Self-Testing of a Single Quantum Device Under Computational Assumptions

Tony Metger
Institute for Theoretical Physics, ETH Zürich, Switzerland
tmetger@ethz.ch

Thomas Vidick
Department of Computing and Mathematical Sciences,
California Institute of Technology, Pasadena, CA, USA
vidick@caltech.edu

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.

2012 ACM Subject Classification Theory of computation → Quantum computation theory

Keywords and phrases Quantum computing, quantum cryptography, device-independence, self-testing, post-quantum cryptography

Digital Object Identifier 10.4230/LIPIcs.ITCS.2021.19

Related Version A full version of the paper is available at `https://arxiv.org/abs/2001.09161`.

Funding *Tony Metger*: supported by the ETH Foundation through the Excellence Scholarship & Opportunity Programme, and the IQIM, an NSF Physics Frontiers Center (NSF Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028).
Thomas Vidick: 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).

Acknowledgements We thank Andrea Coladangelo, Andru Gheorghiu, Anand Natarajan, and Tina Zhang for helpful discussions; Andrea Coladangelo, Andru Gheorghiu, Urmila Mahadev, and Akihiro Mizutani for comments on the manuscript; and Lídia del Rio for pointing out the reference [6]. This work was carried out while Tony Metger was a visiting student researcher at the Department of Computing and Mathematical Sciences at Caltech.

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. [24]) or delegated quantum computation (see e.g. [32]). 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 [24] in the context of proofs of security for quantum key distribution, but the notion was already present in earlier works [34, 29]. 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 [35]. 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 MIP^* [21].

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 [3], for which the non-communication assumption is necessary.[1] 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

[1] Another approach is to base the self-testing statement on non-contextuality inequalities [5, 6]. The violation of non-contextuality inequalities is a uniquely quantum phenomenon that is similar to the violation of Bell inequalities, with the advantage that it only requires a single quantum device and therefore no non-communication assumption. The downside of this approach is that it places additional assumptions, such as memory constraints and compatibility relations between measurements, on the quantum device, limiting its suitability for practical cryptographic applications.

ability to break the Learning with Errors (LWE) assumption [31], 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*:*

1. *the prover's state is* $O(\varepsilon^c)$*-close (in trace distance) to a Bell pair,*
2. *(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 [34, 25].

The proof of our main result builds on techniques introduced in recent works [23, 8, 19] 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 [23], 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.[2] To emphasize the difference, we note that the guarantee of Mahadev's protocol *does not* directly

[2] This comes at the cost that we are only able to certify Bell pairs, while Mahadev's measurement protocol works for measurements on any state.

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. [8], 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 [8] and using techniques from [23], Gheorghiu and Vidick construct a protocol for a task that they call verifiable remote state preparation (RSP) [19]. They consider a set of single-qubit pure states $\{|\psi_1\rangle, \ldots, |\psi_n\rangle\}$.[3] 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 [8] 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. [12], 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 [19] 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 ([27, Section 4]). On a technical level, it requires the certification of compatibility relations

[3] The protocol in [19] is designed for a specific set of ten pure states that are useful for delegated quantum computation, but for the purposes of this overview it is not important which specific states these are.

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 [19], to characterise the prover's measurements; in [27, 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 [35] or [33, 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** (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 $V|\psi\rangle_{AB} = |\phi\rangle_{AB}|\text{Aux}\rangle$ for some ancillary state $|\text{Aux}\rangle$.*

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, 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 [24], and we will see that it can be meaningfully extended to the single-prover setting.

▶ **Definition 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*

1. $V|\psi\rangle_{AB} = |\phi\rangle_{AB}|\text{Aux}\rangle$,
2. $V(P_x^{(a)} \otimes Q_y^{(b)})|\psi\rangle_{AB} = \left((M_x^{(a)} \otimes N_y^{(b)})|\phi\rangle_{AB}\right)|\text{Aux}\rangle$, *for some ancillary state $|\text{Aux}\rangle$.*

The first condition is the same as in Definition 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 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*

1. $V|\psi\rangle = |\phi\rangle|\text{AUX}\rangle$,
2. $VP_{x,y}^{(a,b)}|\psi\rangle = \left(K_{x,y}^{(a,b)}|\phi\rangle\right)|\text{AUX}\rangle$, *for some ancillary state* $|\text{AUX}\rangle \in \mathcal{H}'$.

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 [8] and the verification protocol from [23] can be adapted to this single-round setting [2, 11, 9]. 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 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 $V : \mathcal{H} \to \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}'$ such that*

1. $V|\psi\rangle = |\phi\rangle_{AB}|\text{AUX}\rangle$,
2. $VP_{x,y}^{(a,b)}|\psi\rangle = \left((M_x^{(a)} \otimes N_y^{(b)})|\phi\rangle_{AB}\right)|\text{AUX}\rangle$, *for some ancillary state* $|\text{AUX}\rangle \in \mathcal{H}'$.

Again, this definition is informal for the same reason as for Definition 3. A formal statement of such a single-prover self-testing result with a tensor product structure is given in [27, Theorem 4.38], the main result of our work.

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 [23], building on the construction of noisy trapdoor claw-free function families by Brakerski et al. in [8]. Here, we only give a brief informal description of the main properties of an ENTCF family (see [27, 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:

1. 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 $f_{k,0}(x_0) = f_{k,1}(x_1) = y$.
2. 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*.
3. For fixed $k \in \mathcal{K}_{\mathcal{G}}$, $f_{k,0}$ and $f_{k,1}$ are injective functions with disjoint images.
4. 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*.
5. 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.

2 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 [27, Figure 1].

On a very high level, one can view the protocol as first executing the RSP protocol from [19] 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 θ_1, θ_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

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

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 [23], 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:

$$|\psi_i'\rangle = |b\rangle|x_b\rangle\,, \tag{2}$$

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:

$$|\psi_i'\rangle = \frac{1}{\sqrt{2}}(|0\rangle|x_0\rangle + |1\rangle|x_1\rangle)\,, \tag{3}$$

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

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):

$$|\psi_i''\rangle = \begin{cases} |b\rangle & \text{if } k_i \text{ belongs to an injective family,} \\ \frac{1}{\sqrt{2}}(|0\rangle + (-1)^{d_i\cdot(x_0\oplus x_1)}|1\rangle) & \text{if } k_i \text{ belongs to a claw-free family.} \end{cases} \tag{4}$$

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 σ_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 θ_1 and θ_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 $CZ|\psi_1''\rangle|\psi_2''\rangle$.

2.1 Soundness proof

We now give a brief intuition for the soundness of the protocol; the full soundness proof is given in [27, 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 [19], 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 [30] and is actually one of the main difficulties in our soundness proof, but we leave the details of this for [27, 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 [20]. 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 [20]. 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.

3 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.[4]

[4] The freedom of applying a global change of basis means that the entangled Bell state can be mapped to a product state. However, then the prover's measurements are mapped to entangling measurements, so entanglement is still present.

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., $\mathsf{BQP} = \mathsf{BPP}$.[5] It therefore constitutes a "test of quantumness" in the sense of [8] and differs from proposals for testing quantum supremacy such as [7], which only certify the "quantumness" of a device under the assumption that $\mathsf{BQP} \neq \mathsf{BPP}$.[6]

Existing multi-prover self-testing protocols are typically based on non-local games, e.g., the CHSH game [25]. 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 [13]. 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 [1, 22, 18].

Another approach to constructing single-prover self-tests for a larger class of states might be to strengthen Mahadev's measurement protocol [23] 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 [37]. The concept of a proof of quantum knowledge, first introduced in [10, 15] for the setting of a quantum verifier and extended to the setting of a classical verifier in [37], 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 [32, 14], randomness expansion [16, 36, 28], or secure multi-party quantum computation [17, 4] to a single-prover setting. There are already computationally secure single-prover protocols for delegated quantum computation [23] and randomness expansion [8]; 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) [26]. 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.

---

[5] Note that the LWE assumption is independent of whether $\mathsf{BQP} = \mathsf{BPP}$ or not, since LWE is assumed to be hard for both quantum and classical computers.

[6] Intuitively, the reason for this is the following: in our protocol and in [8], the quantum prover has to be able to compute either a preimage or a pair (u, d) such that $u = d \cdot (x_0 \oplus x_1)$, where (x_0, x_1) forms a claw. If a classical prover was able to correctly compute a preimage or a pair (u, d), it could be rewound to compute both at the same time, contradicting the adaptive hardcore bit property. In a quantum prover, the collapsing nature of quantum measurements prevents us from rewinding the prover.

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.

References

1. William Aiello, Sandeep Bhatt, Rafail Ostrovsky, and S. Raj. 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*, pages 463–474, 2000. doi:10.1007/3-540-45022-X_39.
2. Gorjan Alagic, Andrew M Childs, Alex B Grilo, and Shih-Han Hung. Non-interactive classical verification of quantum computation. *arXiv preprint*, 2019. arXiv:1911.08101.
3. John S. Bell. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika*, 1(3):195–200, November 1964. doi:10.1103/PhysicsPhysiqueFizika.1.195.
4. Michael Ben-Or, Claude Crépeau, Daniel Gottesman, Avinatan Hassidim, and Adam Smith. Secure multiparty quantum computation with (only) a strict honest majority. *IEEE 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 249–260, 2006. doi:10.1109/FOCS.2006.68.
5. Kishor Bharti, Maharshi Ray, Antonios Varvitsiotis, Adán Cabello, and Leong-Chuan Kwek. Local certification of programmable quantum devices of arbitrary high dimensionality. *arXiv preprint*, 2019. arXiv:1911.09448.
6. Kishor Bharti, Maharshi Ray, Antonios Varvitsiotis, Naqueeb Ahmad Warsi, Adán Cabello, and Leong-Chuan Kwek. Robust self-testing of quantum systems via noncontextuality inequalities. *Phys. Rev. Lett.*, 122:250403, June 2019. doi:10.1103/PhysRevLett.122.250403.
7. Adam Bouland, Bill Fefferman, Chinmay Nirkhe, and Umesh Vazirani. On the complexity and verification of quantum random circuit sampling. *Nature Physics*, 15(2):159–163, February 2019. doi:10.1038/s41567-018-0318-2.
8. 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)*, pages 320–331, 2018. doi:10.1109/FOCS.2018.00038.
9. Zvika Brakerski, Venkata Koppula, Umesh Vazirani, and Thomas Vidick. Simpler proofs of quantumness. *arXiv preprint*, 2020. arXiv:2005.04826.
10. Anne Broadbent and Alex B Grilo. Zero-knowledge for QMA from locally simulatable proofs. *arXiv preprint*, 2019. arXiv:1911.07782.
11. Nai-Hui Chia, Kai-Min Chung, and Takashi Yamakawa. Classical verification of quantum computations with efficient verifier. *arXiv preprint*, 2019. arXiv:1912.00990.
12. Alexandru Cojocaru, Léo Colisson, Elham Kashefi, and Petros Wallden. QFactory: Classically-Instructed Remote Secret Qubits Preparation. *Advances in Cryptology - ASIACRYPT 2019, Lecture Notes in Computer Science, Springer*, pages 615–645, 2019. doi:10.1007/978-3-030-34578-5_22.
13. Andrea Coladangelo, Koon Tong Goh, and Valerio Scarani. All pure bipartite entangled states can be self-tested. *Nature Communications*, 8(1):15485, August 2017. doi:10.1038/ncomms15485.
14. Andrea Coladangelo, Alex B. Grilo, Stacey Jeffery, and Thomas 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. doi:10.1007/978-3-030-17659-4_9.
15. Andrea Coladangelo, Thomas Vidick, and Tina Zhang. Non-interactive zero-knowledge arguments for QMA, with preprocessing. In *Annual International Cryptology Conference*, pages 799–828. Springer, 2020.
16. Roger Colbeck. *Quantum and relativistic protocols for secure multi-party computation*. PhD Thesis, University of Cambridge, 2006. arXiv:0911.3814.

17 Claude Crépeau, Daniel Gottesman, and Adam Smith. Secure multi-party quantum computation. *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 643–652, 2002. doi:10.1145/509907.510000.

18 Yevgeniy Dodis, Shai Halevi, Ron D. Rothblum, and Daniel Wichs. Spooky Encryption and Its Applications. *Advances in Cryptology - CRYPTO 2016, Lecture Notes in Computer Science, Springer*, pages 93–122, 2016. doi:10.1007/978-3-662-53015-3_4.

19 Alexandru Gheorghiu and Thomas Vidick. Computationally-secure and composable remote state preparation. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1024–1033. IEEE, 2019.

20 William T. Gowers and Omid Hatami. Inverse and stability theorems for approximate representations of finite groups. *Sbornik: Mathematics*, 208(12):1784, 2017. doi:10.1070/SM8872.

21 Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. $\mathsf{MIP}^* = \mathsf{RE}$. *arXiv preprint*, 2020. arXiv:2001.04383.

22 Yael Tauman Kalai, Ran Raz, and Ron 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)*, pages 485–494, 2014. doi:10.1145/2591796.2591809.

23 Urmila Mahadev. Classical verification of quantum computations. *IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 259–267, 2018. doi:10.1109/FOCS.2018.00033.

24 Dominic Mayers and Andrew Yao. Self testing quantum apparatus. *Quantum Info. Comput.*, 4(4):273–286, July 2004. URL: https://dl.acm.org/doi/10.5555/2011827.2011830.

25 M McKague, T H Yang, and V Scarani. Robust self-testing of the singlet. *Journal of Physics A: Mathematical and Theoretical*, 45(45):455304, October 2012. doi:10.1088/1751-8113/45/45/455304.

26 Tony Metger, Yfke Dulek, Andrea Coladangelo, and Rotem Arnon-Friedman. Device-independent quantum key distribution from computational assumptions. *arXiv preprint*, 2020. arXiv:2010.04175.

27 Tony Metger and Thomas Vidick. Self-testing of a single quantum device under computational assumptions. *CoRR*, 2020. (Full version: arXiv:2001.09161).

28 Carl A. Miller and Yaoyun. Shi. Universal security for randomness expansion from the spot-checking protocol. *SIAM Journal on Computing*, 46(4):1304–1335, 2017. doi:10.1137/15M1044333.

29 Sandu Popescu and Daniel Rohrlich. Which states violate Bell's inequality maximally? *Physics Letters A*, 169(6):411–414, October 1992. doi:10.1016/0375-9601(92)90819-8.

30 Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998. doi:10.1137/S0097539795280895.

31 Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56(6), 2009. doi:10.1145/1568318.1568324.

32 Ben W Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496(7446):456, 2013. doi:10.1038/nature12035.

33 Valerio Scarani. *Bell Nonlocality*. Oxford University Press, 2019.

34 Stephen J. Summers and Reinhard Werner. Maximal violation of Bell's inequalities is generic in quantum field theory. *Communications in Mathematical Physics*, 110(2):247–259, June 1987. doi:10.1007/BF01207366.

35 Ivan Šupić and Joseph Bowles. Self-testing of quantum systems: a review. *Quantum*, 4:337, 2020.

36 Umesh Vazirani and Thomas 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)*, pages 61–76, 2012. doi:10.1145/2213977.2213984.

37 Thomas Vidick and Tina Zhang. Classical proofs of quantum knowledge. *arXiv preprint*, 2020. arXiv:2005.01691.