi_y^{(1,0)}}^2 \right)^{1/2} \cdot \left( \sum_{b, y_1, y_2, d_1, d_2} \left\| Z_{1,y,d}^{(b)} M_y^{(d)} - M_y^{(d)} \tilde{\Pi}_{1,y}^{(b)} \right\|_{\psi_y^{(1,0)}}^2 \right)^{1/2} \\ &\quad + \left( \sum_{b, y_1, y_2, d_1, d_2} \left\| Z_{1,y,d}^{(b)} M_y^{(d)} - M_y^{(d)} \tilde{\Pi}_{1,y}^{(b)} \right\|_{\psi_y^{(1,0)}}^2 \right)^{1/2} \cdot \left( \sum_{b, y_1, y_2, d_1, d_2} \left\| X_{1,y,d} M_y^{(d)} \tilde{\Pi}_{1,y}^{(b)} \right\|_{\psi_y^{(1,0)}}^2 \right)^{1/2}. \end{aligned} \quad (4.65)$$

Using the fact that  $X_{1,y,d}$  is a binary observable and therefore squares to identity, and that the projectors in a measurement sum to identity, it is easy to check that

$$\sum_{b, y_1, y_2, d_1, d_2} \left\| X_{1,y,d} Z_{1,y,d}^{(b)} M_y^{(d)} \right\|_{\psi_y^{(1,0)}}^2 = 1, \quad (4.66)$$

$$\sum_{b, y_1, y_2, d_1, d_2} \left\| X_{1,y,d} M_y^{(d)} \tilde{\Pi}_{1,y}^{(b)} \right\|_{\psi_y^{(1,0)}}^2 \approx_0 1. \quad (4.67)$$

The remaining term is bounded by Lemma 4.19:

$$\left| \chi^{(v_1)} - \xi^{(v_1)} \right| \leq 2 \cdot \left( \sum_{b, y_1, y_2, d_1, d_2} \left\| Z_{1,y,d}^{(b)} M_y^{(d)} - M_y^{(d)} \tilde{\Pi}_{1,y}^{(b)} \right\|_{\psi_y^{(1,0)}}^2 \right)^{1/2} \approx_{\gamma_T(D)^{1/2}} 0. \quad (4.68)$$

$\square$

This lemma enables us to prove the main result of this section, Proposition 4.17, which establishes the anti-commutation of  $X_i$  and  $Z_i$ .

*Proof of Proposition 4.17.* To simplify the notation, we do the proof for the case  $\{Z_1, X_1\}$ . The other case is analogous. By the lifting lemma (Lemma 2.25(iv)) and the indistinguishability of  $\sigma^{(\theta_1, \theta_2)}$  (Lemma 4.14), it suffices to show

$$\mathrm{Tr}\left[\{Z_1, X_1\}^2 \sigma^{(1,0)}\right] \approx_{\gamma_T(D)^{1/2}} 0. \quad (4.69)$$

By the definition of  $\gamma_T$  (Equation (4.19)) and Corollary 4.9, we have

$$X_1 \approx_{\gamma_T(D), \sigma^{(1, v_1; 0, v_2)}} (-1)^{v_1} \mathbb{1}. \quad (4.70)$$

Since  $Z_1$  is a binary observable, we can rewrite  $\{Z_1, X_1\}$  as follows:

$$Z_1 X_1 + X_1 Z_1 = (Z_1^{(0)} - Z_1^{(1)}) X_1 (Z_1^{(0)} + Z_1^{(1)}) + (Z_1^{(0)} + Z_1^{(1)}) X_1 (Z_1^{(0)} - Z_1^{(1)}) \quad (4.71)$$

$$= 2 \cdot Z_1^{(0)} X_1 Z_1^{(0)} - 2 \cdot Z_1^{(1)} X_1 Z_1^{(1)}. \quad (4.72)$$

Squaring this, using  $Z_1 = 2Z_1^{(0)} - \mathbb{1} = \mathbb{1} - 2Z_1^{(1)}$ ,  $X_1^2 = \mathbb{1}$ , and the fact that  $Z_1^{(0)}, Z_1^{(1)}$  are orthogonal projectors, we find

$$\frac{1}{4} \{Z_1, X_1\}^2 = \frac{1}{4} \left( (\mathbb{1} - 2Z_1^{(1)}) X_1 + X_1 (2Z_1^{(0)} - \mathbb{1}) \right)^2 \quad (4.73)$$

$$= \left( X_1 Z_1^{(0)} - Z_1^{(1)} X_1 \right)^2 \quad (4.74)$$

$$= X_1 Z_1^{(0)} X_1 Z_1^{(0)} + Z_1^{(1)} X_1 Z_1^{(1)} X_1. \quad (4.75)$$

Inserting Equation (4.75), we have

$$\frac{1}{4} \mathrm{Tr}\left[\{Z_1, X_1\}^2 \sigma^{(1,0)}\right] = \sum_{v_1, v_2} \mathrm{Tr}\left[\left(X_1 Z_1^{(0)} X_1 Z_1^{(0)} + Z_1^{(1)} X_1 Z_1^{(1)} X_1\right) \sigma^{(1, v_1; 0, v_2)}\right] \quad (4.76)$$

Using Equation (4.70) and the replacement lemma (Lemma 2.21(i)), we can replace the outer  $X_1$  operators by  $(-1)^{v_1} \mathbb{1}$  (the condition of constant operator norm in Lemma 2.21 is clearly satisfied for products of unitaries and projectors):

$$\approx_{\gamma_T(D)^{1/2}} \frac{1}{2} \sum_{v_1, v_2} (-1)^{v_1} \mathrm{Tr}\left[\left(Z_1^{(0)} X_1 Z_1^{(0)} + Z_1^{(1)} X_1 Z_1^{(1)}\right) \sigma^{(1, v_1; 0, v_2)}\right] \quad (4.77)$$

By Lemma 4.21:

$$\approx_{\gamma_T(D)^{1/2}} 0. \quad (4.78)$$

□

## 4.6 Commutation relations

Having shown that operators on the same “qubit” (i.e., with the same  $i$ ) anti-commute, we now turn to commutation relations. We have frequently referred to operators with different  $i$  as acting on different “qubits”, but pointed out in the overview at the start of Section 4 that this intuition is not yet justified, since we do not yet have a characterisation of the prover’s state in terms of qubits. In this section, we make an important step towards showing that the intuition of different  $i$ ’s corresponding to different qubits is indeed valid: we show that observables with different  $i$  (approximately) commute. This is clearly required if we want to think of the observables as acting on different qubits, since observables on different qubits necessarily commute.

**Lemma 4.23.** For any device  $D = (S, \Pi, M, P)$ , the following commutation relations hold:

$$[X_1, X_2] = 0, \quad [Z_1, Z_2] = 0.$$

*Proof.* This follows immediately from Definition 4.4 and the fact that for fixed  $i, j \in \{0, 1\}$ , the elements of  $\{P_{i,j}^{(a_1, a_2)}\}_{a_1, a_2 \in \{0, 1\}}$  are orthogonal projectors.  $\square$

**Proposition 4.24.** For any efficient device  $D = (S, \Pi, M, P)$ , the following approximate commutation relations hold for any  $\theta_1, \theta_2$ :

$$[Z_1, X_2] \approx_{\gamma_T(D), \sigma^{(\theta_1, \theta_2)}} 0, \quad [Z_2, X_1] \approx_{\gamma_T(D), \sigma^{(\theta_1, \theta_2)}} 0. \quad (4.79)$$

*Proof.* For simplicity, we restrict ourselves to proving the first relation; the other one is analogous. Since  $Z_i, X_i$  are efficient, by the lifting lemma (Lemma 2.25(iii)) and the indistinguishability of  $\sigma^{(\theta_1, \theta_2)}$  (Lemma 4.14), it suffices to show the statement for  $\sigma^{(0,1)}$ . We will split this state as  $\sigma^{(0,1)} = \sum_{v_1, v_2} \sigma^{(0, v_1; 1, v_2)}$  and apply Corollary 4.9 to each part, i.e., replace  $X_i$  and  $Z_i$  by  $\pm \mathbb{1}$ .

By the definition of  $\gamma_T$  (Equation (4.19)) and Corollary 4.9, we have for any  $v_1, v_2 \in \{0, 1\}$ :

$$Z_1 \approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} (-1)^{v_1} \mathbb{1}, \quad X_2 \approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} (-1)^{v_2} \mathbb{1}. \quad (4.80)$$

Because  $Z_1^\dagger Z_1 = X_2^\dagger X_2 = \mathbb{1}$ , Lemma 2.18(i) allows us to apply this repeatedly to the product  $Z_1 X_2$ :

$$Z_1 X_2 \approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} (-1)^{v_2} Z_1 \quad (4.81)$$

$$\approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} (-1)^{v_1} (-1)^{v_2} \mathbb{1} \quad (4.82)$$

$$\approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} X_2 \cdot (-1)^{v_1} \mathbb{1} \quad (4.83)$$

$$\approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} X_2 Z_1. \quad (4.84)$$

This holds for every  $v_1, v_2$ . Since  $\sigma^{(0,1)} = \sum_{v_1, v_2} \sigma^{(0, v_1; 1, v_2)}$  and each  $\sigma^{(0, v_1; 1, v_2)}$  is positive, the result follows from Lemma 2.18(ii).  $\square$

*Remark 4.25.* This proof relies on the fact that there is a basis choice for which there is only one accepted answer for both  $Z_1$  and  $X_2$  (or  $Z_2$  and  $X_1$ ). For the tilde observables, we would have to show approximate commutation of  $\tilde{Z}_1, \tilde{Z}_2$ , and  $\tilde{X}_1, \tilde{X}_2$ , since these cases are not covered in Lemma 4.23. However, there are no basis choices for which there is only one accepted answer for both  $\tilde{X}_1$  and  $\tilde{X}_2$ , since for the honest prover, the basis choice  $(1, 1)$  results in an entangled state by application of a  $CZ$  gate. This prevents us from applying this proof to  $[\tilde{X}_1, \tilde{X}_2]$ . Instead, our strategy will be to first show that the non-tilde observables can be rounded to Pauli observables, and then show that tilde and non-tilde observables are approximately equal (on the device's state), implying that the tilde observables can also be rounded to Pauli observables.

## 4.7 Approximate equality of non-tilde observables and Pauli observables

The goal of this section is to show that on the states  $\sigma^{(\theta_1, \theta_2)}$ , the non-tilde observables  $Z_i, X_i$  used by the prover are close to the respective Pauli matrices  $\sigma_Z, \sigma_X$  (under some isometry). The proof of this follows the steps outlined in item (4.) of the introduction to Section 4.

**Definition 4.26.** To simplify the notation, we write

$$A \approx_{R, \psi} B \quad (4.85)$$

if there exists a constant  $c > 0$  such that

$$A \approx_{\gamma_T(D)^c, \psi} B. \quad (4.86)$$

We also use this notation for the other approximate equalities in Definition 2.14.



The convenient feature of this notation is that when we use the replacement lemma (Lemma 2.21), we do not need to change the subscript.

The important results from the preceding sections are the commutation- and anti-commutation relations for the non-tilde observables, which hold for any  $\theta_1, \theta_2 \in \{0, 1\}$  and can be expressed with the shorthand notation as

$$\begin{aligned} \{Z_i, X_i\} &\approx_{\mathbb{R}, \sigma(\theta_1, \theta_2)} 0. && \text{(Proposition 4.17)} \\ [X_1, X_2] &= 0, & [Z_1, Z_2] &= 0. && \text{(Lemma 4.23)} \\ [X_1, Z_2] &\approx_{\mathbb{R}, \sigma(\theta_1, \theta_2)} 0, & [Z_1, X_2] &\approx_{\mathbb{R}, \sigma(\theta_1, \theta_2)} 0. && \text{(Proposition 4.24)} \end{aligned}$$

We now define the ‘‘swap isometry’’. This is the isometry which will map the prover’s states and observables to the desired Bell states and single-qubit Pauli observables.

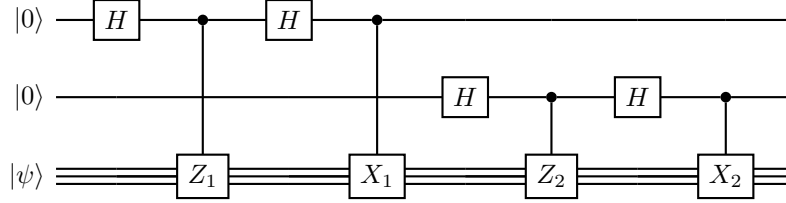
**Definition 4.27** (Swap isometry, [MYS12]). Given a device  $D = (S, \Pi, M, P)$  with Hilbert space  $\mathcal{H}$ , we define the *swap isometry*  $V_S : \mathcal{H} \mapsto \mathbb{C}^4 \otimes \mathcal{H}$  as

$$V_S = \frac{1}{4} \sum_{a, b \in \{0, 1\}} |a, b\rangle \otimes X_2^b (\mathbb{1} + (-1)^b Z_2) X_1^a (\mathbb{1} + (-1)^a Z_1). \quad (4.87)$$

(Note that the superscripts here signify exponents, not projectors.)

**Lemma 4.28.** *For an efficient device  $D = (S, \Pi, M, P)$ , the swap isometry  $V_S$  is efficient.*

*Proof.* It can be verified by a simple calculation that the following circuit implements the swap isometry:



We remark that this circuit is almost identical to that in [MYS12], but instead of applying  $Z_i$  and  $X_i$  to different parts of the state (which only makes sense if we have a Hilbert space with a tensor product structure), we apply all of them in sequence on the same Hilbert space. Also note that the swap isometry introduces an asymmetry between the observables  $Z_1, X_1$ , which are applied first, and the observables  $Z_2, X_2$ , which are applied afterwards.  $\square$

Our goal is to show that under the swap isometry, the prover’s observables  $Z_i, X_i$  are mapped to single-qubit Pauli observables. The following lemma collects the results of conjugating single-qubit Pauli observables by the swap isometry. Informally, the remainder of this section shows that in the state-dependent distance, the right hand sides of the equalities in Lemma 4.29 are close to  $Z_1, X_1, Z_2$ , and  $X_2$ , respectively.

**Lemma 4.29.** *Conjugating Pauli-operators by  $V_S$  yields the following*

$$V_S^\dagger (\sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S = Z_1, \quad (4.88)$$

$$V_S^\dagger (\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S = \frac{1}{4} \sum_{a \in \{0, 1\}} (\mathbb{1} + (-1)^a Z_1) X_1 (\mathbb{1} + (-1)^a Z_1), \quad (4.89)$$

$$V_S^\dagger (\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}) V_S = \frac{1}{4} \sum_{a \in \{0, 1\}} (\mathbb{1} + (-1)^a Z_1) X_1^a Z_2 X_1^a (\mathbb{1} + (-1)^a Z_1), \quad (4.90)$$

$$V_S^\dagger (\mathbb{1}_2 \otimes \sigma_X \otimes \mathbb{1}) V_S = \frac{1}{16} \sum_{a, b \in \{0, 1\}} (\mathbb{1} + (-1)^a Z_1) X_1^a (\mathbb{1} + (-1)^b Z_2) X_2 (\mathbb{1} + (-1)^b Z_2) X_1^a (\mathbb{1} + (-1)^a Z_1). \quad (4.91)$$

*Proof.* Inserting the definition of  $V_S$ , these can be verified by direct calculation.  $\square$

The following lemma shows item 4(ii) from the overview at the start of Section 4, namely that the swap isometry maps the prover's observable  $X_1$  to a Pauli  $\sigma_X$  observable on the first qubit.

**Lemma 4.30.** *For any efficient perfect device  $D = (S, \Pi, M, P)$ , the following holds for any  $\theta_1, \theta_2 \in \{0, 1\}$ :*

$$V_S^\dagger(\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1})V_S \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} X_1. \quad (4.92)$$

*Proof.* By Lemma 2.16, it suffices to show

$$\text{Tr} \left[ V_S^\dagger(\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1})V_S X_1 \sigma^{(\theta_1, \theta_2)} \right] \approx_{\mathbb{R}} 1. \quad (4.93)$$

We expand  $V_S^\dagger(\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1})V_S$  (Equation (4.89)) and get

$$\text{Tr} \left[ V_S^\dagger(\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1})V_S X_1 \sigma^{(\theta_1, \theta_2)} \right] = \frac{1}{4} \sum_{a \in \{0, 1\}} \text{Tr} \left[ (\mathbb{1} + (-1)^{\bar{a}} Z_1) X_1 (\mathbb{1} + (-1)^a Z_1) X_1 \sigma^{(\theta_1, \theta_2)} \right]. \quad (4.94)$$

At this point we would like to use  $\{Z_1, X_1\} \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} 0$  (Proposition 4.17) to anti-commute the right-most  $X_1$  and  $Z_1$ . Since  $(\mathbb{1} + (-1)^{\bar{a}} Z_1) X_1$  has constant operator norm, this can be achieved using the replacement lemma (Lemma 2.21(i)):

$$(4.94) \approx_{\mathbb{R}} \frac{1}{4} \sum_{a \in \{0, 1\}} \text{Tr} \left[ (\mathbb{1} + (-1)^{\bar{a}} Z_1) X_1 X_1 (\mathbb{1} + (-1)^{\bar{a}} Z_1) \sigma^{(\theta_1, \theta_2)} \right] \quad (4.95)$$

$$= 1. \quad (4.96)$$

In the last line we used  $\sum_{a \in \{0, 1\}} (\mathbb{1} + (-1)^{\bar{a}} Z_1) X_1 X_1 (\mathbb{1} + (-1)^{\bar{a}} Z_1) = 4$ , which follows from  $X_1^2 = Z_1^2 = \mathbb{1}$ .  $\square$

Having established a characterisation of the prover's operators  $Z_1$  and  $X_1$ , we now use this to partially characterise the prover's state. In particular, we will show that in the test case, the swap isometry maps the prover's state to a product state, where the first qubit is in the computational or Hadamard basis, depending on the verifier's basis choice (Lemma 4.31). We will then show that the auxiliary states that the prover holds in addition to the first qubit must be computationally indistinguishable to the prover (Lemma 4.32). This is similar to the result of [GV19] (but with fewer different single-qubit states).

**Lemma 4.31.** *Let  $D = (S, \Pi, M, P)$  be an efficient perfect device. Then, for  $v_1, v_2 \in \{0, 1\}$ , there exist positive matrices  $\alpha^{(0, v_1; 1, v_2)}$  and  $\alpha^{(1, v_1; 0, v_2)}$  such that the following holds:*

$$V_S \sigma^{(1, v_1; 0, v_2)} V_S^\dagger \approx_{\mathbb{R}} |(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \alpha^{(1, v_1; 0, v_2)}, \quad (4.97)$$

$$V_S \sigma^{(0, v_1; 1, v_2)} V_S^\dagger \approx_{\mathbb{R}} |v_1\rangle\langle v_1| \otimes \alpha^{(0, v_1; 1, v_2)}. \quad (4.98)$$

*Proof.* We show the first relation. The second one is analogous (but simpler, because we have  $V_S^\dagger(\sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1})V_S = Z_1$  (Equation (4.88)), whereas the corresponding statement for  $X_1$  only holds approximately (Lemma 4.30)).

By Lemmas 4.30 and 2.18(ii), we have

$$X_1 \approx_{\mathbb{R}, \sigma^{(1, v_1; 0, v_2)}} V_S^\dagger(\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1})V_S. \quad (4.99)$$

By Lemma 2.24, this implies

$$X_1^{(v_1)} \approx_{\mathbb{R}, \sigma^{(1, v_1; 0, v_2)}} V_S^\dagger(|(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \mathbb{1}_2 \otimes \mathbb{1})V_S. \quad (4.100)$$

Using this, the definition of  $\gamma_T$  (Equation (4.19)), and the replacement lemma (Lemma 2.21(i)) we get

$$\sum_{v_1, v_2} \text{Tr} \left[ (|(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S \sigma^{(1, v_1; 0, v_2)} V_S^\dagger \right] \approx_{\mathbb{R}} 1. \quad (4.101)$$

Using (a simple extension of) Corollary 4.9 combined with Lemma 2.20, this means that

$$|(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \mathbb{1}_2 \otimes \mathbb{1} \approx_{\mathbb{R}, V_S \sigma^{(1, v_1; 0, v_2)} V_S^\dagger} \mathbb{1}. \quad (4.102)$$

Hence, by Lemma 2.22, we get

$$V_S \sigma^{(1, v_1; 0, v_2)} V_S^\dagger \approx_{\mathbb{R}} (|(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S \sigma^{(1, v_1; 0, v_2)} V_S^\dagger (|(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \mathbb{1}_2 \otimes \mathbb{1}). \quad (4.103)$$

Defining

$$\alpha^{(1, v_1; 0, v_2)} = (|(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S \sigma^{(1, v_1; 0, v_2)} V_S^\dagger (|(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \mathbb{1}_2 \otimes \mathbb{1}) \quad (4.104)$$

yields the result.  $\square$

**Lemma 4.32.** *Let  $D = (S, \Pi, M, P)$  be an efficient perfect device. Then, there exists a (normalised) state  $\alpha$  such that the following holds for  $v_1 \in \{0, 1\}$ :*

$$\sum_{v_2} V_S \sigma^{(1, v_1; 0, v_2)} V_S^\dagger \stackrel{c}{\approx}_{\mathbb{R}} \frac{1}{2} |(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \alpha, \quad (4.105)$$

$$\sum_{v_2} V_S \sigma^{(0, v_1; 1, v_2)} V_S^\dagger \stackrel{c}{\approx}_{\mathbb{R}} \frac{1}{2} |v_1\rangle\langle v_1| \otimes \alpha. \quad (4.106)$$

*Proof.* We first prove the first relation, the proof of the second one is analogous and we briefly comment on it at the end.

**Proof of the first relation.** We need to show that  $\{\sum_{v_2} \alpha^{(1, v_1; 0, v_2)}\}_{v_1}$  are computationally indistinguishable up to  $O(\gamma_T(D)^c)$  for some constant  $c$ . Combined with Lemma 4.31, this proves the desired statement.

By Lemma 4.31, there exists a  $c > 0$  and an  $\varepsilon = O(\gamma_T(D)^c)$  such that for any  $v_1, v_2$ :

$$\begin{aligned} \left\| V_S \sigma^{(0, v_1; 1, v_2)} V_S^\dagger - |v_1\rangle\langle v_1| \otimes \alpha^{(0, v_1; 1, v_2)} \right\|_1^2 &\leq \varepsilon, \\ \left\| V_S \sigma^{(1, v_1; 0, v_2)} V_S^\dagger - |(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \alpha^{(1, v_1; 0, v_2)} \right\|_1^2 &\leq \varepsilon. \end{aligned} \quad (4.107)$$

For the sake of contradiction, assume that there is an efficient POVM  $\{\Lambda^{(v_1)}\}_{v_1 \in \{0, 1\}}$  such that

$$\sum_{v_2} \text{Tr} \left[ \Lambda^{(0)} \left( \alpha^{(1, 0; 0, v_2)} - \alpha^{(1, 1; 0, v_2)} \right) \right] \geq 24 \varepsilon^{1/2} + 2 \mu(\lambda), \quad (4.108)$$

where  $\mu(\lambda)$  is non-negligible. Define

$$\Gamma = V_S^\dagger \left( |+\rangle\langle+| \otimes \Lambda^{(0)} \right) V_S. \quad (4.109)$$

The outcome of the POVM  $\{\Gamma, \mathbb{1} - \Gamma\}$  can be efficiently estimated as follows: apply  $V_S$ ; measure the first register in the Hadamard basis, obtaining a bit  $a$ ; measure  $\{\Lambda^{(u)}\}$  on the remaining registers, obtaining a bit  $u$ ; output 0 if and only if  $a = u = 0$ .

We now show that the POVM  $\{\Gamma, \mathbb{1} - \Gamma\}$  enables us to construct a distinguisher that can distinguish  $\sigma^{(1, 0)}$  from  $\sigma^{(0, 1)}$  with non-negligible advantage, contradicting Lemma 4.14. Given  $\sigma^{(\bar{b}, b)}$  for  $b \in \{0, 1\}$ , the

distinguisher measures the POVM  $\{\Gamma, \mathbb{1} - \Gamma\}$ . If the result of the measurement is “ $\Gamma$ ”, the distinguisher guesses “0”; if the result is “ $\mathbb{1} - \Gamma$ ”, the distinguisher guesses “1”. We calculate the distinguishing advantage:

$$\begin{aligned} & \Pr[\text{Guess} = 0|b = 0] - \Pr[\text{Guess} = 0|b = 1] \\ &= \text{Tr}\left[\Gamma\sigma^{(1,0)}\right] - \text{Tr}\left[\Gamma\sigma^{(0,1)}\right] \\ &= \text{Tr}\left[\left(|+\rangle\langle+| \otimes \Lambda^{(0)}\right) \left(V_S\sigma^{(1,0)}V_S^\dagger - V_S\sigma^{(0,1)}V_S^\dagger\right)\right] \end{aligned}$$

Using Equation (4.107) and the replacement lemma (Lemma 2.21(ii); the non-asymptotic statement used here is easily seen to hold from the proof) to replace the state inside the trace:

$$\begin{aligned} & \geq \sum_{v_1, v_2} \text{Tr}\left[\left(|+\rangle\langle+| \otimes \Lambda^{(0)}\right) \left(|(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \alpha^{(1, v_1; 0, v_2)} - |v_1\rangle\langle v_1| \otimes \alpha^{(0, v_1; 1, v_2)}\right)\right] - 8\varepsilon^{1/2} \\ &= \sum_{v_2} \text{Tr}\left[\Lambda^{(0)}\alpha^{(1,0; 0, v_2)}\right] - \frac{1}{2} \sum_{v_1, v_2 \in \{0,1\}} \text{Tr}\left[\Lambda^{(0)}\alpha^{(0, v_1; 1, v_2)}\right] - 8\varepsilon^{1/2} \end{aligned}$$

Using that  $\sum_{v_2 \in \{0,1\}} \alpha^{(1,0; 0, v_2)} = \frac{1}{2} \sum_{v_2 \in \{0,1\}} (\alpha^{(1,0; 0, v_2)} - \alpha^{(1,1; 0, v_2)}) + \frac{1}{2} \sum_{v_1, v_2} \alpha^{(1, v_1; 0, v_2)}$ :

$$= \frac{1}{2} \sum_{v_2} \text{Tr}\left[\Lambda^{(0)}\left(\alpha^{(1,0; 0, v_2)} - \alpha^{(1,1; 0, v_2)}\right)\right] - \frac{1}{2} \sum_{v_1, v_2 \in \{0,1\}} \text{Tr}\left[\Lambda^{(0)}\left(\alpha^{(0, v_1; 1, v_2)} - \alpha^{(1, v_1; 0, v_2)}\right)\right] - 8\varepsilon^{1/2}$$

By the assumption in Equation (4.108), the first term is at least  $12\varepsilon^{1/2} + \mu(\lambda)$ . Expanding the second term:

$$\geq -\frac{1}{2} \sum_{v_1, v_2 \in \{0,1\}} \text{Tr}\left[\left(\mathbb{1} \otimes \Lambda^{(0)}\right) \left(|v_1\rangle\langle v_1| \otimes \alpha^{(0, v_1; 1, v_2)} - |(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \alpha^{(1, v_1; 0, v_2)}\right)\right] + 4\varepsilon^{1/2} + \mu(\lambda)$$

Using Equation (4.107) and the replacement lemma (Lemma 2.21(ii) as before:

$$\begin{aligned} & \geq -\frac{1}{2} \sum_{v_1, v_2 \in \{0,1\}} \text{Tr}\left[\left(\mathbb{1} \otimes \Lambda^{(0)}\right) \left(V_S\sigma^{(0,1)}V_S^\dagger - V_S\sigma^{(1,0)}V_S^\dagger\right)\right] + \mu(\lambda) \\ &= -\frac{1}{2} \sum_{v_1, v_2 \in \{0,1\}} \text{Tr}\left[V_S^\dagger\left(\mathbb{1} \otimes \Lambda^{(0)}\right)V_S\left(\sigma^{(0,1)} - \sigma^{(1,0)}\right)\right] + \mu(\lambda). \end{aligned}$$

Since  $\{V_S^\dagger(\mathbb{1} \otimes \Lambda^{(0)})V_S, \mathbb{1} - V_S^\dagger(\mathbb{1} \otimes \Lambda^{(0)})V_S\}$  is an efficient measurement on account of  $V_S$  and  $\{\Lambda^{(0)}, \Lambda^{(1)}\}$  being efficient, the first term must be negligibly close to 0 by Lemma 4.14. Therefore, we find that the distinguishing advantage is non-negligibly greater than 0, contradicting Lemma 4.14 and finishing the proof.

**Proof of the second relation.** The proof is analogous to that of the first relation. Assuming a measurement  $\{\Lambda^{(0)}, \Lambda^{(1)}\}$  for distinguishing  $\{\sum_{v_2} \alpha^{(0, v_1; 0, v_2)}\}_{v_1}$ , one chooses

$$\Gamma = V_S^\dagger\left(|0\rangle\langle 0| \otimes \Lambda^{(0)}\right) \quad (4.110)$$

and calculates the distinguishing advantage analogously to the first part.  $\square$

At this point, we have established a characterisation of the prover’s observables  $Z_1$  and  $X_1$  as well as of its state in the test case. As outlined in the introduction to Section 4, we will now use these results to show that the prover’s observables  $Z_2$  and  $X_2$  are also close to Pauli observables (under the swap isometry  $V_S$ ).

**Lemma 4.33.** *For any efficient perfect device  $D = (S, \Pi, M, P)$ , the following holds for any  $\theta_1, \theta_2 \in \{0, 1\}$ :*

$$V_S^\dagger(\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1})V_S \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} Z_2, \quad (4.111)$$

$$V_S^\dagger(\mathbb{1}_2 \otimes \sigma_X \otimes \mathbb{1})V_S \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} X_2. \quad (4.112)$$

*Proof.* We prove each relation in turn.

**Proof of the first relation.** By Lemmas 4.14 and 4.32, we have

$$V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger \stackrel{c}{\approx}_0 \sum_{v_1, v_2} V_S \sigma^{(0, v_1; 1, v_2)} V_S^\dagger \stackrel{c}{\approx}_R \frac{1}{2} \mathbb{1}_2 \otimes \alpha. \quad (4.113)$$

This implies

$$\sigma^{(\theta_1, \theta_2)} = \sum_{v_1, v_2} \sigma^{(0, v_1; 1, v_2)} \stackrel{c}{\approx}_R V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) V_S, \quad (4.114)$$

since any distinguisher  $D'$  for the latter problem can be used to construct a distinguisher  $D$  for the former problem: given  $V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger$  or  $\frac{1}{2} \mathbb{1}_2 \otimes \alpha$ ,  $D$  first applies the inverse of the unitary extension of  $V_S$ .  $D$  then measures the ancillary registers used for the unitary extension: if the result is not the all-zero string,  $D$  guesses that it was given  $\frac{1}{2} \mathbb{1}_2 \otimes \alpha$ ; if the result is the all-zero string,  $D$  runs  $D'$  on the post-measurement state. Then, the advantage of  $D$  in distinguishing the states in Equation (4.113) will be at least the advantage of  $D'$  in distinguishing the state in Equation (4.114).

The binary observables  $\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}$  and  $Z_2$  as well as the isometry  $V_S$  are all efficient. Therefore, by Lemma 2.25(v) we can replace  $\sigma^{(\theta_1, \theta_2)}$  by  $V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) V_S$  in the statement of this lemma, so we need to show

$$V_S^\dagger (\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}) V_S \approx_{R, V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) V_S} Z_2. \quad (4.115)$$

By Lemma 2.16, it suffices to show

$$\text{Tr} \left[ Z_2 V_S^\dagger (\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}) V_S V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) V_S \right] \approx_R 1. \quad (4.116)$$

Using the cyclicity of the trace,  $V_S^\dagger V_S = \mathbb{1}$ , and Equation (4.90) to expand  $V_S^\dagger (\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}) V_S$ , we get:

$$\begin{aligned} & \text{Tr} \left[ Z_2 V_S^\dagger (\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}) V_S V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) V_S \right] \\ &= \frac{1}{4} \sum_{a \in \{0, 1\}} \text{Tr} \left[ V_S Z_2 (\mathbb{1} + (-1)^a Z_1) X_1^a Z_2 X_1^a V_S^\dagger V_S (\mathbb{1} + (-1)^a Z_1) V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right] \end{aligned} \quad (4.117)$$

At this point, we would like to replace the right-most operator  $V_S (\mathbb{1} + (-1)^a Z_1) V_S^\dagger$  by  $\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}$ . For this, we need

$$\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1} \approx_{R, \frac{1}{2} \mathbb{1}_2 \otimes \alpha} V_S (\mathbb{1} + (-1)^a Z_1) V_S^\dagger. \quad (4.118)$$

We delay the proof of this statement until Lemma 4.34 and continue with the main argument here. Because for any  $a \in \{0, 1\}$ , the operator norm of the other operators inside the trace is constant, we can use Equation (4.118) and the replacement lemma (Lemma 2.21(i)) to obtain:

$$(4.117) \approx_R \frac{1}{4} \sum_{a \in \{0, 1\}} \text{Tr} \left[ V_S Z_2 (\mathbb{1} + (-1)^a Z_1) X_1^a Z_2 X_1^a V_S^\dagger (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right]$$

The right-most operator only acts non-trivially on the first qubit, so we can commute it past the state and use the cyclicity of the trace.

$$= \frac{1}{4} \sum_{a \in \{0, 1\}} \text{Tr} \left[ (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S Z_2 (\mathbb{1} + (-1)^a Z_1) X_1^a V_S^\dagger V_S Z_2 X_1^a V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right]. \quad (4.119)$$

Now we would like to commute the right-most  $X_1^a$  with  $Z_2$ . We need

$$V_S [Z_2, X_1] V_S^\dagger \approx_{R, \frac{1}{2} \mathbb{1}_2 \otimes \alpha} 0, \quad (4.120)$$

the proof of which is given in Lemma 4.34.

We therefore have (using Lemma 2.21 as above):

$$(4.119) \approx_{\mathbb{R}} \frac{1}{4} \sum_{a \in \{0,1\}} \text{Tr} \left[ (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S Z_2 (\mathbb{1} + (-1)^a Z_1) X_1^a X_1^a Z_2 V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right] \quad (4.121)$$

Using  $X_1^2 = \mathbb{1}$ ,  $[Z_1, Z_2] = 0$ , and  $Z_2^2 = \mathbb{1}$ :

$$= \frac{1}{4} \sum_{a \in \{0,1\}} \text{Tr} \left[ (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S (\mathbb{1} + (-1)^a Z_1) V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right] \quad (4.122)$$

Using Equation (4.118) in the same manner as above:

$$\approx_{\mathbb{R}} \frac{1}{4} \sum_{a \in \{0,1\}} \text{Tr} \left[ (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1})^2 \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right] \quad (4.123)$$

Using  $\sum_a (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1})^2 = 4 \cdot \mathbb{1}$  and the normalisation of  $\alpha$ :

$$= 1. \quad (4.124)$$

This proves the first relation in the lemma,  $V_S^\dagger (\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}) V_S \approx_{\mathbb{R}, \sigma(\theta_1, \theta_2)} Z_2$ .

**Proof of the second relation.** The proof is similar to the first case. As in the first case, the proof reduces to showing

$$\text{Tr} \left[ X_2 V_S^\dagger (\mathbb{1}_2 \otimes \sigma_X \otimes \mathbb{1}) V_S V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) V_S \right] \approx_{\mathbb{R}} 1. \quad (4.125)$$

For the proof, we will need the relation

$$\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1} \approx_{\mathbb{R}, \frac{1}{2} \mathbb{1}_2 \otimes \alpha} V_S X_1 V_S^\dagger, \quad (4.126)$$

which is analogous to Equation (4.118) and will be shown in Lemma 4.34. We proceed with proving Equation (4.125). Throughout the proof, we will always use the replacement lemma (Lemma 2.21(i)) to replace operators with one another. Expanding  $V_S^\dagger (\mathbb{1}_2 \otimes \sigma_X \otimes \mathbb{1}) V_S$  using Equation (4.91):

$$\begin{aligned} & \text{Tr} \left[ X_2 V_S^\dagger (\mathbb{1}_2 \otimes \sigma_X \otimes \mathbb{1}) V_S V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) V_S \right] \\ &= \frac{1}{16} \sum_{a, b \in \{0,1\}} \text{Tr} \left[ V_S X_2 (\mathbb{1} + (-1)^a Z_1) X_1^a (\mathbb{1} + (-1)^{\bar{b}} Z_2) X_2 (\mathbb{1} + (-1)^b Z_2) X_1^a (\mathbb{1} + (-1)^a Z_1) V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right] \end{aligned}$$

Using Equation (4.118) to replace  $V_S (\mathbb{1} + (-1)^a Z_1) V_S^\dagger$  with  $(\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1})$  and commuting it past the state:

$$\approx_{\mathbb{R}} \frac{1}{16} \sum_{a, b \in \{0,1\}} \text{Tr} \left[ (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S X_2 (\mathbb{1} + (-1)^a Z_1) X_1^a (\mathbb{1} + (-1)^{\bar{b}} Z_2) X_2 (\mathbb{1} + (-1)^b Z_2) X_1^a V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right]$$

Using Equation (4.126) to replace  $V_S X_1 V_S^\dagger$  with  $\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1}$  and commuting it past the state:

$$\approx_{\mathbb{R}} \frac{1}{16} \sum_{a, b \in \{0,1\}} \text{Tr} \left[ (\sigma_X^a \otimes \mathbb{1}_2 \otimes \mathbb{1}) (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S X_2 (\mathbb{1} + (-1)^a Z_1) X_1^a (\mathbb{1} + (-1)^{\bar{b}} Z_2) X_2 (\mathbb{1} + (-1)^b Z_2) V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right]$$

Anti-commuting  $Z_2$  and  $X_2$  (this can be shown analogously to Equation (4.120), making use of Proposition 4.17):

$$\approx_{\mathbb{R}} \frac{1}{16} \sum_{a,b \in \{0,1\}} \text{Tr} \left[ (\sigma_X^a \otimes \mathbb{1}_2 \otimes \mathbb{1}) (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S X_2 (\mathbb{1} + (-1)^a Z_1) X_1^a \right. \\ \left. (\mathbb{1} + (-1)^{\bar{b}} Z_2)^2 X_2 V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right]$$

Since  $Z_2$  is a binary observable,  $\sum_b (\mathbb{1} + (-1)^{\bar{b}} Z_2)^2 = 4 \cdot \mathbb{1}$ . Therefore, summing over  $b$  yields:

$$= \frac{1}{4} \sum_{a \in \{0,1\}} \text{Tr} \left[ (\sigma_X^a \otimes \mathbb{1}_2 \otimes \mathbb{1}) (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S X_2 (\mathbb{1} + (-1)^a Z_1) X_1^a X_2 V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right]$$

Commuting  $X_2$  and  $X_1^a$ , then replacing  $V_S X_1^a V_S^\dagger$  with  $\sigma_X^a \otimes \mathbb{1}_2 \otimes \mathbb{1}$  using Equation (4.126) and commuting it past the state:

$$\approx_{\mathbb{R}} \frac{1}{4} \sum_{a \in \{0,1\}} \text{Tr} \left[ (\sigma_X^a \otimes \mathbb{1}_2 \otimes \mathbb{1})^2 (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S X_2 (\mathbb{1} + (-1)^a Z_1) X_2 V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right]$$

We have  $\sigma_X^2 = \mathbb{1}$ . We can also commute  $X_2$  and  $Z_1$  using the analogous statement of Equation (4.120) with reversed indices. Then we obtain (using  $X_2^2 = \mathbb{1}$ ):

$$\approx_{\mathbb{R}} \frac{1}{4} \sum_{a \in \{0,1\}} \text{Tr} \left[ (\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}) V_S (\mathbb{1} + (-1)^a Z_1) V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right].$$

This expression is identical to Equation (4.122), so the result follows.  $\square$

The following lemma shows the steps that we skipped in the proof above.

**Lemma 4.34.** *With the notation from the proof of Lemma 4.33, we show the following statements:*

(i) Equation (4.125):

$$\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1} \approx_{\mathbb{R}, \frac{1}{2} \mathbb{1}_2 \otimes \alpha} V_S X_1 V_S^\dagger. \quad (4.127)$$

(ii) Equation (4.118):

$$\mathbb{1} + (-1)^a \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1} \approx_{\mathbb{R}, \frac{1}{2} \mathbb{1}_2 \otimes \alpha} V_S (\mathbb{1} + (-1)^a Z_1) V_S^\dagger. \quad (4.128)$$

(iii) Equation (4.120):

$$V_S [Z_2, X_1] V_S^\dagger \approx_{\mathbb{R}, \frac{1}{2} \mathbb{1}_2 \otimes \alpha^{(\theta_1, \theta_2)}} 0. \quad (4.129)$$

*Proof.*

(i) From Lemma 4.30, we have

$$V_S^\dagger (\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} X_1. \quad (4.130)$$

By Lemma 2.23, this implies

$$\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1} \approx_{\mathbb{R}, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} V_S X_1 V_S^\dagger. \quad (4.131)$$

Recall that by Equation (4.113), we have

$$V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger \stackrel{c}{\approx}_{\mathbb{R}} \frac{1}{2} \mathbb{1}_2 \otimes \alpha.$$

The result follows by the lifting lemma (Lemma 2.25(vi)).

(ii) First note that by the triangle inequality for the state-dependent norm, it suffices to show the following two relations individually:

$$V_S V_S^\dagger \approx_{\mathbb{R}, \frac{1}{2} \mathbb{1}_2 \otimes \alpha} \mathbb{1}, \quad (4.132)$$

$$\sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1} \approx_{\mathbb{R}, \frac{1}{2} \mathbb{1}_2 \otimes \alpha} V_S Z_1 V_S^\dagger. \quad (4.133)$$

The first relation follows from  $V_S V_S^\dagger \approx_{0, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} \mathbb{1}$  by Equation (4.113) and the lifting lemma (Lemma 2.25(vi)). The second relation follows by the same reasoning used for (i), making use of Equation (4.88).

(iii) We have

$$\mathrm{Tr} \left[ \left( V_S [Z_2, X_1] V_S^\dagger \right)^\dagger \left( V_S [Z_2, X_1] V_S^\dagger \right) \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) \right] = \mathrm{Tr} \left[ [Z_2, X_1]^\dagger [Z_2, X_1] V_S^\dagger \left( \frac{1}{2} \mathbb{1}_2 \otimes \alpha \right) V_S \right]$$

By Equation (4.114) and the lifting lemma (Lemma 2.25(iii)):

$$\approx_{\mathbb{R}} \mathrm{Tr} \left[ [Z_2, X_1]^\dagger [Z_2, X_1] \sigma^{(\theta_1, \theta_2)} \right]$$

By Proposition 4.24

$$\approx_{\mathbb{R}} 0.$$

□

## 4.8 Approximate equality of tilde observables and Pauli observables

The preceding section establishes that on the states  $\sigma^{(\theta_1, \theta_2)}$ , the non-tilde operators are approximately equal to the corresponding Pauli operators. However, to certify Bell states, we need the prover to perform measurements where its two “qubits” are measured in different bases, i.e., use measurement operators from  $P_{0,1}$  and  $P_{1,0}$ . The observables associated to these mixed-basis projectors are the tilde observables. Recall that for the tilde observables, we cannot get the required commutation relations, as explained in Section 4.6. This prevents us from using the argument that we used for the non-tilde observables. Instead, we will show that on the state, the tilde and non-tilde observables are approximately equal. Using the triangle inequality for the state-dependent distance, we can then conclude that the tilde observables are also close to Pauli observables (under the same isometry  $V_S$  and in the state-dependent distance).

**Lemma 4.35.** *For any efficient device  $D = (S, \Pi, M, P)$ , the following holds for any  $\theta_1, \theta_2 \in \{0, 1\}$  and  $i \in \{1, 2\}$ :*

$$\tilde{Z}_i \approx_{\gamma_T(D), \sigma^{(\theta_1, \theta_2)}} Z_i, \quad \tilde{X}_i \approx_{\gamma_T(D), \sigma^{(\theta_1, \theta_2)}} X_i. \quad (4.134)$$

*Proof.* We show  $\tilde{Z}_1 \approx_{\gamma_T(D), \sigma^{(\theta_1, \theta_2)}} Z_1$ , the other cases are analogous. Since  $Z_1$  and  $\tilde{Z}_1$  are both efficient, by the lifting lemma (Lemma 2.25(ii)) and the indistinguishability of  $\sigma^{(\theta_1, \theta_2)}$  (Lemma 4.14), it suffices to show this for  $(\theta_1, \theta_2) = (0, 1)$ . We can split  $\sigma^{(0,1)}$  as  $\sigma^{(0,1)} = \sum_{v_1, v_2} \sigma^{(0, v_1; 1, v_2)}$ , so by Lemma 2.18(ii) it is sufficient to show that for all  $v_1, v_2 \in \{0, 1\}$ ,  $\tilde{Z}_1 \approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} Z_1$ . By the same reasoning as in Proposition 4.24, where we showed the commutation relations, we have

$$Z_1 \approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} (-1)^{v_1} \mathbb{1}, \quad \tilde{Z}_1 \approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} (-1)^{v_1} \mathbb{1} \quad (4.135)$$

Therefore, by the triangle inequality, we also have

$$\tilde{Z}_1 \approx_{\gamma_T(D), \sigma^{(0, v_1; 1, v_2)}} Z_1 \quad (4.136)$$

as desired. □



**Corollary 4.36.** *For any efficient perfect device  $D = (S, \Pi, M, P)$ , the following holds for any  $\theta_1, \theta_2 \in \{0, 1\}$ :*

$$\tilde{Z}_1 \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} V_S^\dagger (\sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S, \quad (4.137)$$

$$\tilde{X}_1 \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} V_S^\dagger (\sigma_X \otimes \mathbb{1}_2 \otimes \mathbb{1}) V_S, \quad (4.138)$$

$$\tilde{Z}_2 \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} V_S^\dagger (\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}) V_S, \quad (4.139)$$

$$\tilde{X}_2 \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} V_S^\dagger (\mathbb{1}_2 \otimes \sigma_X \otimes \mathbb{1}) V_S. \quad (4.140)$$

*Proof.* Using the triangle inequality, these follow from Lemma 4.35 with Equation (4.88) for  $\tilde{Z}_1$ , Lemma 4.30 for  $\tilde{X}_1$ , and Lemma 4.33 for  $\tilde{Z}_2, \tilde{X}_2$ .  $\square$

## 4.9 Products of observables

We have shown in Corollary 4.36 that on the state, the tilde observables are approximately equal to the corresponding Pauli matrices, under the isometry  $V_S$ . To certify that the prover has a Bell state, we want to show that the prover must possess, up to the isometry  $V_S$ , a joint eigenstate of  $\sigma_Z \otimes \sigma_X$  and  $\sigma_Z \otimes \sigma_X$ , since the only such eigenstates are the Bell states (with the second qubit in the Hadamard basis). Therefore, we have to be able to “round” not just individual tilde observables to Pauli matrices, but also products of tilde observables. That is, we have to show, that e.g.  $\tilde{Z}_1 \tilde{X}_2$  is approximately equal to  $\sigma_Z \otimes \sigma_X \otimes \mathbb{1}$  under the isometry  $V_S$ . Note that this does not directly follow from the above, since we can only round operators next to the state. For example, we know from the previous section (ignoring the isometry for this explanation) that  $\tilde{Z}_1 \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}$  and  $\tilde{X}_2 \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} \mathbb{1}_2 \otimes \sigma_X \otimes \mathbb{1}$ . To deal with the product  $\tilde{Z}_1 \tilde{X}_2$ , we can use Lemma 2.18(i) to obtain

$$\tilde{Z}_1 \tilde{X}_2 \approx_{\mathbb{R}, \sigma^{(\theta_1, \theta_2)}} \tilde{Z}_1 (\mathbb{1}_2 \otimes \sigma_X \otimes \mathbb{1}).$$

However, now we cannot make use of Lemma 2.18(i) again, since the operator  $\tilde{Z}_1$  that we wish to round is multiplied *on the right* by another operator  $\mathbb{1}_2 \otimes \sigma_X \otimes \mathbb{1}$ , which, if one writes out the definition of the state-dependent distance, effectively sits in between the state and  $\tilde{Z}_1$ . In other words, we would require a version of Lemma 2.18(i) where the operator  $C$  in that lemma is multiplied on the right of  $A$  and  $B$ , but this does not hold.

To overcome this problem, we rely on a characterisation of the states  $\sigma^{(\theta_1, \theta_2)}$  that we can deduce from the relations for the non-tilde observables derived in the previous sections. The following lemmas, which are extensions of Lemmas 4.31 and 4.32, establish this characterisation.

**Lemma 4.37.** *Let  $D = (S, \Pi, M, P)$  be an efficient perfect device. Then, for  $v_1, v_2 \in \{0, 1\}$ , there exist positive matrices  $\tilde{\alpha}^{(0, v_1; 1, v_2)}$  and  $\tilde{\alpha}^{(1, v_1; 0, v_2)}$  such that the following holds:*

$$V_S \sigma^{(1, v_1; 0, v_2)} V_S^\dagger \approx_{\mathbb{R}} |(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes |v_2\rangle\langle v_2| \otimes \tilde{\alpha}^{(1, v_1; 0, v_2)}, \quad (4.141)$$

$$V_S \sigma^{(0, v_1; 1, v_2)} V_S^\dagger \approx_{\mathbb{R}} |v_1\rangle\langle v_1| \otimes |(-)^{v_2}\rangle\langle(-)^{v_2}| \otimes \tilde{\alpha}^{(0, v_1; 1, v_2)}. \quad (4.142)$$

*Proof.* We give the proof for the first relation, the second one can be shown analogously. Most of the proof is analogous to that of Lemma 4.31, and we only sketch it here.

Starting from Lemma 4.33, by the same steps as in Lemma 4.31 we get

$$\mathbb{1}_2 \otimes |v_2\rangle\langle v_2| \otimes \mathbb{1} \approx_{\mathbb{R}, V_S \sigma^{(1, v_1; 0, v_2)} V_S^\dagger} \mathbb{1}. \quad (4.143)$$

By Lemma 4.31 and the replacement lemma (Lemma 2.21(ii); to see that this also allows us to replace the state in this case, it is enough to write out the definition of the state-dependent distance as a trace):

$$\mathbb{1}_2 \otimes |v_2\rangle\langle v_2| \otimes \mathbb{1} \approx_{\mathbb{R}, |(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \alpha^{(1, v_1; 0, v_2)}} \mathbb{1}. \quad (4.144)$$

Therefore, by Lemma 2.22 we have:

$$V_S \sigma^{(1, v_1; 0, v_2)} V_S^\dagger \approx_{\mathbb{R}} \frac{1}{4} |(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \alpha^{(1, v_1; 0, v_2)} \quad (4.145)$$

$$\approx_{\mathbb{R}} (\mathbb{1}_2 \otimes |v_2\rangle\langle v_2| \otimes \mathbb{1}) \left( |(-)^{v_1}\rangle\langle(-)^{v_1}| \otimes \alpha^{(1, v_1; 0, v_2)} \right) (\mathbb{1}_2 \otimes |v_2\rangle\langle v_2| \otimes \mathbb{1}) \quad (4.146)$$

Defining

$$\tilde{\alpha}^{(1,v_1; 0,v_2)} = (|v_2\rangle \otimes \mathbf{1}) \alpha^{(1,v_1; 0,v_2)} (|v_2\rangle \otimes \mathbf{1}) \quad (4.147)$$

yields the result.  $\square$

*Remark 4.38.* Comparing the statement of Lemma 4.31 with that of Lemma 4.37 makes the connection between  $\alpha^{(1,v_1; 0,v_2)}$  and  $\tilde{\alpha}^{(1,v_1; 0,v_2)}$  explicit:

$$\alpha^{(1,v_1; 0,v_2)} \approx_{\mathbb{R}} |v_2\rangle\langle v_2| \otimes \tilde{\alpha}^{(1,v_1; 0,v_2)}. \quad (4.148)$$

The analogous relation of course also holds for  $\alpha^{(0,v_1; 1,v_2)}$  and  $\tilde{\alpha}^{(0,v_1; 1,v_2)}$ :

$$\alpha^{(0,v_1; 1,v_2)} \approx_{\mathbb{R}} |(-)^{v_2}\rangle\langle (-)^{v_2}| \otimes \tilde{\alpha}^{(0,v_1; 1,v_2)}. \quad (4.149)$$

The proof of the following lemma is very similar to that of Lemma 4.32, but we give the full proof for the sake of completeness.

**Lemma 4.39.** *Let  $D = (S, \Pi, M, P)$  be an efficient perfect device. Then, there exists a (normalised) state  $\tilde{\alpha}$  such that the following holds for  $v_1, v_2 \in \{0, 1\}$ :*

$$V_S \sigma^{(1,v_1; 0,v_2)} V_S^\dagger \stackrel{c}{\approx}_{\mathbb{R}} \frac{1}{4} |(-)^{v_1}\rangle\langle (-)^{v_1}| \otimes |v_2\rangle\langle v_2| \otimes \tilde{\alpha}, \quad (4.150)$$

$$V_S \sigma^{(0,v_1; 1,v_2)} V_S^\dagger \stackrel{c}{\approx}_{\mathbb{R}} \frac{1}{4} |v_1\rangle\langle v_1| \otimes |(-)^{v_2}\rangle\langle (-)^{v_2}| \otimes \tilde{\alpha}. \quad (4.151)$$

*Proof.* We focus on the first relation, the second one is analogous (and once we have established the two relations individually, the fact that we can use the same  $\tilde{\alpha}$  for both immediately follows from the indistinguishability of  $\sigma^{(1,0)} = \sum_{v_1, v_2} \sigma^{(1,v_1; 0,v_2)}$  and  $\sigma^{(0,1)} = \sum_{v_1, v_2} \sigma^{(0,v_1; 1,v_2)}$ ). We need to show that  $\{\tilde{\alpha}^{(1,v_1; 0,v_2)}\}_{v_1, v_2}$  from Lemma 4.37 are computationally indistinguishable, i.e. that  $\tilde{\alpha}^{(1,v_1; 0,v_2)} \stackrel{c}{\approx}_{\mathbb{R}} \tilde{\alpha}^{(1,v'_1; 0,v'_2)}$  for any  $v_1, v_2, v'_1, v'_2$ .

From Lemma 4.32, we already know that  $\{\sum_{v_2} \alpha^{(1,v_1; 0,v_2)}\}_{v_1}$  are computationally indistinguishable, and by Equation (4.148) this implies that  $\{\sum_{v_2} \tilde{\alpha}^{(1,v_1; 0,v_2)}\}_{v_1}$  are computationally indistinguishable, too, as the latter is obtained from the former by tracing out the first qubit, which cannot increase distinguishability. It therefore suffices to show that for any fixed  $v_1$ , no efficient distinguisher can distinguish  $\{\tilde{\alpha}^{(1,v_1; 0,v_2)}\}_{v_2}$ .

Fix  $v_1$ . By Lemma 4.37 and Lemma 4.32, there exists a  $c > 0$  and an  $\varepsilon = O(\gamma_T(D)^c)$  such that for any  $v_2$  and any efficient measurement  $\{M^{(0)}, M^{(1)}\}$ :

$$\begin{aligned} & \left\| V_S \sigma^{(1,v_1; 0,v_2)} V_S^\dagger - |(-)^{v_1}\rangle\langle (-)^{v_1}| \otimes |v_2\rangle\langle v_2| \otimes \tilde{\alpha}^{(1,v_1; 0,v_2)} \right\|_1^2 \leq \varepsilon, \\ & \left\| V_S \sigma^{(0,v_1; 1,v_2)} V_S^\dagger - |v_1\rangle\langle v_1| \otimes |(-)^{v_2}\rangle\langle (-)^{v_2}| \otimes \tilde{\alpha}^{(0,v_1; 1,v_2)} \right\|_1^2 \leq \varepsilon, \\ & \left| \sum_{v'_2} \text{Tr} \left[ M^{(0)} \left( \tilde{\alpha}^{(1,0; 0,v'_2)} - \tilde{\alpha}^{(1,1; 0,v'_2)} \right) \right] \right| \leq 4\varepsilon^{1/2}. \end{aligned} \quad (4.152)$$

For the sake of contradiction, assume that there is an efficient POVM  $\{\Lambda^{(0)}, \Lambda^{(1)}\}$  such that

$$\text{Tr} \left[ \Lambda^{(0)} \left( \tilde{\alpha}^{(1,v_1; 0,0)} - \tilde{\alpha}^{(1,v_1; 0,1)} \right) \right] \geq 20\varepsilon^{1/2} + 2\mu(\lambda), \quad (4.153)$$

where  $\mu(\lambda)$  is non-negligible. Define

$$\Gamma = V_S^\dagger \left( |(-)^{v_1}\rangle\langle (-)^{v_1}| \otimes |0\rangle\langle 0| \otimes \Lambda^{(0)} \right) V_S, \quad (4.154)$$

which is efficient by the same reasoning as in Lemma 4.32.

Constructing a distinguisher as in Lemma 4.32, it suffices to show that  $\text{Tr}[\Gamma(\sigma^{(1,0)} - \sigma^{(0,1)})]$  is non-negligible. This can be shown by a similar calculation as in Lemma 4.32:

$$\text{Tr}[\Gamma(\sigma^{(1,0)} - \sigma^{(0,1)})]$$

$$= \text{Tr}\left[\left(|(-)^{v_1}\langle(-)^{v_1}| \otimes |0\rangle\langle 0| \otimes \Lambda^{(0)}\right) \left(V_S \sigma^{(1,0)} V_S^\dagger - V_S \sigma^{(0,1)} V_S^\dagger\right)\right]$$

Using Equation (4.152) and the replacement lemma (Lemma 2.21(ii); the non-asymptotic statement used here is easily seen to hold from the proof) to replace the state inside the trace:

$$\begin{aligned} &\geq \sum_{v'_1, v'_2} \text{Tr}\left[\left(|(-)^{v_1}\langle(-)^{v_1}| \otimes |0\rangle\langle 0| \otimes \Lambda^{(0)}\right) \right. \\ &\quad \left. \left(|(-)^{v'_1}\langle(-)^{v'_1}| \otimes |v'_2\rangle\langle v'_2| \otimes \tilde{\alpha}^{(1, v'_1; 0, v'_2)} - |v'_1\rangle\langle v'_1| \otimes |(-)^{v'_2}\langle(-)^{v'_2}| \otimes \tilde{\alpha}^{(0, v'_1; 1, v'_2)}\right)\right] - 8\varepsilon^{1/2} \\ &= \text{Tr}\left[\Lambda^{(0)} \tilde{\alpha}^{(1, v_1; 0, 0)}\right] - \frac{1}{4} \sum_{v'_1, v'_2 \in \{0, 1\}} \text{Tr}\left[\Lambda^{(0)} \tilde{\alpha}^{(0, v'_1; 1, v'_2)}\right] - 8\varepsilon^{1/2} \\ &= \frac{1}{2} \text{Tr}\left[\Lambda^{(0)} \left(\tilde{\alpha}^{(1, v_1; 0, 0)} - \tilde{\alpha}^{(1, v_1; 0, 1)}\right)\right] + \frac{1}{2} \sum_{v'_2} \text{Tr}\left[\Lambda^{(0)} \tilde{\alpha}^{(1, v_1; 0, v'_2)}\right] - \frac{1}{4} \sum_{v'_1, v'_2 \in \{0, 1\}} \text{Tr}\left[\Lambda^{(0)} \tilde{\alpha}^{(0, v'_1; 1, v'_2)}\right] - 8\varepsilon^{1/2} \end{aligned}$$

By assumption, the first term is at least  $10\varepsilon^{1/2} + \mu(\lambda)$ . Since  $\Lambda^{(0)}$  is efficient, by Equation (4.152) we have  $\sum_{v'_2} \text{Tr}\left[\Lambda^{(0)} \tilde{\alpha}^{(1, v_1; 0, v'_2)}\right] \geq \frac{1}{2} \sum_{v'_1, v'_2} \text{Tr}\left[\Lambda^{(0)} \tilde{\alpha}^{(1, v'_1; 0, v'_2)}\right] - 2\varepsilon^{1/2}$ . Inserting this:

$$= -\frac{1}{4} \sum_{v'_1, v'_2 \in \{0, 1\}} \text{Tr}\left[\Lambda^{(0)} \left(\tilde{\alpha}^{(0, v'_1; 1, v'_2)} - \tilde{\alpha}^{(1, v'_1; 0, v'_2)}\right)\right] + \varepsilon^{1/2} + \mu(\lambda)$$

The remaining term is the same one we had in Lemma 4.32. Applying the same steps as in that proof yields:

$$\geq \mu(\lambda) - \text{negl}(\lambda).$$

Therefore,  $\text{Tr}[\Gamma(\sigma^{(1,0)} - \sigma^{(0,1)})]$  is non-negligibly larger than 0, contradicting Lemma 4.14 and finishing the proof.  $\square$

**Corollary 4.40.** *Let  $D = (S, \Pi, M, P)$  be an efficient perfect device. Then, there exists a (normalised) state  $\tilde{\alpha}$  such that for any  $\theta_1, \theta_2 \in \{0, 1\}$ :*

$$V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger \stackrel{c}{\approx}_{\mathbb{R}} \frac{1}{4} \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \tilde{\alpha}. \quad (4.155)$$

*Proof.* Summing over  $v_1, v_2$  in Lemma 4.39 yields the statement for  $(\theta_1, \theta_2) = (0, 1)$  and  $(\theta_1, \theta_2) = (1, 0)$ . By Lemma 4.14, the states  $\sigma^{(\theta_1, \theta_2)}$  are computationally indistinguishable. Since  $V_S$  is efficient, the states  $V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger$  can be computed efficiently from  $\sigma^{(\theta_1, \theta_2)}$ . Therefore, they must also be computationally indistinguishable. The transitivity of computational indistinguishability then allows us to extend the statement to any  $\theta_1, \theta_2$ .  $\square$

We are now in a position to prove the statement about products of observables that we mentioned at the beginning of this section.

**Lemma 4.41.** *For any efficient perfect device  $D = (S, \Pi, M, P)$ , the following holds for any  $\theta_1, \theta_2 \in \{0, 1\}$ :*

$$V_S Z_1 Z_2 V_S^\dagger \approx_{\mathbb{R}, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} \sigma_Z \otimes \sigma_Z \otimes \mathbb{1}, \quad (4.156)$$

$$V_S X_1 X_2 V_S^\dagger \approx_{\mathbb{R}, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} \sigma_X \otimes \sigma_X \otimes \mathbb{1}, \quad (4.157)$$

$$V_S \tilde{Z}_1 \tilde{X}_2 V_S^\dagger \approx_{\mathbb{R}, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} \sigma_Z \otimes \sigma_X \otimes \mathbb{1}, \quad (4.158)$$

$$V_S \tilde{X}_1 \tilde{Z}_2 V_S^\dagger \approx_{\mathbb{R}, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} \sigma_X \otimes \sigma_Z \otimes \mathbb{1}. \quad (4.159)$$

*Proof.* We show the first relation, the others are analogous (using Corollary 4.36 for the tilde observables). First note that because  $Z_1$  and  $Z_2$  are efficient binary observables, and  $[Z_1, Z_2] = 0$  (Lemma 4.23),  $Z_1 Z_2$  is also an efficient binary observable by Lemma 2.5. Since  $V_S, Z_1 Z_2$ , and  $\sigma_Z$  are efficient, we can use the lifting lemma (Lemma 2.25(vi)) and Corollary 4.40 to replace  $V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger$  by  $\frac{1}{4} \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \tilde{\alpha}$ , so the statement that we need to show is

$$V_S Z_1 Z_2 V_S^\dagger \approx_{\mathbb{R}, \frac{1}{4} \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \tilde{\alpha}} \sigma_Z \otimes \sigma_Z \otimes \mathbb{1}. \quad (4.160)$$

Using Lemma 2.16, this reduces to showing

$$\text{Tr} \left[ (\sigma_Z \otimes \sigma_Z \otimes \mathbb{1}) V_S Z_1 Z_2 V_S^\dagger \left( \frac{1}{4} \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \tilde{\alpha} \right) \right] \approx_{\mathbb{R}} 1. \quad (4.161)$$

By Lemma 2.23, we have (loosening the bound on  $Z_1$  to match that of  $Z_2$ ):

$$V_S Z_1 V_S^\dagger \approx_{\mathbb{R}, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1} \quad (\text{from Equation (4.88)}), \quad (4.162)$$

$$V_S Z_2 V_S^\dagger \approx_{\mathbb{R}, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} \mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1} \quad (\text{from Lemma 4.33}). \quad (4.163)$$

Since  $V_S, Z_1, Z_2$ , and  $\sigma_Z$  are efficient, we can again use the lifting property (Lemma 2.25(vi)) and Corollary 4.40 to replace  $V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger$  by  $\frac{1}{4} \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \tilde{\alpha}$ , so we have:

$$V_S Z_1 V_S^\dagger \approx_{\mathbb{R}, \frac{1}{4} \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \tilde{\alpha}} \sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}, \quad (4.164)$$

$$V_S Z_2 V_S^\dagger \approx_{\mathbb{R}, \frac{1}{4} \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \tilde{\alpha}} \mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}. \quad (4.165)$$

We can use this to show Equation (4.161) by the same method that we used in the proof of Lemma 4.33. Using Equation (4.165), the replacement lemma (Lemma 2.21(i)), and commuting  $\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}$  past the state:

$$\text{Tr} \left[ (\sigma_Z \otimes \sigma_Z \otimes \mathbb{1}) V_S Z_1 Z_2 V_S^\dagger \left( \frac{1}{4} \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \tilde{\alpha} \right) \right] \quad (4.166)$$

$$\approx_{\mathbb{R}} \text{Tr} \left[ (\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}) (\sigma_Z \otimes \sigma_Z \otimes \mathbb{1}) V_S Z_1 V_S^\dagger \left( \frac{1}{4} \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \tilde{\alpha} \right) \right] \quad (4.167)$$

Similarly using Equation (4.164) and commuting  $\sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}$  past the state:

$$\approx_{\mathbb{R}} \text{Tr} \left[ (\sigma_Z \otimes \mathbb{1}_2 \otimes \mathbb{1}) (\mathbb{1}_2 \otimes \sigma_Z \otimes \mathbb{1}) (\sigma_Z \otimes \sigma_Z \otimes \mathbb{1}) \left( \frac{1}{4} \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \tilde{\alpha} \right) \right] \quad (4.168)$$

$$= 1. \quad (4.169)$$

□

## 4.10 Certifying Bell states

Having established this characterisation of products of observables, we can show the main result of this paper, namely that the prover's states and measurements in the Bell case must be close to Bell states and single-qubit Pauli measurements.

**Theorem 4.42.** *We label the (Hadamard-rotated) Bell states as follows:*

$$|\phi_H^{(a,b)}\rangle = (\sigma_X^a \otimes \sigma_X^b)(|0\rangle|+\rangle + |1\rangle|-\rangle). \quad (4.170)$$

For  $b \in \{0, 1\}$ , we also use the notation

$$|b_0\rangle = |b\rangle, \quad |b_1\rangle = |(-)^b\rangle. \quad (4.171)$$

Let  $D = (S, \Pi, M, P)$  be an efficient device, with  $\gamma_P(D), \gamma_T(D)$ , and  $\gamma_B(D)$  as in Lemma 4.10. Let  $\mathcal{H}$  be the private Hilbert space of the device and let  $\sigma^{(1, s_1; 1, s_2)}$  be as in Definition 4.6.

Then there exists a Hilbert space  $\mathcal{H}'$ , an isometry  $V : \mathcal{H} \rightarrow \mathbb{C}^4 \otimes \mathcal{H}'$ , and a constant  $c > 0$ , such that there are states  $\xi^{(s_1, s_2)} \in \mathcal{D}(\mathcal{H}')$  for  $s_1, s_2 \in \{0, 1\}$  satisfying the following:

(i) Under the isometry  $V$ , the state of the prover in a Hadamard round is close to a Bell state:

$$V\sigma^{(1,s_1; 1,s_2)}V^\dagger \approx_{\gamma_P(D)^c + \gamma_T(D)^c + \gamma_B(D)^c} \frac{1}{4} |\phi_H^{(s_1,s_2)}\rangle\langle\phi_H^{(s_1,s_2)}| \otimes \xi_{\mathcal{H}'}^{(s_1,s_2)}, \quad (4.172)$$

and the different  $\xi_{\mathcal{H}'}^{(s_1,s_2)}$  are computationally indistinguishable.

(ii) Under the isometry  $V$ , the prover's measurements  $P_{q_1,q_2}^{(a,b)}$  act on  $\sigma^{(1,s_1; 1,s_2)}$  in the same way that single-qubit measurements in the computational or Hadamard basis act on a Bell state:

$$\begin{aligned} VP_{q_1,q_2}^{(a,b)}\sigma^{(1,s_1; 1,s_2)}P_{q_1,q_2}^{(a,b)}V^\dagger \\ \approx_{\gamma_P(D)^c + \gamma_T(D)^c + \gamma_B(D)^c} \frac{1}{4} (|a_{q_1}, b_{q_2}\rangle\langle a_{q_1}, b_{q_2}|) |\phi_H^{(s_1,s_2)}\rangle\langle\phi_H^{(s_1,s_2)}| (|a_{q_1}, b_{q_2}\rangle\langle a_{q_1}, b_{q_2}|) \otimes \xi_{\mathcal{H}'}^{(s_1,s_2)} \end{aligned}$$

*Proof.* By Lemma 4.13, up to an additional error  $O(\gamma_P(D)^{1/2})$ , we can assume that the device  $D$  is perfect. (Note that the statement in Lemma 4.13 implies an analogous statement for  $\sigma^{(1,s_1; 1,s_2)}$  by the fact that the application of a CPTP map cannot increase the trace distance.)

(i) Take  $V$  to be the swap isometry  $V_S$  defined in Definition 4.27. From Lemmas 4.41, 2.18(ii), and 2.24, we have for  $a \in \{0, 1\}$ :

$$(\tilde{Z}_1 \tilde{X}_2)^{(a)} \approx_{\mathbb{R}, \sigma^{(1,s_1; 1,s_2)}} V_S^\dagger (\sigma_Z \otimes \sigma_X \otimes \mathbf{1})^{(a)} V_S, \quad (4.173)$$

$$(\tilde{X}_1 \tilde{Z}_2)^{(a)} \approx_{\mathbb{R}, \sigma^{(1,s_1; 1,s_2)}} V_S^\dagger (\sigma_X \otimes \sigma_Z \otimes \mathbf{1})^{(a)} V_S, \quad (4.174)$$

Using the definition of  $\gamma_B$  (Equation (4.22)) and Lemma 4.8, we have

$$\text{Tr} \left[ (\tilde{Z}_1 \tilde{X}_2)^{(s_1)} \sigma^{(1,s_1; 1,s_2)} \right] \approx_{\gamma_B(T)} \text{Tr} \left[ \sigma^{(1,s_1; 1,s_2)} \right], \quad (4.175)$$

$$\text{Tr} \left[ (\tilde{X}_1 \tilde{Z}_2)^{(s_2)} \sigma^{(1,s_1; 1,s_2)} \right] \approx_{\gamma_B(T)} \text{Tr} \left[ \sigma^{(1,s_1; 1,s_2)} \right]. \quad (4.176)$$

By the replacement lemma (Lemma 2.21(i)), Equations (4.173) and (4.174), and  $V_S^\dagger V_S = \mathbf{1}$ , we therefore have

$$\text{Tr} \left[ (\sigma_Z \otimes \sigma_X \otimes \mathbf{1}_{\mathcal{H}'})^{(s_1)} V_S \sigma^{(1,s_1; 1,s_2)} V_S^\dagger \right] \approx_{\mathbb{R}} \text{Tr} \left[ V_S \sigma^{(1,s_1; 1,s_2)} V_S^\dagger \right], \quad (4.177)$$

$$\text{Tr} \left[ (\sigma_X \otimes \sigma_Z \otimes \mathbf{1}_{\mathcal{H}'})^{(s_2)} V_S \sigma^{(1,s_1; 1,s_2)} V_S^\dagger \right] \approx_{\mathbb{R}} \text{Tr} \left[ V_S \sigma^{(1,s_1; 1,s_2)} V_S^\dagger \right]. \quad (4.178)$$

Using Lemmas 2.19 and 2.20, this implies

$$(\sigma_Z \otimes \sigma_X \otimes \mathbf{1}_{\mathcal{H}'})^{(a)} \approx_{\mathbb{R}, V_S \sigma^{(1,s_1; 1,s_2)} V_S^\dagger} \delta_{a,s_1} \mathbf{1}, \quad (4.179)$$

$$(\sigma_X \otimes \sigma_Z \otimes \mathbf{1}_{\mathcal{H}'})^{(b)} \approx_{\mathbb{R}, V_S \sigma^{(1,s_1; 1,s_2)} V_S^\dagger} \delta_{b,s_2} \mathbf{1}. \quad (4.180)$$

where  $\delta_{a,s_1}$  is the Kronecker- $\delta$ . Expanding  $\sigma^{(1,1)} = \sum_{s_1, s_2} \sigma^{(1,s_1; 1,s_2)}$  and using Lemma 2.22 twice, we get:

$$\begin{aligned} V_S \sigma^{(1,s_1; 1,s_2)} V_S^\dagger \approx_{\mathbb{R}} (\sigma_X \otimes \sigma_Z \otimes \mathbf{1}_{\mathcal{H}'})^{(s_2)} (\sigma_Z \otimes \sigma_X \otimes \mathbf{1}_{\mathcal{H}'})^{(s_1)} (V_S \sigma^{(1,1)} V_S^\dagger) \\ (\sigma_Z \otimes \sigma_X \otimes \mathbf{1}_{\mathcal{H}'})^{(s_1)} (\sigma_X \otimes \sigma_Z \otimes \mathbf{1}_{\mathcal{H}'})^{(s_2)}. \end{aligned} \quad (4.181)$$

A direct calculation shows

$$(\sigma_Z \otimes \sigma_X \otimes \mathbf{1}_{\mathcal{H}'})^{(s_1)} (\sigma_X \otimes \sigma_Z \otimes \mathbf{1}_{\mathcal{H}'})^{(s_2)} = |\phi_H^{(s_1,s_2)}\rangle\langle\phi_H^{(s_1,s_2)}| \otimes \mathbf{1}_{\mathcal{H}'}. \quad (4.182)$$

Using Corollary 4.40 and the two equations above, we get

$$V_S \sigma^{(1, s_1; 1, s_2)} V_S^\dagger \approx_{\mathbb{R}} (|\phi_H^{(s_1, s_2)}\rangle\langle\phi_H^{(s_1, s_2)}| \otimes \mathbb{1}_{\mathcal{H}'}) V_S \sigma^{(1, 1)} V_S^\dagger (|\phi_H^{(s_1, s_2)}\rangle\langle\phi_H^{(s_1, s_2)}| \otimes \mathbb{1}_{\mathcal{H}'}) \quad (4.183)$$

$$\stackrel{c}{\approx}_{\mathbb{R}} \frac{1}{4} |\phi_H^{(s_1, s_2)}\rangle\langle\phi_H^{(s_1, s_2)}| \otimes \tilde{\alpha}_{\mathcal{H}'}. \quad (4.184)$$

Defining  $\xi_{\mathcal{H}'}^{(s_1, s_2)}$  to be the renormalised version of  $(|\phi_H^{(s_1, s_2)}\rangle\langle\phi_H^{(s_1, s_2)}| \otimes \mathbb{1}_{\mathcal{H}'}) V_S \sigma^{(1, 1)} V_S^\dagger (|\phi_H^{(s_1, s_2)}\rangle\langle\phi_H^{(s_1, s_2)}| \otimes \mathbb{1}_{\mathcal{H}'})$ , Equation (4.183) implies that  $V_S \sigma^{(1, s_1; 1, s_2)} V_S^\dagger$  has the desired form, and Equation (4.184) implies the normalisation of  $\frac{1}{4}$  and the computational indistinguishability of the different  $\xi_{\mathcal{H}'}^{(s_1, s_2)}$ .

(ii) To simplify the notation, we show this for  $(q_1, q_2) = (0, 0)$ . The other cases are analogous.

We can write  $P_{0,0}^{(a,b)}$  as

$$P_{0,0}^{(a,b)} = Z_1^{(a)} Z_2^{(b)}. \quad (4.185)$$

By Lemma 2.24, Equations (4.162) and (4.163) imply

$$V_S Z_1^{(a)} V_S^\dagger \approx_{\mathbb{R}, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} |a\rangle\langle a| \otimes \mathbb{1}_2 \otimes \mathbb{1}_{\mathcal{H}'}, \quad (4.186)$$

$$V_S Z_2^{(b)} V_S^\dagger \approx_{\mathbb{R}, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} \mathbb{1}_2 \otimes |b\rangle\langle b| \otimes \mathbb{1}_{\mathcal{H}'}. \quad (4.187)$$

The argument now proceeds similarly to that in the proof of Lemma 4.41. We need to replace  $V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger$  by  $\frac{1}{4} \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \tilde{\alpha}$ . This can be shown analogously to the lifting lemma (Lemma 2.25(vi), which only deals with binary observables, not projectors), by noting that  $V V^\dagger \approx_{0, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} \mathbb{1}$ . Using the trick of commuting Pauli operators past the state in the same manner as before, we obtain

$$V_S P_{0,0}^{(a,b)} V_S^\dagger = V_S Z_1^{(a)} Z_2^{(b)} V_S^\dagger \approx_{\mathbb{R}, V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger} |a\rangle\langle a| \otimes |b\rangle\langle b| \otimes \mathbb{1}_{\mathcal{H}'}. \quad (4.188)$$

Since each  $\sigma^{(\theta_1, v_1; \theta_2, v_2)}$  is positive, this also holds if we replace  $V_S \sigma^{(\theta_1, \theta_2)} V_S^\dagger$  by  $V_S \sigma^{(\theta_1, v_1; \theta_2, v_2)} V_S^\dagger$  (Lemma 2.18(ii)). Therefore, using Lemma 2.22 we have

$$V_S P_{q_1, q_2}^{(a,b)} \sigma^{(1, v_1; 1, v_2)} P_{x, y}^{(a,b)} V_S^\dagger = V_S P_{q_1, q_2}^{(a,b)} V_S^\dagger V_S \sigma^{(1, v_1; 1, v_2)} V_S^\dagger V_S P_{x, y}^{(a,b)} V_S^\dagger \quad (4.189)$$

$$\approx_{\mathbb{R}} |a\rangle\langle a| \otimes |b\rangle\langle b| \otimes \mathbb{1}_{\mathcal{H}'} \left( V_S \sigma^{(1, v_1; 1, v_2)} V_S^\dagger \right) |a\rangle\langle a| \otimes |b\rangle\langle b| \otimes \mathbb{1}_{\mathcal{H}'}. \quad (4.190)$$

Since applying projectors cannot increase the trace distance, we can replace  $V_S \sigma^{(1, v_1; 1, v_2)} V_S^\dagger$  by  $\frac{1}{4} |\phi_H^{(s_1, s_2)}\rangle\langle\phi_H^{(s_1, s_2)}| \otimes \xi_{\mathcal{H}'}^{(s_1, s_2)}$  using part (i). This completes the proof. □

## References

- [ABOR00] W. Aiello, S. Bhatt, R. Ostrovsky, and S. R. Rajagopalan. “Fast Verification of Any Remote Procedure Call: Short Witness-Indistinguishable One-Round Proofs for NP”, *Automata, Languages and Programming - ICALP 2000, Lecture Notes in Computer Science*, Springer, 463-474 (2000).
- [ACGH19] G. Alagic, A. M. Childs, A. B. Grilo, and S.-H. Hung. “Non-interactive classical verification of quantum computation”, *Preprint* (2019). [arXiv:1911.08101](https://arxiv.org/abs/1911.08101).
- [BCG<sup>+</sup>06] M. Ben-Or, C. Crépeau, D. Gottesman, A. Hassidim, and A. Smith. “Secure Multiparty Quantum Computation with (Only) a Strict Honest Majority”, *IEEE 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 249-260 (2006).
- [BCM<sup>+</sup>18] Z. Brakerski, P. Christiano, U. Mahadev, U. Vazirani, and T. Vidick. “A Cryptographic Test of Quantumness and Certifiable Randomness from a Single Quantum Device”, *IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, 320-331 (2018). [arXiv:1804.00640v3](https://arxiv.org/abs/1804.00640v3).
- [Bel64] J. S. Bell. “On the Einstein Podolsky Rosen paradox”, *Physics Physique Fizika* 1, 195–200 (1964).
- [BFNV19] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani. “On the complexity and verification of quantum random circuit sampling”, *Nature Physics* 15, 159–163 (2019).
- [BG20] A. Broadbent and A. B. Grilo. “QMA-hardness of consistency of local density matrices with applications to quantum zero-knowledge”, *IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, 196–205 (2020).
- [BKVV20] Z. Brakerski, V. Koppula, U. Vazirani, and T. Vidick. “Simpler Proofs of Quantumness”, *Preprint* (2020). [arXiv:2005.04826](https://arxiv.org/abs/2005.04826).
- [BRV<sup>+</sup>19] K. Bharti, M. Ray, A. Varvitsiotis, N. A. Warsi, A. Cabello, and L.-C. Kwek. “Robust Self-Testing of Quantum Systems via Noncontextuality Inequalities”, *Phys. Rev. Lett.* 122, 250403 (2019). [arXiv:1812.07265](https://arxiv.org/abs/1812.07265).
- [CCKW19] A. Cojocaru, L. Colisson, E. Kashefi, and P. Wallden. “QFactory: Classically-Instructed Remote Secret Qubits Preparation”, *Advances in Cryptology - ASIACRYPT 2019, Lecture Notes in Computer Science*, Springer, 615-645 (2019). [arXiv:1904.06303](https://arxiv.org/abs/1904.06303).
- [CCY19] N.-H. Chia, K.-M. Chung, and T. Yamakawa. “Classical Verification of Quantum Computations with Efficient Verifier”, *Preprint* (2019). [arXiv:1912.00990](https://arxiv.org/abs/1912.00990).
- [CGJV19] A. Coladangelo, A. B. Grilo, S. Jeffery, and T. Vidick. “Verifier-on-a-leash: New schemes for verifiable delegated quantum computation, with quasilinear resources”, *Advances in Cryptology - EUROCRYPT 2019, Lecture Notes in Computer Science*, Springer 11478 LNCS, 247-277 (2019). [arXiv:1708.07359](https://arxiv.org/abs/1708.07359).
- [CGS02] C. Crépeau, D. Gottesman, and A. Smith. “Secure Multi-Party Quantum Computation”, *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, 643-652 (2002).
- [CGS17] A. Coladangelo, K. T. Goh, and V. Scarani. “All pure bipartite entangled states can be self-tested”, *Nature Communications* 8, 15485 (2017). [arXiv:1611.08062](https://arxiv.org/abs/1611.08062).
- [Col06] R. Colbeck. *Quantum and relativistic protocols for secure multi-party computation*, PhD Thesis, University of Cambridge (2006). [arXiv:0911.3814](https://arxiv.org/abs/0911.3814).
- [CVZ20] A. Coladangelo, T. Vidick, and T. Zhang. “Non-interactive zero-knowledge arguments for QMA, with preprocessing”, *Annual International Cryptology Conference (CRYPTO)*, 799–828 (2020).
- [DHRW16] Y. Dodis, S. Halevi, R. D. Rothblum, and D. Wichs. “Spooky Encryption and Its Applications”, *Advances in Cryptology - CRYPTO 2016, Lecture Notes in Computer Science*, Springer, 93-122 (2016).

- [GH17] W. T. Gowers and O. Hatami. “Inverse and stability theorems for approximate representations of finite groups”, *Sbornik: Mathematics* 208, 1784 (2017).
- [GV19] A. Gheorghiu and T. Vidick. “Computationally-secure and composable remote state preparation”, *IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, 1024–1033 (2019).
- [JNV+20] Z. Ji, A. Natarajan, T. Vidick, J. Wright, and H. Yuen. “MIP\* = RE”, *Preprint* (2020). [arXiv:2001.04383](https://arxiv.org/abs/2001.04383).
- [KRR14] Y. T. Kalai, R. Raz, and R. D. Rothblum. “How to Delegate Computations: The Power of No-Signaling Proofs”, *Proceedings of the 46th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 485-494 (2014).
- [Mah18] U. Mahadev. “Classical Verification of Quantum Computations”, *IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, 259-267 (2018). [arXiv:1804.01082v2](https://arxiv.org/abs/1804.01082v2).
- [MDCAF20] T. Metger, Y. Dulek, A. Coladangelo, and R. Arnon-Friedman. “Device-independent quantum key distribution from computational assumptions”, *Preprint* (2020). [arXiv:2010.04175](https://arxiv.org/abs/2010.04175).
- [MS17] C. A. Miller and Y. Shi. “Universal Security for Randomness Expansion from the Spot-Checking Protocol”, *SIAM Journal on Computing* 46, 1304-1335 (2017). [arXiv:1411.6608](https://arxiv.org/abs/1411.6608).
- [MY04] D. Mayers and A. Yao. “Self Testing Quantum Apparatus”, *Quantum Info. Comput.* 4, 273-286 (2004). [arXiv:quant-ph/0307205](https://arxiv.org/abs/quant-ph/0307205).
- [MYS12] M. McKague, T. H. Yang, and V. Scarani. “Robust self-testing of the singlet”, *Journal of Physics A: Mathematical and Theoretical* 45, 455304 (2012).
- [NW19] A. Natarajan and J. Wright. “NEEXP in MIP\*”, *IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, 510-518 (2019). [arXiv:1904.05870](https://arxiv.org/abs/1904.05870).
- [PR92] S. Popescu and D. Rohrlich. “Which states violate Bell’s inequality maximally?”, *Physics Letters A* 169, 411–414 (1992).
- [Raz98] R. Raz. “A parallel repetition theorem”, *SIAM Journal on Computing* 27, 763–803 (1998).
- [Reg09] O. Regev. “On Lattices, Learning with Errors, Random Linear Codes, and Cryptography”, *J. ACM* 56 (2009).
- [RUV13] B. W. Reichardt, F. Unger, and U. Vazirani. “Classical command of quantum systems”, *Nature* 496, 456 (2013). [arXiv:1209.0449](https://arxiv.org/abs/1209.0449).
- [ŠB19] I. Šupić and J. Bowles. “Self-testing of quantum systems: a review”, *Preprint* (2019). [arXiv:1904.10042](https://arxiv.org/abs/1904.10042).
- [Sca19] V. Scarani. *Bell Nonlocality*, Oxford University Press (2019).
- [SW87] S. J. Summers and R. Werner. “Maximal violation of Bell’s inequalities is generic in quantum field theory”, *Communications in Mathematical Physics* 110, 247–259 (1987).
- [Vid11] T. Vidick. *The complexity of entangled games*. PhD thesis, 2011.
- [VV12] U. Vazirani and T. Vidick. “Certifiable Quantum Dice: Or, True Random Number Generation Secure against Quantum Adversaries”, *Proceedings of the 44th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 61-76 (2012). [arXiv:1111.6054](https://arxiv.org/abs/1111.6054).
- [VZ20] T. Vidick and T. Zhang. “Classical proofs of quantum knowledge”, *Preprint* (2020). [arXiv:2005.01691](https://arxiv.org/abs/2005.01691).
- [Wil11] M. Wilde. “From Classical to Quantum Shannon Theory”, *Preprint* (2011). [arXiv:1106.1445v8](https://arxiv.org/abs/1106.1445v8).