

A universal scheme for robust self-testing in the prepare-and-measure scenario

Nikolai Miklin^{1,2} and Michał Oszmaniec³

¹Institute of Theoretical Physics and Astrophysics, National Quantum Information Center, Faculty of Mathematics, Physics and Informatics, University of Gdansk, 80-306 Gdańsk, Poland

²International Centre for Theory of Quantum Technologies (ICTQT), University of Gdansk, 80-308 Gdańsk, Poland

³Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warszawa, Poland

We consider the problem of certification of arbitrary ensembles of pure states and projective measurements solely from the experimental statistics in the prepare-and-measure scenario assuming the upper bound on the dimension of the Hilbert space. To this aim, we propose a universal and intuitive scheme based on establishing perfect correlations between target states and suitably-chosen projective measurements. The method works in all finite dimensions and allows for robust certification of the overlaps between arbitrary preparation states and between the corresponding measurement operators. Finally, we prove that for qubits, our technique can be used to robustly self-test arbitrary configurations of pure quantum states and projective measurements. These results pave the way towards the practical application of the prepare-and-measure paradigm to certification of quantum devices.

1 Introduction

Quantum devices are becoming more and more complex and the possibilities of their precise control and manipulation keep increasing. Recently reported demonstration of quantum computational advantage by Google [1] is only an intermediate milestone and quantum technologies have a potential real-life applications in fields such as quantum sensing [2], simulation of quantum systems [3], efficient computation [4] and machine learning [5, 6].

With the increasing complexity of quantum systems, there is a growing need for certification and verification of their performance. This task is usually realized via the combination of quantum tomography and various benchmarking schemes (see [7] for a recent review). However, these methods, despite being powerful and universally applicable, depend on the assumptions about the inner workings of quantum systems, such as perfect measurements or uncorrelated and independent errors. In contrast to these approaches *self-testing* is a method which aims at proving the uniqueness of the implemented states or measurements based solely on the observed statistics and under minimal physical assumptions.

The paradigm of self-testing was first introduced in the context of quantum cryptography [8], with the aim to obtain trust in cryptographic devices (see [9] for a recent review).

Nikolai Miklin: nikolai.miklin@ug.edu.pl

Michał Oszmaniec: oszmaniec@cft.edu.pl

Initially, it was applied to correlations observed in the Bell scenario [10] (see e.g., [8, 11–13]). The most know result in this area is certification of the singlet state in the case of maximal violation of the Clauser-Horne-Shimony-Holt Bell inequality [14].

With the growing number of results on self-testing, more and more attention is being drawn to prepare-and-measure scenarios, that are more experimentally appealing as compared to the one of Bell (see e.g., [15–18]). Therein, one does not need to ensure space-like separation of the measurement events by two parties; in contrast, one party, Alice, communicates some of her states to Bob, who measures them in some basis of his choice. In order to get meaningful certification results further assumptions are needed. In the most commonly studied *semi-device-independent* (SDI) scenario [19], one assumes that the dimension of the quantum system used for transmitting information is bounded from above. There exist, however, alternative approaches based on other constraints like minimal overlap [16], mean energy constraint [20] or entropy constraint [21].

It is important to point out that without the assumption on the dimension, or an equivalent, no certification of quantum systems is possible in the prepare-and-measure scenario. The main reason is the fact that any observed distribution resulting from quantum preparations and quantum measurements can also be obtained through classical communication if the latter is unconstrained (see, e.g., Ref. [22]). In this work, we chose the assumption of the bounded dimension as we find it physically motivated. For example, if photons' polarization is used as a carrier of quantum information, it is natural to assume this upper bound to be equal to two. By ensuring that no other degree of freedom of photons is correlated with the choice of her preparation, Alice can gain trust that no additional information can be leaked to Bob or an adversary.

Let us briefly recap on what has been done so far in the area of SDI self-testing. First, self-testing results were proven for mutually unbiased bases (MUBs) in dimension 2, for both state ensembles and measurements [23]. This was further generalised to SDI certification of mutual unbiasedness of pairs of bases in an arbitrary dimension in Ref. [24]. Methods for self-testing of extremal qubit positive operator-valued measures (POVMs) were proposed in [25, 26] and further extended to symmetric-informationally complete (SIC) POVMs [27]. Importantly, all of the above results either rely on numerical approaches, for general state preparations and POVMs, or work only for special scenarios that exhibit many symmetries.

In this work we propose a simple analytical method allowing to certify overlaps between preparations of arbitrary pure states and arbitrary projective measurements in *qudit* systems. The scheme relies on establishing perfect correlations between preparation states and outcomes of suitably-chosen projective measurements. The method is universally applicable and robust to experimental noise. We prove that for qubits our SDI certification method can be used to obtain a robust self-testing result for arbitrary preparations of pure qubit states and corresponding projective measurements. While for higher dimensions we do not show self-testing, our scheme allows for SDI certification of numerous inherently quantum properties. The examples include, but are not limited to: *arbitrary* overlaps between any number of measurement bases, MUB conditions, information-completeness of measurements, and SIC relations among measurement effects.

We believe that our findings greatly extend the applicability of the paradigm of self-testing in the SDI setting. Our approach should be of interest to the researchers working on certification of near-term quantum computers [7] since our scheme can be used to characterize state-preparation and measurement errors, which is one of the major problems in the certification of quantum devices [28]. We also expect possible cryptographic applications as our setup is very similar to the one of textbook quantum key distribution schemes

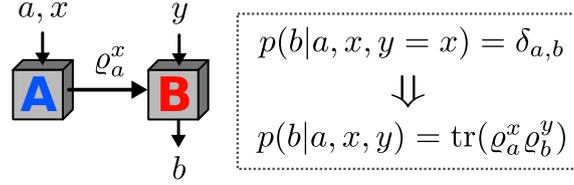


Figure 1: The idea of the certification scheme for the overlaps between states. Alice chooses her inputs a, x and Bob his input y . Alice sends states ϱ_a^x to Bob, who produces his outcome b . After establishing perfect correlations between preparation states and measurements for $y = x$, the rest of the statistics can be used to compute the overlaps between the preparation states. This can be used to directly certify presence of coherences between Alice's input states and between effects of measurements implemented by Bob.

[29, 30]. The perfect correlations can be utilized for generation of the secret key, while the rest can be used to estimate the security. Thus, our methods can be directly applied for certification of quantum devices implementing protocols such as BB84 [29], which is normally achieved by introducing additional preparation states or measurement bases [31].

Before we proceed, let us first fix some of the notations used in this paper. Let X be a linear operator acting on a finite-dimensional Hilbert space \mathcal{H} . Throughout the paper we use $\|X\|$, $\|X\|_F$ to denote operator norm and Frobenius norm of X . We will also use $|\mathbf{n}|$ to denote the Euclidean norm of $\mathbf{n} \in \mathbb{R}^3$, and $[n]$ to denote an n -element set $\{1, 2, \dots, n\}$.

2 Description of the scenario

We consider the prepare-and-measure scenario in which in each run of the experiment Alice prepares a quantum d -level system in one of the states from a finite set of preparations ϱ_a^x , for which we use two indexes $x \in [n]$ and $a \in [d]$. The purpose of the double index will become clear later. Subsequently, Bob performs a measurement on this state with a finite choice of measurement settings $y \in [n]$ having the possible outcomes $b \in [d]$. We assume that the parties do not communicate in any other way and do not have access to any entangled states or shared randomness [32]. This implies the standard i.i.d. assumption that in all the experiment runs, the same choices of preparations lead to the identical and uncorrelated quantum states prepared by Alice's device. The equivalent assumption is placed on Bob's measuring device. The role of these assumptions is discussed later in the text. This also implies that the observed statistics $p(b|a, x, y)$ are given by the Born rule i.e., $p(b|a, x, y) = \text{tr}(\varrho_a^x M_b^y)$, where $\mathbf{M}^y = (M_1^y, M_2^y, \dots, M_d^y)$ is a quantum measurement (POVM) performed by Bob upon the choice of the setting y . The goal of SDI certification is then to identify the preparation states and the measurements, or their properties, based solely on the observed statistics $p(b|a, x, y)$ assuming the upper bound on the dimension d and the validity of Born's rule. We say that certain states ϱ_a^x and measurements \mathbf{M}^y can be *self-tested* if the observed statistics specify these objects uniquely up to a unitary transformation and, perhaps, a global transposition.

3 Certification of overlaps

We start with a presentation of our scheme for certifying pairwise overlaps between *pure qudit* preparation states and between the corresponding *projective* measurements. By "corresponding" we mean that in our scheme we set these states and measurements to be "equal" in a sense that $\varrho_a^x = (\varrho_a^x)^2 = M_a^x$ for all $a \in [d]$ and $x \in [n]$. This is the

reason why we chose to enumerate the preparation states with two indexes. The choice of d and n is arbitrary as long as $d > 1$ and $n \geq 1$. In what follows we will refer to these objects as target pure states and target measurements respectively. Their experimental counterparts we denote as $\tilde{\varrho}_a^x$, and $\tilde{\mathbf{M}}^y$ respectively. By “experimental” we mean any states and measurements defined over a Hilbert space of dimension at most d and that reproduce the *observed statistics* i.e., $\tilde{p}(b|a, x, y) = \text{tr}(\tilde{\varrho}_a^x \tilde{\mathbf{M}}_b^y)$. Clearly, we *do not* assume that the experimental states and measurement have to be “equal”. Neither do we assume any properties of experimental states or measurements like the purity of the projective character.

The idea of our certification scheme is very intuitive, yet powerful (see Fig. 1). Assume that Alice and Bob prepared their devices in a way that $\tilde{p}(b|a, x, y) = 1$, whenever $y = x$ and $b = a$. In other words, outcomes of Bob’s measurement are *perfectly correlated* with the preparations of Alice (whenever $x = y$). Since we assumed the upper bound of d on the Hilbert space’s dimension in which experimental states and measurements are defined, and in this space, we have a set of d quantum states and a POVM with d effects which produce perfect correlations, we can conclude that, first, the Hilbert space’s dimension must be exactly d , and second that $\tilde{\varrho}_a^x = \tilde{M}_a^x$, for all $a \in [d]$ and $x \in [n]$. Clearly, after these perfect correlations are established, the “cross-terms”, can be used to compute the overlaps between the preparation states and between measurement operators: $\tilde{p}(b|a, x, y) = \text{tr}(\tilde{\varrho}_a^x \tilde{\mathbf{M}}_b^y) = \text{tr}(\tilde{\varrho}_a^x \tilde{\varrho}_b^y) = \text{tr}(\tilde{M}_a^x \tilde{M}_b^y)$. Therefore, if the experimental statistics $\tilde{p}(b|a, x, y)$ match the target statistics $p(b|a, x, y)$, we can certify that overlaps between experimental states match those of the target states. The same holds for the corresponding measurement operators.

Our method can be also applied when experimental statistics do not match exactly the target ones.

Theorem 1 (Robust SDI certification of overlaps). *Consider pure target qudit preparation states ϱ_a^x and target projective measurements \mathbf{M}^y , where $a \in [d]$ and $x, y \in [n]$. Assume that $\varrho_a^x = M_a^x$ for all a, x and furthermore that experimental states $\tilde{\varrho}_a^x$ and measurements $\tilde{\mathbf{M}}^y$ act on Hilbert space of dimension at most d and generate statistics $\tilde{p}(b|a, x, y) = \text{tr}(\tilde{\varrho}_a^x \tilde{\mathbf{M}}_b^y)$ such that $|\tilde{p}(b|a, x, y) - \text{tr}(\varrho_a^x M_b^y)| \leq \varepsilon$, for all a, b, x, y . Then, input states $\tilde{\varrho}_a^x$ are almost pure and measurements $\tilde{\mathbf{M}}^y$ are almost projective in the sense that*

$$\text{for all } x \quad \sum_{a=1}^d \|\tilde{\varrho}_a^x\| \geq d(1 - 2\varepsilon) , \quad (1)$$

$$\text{for all } y \quad \sum_{b=1}^d \|\tilde{\mathbf{M}}_b^y\| \geq d(1 - \varepsilon) . \quad (2)$$

Moreover, for all $x \neq x'$, $a \neq a'$, $y \neq y'$, and $b \neq b'$, we have

$$\begin{aligned} |\text{tr}(\tilde{\varrho}_a^x \tilde{\varrho}_{a'}^{x'}) - \text{tr}(\varrho_a^x \varrho_{a'}^{x'})| &\leq \varepsilon + \sqrt{2\varepsilon + d^2\varepsilon^2}, \\ |\text{tr}(\tilde{\mathbf{M}}_b^y \tilde{\mathbf{M}}_{b'}^{y'}) - \text{tr}(M_b^y M_{b'}^{y'})| &\leq \varepsilon + (1 + d\varepsilon)\sqrt{2\varepsilon + d^2\varepsilon^2} . \end{aligned} \quad (3)$$

The parameter ε accounts for all possible experimental errors, including the finite statistics’ error that can be estimated using standard tools of statistical analysis.

Proof. We start with a straightforward proof of Eq.(2). Since $\|\tilde{M}_a^x\| \geq \text{tr}(\tilde{M}_a^x \varrho)$ for any state ϱ , the relation $\text{tr}(\tilde{\varrho}_a^x \tilde{M}_a^x) \geq 1 - \varepsilon$, valid for all a, x , implies $\|\tilde{M}_a^x\| \geq 1 - \varepsilon$, for all a, x , and hence for all x we have $\sum_{a=1}^d \|\tilde{M}_a^x\| \geq d - d\varepsilon$.

To prove the bounds on the norms of experimental states in Eq. (1), we fix a setting x and use a decomposition:

$$\tilde{M}_a^x = \lambda_a |\phi_a\rangle\langle\phi_a| + \text{Rest}_a, \quad (4)$$

where λ_a is the largest eigenvalue of \tilde{M}_a^x , $|\phi_a\rangle\langle\phi_a|$ is the corresponding eigenvector and, Rest_a is a positive-semi-definite operator satisfying $\langle\phi_a|\text{Rest}_a|\phi_a\rangle = 0$ (to keep the derivations legible, we omit the index x). Using the fact that observed statistics are ε close to the target ones we get:

$$1 - \varepsilon \leq \text{tr}(\tilde{\varrho}_a^x \tilde{M}_a^x) = \lambda_a \langle\phi_a|\tilde{\varrho}_a^x|\phi_a\rangle + \text{tr}(\tilde{\varrho}_a^x \text{Rest}_a) \leq \|\tilde{\varrho}_a^x\| + \text{tr}(\text{Rest}_a). \quad (5)$$

By taking the sum over a we obtain $\sum_{a=1}^d \|\tilde{\varrho}_a^x\| \geq d(1 - \varepsilon) - \sum_{a=1}^d \text{tr}(\text{Rest}_a)$. We can now use the identity $d = \sum_{a=1}^d \text{tr}(\tilde{M}_a^x)$, which follows from the fact that operators M_a^x form a POVM in \mathbb{C}^d . This equality allows us to give a bound:

$$d = \sum_{a=1}^d \lambda_a + \sum_{a=1}^d \text{tr}(\text{Rest}_a) \leq d(1 - \varepsilon) + \sum_{a=1}^d \text{tr}(\text{Rest}_a), \quad (6)$$

which is equivalent to $\sum_{a=1}^d \text{tr}(\text{Rest}_a) \leq d\varepsilon$. Inserting this to $\sum_{a=1}^d \|\tilde{\varrho}_a^x\| \geq d(1 - \varepsilon) - \sum_{a=1}^d \text{tr}(\text{Rest}_a)$ we obtain Eq. (1).

We now proceed to the proof of Eqs. (3). We start with the one for the overlaps between preparation states. The proof is given by the following sequence of inequalities which hold for every $a \neq a'$, $x \neq x'$:

$$\begin{aligned} |\text{tr}(\tilde{\varrho}_a^x \tilde{\varrho}_{a'}^{x'}) - \text{tr}(\varrho_a^x \varrho_{a'}^{x'})| &\leq |\text{tr}(\tilde{\varrho}_a^x \tilde{\varrho}_{a'}^{x'}) - \text{tr}(\tilde{M}_a^x \tilde{\varrho}_{a'}^{x'})| + |\text{tr}(\tilde{M}_a^x \tilde{\varrho}_{a'}^{x'}) - \text{tr}(M_a^x \varrho_{a'}^{x'})| \\ &\leq \|\tilde{\varrho}_a^x - \tilde{M}_a^x\| + \varepsilon \leq \|\tilde{\varrho}_a^x - \tilde{M}_a^x\|_F + \varepsilon = \sqrt{\text{tr}((\tilde{\varrho}_a^x)^2) + \text{tr}((\tilde{M}_a^x)^2) - 2\text{tr}(\tilde{\varrho}_a^x \tilde{M}_a^x)} + \varepsilon \\ &\leq \sqrt{1 + (1 + d^2\varepsilon^2) - 2(1 - \varepsilon)} = \varepsilon + \sqrt{2\varepsilon + d^2\varepsilon^2}. \end{aligned}$$

All of the above inequalities are pretty straightforward apart from $\text{tr}((\tilde{M}_a^x)^2) \leq 1 + d^2\varepsilon^2$ that we prove below. For this we again use the partial spectral decomposition from Eq. (4) (without writing the superscript x as above) which implies:

$$\text{tr}((\tilde{M}_a^x)^2) = \lambda_a^2 + \text{tr}(\text{Rest}_a^2) \leq 1 + (\text{tr}(\text{Rest}_a))^2 \leq 1 + d^2\varepsilon^2. \quad (7)$$

The proof for the overlaps of POVM effects is very similar to the one for states with the only difference being the following inequality:

$$|\text{tr}(\tilde{M}_a^x \tilde{M}_{a'}^{x'}) - \text{tr}(\tilde{M}_a^x \varrho_{a'}^{x'})| \leq (1 + d\varepsilon) \|\tilde{\varrho}_a^x - \tilde{M}_a^x\|, \quad (8)$$

that is used in the second step in the proof in Eq. (7). One can easily verify the validity of this inequality by writing once more the decomposition $\tilde{M}_{a'}^{x'} = \lambda_{a'} |\phi_{a'}\rangle\langle\phi_{a'}| + \text{Rest}_{a'}$ and remembering that $\text{tr}(\text{Rest}_a) \leq d\varepsilon$. \square

The above result states that if the statistics observed in our certification scheme vary just a little bit from the target ones, the overlaps of the experimental states are also close to the overlaps between target states (and analogously for measurements). To our best knowledge analogous results have been previously known only for very special symmetric target states and measurements forming MUBs [23, 24].

In Appendix B (Lemma 1) we improve the above bounds for the case of qubit systems. Moreover, in Appendix A we prove, by giving explicit examples, that the bounds in Eq. (3) are tight in the first orders in $\sqrt{\varepsilon}$ and d .

4 Self-testing of qubits

We now show that certification of overlaps allows to prove robust self-testing result for *arbitrary* pure qubit preparations and projective measurements appearing in our certification scheme.

Theorem 2 (Robust self-testing of qubit systems). *Consider target pure qubit states ϱ_a^x and projective measurements \mathbf{M}^y , where $a = 1, 2$, $x, y \in [n]$, and $\varrho_a^x = M_a^x$ for all a, x . Assume that experimental qubit states $\tilde{\varrho}_a^x$ and measurements $\tilde{\mathbf{M}}^y$ generate statistics $\tilde{p}(b|a, x, y) = \text{tr}(\tilde{\varrho}_a^x \tilde{M}_b^y)$ such that $|\tilde{p}(b|a, x, y) - \text{tr}(\varrho_a^x M_b^y)| \leq \varepsilon$, for all $a, b = 1, 2$ and $x, y \in [n]$. Then, there exist ε_0 such that for $\varepsilon \leq \varepsilon_0$ there exist a qubit unitary U such that*

$$\begin{aligned} \frac{1}{2n} \sum_{a,x} \text{tr}(U(\tilde{\varrho}_a^x)^{(T)}U^\dagger \varrho_a^x) &\geq 1 - f(\varepsilon), \\ \frac{1}{2n} \sum_{b,y} \text{tr}(U(\tilde{M}_b^y)^{(T)}U^\dagger M_b^y) &\geq 1 - g(\varepsilon), \end{aligned} \quad (9)$$

where $(\cdot)^{(T)}$ is a potential transposition with respect to a fixed basis in \mathbb{C}^2 that may have to be applied to all experimental states and measurements at the same time. Moreover, functions $f, g : [0, \varepsilon_0) \rightarrow \mathbb{R}_+$ depend solely on the target states and measurements and, for small ε , have the asymptotics $f(\varepsilon) \propto \varepsilon$, $g(\varepsilon) \propto \varepsilon$.

In the above we used the fidelity $F(\varrho, \sigma) = \text{tr}(\varrho\sigma)$ to indicate the closeness between *rotated* experimental states and target pure states (and analogously for measurements), following existing literature. The case $\varepsilon = 0$ corresponds to ideal reconstruction of target states and measurements after applying a suitable isometry. We remark that we allow only unitary operations (and possible transposition), as opposed to general channels [23, 26], to be applied to the experimental states in order to approximate the target states *as well as possible*.

In Appendix B we give a formal version of the above result and its proof. Moreover, we present there robustness bounds expressed in terms of the trace distance and its analogue for measurements [33]. We remark that the functions f, g become unbounded once Bloch vectors of target qubit states become singular, i.e., once the target states are close to being aligned in a space of smaller dimension.

Remark. *Certification of overlaps between pure states in general does not allow for their self-testing in higher-dimensional systems. This is e.g., due to the existence of unitary inequivalent sets of SIC-POVMs for $d = 3$ [34] and MUBs for $d = 4$ [35] (even if we allow for complex conjugation).*

Sketch of the proof. We give a full proof for the ideal case ($\varepsilon = 0$) below. From Theorem 1 it follows that for all x, x', a, a' we have $\tilde{\varrho}_a^x = \tilde{M}_a^x$ and $\text{tr}(\tilde{\varrho}_a^x \tilde{\varrho}_a^{x'}) = \text{tr}(\varrho_a^x \varrho_a^{x'})$. Using the Bloch representation, $\varrho_a^x = \frac{1}{2}(\mathbb{1} + \mathbf{n}_a^x \cdot \boldsymbol{\sigma})$, we can conclude that also $\tilde{\mathbf{n}}_a^x \cdot \tilde{\mathbf{n}}_a^{x'} = \mathbf{n}_a^x \cdot \mathbf{n}_a^{x'}$, where \mathbf{n}_a^x and $\tilde{\mathbf{n}}_a^x$ are Bloch vectors of ϱ_a^x and $\tilde{\varrho}_a^x$ respectively.

Assume now that the vectors $\mathbf{n}_1^1, \mathbf{n}_1^2, \mathbf{n}_1^3$ are linear independent. Let O be the linear transformation defined by $O\mathbf{n}_1^x = \tilde{\mathbf{n}}_1^x$, $x = 1, 2, 3$, and let L be the matrix whose rows are the vectors \mathbf{n}_1^x , $x = 1, 2, 3$. Then, we have $LO^T O L^T = LL^T$, and consequently, since L is invertible by the construction, $O^T O = \mathbb{1}_3$, i.e., O is an orthogonal transformation in \mathbb{R}^3 . It is well-known [36] that if $\det(O) = 1$, there exist a unitary matrix U such that $\text{tr}(U(\tilde{\varrho}_a^x)U^\dagger \varrho_a^x) = 1$ for $x = 1, 2, 3$ and $a = 1$. By our assumption all remaining states ϱ_a^x can be decomposed in the basis $\{\mathbb{1}, \varrho_1^1, \varrho_1^2, \varrho_1^3\}$, with the coefficients depending solely on

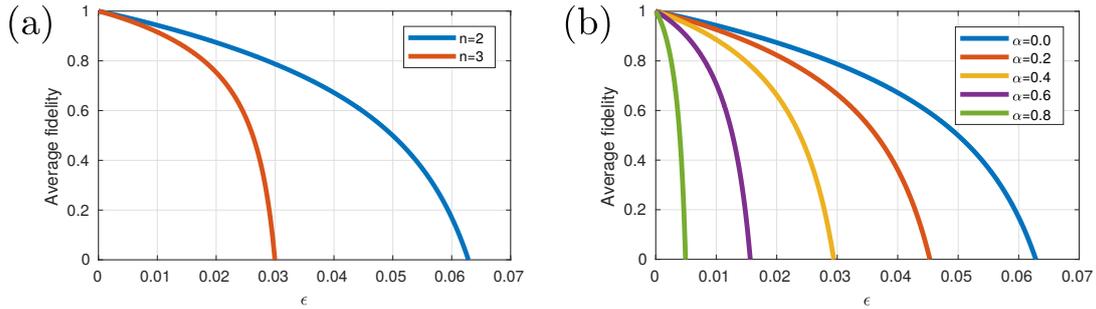


Figure 2: Lower bounds for the average fidelity between the experimental and target states. Part (a) presents results for $n = 2, 3$ qubit MUBs. Part (b) presents results for two qubit bases for different degree of bias α . Value $\alpha = 0$ corresponds to two MUBs while $\alpha = 1$ gives two identical bases.

the overlaps $\text{tr}(\varrho_a^x \varrho_{a'}^{x'})$. The, the same U that maps $\tilde{\varrho}_1^x$ to ϱ_1^x , for $x = 1, 2, 3$, also connects the remaining pairs of states. Finally, if $\det(O) = -1$, the transformation O corresponds to application of the transposition in the standard basis of \mathbb{C}^2 followed by application of a unitary operation U , determined by O [37].

For the general case $\epsilon > 0$ we consider Cholesky factorisations of Gram matrices, denoted by Γ and $\tilde{\Gamma}$, of Bloch vectors $\mathbf{n}_1^1, \mathbf{n}_1^2, \mathbf{n}_1^3$ and their experimental counterparts respectively. Theorem 1 is then used to bound $\|\Gamma - \tilde{\Gamma}\|_F$ and utilize results of Ref. [38] to gauge how much the Cholesky decompositions of Γ and $\tilde{\Gamma}$ differ in the Frobenius norm. The latter can be connected to the average fidelity between the selected target and experimental states. The robustness for the remaining states follows from the fact that they can be decomposed (as operators) using the three initially chosen target states and the identity.

If vectors from the set $\{\mathbf{n}_a^x\}$ span two dimensional space then the same arguments for both $\epsilon = 0$ and $\epsilon > 0$ can be repeated for the considered subspace. Importantly, in this case additional transposition is not necessary. \square

5 Examples

We now apply the quantitative formulation of Theorem 2 to lower-bound average fidelities for different configurations of target states as a function of the allowed error ϵ . In Figure 2 we present results for eigenstates of $n = 2$ and $n = 3$ Pauli matrices (i.e., states forming $n = 2$ and $n = 3$ qubit MUBs), and states belonging to two biased projective measurements satisfying $\text{tr}(\varrho_a^1 \varrho_a^2) = \frac{1+\alpha}{2}$, $a = 1, 2$ where we take $\alpha \in [0, 1]$.

For $n = 2$ MUBs we compare our results with Ref. [23] that focuses on self-testing of qubit MUBs. The results of Ref. [23] give the upper bound on average fidelity equal 0.75 for the deviation of $\simeq 0.1$ in the figure of merit. In our scheme this happens for $\epsilon \simeq 0.033$ as shown on Fig. 2. To test the versatility of our scheme, we also applied it to $n = 3$ MUBs, trine and tetrahedral configurations of qubit states. Quantitative results concerning these examples are listed in Table 1, while detailed derivations are given in Appendix D. For trine and tetrahedral configurations the robustness is obtained via the so-called Procrustes problem, described in Appendix C, based on Ref. [39]. For cases other than two MUBs we cannot make any comparison with the existing literature since, to our best knowledge, these case have not been studied previously.

Configuration	ε_0	C
2 MUBs	≈ 0.062	$\frac{7}{2} + \sqrt{2}$
3 MUBs	≈ 0.030	6
2 biased bases	$\lesssim \frac{4-3\alpha-\sqrt{7-6\alpha}}{18}$	$2 + \left(1 + \frac{\sqrt{1+\alpha}}{\sqrt{2(1-\alpha)}}\right)^2$
Trine	≈ 0.058	$\frac{19}{3}$
Tetrahedron	≈ 0.037	10

Table 1: Results of quantitative variant of Theorem 2 applied to different configurations of target quantum states. The threshold ε_0 sets the maximal noise level which is tolerated by our scheme. The constant C is defined via the relation $f(\varepsilon) \stackrel{\varepsilon \rightarrow 0}{\approx} C\varepsilon$, where $1 - f$ is the lower bound on the average fidelity from Eq. (9).

6 Shared randomness

Throughout the article we assumed that preparation and measurement devices are uncorrelated, which is often true in practice. However, one may also consider a situation in which Alice and Bob share a random variable λ , that they can use to coordinate their actions in a given round of the experiment. The most general statistics that can be generated in such a scenario can be expressed as $p(b|a, x, y) = \int d\lambda p(\lambda) \text{tr}(\varrho_a^x(\lambda) M_b^y(\lambda))$, where $p(\lambda)$ denotes the probability distribution of λ .

Interestingly, the presence of shared randomness makes our main results, as stated by Theorems 1 and 2, inapplicable. The easiest way to see it is to consider the following simple example of $n = d = 2$, one bit of shared randomness $\lambda \in \{1, 2\}$, and $\tilde{\varrho}_a^{1,2} = |a\rangle\langle a|$, $\tilde{M}_b^{1,2} = |b\rangle\langle b|$. This clearly satisfies the requirement on $\tilde{p}(a|a, x, x) = 1$. Now, Alice's and Bob's devices can decide to "flip" their preparations and measurement operators whenever $x = 2, y = 2$ and $\lambda = 2$ (note that λ can be distributed in an arbitrary way). This procedure does not affect the correlations $\tilde{p}(a|a, x, y = x)$, but it can be used to set $\tilde{p}(a|a, 1, 2)$ to an arbitrary value.

Additionally, we believe that in the presence of shared randomness, the usual notion of self-testing in the SDI setting has to be reconsidered. Specifically, for *any* quantum realisation giving the statistics $p(b|a, x, y) = \text{tr}(\varrho_a^x M_b^y)$, one can always consider a strategy in which with probability $p(\lambda)$ Alice prepares states $\varrho_a^x(\lambda) = U_\lambda \varrho_a^x U_\lambda^\dagger$ and Bob implements measurements $M_b^y(\lambda) = U_\lambda M_b^y U_\lambda^\dagger$, where U_λ is some arbitrary unitary transformation depending on λ . Clearly, such strategy reproduces the original statistics, and makes it impossible to find a single unitary that connects the target and experimental states (or measurements).

A very similar argument as above can be applied to situations of non-identical preparations of states and non-identical measurements across different runs of the experiment. Indeed, the value of a given run of the experiment is accessible to both Alice and Bob and can be considered a shared random variable λ that controls both Alice's and Bob's devices. Once again, the existence of a single unitary that connects the target and experimental states can never be guaranteed.

While, we do not consider the assumption on no shared randomness strong, below we propose a modification of our SDI certification scheme that allows to certify overlaps between arbitrary pure states even in the presence of shared randomness. The idea is to introduce, for every pair of non-orthogonal states, a suitable *intermediate state* [40], that enforces fixed overlaps between experimental states in every round of the experiment.

Recall that in our certification scheme we consider pure target qudit states ϱ_a^x and

target projective measurements \mathbf{M}^y , where $a \in [d]$ and $x, y \in [n]$. We extend this scheme by introducing an additional intermediate target state ϱ_z for every two pairs of Alice's input (x, a) , (x', a') , where $a \neq a'$ and $x \neq x'$. The state ϱ_z is chosen as the unique state satisfying:

$$\text{tr} \left(\varrho_z (\varrho_a^x + \varrho_{a'}^{x'}) \right) = 1 + \sqrt{\text{tr}(\varrho_a^x \varrho_{a'}^{x'})}. \quad (10)$$

Let $\tilde{p}(b|z, y)$ denote the experimental statistics, corresponding to inputs z, y and outcome b . In the ideal scenario, in which experimental statistics satisfy assumptions of Theorem 1 with $\varepsilon = 0$, assume additionally that $\tilde{p}(a|z, x) + \tilde{p}(a'|z, x') = 1 + \sqrt{\tilde{p}(a'|x, a, x')}$. Below we prove that under the above assumptions for all λ we have $\text{tr}(\tilde{\varrho}_a^x(\lambda)\tilde{\varrho}_{a'}^{x'}(\lambda)) = \text{tr}(\varrho_a^x\varrho_{a'}^{x'})$. Since in the ideal case overlaps between preparation states *do not depend* on λ and match the target value, our modified certification scheme works also in the presence of shared randomness. In the noisy scenario one can generally expect to be able to upper-bound the probability (over the choice of λ) of the deviations from perfect correlations between states and measurements, as a function of the error ε . We postpone the discussion of this interesting point to the future work.

Proof of soundness of certification in the presence of shared randomness. Consider a general physical realisation of experimental statistics $\tilde{p}(b|x, a, y)$, $\tilde{p}(b|z, y)$ via classically correlated preparations and measurements on \mathbb{C}^d :

$$\tilde{p}(b|x, a, y) = \int d\lambda p(\lambda) \text{tr}(\tilde{\varrho}_a^x(\lambda) \tilde{M}_b^y(\lambda)), \quad (11)$$

$$\tilde{p}(b|z, y) = \int d\lambda p(\lambda) \text{tr}(\tilde{\varrho}_z(\lambda) \tilde{M}_b^y(\lambda)), \quad (12)$$

where $x, y \in [n]$, $a, b \in [d]$ and the variable (input) z labels elements in the set of unordered pairs $\{(x, a), (x', a')\}$, where $x \neq x'$ and $a \neq a'$. Assume now that the above statistics match *exactly* the ones required by our scheme and are compatible with Eq. (10) in the sense that:

$$\tilde{p}(b|x, a, y) = \text{tr}(\varrho_a^x \varrho_b^y), \quad (13)$$

$$\tilde{p}(a|z, x) + \tilde{p}(a'|z, x') = 1 + \sqrt{\tilde{p}(a'|x, a, x')}. \quad (14)$$

In what follows we prove that if the above constraints are satisfied, then for all¹:

$$\tilde{\varrho}_a^x(\lambda) = \tilde{M}_a^x(\lambda), \quad (15)$$

$$\text{tr}(\varrho_a^x \tilde{\varrho}_{a'}^{x'}) = \text{tr}(\varrho_a^x(\lambda) \tilde{\varrho}_{a'}^{x'}(\lambda)), \quad (16)$$

where $x, x' \in [n]$, $a, a' \in [d]$. The above equation means that the overlaps between preparation states (and measurements) do not depend on the value of the shared random variable λ .

The proof of Eq. (15) is straightforward. Namely, from Eq. (13) it follows that:

$$\int d\lambda p(\lambda) \text{tr} \left(\tilde{\varrho}_a^x(\lambda) \tilde{M}_a^x(\lambda) \right) = 1. \quad (17)$$

Since $\text{tr}(\tilde{\varrho}_a^x(\lambda)\tilde{M}_a^x(\lambda)) \leq 1$, we get that for all λ we have $\text{tr}(\tilde{\varrho}_a^x(\lambda)\tilde{M}_a^x(\lambda)) = 1$. Since this reasoning can be repeated for all $a \in [d]$ (for the fixed value of x) and since operators \tilde{M}_a^x form a POVM on \mathbb{C}^d , we finally get Eq. (15).

¹Technically speaking this equality holds for *almost all* λ .

The proof of Eq. (16) is more involved and relies on both Eqs. (13) and (14). Specifically, from Eq. (13) and the already established identity $\tilde{\varrho}_a^x = \tilde{M}_a^x$, it follows that Eq. (14) is equivalent to:

$$\int d\lambda p(\lambda) \text{tr} \left(\tilde{\varrho}_z(\lambda) [\tilde{\varrho}_a^x(\lambda) + \tilde{\varrho}_{a'}^{x'}(\lambda)] \right) = 1 + \sqrt{\text{tr}(\varrho_a^x \varrho_{a'}^{x'})}, \quad (18)$$

where moreover:

$$\int d\lambda p(\lambda) \text{tr} \left(\tilde{\varrho}_a^x(\lambda) \tilde{\varrho}_{a'}^{x'}(\lambda) \right) = \text{tr}(\varrho_a^x \varrho_{a'}^{x'}). \quad (19)$$

Using Bloch representation of qubit states (states $\tilde{\varrho}_a^x(\lambda)$, $\tilde{\varrho}_{a'}^{x'}(\lambda)$ are pure and hence span a two dimensional subspace of \mathbb{C}^d) it is straightforward to obtain the bound:

$$\text{tr} \left(\tilde{\varrho}_z(\lambda) [\tilde{\varrho}_a^x(\lambda) + \tilde{\varrho}_{a'}^{x'}(\lambda)] \right) \leq 1 + \sqrt{\text{tr}(\tilde{\varrho}_a^x(\lambda) \tilde{\varrho}_{a'}^{x'}(\lambda))}. \quad (20)$$

After setting $g(\lambda) = \sqrt{\text{tr}(\tilde{\varrho}_a^x(\lambda) \tilde{\varrho}_{a'}^{x'}(\lambda))}$ and using Eq. (19) we obtain:

$$\int d\lambda p(\lambda) g(\lambda) \geq \sqrt{\int d\lambda p(\lambda) g(\lambda)^2}. \quad (21)$$

Using Chauchy-Schwartz inequality for the left hand side of the above inequality we finally get:

$$\int d\lambda p(\lambda) g(\lambda) = \sqrt{\int d\lambda p(\lambda) g(\lambda)^2}, \quad (22)$$

which is equivalent to saying that the variance of the random variable $g(\lambda)$ vanishes. Therefore, $g(\lambda) = \alpha$, where α is some numerical constant. We conclude the proof by noting that from Eq. (19) it follows that $\alpha^2 = \text{tr}(\varrho_a^x \varrho_{a'}^{x'})$ which implies Eq. (16). \square

7 Extension of the scheme to SDI characterization of rank-1 non-projective measurements

In our work we focus on self-testing of arbitrary configurations of preparation states and projective measurements. In this section we discuss possible extension of our results to certification of properties of general rank-1 measurements in arbitrary dimension d . To that end, we introduce additional states coming from suitably-chosen orthonormal bases per every effect of the measurement we want to certify. With the help of these extra states (and corresponding projective measurements) we can e.g., certify the following properties of POVMs in arbitrary dimension d : (i) information completeness, (ii) extremality of rank-1 POVMs, and (iii) SIC-property.

The main idea of the extension is the following. Suppose we would like to self-test an m -outcome POVM $\mathbf{N} = (N_1, \dots, N_m)$ acting on a Hilbert space of dimension $d < m$ with each effect N_b being of rank 1 i.e., $N_b = \alpha_b \sigma_b$, for some pure states σ_b and positive scalars α_b . We consider m sets of preparation states $\{\varrho_a^x\}$, $a \in [d]$, $x \in [m]$ and the corresponding projective measurements \mathbf{M}^y , $y \in [m]$. Assume now that in the idealized case we observe statistics satisfying:

$$\begin{aligned} \text{tr}(\varrho_a^b N_b) &= 0, \quad \forall b \in [m], a \neq 1, \\ \text{tr}(\varrho_a^x M_b^x) &= \delta_{a,b}, \quad \forall x \in [m]. \end{aligned} \quad (23)$$

From the second set of conditions we conclude that $\{\varrho_a^x\}_a$ forms, as before, a basis of pure states for each x . From the first condition in Eq. (23) we get that \mathbf{N} has to be a rank-1 POVM and, moreover, that $\varrho_1^b = \sigma_b$ (N_b is orthogonal to all of the states except the one with $a = 1$). Consequently, we have $\text{tr}(\varrho_1^b N_b) = \alpha_b$ for all $b \in [m]$. This fact together with the measured overlaps $\text{tr}(\varrho_1^b \varrho_1^{b'}) = \text{tr}(\varrho_1^b M_1^{b'})$ gives the information about the overlaps $\text{tr}(N_b N_{b'})$ between effects of the POVM \mathbf{N} .

The knowledge of all overlaps $\text{tr}(N_b N_{b'})$ (including $b' = b$) and the fact that \mathbf{N} is a rank-1 POVM allow to certify many interesting properties of \mathbf{N} . First, as long as $m \leq d^2$ we can certify extremality of \mathbf{N} . This follows from the fact that rank-1 POVMs are extremal if and only if their effects are linearly independent (see e.g., Ref. [41]). Linear independence can be directly inferred from the experimentally accessible moment matrix $\Gamma_{b,b'} = \text{tr}(N_b N_{b'})$. Specifically, Γ is non-singular if and only if operators $\{N_b\}_b$ are linear independent. Now, if $m = d^2$ we can use the same reasoning to infer information completeness of \mathbf{N} . Finally, from the moment matrix Γ we can directly verify the SIC property i.e., the condition $\text{tr}(N_b N_{b'}) = \frac{1+\delta_{b,b'}}{d^2(d+1)}, \forall b, b'$.

8 Discussion

We have presented a systematic analytical scheme for noise-resilient certification of overlaps between arbitrary configurations of pure quantum states and rank-1 projective measurements in the prepare-and-measure scenario. For qubits our scheme can be used to robustly self-test general ensembles of pure quantum states and the corresponding projective measurements. We believe that these findings pave the way towards systematic certification of general quantum systems in the semi-device-independent paradigm. This is supported by the concrete qubit results from Table 1 and by the universality of our protocol for certifying overlaps of quantum states in arbitrary dimension.

In our main claims, we state that the scheme can be applied to an arbitrary set of preparation states. However, the sets of states in Theorems 1 and 2 always come in a collection of bases. These two statements are not contradicting one another since any given constellation of states can always be completed to a set of bases by adding extra preparation states. In practice, one should aim at aligning each basis of preparation states and the corresponding measurement to the best degree allowed by the experimental setup.

There are many exciting research directions that stem from our work. One of them is a systematic SDI characterization of generalized measurements beyond qubits. Another research direction is the robustness analysis of our scheme in the presence of shared randomness. An important contribution to the field of self-testing would be a generalization of our self-testing results for qubit to higher-dimensional systems. Finally, it is interesting to combine our methods with recent findings relating state discrimination games and resource theories of measurements and channels [42–44].

We would also like to draw reader’s attention to a related work [45].

Acknowledgements

We thank Marcin Pawłowski, Jędrzej Kaniewski and Tanmoy Biswas for interesting discussions and comments. NM acknowledges the financial support by First TEAM Grant No. 2016-1/5. We acknowledge partial support by the Foundation for Polish Science (IRAP project, ICTQT, contract no. 2018/MAB/5, co-financed by EU within Smart Growth Operational Programme). We acknowledge the support from International Research Agendas

Programme of the Foundation for Polish Science funded by the European Regional Development Fund. MO acknowledges the financial support by TEAM-NET project (contract no. POIR.04.04.00-00-17C1/18-00).

Appendix

In this Appendix we provide technical details that were omitted in the main text. First, in Appendix A we prove statements regarding saturation of the bounds in Theorem 1. Second, in Appendix B we formulate and prove a quantitative version of Theorem 2, which concerns robustness of our self-testing protocol expressed in terms of average fidelities. We also prove there Theorem 3 that gives analogous results expressed in terms of trace distance. In Appendix C we provide alternative robustness analysis based on orthogonal Procrustes problem. In Appendix D we provide technical details about the considered examples in Table 1. Finally, in Appendix E we give proofs of two auxiliary Lemmas needed in the proof of the technical version of Theorem 2.

A Explicit form of states and measurements saturating the bounds in Theorem 1

We start by giving an example of four states and two qubit measurements ($n = 2, d = 2$), for which the bound in Eq. (3) on the deviation of the overlaps scales like $\sqrt{\varepsilon}$. Their explicit form is given below:

$$\varrho_1^1 = |\tilde{0}\rangle\langle\tilde{0}|, \quad |\tilde{0}\rangle = \sqrt{1-\varepsilon}|0\rangle + \sqrt{\varepsilon}|1\rangle, \quad \varrho_1^2 = |+\rangle\langle+|, \quad (24)$$

where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and as required by our scheme $\varrho_2^x = \mathbb{1} - \varrho_1^x$, $x = 1, 2$ and $M_a^x = \varrho_a^x$, for all x, a . Let the experimental states and measurements be the following:

$$\begin{aligned} \tilde{\varrho}_1^1 &= |0\rangle\langle 0|, & \tilde{\varrho}_2^1 &= |1\rangle\langle 1|, & \tilde{\varrho}_1^2 &= |+\rangle\langle+|, & \tilde{\varrho}_2^2 &= |-\rangle\langle-|, \\ \tilde{M}_1^1 &= |\tilde{0}\rangle\langle\tilde{0}|, & \tilde{M}_1^2 &= |\tilde{+}\rangle\langle\tilde{+}|, & |\tilde{+}\rangle &= \sqrt{1-\varepsilon}|+\rangle + \sqrt{\varepsilon}|-\rangle. \end{aligned} \quad (25)$$

First, we need to show that the experimental statistics is deviated from the target one by at most ε , i.e., $|\text{tr}(\tilde{\varrho}_a^x \tilde{M}_b^y) - \text{tr}(\varrho_a^x \varrho_b^y)| \leq \varepsilon, \forall a, b, x, y$. Since $\tilde{\varrho}_1^x + \tilde{\varrho}_2^x = \mathbb{1}$ for both $x = 1, 2$, it is sufficient to consider the cases of $a = 1, b = 1$. Calculating the statistics we can see that $\text{tr}(\tilde{\varrho}_1^1 \tilde{M}_1^1) = \text{tr}(\tilde{\varrho}_1^2 \tilde{M}_1^2) = 1 - \varepsilon$, and $\text{tr}(\tilde{\varrho}_1^1 \tilde{M}_1^2) = \text{tr}(\tilde{\varrho}_1^2 \tilde{M}_1^1) = \frac{1}{2} + \sqrt{\varepsilon - \varepsilon^2}$ and the latter is just the same as the target probability, $\text{tr}(\varrho_1^1 \varrho_1^2)$. Finally, calculation of the overlap between the states yields:

$$|\text{tr}(\varrho_1^1 \varrho_1^2) - \text{tr}(\tilde{\varrho}_1^1 \tilde{\varrho}_1^2)| = \sqrt{\varepsilon - \varepsilon^2}, \quad (26)$$

which confirms the scaling of the first order of $\sqrt{\varepsilon}$.

Now, let us show that the bounds from Eq. (3) are also tight in the first order of d . We still consider the case of $n = 2$, but now d is arbitrary. Let us consider the following POVM:

$$M_1 = |1\rangle\langle 1| + \varepsilon(d-1)|+\rangle\langle+|, \quad M_a = (|a\rangle - \delta|+\rangle)\langle a| - \delta\langle+| \quad a = 2, 3, \dots, d, \quad (27)$$

where $\delta = \frac{1}{\sqrt{d-1}} + \sqrt{\frac{1}{d-1} - \varepsilon}$, $\{|a\rangle\}_{a=1}^d$ is the computational basis of \mathbb{C}^d and $|+\rangle = \frac{1}{\sqrt{d-1}} \sum_{a=2}^d |a\rangle$ is the "maximally coherent" state in the subspace spanned by $\{|i\rangle\}_{i=2}^d$.

In the proof of Theorem 1 the quantity $|\text{tr}(\tilde{\varrho}_a^x \tilde{\varrho}_a^{x'}) - \text{tr}(\varrho_a^x \varrho_a^{x'})|$ is upper-bounded by $\varepsilon + \|\tilde{\varrho}_a^x - \tilde{M}_a^x\|$, for which there exist a state $\tilde{\varrho}_a^{x'}$ and an effect $\tilde{M}_a^{x'}$ reaching the bound. Now if we take $\tilde{\mathbf{M}}^x$ to be the POVM we just introduced, and $\tilde{\varrho}_1^x = |1\rangle\langle 1|$, $\tilde{\varrho}_a^x = \frac{M_a}{\text{tr}(M_a)}$, $a = 2, 3, \dots, d$, the conditions of the Theorem 1 will be satisfied and the resulting bound will be $d\varepsilon$.

B Quantitative statement of the robust self-testing

In this part we first formulate Theorem 3 in which we present robustness of our self-testing scheme in terms of trace distance² for states and operator norm for measurements. Then, we give a quantitative statement of Theorem 2, that concerns robustness of our self-testing scheme in terms of average fidelities. We then proceed with proofs of both results.

In what follows we will need the following definition.

Definition 1. Let $\{\mathbf{n}_i\}_{i \in [n]}$ be a set of n vectors in \mathbb{R}^l . Matrix $\Gamma \in \mathbb{R}^{n \times n}$ is called the Gram matrix of the set $\{\mathbf{n}_i\}_{i \in [n]}$, if its elements are given by $\Gamma_{i,j} = \mathbf{n}_i \cdot \mathbf{n}_j$, $i, j \in [n]$.

Theorem 3 (Robust self-testing for qubits via trace distance and operator norm). Consider pure target qubit preparation states ϱ_a^x and target projective measurements \mathbf{M}^y , where $a = 1, 2$ and $x, y \in [n]$. Assume that $\varrho_a^x = M_a^x$ for all a, x and, furthermore, that experimental states $\tilde{\varrho}_a^x$ and measurements $\tilde{\mathbf{M}}^y$ act on Hilbert space of dimension at most d and generate statistics $\tilde{p}(b|a, x, y) = \text{tr}(\tilde{\varrho}_a^x \tilde{M}_b^y)$ such that $|\tilde{p}(b|a, x, y) - \text{tr}(\varrho_a^x M_b^y)| \leq \varepsilon$, for all a, b, x, y .

Let $k \in \{2, 3\}$ be the cardinality of the maximal set of linearly independent Bloch vectors of states ϱ_1^x . Fix a set $S \subset [n]$ of k linearly independent vectors $\{\mathbf{n}_1^x\}_{x \in S}$, construct their Gram matrix Γ_S , and let L_S be a Cholesky factor of Γ_S (i.e., $\Gamma_S = L_S L_S^T$ and L_S is lower triangular). For every $x \in [n] \setminus S$ and both $a = 1, 2$ let $\mathbf{c}_S^{x,a}$, denote the coefficients of decomposition of \mathbf{n}_a^x as a linear combination of $\{\mathbf{n}_1^x\}_{x \in S}$. Finally, let us define three auxiliary functions

$$\begin{aligned} F_k(\varepsilon) &:= \sqrt{\varepsilon} \sqrt{4k(k-1)} \sqrt{1 + 2\sqrt{\varepsilon} + \frac{k+3}{k-1}\varepsilon}, \\ O_k(\varepsilon) &:= 2((k-1)\sqrt{\varepsilon} + (k+1)\varepsilon), \\ E_{S,k}(\varepsilon) &:= \frac{1}{2\sqrt{2}} \frac{\|\Gamma_S^{-1}\| F_k(\varepsilon)}{\sqrt{1 - \|\Gamma_S^{-1}\| O_k(\varepsilon)}} \min \left[\frac{\|\Gamma_S\|}{\|\Gamma_S\|_F}, \frac{\|L_S\|}{\sqrt{k}} \right]. \end{aligned} \quad (28)$$

Then, there is a region of $\varepsilon \in [0, \varepsilon_0)$ determined by

$$\|\Gamma_S^{-1}\| O_k(\varepsilon) \leq 1, \quad (29)$$

²Recall that the trace distance is defined via $d_{\text{tr}}(\sigma, \varrho) = \frac{1}{2} \|\sigma - \varrho\|_1 = \frac{1}{2} \text{tr} \sqrt{(\sigma - \varrho)^\dagger (\sigma - \varrho)}$ and it has a neat operational interpretation in term of optimal success probability p of distinguishing ϱ and σ via the most general quantum measurements: $p = \frac{1}{2}(1 + d_{\text{tr}}(\sigma, \varrho))$.

for which there exist a qubit unitary matrix U such that

$$\frac{1}{k} \sum_{x \in S} \text{d}_{\text{tr}}(U(\tilde{\varrho}_1^x)^{(T)}U^\dagger, \varrho_1^x) \leq E_{S,k}(\varepsilon), \quad (30a)$$

$$\frac{1}{k} \sum_{x \in S} \text{d}_{\text{tr}}(U(\tilde{\varrho}_2^x)^{(T)}U^\dagger, \varrho_2^x) \leq E_{S,k}(\varepsilon) + 2\sqrt{\varepsilon}, \quad (30b)$$

$$\text{d}_{\text{tr}}(U(\tilde{\varrho}_a^x)^{(T)}U^\dagger, \varrho_a^x) \leq \sqrt{k} |\mathbf{c}_S^{a,x}| E_{S,k}(\varepsilon) + \frac{\sqrt{k}}{2} \left(|\mathbf{c}_S^{a,x}| + \frac{\sqrt{k}}{k-1} \right) \frac{\|\Gamma_S^{-1}\| O_k(\varepsilon)}{1 - \|\Gamma_S^{-1}\| O_k(\varepsilon)}, \quad (30c)$$

for $x \notin S$, $a = 1, 2$,

$$\frac{1}{k} \sum_{y \in S} \|U(\tilde{M}_1^y)^{(T)}U^\dagger - M_1^y\| \leq E_{S,k}(\varepsilon) + \sqrt{\varepsilon}, \quad (30d)$$

$$\|U(\tilde{M}_1^y)^{(T)}U^\dagger - M_1^y\| \leq \text{d}_{\text{tr}}(U(\tilde{\varrho}_1^y)^{(T)}U^\dagger, \varrho_1^y) + \sqrt{\varepsilon}, \quad \text{for } y \notin S, \quad (30e)$$

where $(\cdot)^{(T)}$ is the transposition (with respect a fixed basis in \mathbb{C}^2) that may have to be applied to all experimental states and measurements simultaneously.

The reason for such a formulation of the theorem lays in the proof techniques that were used to derive it. Namely, we use results on the stability of the Cholesky factorization. Since the result employed by us (Ref. [38]) is valid for positive-definite matrices we cannot apply it to the Gram matrix of all of the vectors \mathbf{n}_a^x due to their linear dependence. As a result, our bounds depend on the particular choice of the set S of states whose Bloch vectors are linearly independent. In what follows, without the loss of generality we take $S = \{1, 2\}$ for $k = 2$ and $S = \{1, 2, 3\}$ for $k = 3$ in the proof. In practice, however, one should consider all subsets S of states that give non-singular Gram matrices having in mind a specific application.

Outline of the proof. – In what follows we will consider a particular subset $\{\mathbf{n}_1^x\}_{x \in S}$ of k linearly independent Bloch vectors. For simplify we will omit the subscript S whenever possible. The main idea of the proof is to use a Cholesky factors of $k \times k$ the Gram matrices Γ and $\tilde{\Gamma}$ of target $(\{\mathbf{n}_1^x\}_{x \in S})$ and experimental $(\{\tilde{\mathbf{n}}_1^x\}_{x \in S})$ Bloch vectors respectively. We then make use of the result of Ref. [38] on the stability of Cholesky factorization which, in simple terms, states that if Γ and $\tilde{\Gamma}$ are close to each other, so are their Cholesky factors L and \tilde{L} . Specifically, this result, which we quote bellow (see Theorem [Sun 1991]), sets an upper bound on the Frobenius norm of $\Delta L = \tilde{L} - L$ in terms of the Frobenius norm of the perturbation $\Delta \Gamma = \tilde{\Gamma} - \Gamma$. The Frobeniuous norm of the perturbation ΔL can be connected to the trace distance between states ϱ_1^x and the rotated states $\tilde{\varrho}_1^x$ in the selected subset S . On the other hand, the bound on $\|\Delta \Gamma\|_F$ can be estimated from our assumption: $|p(b|a, x, y) - \tilde{p}(b|a, x, y)| \leq \varepsilon$. This bound follows directly from the results of Theorem 1, but here we use an improved qubit stability bounds, which are given by Lemma 1. This, in turn, leads to stronger estimates for the norms of $\Delta \Gamma$ in Lemma 2 bellow. Combining all these results produces bounds in Eq. (30a).

In the next part of the proof we determine the trace distance between states ϱ_2^x and $\tilde{\varrho}_2^x$ for $x \in S$ based solely on the fact that $\varrho_2^x = \mathbb{1} - \varrho_1^x$ and $\tilde{\varrho}_2^x$ are close to $\mathbb{1} - \tilde{\varrho}_1^x$. This gives the bounds in Eq. (30b). For states $x \notin S$ we use the fact that they can be decomposed in the basis of the states in S . Since this linear decomposition can be different for target and experimental states, we use the result on stability of linear systems (see Theorem [Higham 2002], which we also quote below). The bounds for the states $x \notin S$ are given by Eq. (30c). Finally, we connect the bounds for the distance between the target and

experimental measurements and the distance between the corresponding states resulting in Eq. (30d) and (30e).

Theorem 2 (Quantitative formulation of robust self-testing for qubits). *Consider pure target qubit preparation states ϱ_a^x and target projective measurements \mathbf{M}^y , where $a = 1, 2$ and $x, y \in [n]$. Assume that $\varrho_a^x = M_a^x$ for all a, x and, furthermore, that experimental states $\tilde{\varrho}_a^x$ and measurements $\tilde{\mathbf{M}}^y$ act on Hilbert space of dimension at most d and generate statistics $\tilde{p}(b|a, x, y) = \text{tr}(\tilde{\varrho}_a^x \tilde{M}_b^y)$ such that $|\tilde{p}(b|a, x, y) - \text{tr}(\varrho_a^x M_b^y)| \leq \varepsilon$, for all a, b, x, y .*

Let $k \in \{2, 3\}$ be the cardinality of the maximal set of linearly independent Bloch vectors of states ϱ_1^x . Fix a set $S \subset [n]$ of k linearly independent vectors $\{\mathbf{n}_1^x\}_{x \in S}$, construct their Gram matrix Γ_S , and let L_S be a Cholesky factor of Γ_S (i.e., $\Gamma_S = L_S L_S^T$ and L_S is lower triangular). For every $x \in [n] \setminus S$ and both $a = 1, 2$ let $\mathbf{c}_S^{x,a}$, denote the coefficients of decomposition of \mathbf{n}_a^x as a linear combination of $\{\mathbf{n}_1^x\}_{x \in S}$. Finally, let us define three auxiliary functions

$$\begin{aligned} F_k(\varepsilon) &:= \sqrt{\varepsilon} \sqrt{4k(k-1)} \sqrt{1 + 2\sqrt{\varepsilon} + \frac{k+3}{k-1}\varepsilon}, \\ O_k(\varepsilon) &:= 2((k-1)\sqrt{\varepsilon} + (k+1)\varepsilon), \\ E_{S,k}(\varepsilon) &:= \frac{1}{2\sqrt{2}} \frac{\|\Gamma_S^{-1}\| F_k(\varepsilon)}{\sqrt{1 - \|\Gamma_S^{-1}\| O_k(\varepsilon)}} \min \left[\frac{\|\Gamma_S\|}{\|\Gamma_S\|_F}, \frac{\|L_S\|}{\sqrt{k}} \right]. \end{aligned} \quad (31)$$

Then, there is a region of $\varepsilon \in [0, \varepsilon_0)$ determined by

$$\|\Gamma_S^{-1}\| O_k(\varepsilon) \leq 1, \quad (32)$$

for which there exist a qubit unitary matrix U such that

$$\frac{1}{k} \sum_{x \in S} \text{tr}(U(\tilde{\varrho}_1^x)^{(T)} U^\dagger \varrho_1^x) \geq 1 - \frac{\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2} - E_{S,k}(\varepsilon)^2, \quad (33a)$$

$$\frac{1}{k} \sum_{x \in S} \text{tr}(U(\tilde{\varrho}_2^x)^{(T)} U^\dagger \varrho_2^x) \geq 1 - \frac{\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2} - (E_{S,k}(\varepsilon) + 2\sqrt{\varepsilon})^2, \quad (33b)$$

$$\text{tr}(U(\tilde{\varrho}_a^x)^{(T)} U^\dagger \varrho_a^x) \geq 1 - \frac{\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2} - \left(\text{d}_{\text{tr}}(U(\tilde{\varrho}_a^x)^{(T)} U^\dagger, \varrho_a^x) \right)^2, \quad \text{for } x \notin S, \quad a = 1, 2, \quad (33c)$$

$$\frac{1}{k} \sum_{y \in S} \text{tr}(U(\tilde{M}_1^y)^{(T)} U^\dagger M_1^y) \geq 1 - \frac{5}{2}\varepsilon - \left(\sqrt{2}E_{S,k}(\varepsilon) + \sqrt{\varepsilon} \right)^2, \quad (33d)$$

$$\text{tr}(U(\tilde{M}_1^y)^{(T)} U^\dagger M_1^y) \geq 1 - \varepsilon - \left(\text{d}_{\text{tr}}(U(\tilde{\varrho}_1^y)^{(T)} U^\dagger, \varrho_1^y) + \sqrt{\varepsilon} \right)^2, \quad \text{for } y \notin S, \quad (33e)$$

where $(\cdot)^{(T)}$ is the transposition (with respect a fixed basis in \mathbb{C}^2) that may have to be applied to all experimental states and measurements simultaneously. Note that in formulas (33c) and (30e) we have used trace distance $\text{d}_{\text{tr}}(U(\tilde{\varrho}_a^x)^{(T)} U^\dagger, \varrho_a^x)$ in order to simplify the resulting formulas. The bound on this quantity is given in Eq. (30c).

Remark. Results of Theorem 2 are obtained using a similar reasoning to that given in the outline of the proof of Theorem 3. However, the proof steps are supplemented by bounds connecting fidelity to the trace distance (for states) and operator norm (for measurement operators). The proof of Theorem 2 is given at the end of this part of the Appendix.

Before we proceed we state two theorems from the literature that are needed for our proof. First, we repeat the statement of the Theorem 1.4 from Ref. [38] for real-valued matrices. Second, we state a result concerning stability of systems of linear equation, which we borrow from Ref. [46] (Theorem 7.2). We changed the notation used in these papers in accordance with our paper.

Theorem (Sun 1991 - Stability of Cholesky factorisation). *Let Γ be an $k \times k$ positive definite matrix and $\Gamma = LL^T$ its Cholesky factorization. If $\Delta\Gamma$ is an $k \times k$ symmetric matrix satisfying*

$$\|\Gamma^{-1}\| \|\Delta\Gamma\| \leq 1, \quad (34)$$

then there is a unique Cholesky factorization

$$\Gamma + \Delta\Gamma = (L + \Delta L)(L + \Delta L)^T,$$

and

$$\|\Delta L\|_F \leq \frac{\|\Gamma^{-1}\| \|\Delta\Gamma\|_F}{\sqrt{2(1 - \|\Gamma^{-1}\| \|\Delta\Gamma\|)}} \min \left[\frac{\|L\|_F \|\Gamma\|}{\|\Gamma\|_F}, \|L\| \right]. \quad (35)$$

Remark. *This theorem dates back to 1991, and, of course there have been attempts to improve this result. However, the recent review [47] on this topic suggests that the bound given in the Theorem above is the most appropriate in our case (Remark 3.2, Ref. [47]).*

Theorem (Higham 2002 - Stability of systems of linear equations). *Let \mathbf{c} be a solution of a system of linear equations $\Gamma\mathbf{c} = \mathbf{g}$, where $\mathbf{c}, \mathbf{g} \in \mathbb{R}^k$ and $\Gamma \in \text{Mat}_{k \times k}(\mathbb{R})$. Let now $\tilde{\mathbf{c}}$ be a solution to $(\Gamma + \Delta\Gamma)\tilde{\mathbf{c}} = \mathbf{g} + \Delta\mathbf{g}$, where $\Delta\Gamma \in \text{Mat}_{k \times k}(\mathbb{R})$, $\Delta\mathbf{g} \in \mathbb{R}^k$. Assume that there exists $\delta' > 0$ such that $\|\Delta\Gamma\| \leq \delta' \|E\|$ and $|\Delta\mathbf{g}| \leq \delta' |\mathbf{f}|$ (for some $E \in \text{Mat}_{k \times k}(\mathbb{R})$, $\mathbf{f} \in \mathbb{R}^k$), and that $\delta' \|\Gamma^{-1}\| \|E\| \leq 1$. Then, we have*

$$\frac{|\mathbf{c} - \tilde{\mathbf{c}}|}{|\mathbf{c}|} \leq \frac{\delta'}{1 - \delta' \|\Gamma^{-1}\| \|E\|} \left(\frac{\|\Gamma^{-1}\| |\mathbf{f}|}{|\mathbf{c}|} + \|\Gamma^{-1}\| \|E\| \right). \quad (36)$$

Lemma 1. *Under the conditions of Theorem 1 qubit states and measurements $\{\varrho_a^x\}_{a,x}$, $\{\mathbf{M}^y\}_y$ and $\{\tilde{\varrho}_a^x\}_{a,x}$, $\{\tilde{\mathbf{M}}^y\}_y$ satisfy*

$$\begin{aligned} \|\tilde{\varrho}_a^x\| &\geq \frac{1 - 2\varepsilon}{1 - \varepsilon}, \quad \forall a, x, \\ |\text{tr}(\tilde{\varrho}_a^x \tilde{\varrho}_{a'}^{x'}) - \text{tr}(\varrho_a^x \varrho_{a'}^{x'})| &\leq \varepsilon + \sqrt{\varepsilon}, \\ |\text{tr}(\tilde{M}_b^y \tilde{M}_{b'}^{y'}) - \text{tr}(M_b^y M_{b'}^{y'})| &\leq \varepsilon + (1 + \varepsilon)\sqrt{\varepsilon}, \end{aligned}$$

$\forall x \neq x', a \neq a'$ and $\forall y \neq y', b \neq b'$ respectively, and $\varepsilon \leq \frac{1}{3}$.

Lemma 2. *Let Γ be the Gram matrix of Bloch vectors of target states $\{\varrho_1^x\}_{x \in S}$. Likewise, let $\tilde{\Gamma}$ be the Gram matrix of Bloch vectors of experimental states $\{\tilde{\varrho}_1^x\}_{x \in S}$. Let $|p(b|a, x, y) - \tilde{p}(b|a, x, y)| \leq \varepsilon$ for all b, a, x, y and let $k = |S|$ be the cardinality of the set S . Then, we have the following bounds on norms of $\Delta\Gamma = \tilde{\Gamma} - \Gamma$*

$$\begin{aligned} \|\Delta\Gamma\|_F &\leq F_k(\varepsilon), \\ \|\Delta\Gamma\| &\leq O_k(\varepsilon), \end{aligned} \quad (37)$$

where

$$\begin{aligned} F_k(\varepsilon) &= \sqrt{\varepsilon} \sqrt{4k(k-1)} \sqrt{1 + 2\sqrt{\varepsilon} + \frac{k+3}{k-1}\varepsilon}, \\ O_k(\varepsilon) &= 2((k-1)\sqrt{\varepsilon} + (k+1)\varepsilon). \end{aligned}$$

Proof of Theorem 3. Let $\tilde{\Gamma}$ be the Gram matrix of Bloch vectors $\{\tilde{\mathbf{n}}_1^x\}_{x \in S}$ of the experimental states for $x \in S$. Let us for now assume that the Cholesky factorization $\tilde{\Gamma} = \tilde{L}\tilde{L}^T$ exists and let $\Delta L = \tilde{L} - L$.

We start our *first part of the proof* regarding the states $x \in S$, $a = 1$ by connecting the square of the norm $\|\Delta L\|_F$ with trace distance between the states $\{\tilde{\varrho}_1^x\}_{x \in S}$ and $\{\varrho_1^x\}_{x \in S}$. From the sketch of the proof of Theorem 3 it follows that the Bloch vectors $\{\mathbf{n}_1^x\}_{x \in S}$ of the target states and the vectors $\{\mathbf{1}^x\}_{x \in S}$, transpose of which are rows of L , are connected via an orthogonal transformation, which we denote as O , i.e., $\mathbf{1}^x = O\mathbf{n}_1^x$, $x \in S$. Similarly, for the Bloch vectors $\{\tilde{\mathbf{n}}_1^x\}_{x \in S}$ of the experimental states and the vectors $\{\tilde{\mathbf{1}}^x\}_{x \in S}$ that form \tilde{L} there exists an orthogonal transformation \tilde{O} , such that $\tilde{\mathbf{1}}^x = \tilde{O}\tilde{\mathbf{n}}_1^x$, $x \in S$. Let us define states $\tau^x = \frac{1}{2}(\mathbb{1} + \mathbf{1}^x \cdot \boldsymbol{\sigma})$, and $\tilde{\tau}^x = \frac{1}{2}(\mathbb{1} + \tilde{\mathbf{1}}^x \cdot \boldsymbol{\sigma})$, $x \in S$. We know that there exist unitary transformations V and \tilde{V} such that: $\tau^x = V(\varrho_1^x)^{(T)}V^\dagger$, and $\tilde{\tau}^x = \tilde{V}(\tilde{\varrho}_1^x)^{(T)}\tilde{V}^\dagger$ for $x \in S$ (the optional transposition $(\cdot)^{(T)}$ is reserved for the cases $\det(O) = -1$ and $\det(\tilde{O}) = -1$). We have the following sequence of equalities valid for $x \in S$:

$$\begin{aligned} d_{\text{tr}}(\tilde{\tau}^x, \tau^x) &= d_{\text{tr}}(\tilde{V}(\tilde{\varrho}_1^x)^{(T)}\tilde{V}^\dagger, V(\varrho_1^x)^{(T)}V^\dagger) = d_{\text{tr}}(V^\dagger\tilde{V}(\tilde{\varrho}_1^x)^{(T)}\tilde{V}^\dagger V, (\varrho_1^x)^{(T)}) \\ &= d_{\text{tr}}(U(\tilde{\varrho}_1^x)^{(T)}U^\dagger, \varrho_1^x), \end{aligned} \quad (38)$$

where we used the invariance of trace distance under unitary evolution and transposition, and also denoted the resulting unitary $V^\dagger\tilde{V}$ as U . It is clear that one transposition in the final formula in Eq. (38) is enough as the case $\det(\tilde{O}) = \det(O) = -1$ is equivalent to having $\det(\tilde{O}) = \det(O) = 1$ and changing U to U^T .

We need to show now that this unitary U satisfies the claim of Theorem 2 for the states $x \in S$. For qubits, trace distance can be expressed directly via Euclidean distance between Bloch vectors:

$$d_{\text{tr}}(\tilde{\tau}^x, \tau^x) = \frac{1}{2}|\tilde{\mathbf{1}}^x - \mathbf{1}^x|. \quad (39)$$

Using this fact and Eq. (38), it is straightforward to derive the following upper bound:

$$\frac{1}{k} \sum_{x=1}^k d_{\text{tr}}(U(\tilde{\varrho}_1^x)^{(T)}U^\dagger, \varrho_1^x) \leq \frac{1}{2\sqrt{k}} \|\Delta L\|_F. \quad (40)$$

We now use Theorem [Sun 1991] (Ref. [38]) to upper-bound $\|\Delta L\|_F$. Specifically, we apply it to the Gram matrix of the Bloch vectors of states $\{\varrho_1^x\}_{x \in S}$. We have $\|L\|_F = \sqrt{k}$, since the target states are assumed to be pure. The direct substitution of the bound in Eq. (35) into Eq. (40) gives the bound in Eq. (30a), and also the condition in Eq. (29) when the Theorem 3 (and also Theorem 2) applies.

In the beginning of the proof we assumed that the Cholesky factorization $\tilde{\Gamma} = \tilde{L}\tilde{L}^T$ exists. The condition in Eq. (29) gives the sufficient condition for this to hold. We conclude this part of the proof by noting that a similar statement in terms of fidelity (Theorem 2, Eq. (33a)) follows from Eq. (40) by connecting fidelity to the trace distance as described in Eq. (54).

In the *second part of the proof* we derive upper bounds on the trace distances between the states corresponding to $x \in S$ and $a = 2$. As we will see, those can be connected to the bounds for the states with $x \in S$ and $a = 1$. Indeed, we can write the following for every $x \in S$:

$$\begin{aligned} 2d_{\text{tr}}(U\tilde{\varrho}_2^x U^\dagger, \varrho_2^x) &= \left\| \varrho_2^x - U\tilde{\varrho}_2^x U^\dagger \right\|_1 \leq \left\| \mathbb{1} - \varrho_1^x - U(\mathbb{1} - \tilde{\varrho}_1^x)U^\dagger \right\|_1 + \left\| U(\mathbb{1} - \tilde{\varrho}_1^x - \tilde{\varrho}_2^x)U^\dagger \right\|_1 \\ &\leq \left\| \varrho_1^x - U\tilde{\varrho}_1^x U^\dagger \right\|_1 + \left\| \mathbb{1} - \tilde{\varrho}_1^x + \tilde{M}_2^x \right\|_1 - \left\| \tilde{M}_2^x - \tilde{\varrho}_2^x \right\|_1 \\ &= 2d_{\text{tr}}(U\tilde{\varrho}_1^x U^\dagger, \varrho_1^x) + \sum_{a=1,2} \left\| \tilde{\varrho}_a^x - \tilde{M}_a^x \right\|_1. \end{aligned} \quad (41)$$

Exactly the same reasoning can be applied to upper bound the trace distance $d_{\text{tr}}(U(\tilde{\rho}_2^x)^T U^\dagger, \rho_2^x)$, in which case the transposition will propagate to $d_{\text{tr}}(U(\tilde{\rho}_1^x)^T U^\dagger, \rho_1^x)$ but would not affect the term $\sum_{a=1,2} \left\| \tilde{\rho}_a^x - \tilde{M}_a^x \right\|_1$ (as we can take $(\tilde{M}_a^x)^T$).

To finish the calculations we need to upper-bound the second summand in Eq. (41). We do it by writing $\left\| \tilde{\rho}_a^x - \tilde{M}_a^x \right\|_1 \leq 2 \left\| \tilde{\rho}_a^x - \tilde{M}_a^x \right\| \leq 2\sqrt{\varepsilon}$, $\forall a, x$, where the last inequality is proven in Lemma 1 (see Eq. (77)). From here, it is easy to obtain the bound in Eq. (30b). The reason why we do not simply apply the same reasoning to the states corresponding to $a = 2$ and $x \in S$ as for the states with $a = 1$ is because we want the isometry U to be the same for all of the states.

In *the third part of the proof* we derive the bounds for the preparations states corresponding to $x \notin S$ and both $a = 1, 2$. We start by reminding ourselves that the set of states $\{\rho_1^x\}_{x \in S}$ is assumed to be tomographically complete. Is equivalent to the assumption that the vectors $\{\mathbf{I}^x\}_{x \in S}$, defined above, are linearly independent and span \mathbb{R}^3 for $k = 3$, or the considered subspace for $k = 2$. The condition in Eq. (29) of the Theorem 3 (also stated in Eq. (34)) ensures that the same holds for the vectors $\{\tilde{\mathbf{I}}^x\}_{x \in S}$. If so, let us expand the Bloch vectors of the states $\rho_{a'}^{x'}$ and $U\tilde{\rho}_{a'}^{x'}U^\dagger$, $x' \notin S$ in terms of $\{\mathbf{I}^x\}_{x \in S}$ and $\{\tilde{\mathbf{I}}^x\}_{x \in S}$ respectively. Let us denote the coefficients of these linear expansions as $\{c_x^{a', x'}\}_{x \in S}$ and $\{\tilde{c}_x^{a', x'}\}_{x \in S}$ respectively. These coefficients will, of course, depend on x' and a' as well as on the choice of the set S . However, for simplicity of the derivations we are going omit the subscripts x' and a' until we present the final result. We will also assume, without loss of generality, that $S = \{1, 2, 3\}$ (or $S = \{1, 2\}$ for $k = 2$).

It is clear that the coefficients $\{c_x\}_{x \in S}$ and $\{\tilde{c}_x\}_{x \in S}$ satisfy the following respective systems of linear equations:

$$\Gamma \mathbf{c} = \mathbf{g}, \quad \tilde{\Gamma} \tilde{\mathbf{c}} = \tilde{\mathbf{g}}, \quad (42)$$

where $\mathbf{c} = (c_1, c_2, c_3)$, $\mathbf{g} = (2\text{tr}(\rho_{a'}^{x'} \rho_1^1) - 1, 2\text{tr}(\rho_{a'}^{x'} \rho_1^2) - 1, 2\text{tr}(\rho_{a'}^{x'} \rho_1^3) - 1)$ and analogously $\tilde{\mathbf{c}} = (\tilde{c}_1, \tilde{c}_2, \tilde{c}_3)$, $\tilde{\mathbf{g}} = (2\text{tr}(\tilde{\rho}_{a'}^{x'} \tilde{\rho}_1^1) - 1, 2\text{tr}(\tilde{\rho}_{a'}^{x'} \tilde{\rho}_1^2) - 1, 2\text{tr}(\tilde{\rho}_{a'}^{x'} \tilde{\rho}_1^3) - 1)$ whenever the cardinality k of the set S is 3. For $k = 2$ and $\mathbf{c}, \tilde{\mathbf{c}}, \mathbf{g}, \tilde{\mathbf{g}} \in \mathbb{R}^2$ their definition is analogous. We again omitted the subscripts a', x' for the vectors \mathbf{g} and $\tilde{\mathbf{g}}$ for simplicity. The matrices Γ and $\tilde{\Gamma}$ are still the moment matrices for the states $x \in S$, $a = 1$.

In complete analogy to Eq. (39) we can write:

$$d_{\text{tr}}(U(\tilde{\rho}_{a'}^{x'})^T U^\dagger, \rho_{a'}^{x'}) = \frac{1}{2} \left| \sum_{x \in S} (\tilde{c}_x \tilde{\mathbf{I}}^x - c_x \mathbf{I}^x) \right|. \quad (43)$$

We then can upper-bound the latter norm as follows:

$$\left| \sum_{x \in S} (\tilde{c}_x \tilde{\mathbf{I}}^x - c_x \mathbf{I}^x) \right| \leq \left| \sum_{x \in S} c_x (\tilde{\mathbf{I}}^x - \mathbf{I}^x) \right| + \left| \sum_{x \in S} (\tilde{c}_x - c_x) \tilde{\mathbf{I}}^x \right| \leq |\mathbf{c}| \|\Delta L\|_F + \sum_{x \in S} |\tilde{c}_x - c_x| |\tilde{\mathbf{I}}^x|. \quad (44)$$

In the equation above we used the following relation:

$$\left| \sum_{x \in S} c_x (\tilde{\mathbf{I}}^x - \mathbf{I}^x) \right| = \sqrt{\sum_{i=1}^3 \left(\sum_{x \in S} c_x (\tilde{l}_i^x - l_i^x) \right)^2} \leq \sqrt{\sum_{x' \in S} c_{x'}^2 \sum_{i=1}^3 \sum_{x \in S} (\tilde{l}_i^x - l_i^x)^2} = |\mathbf{c}| \|\Delta L\|_F. \quad (45)$$

Also, since the norms $|\tilde{\mathbf{I}}^x|$ can be upper-bounded by 1 for all x , the resulting upper bound can be simplified further to be $|\mathbf{c}| \|\Delta L\|_F + \sum_{x \in S} |\tilde{c}_x - c_x|$.

To estimate the deviation $|\tilde{c}_x - c_x|$ we apply Theorem [Higham 2002] by taking $\Delta\Gamma = \tilde{\Gamma} - \Gamma$ and $\Delta\mathbf{g} = \tilde{\mathbf{g}} - \mathbf{g}$. The bound on the operator norm $\|\Delta\Gamma\|$ is given by Lemma 2.

At the same time $|\Delta \mathbf{g}| = 2\sqrt{\sum_{x \in S} (\text{tr}(\rho_{a'}^{x'} \rho_1^x) - \text{tr}(\tilde{\rho}_{a'}^{x'} \tilde{\rho}_1^x))^2} \leq 2\sqrt{k}(\sqrt{\varepsilon} + \varepsilon)$. If we take $\delta' = 2((k-1)\sqrt{\varepsilon} + (k+1)\varepsilon)$, some matrix E with $\|E\| = 1$, vector \mathbf{f} such that $|\mathbf{f}| = \frac{\sqrt{k}}{k-1}$, we satisfy the conditions of the Theorem [Higham 2002] and get the following bound:

$$|\mathbf{c} - \tilde{\mathbf{c}}| \leq \left(|\mathbf{c}| + \frac{\sqrt{k}}{k-1} \right) \frac{\|\Gamma^{-1}\| \|\Delta\Gamma\|}{1 - \|\Gamma^{-1}\| \|\Delta\Gamma\|}, \quad (46)$$

As the final step we need to connect the bound in Eq. (44) with the one in Eq. (46) by the relation between 1-norm and the Euclidean norm in \mathbb{R}^k , which effectively adds a factor of \sqrt{k} to the bound in Eq. (46). Combining everything together we obtain the following:

$$\begin{aligned} d_{\text{tr}}(U(\tilde{\rho}_{a'}^{x'})^{(T)}U^\dagger, \rho_{a'}^{x'}) &\leq \frac{1}{2}|\mathbf{c}| \|\Delta L\|_F + \frac{\sqrt{k}}{2} \left(|\mathbf{c}| + \frac{\sqrt{k}}{k-1} \right) \frac{\|\Gamma^{-1}\| \|\Delta\Gamma\|}{1 - \|\Gamma^{-1}\| \|\Delta\Gamma\|}, \\ x' &\notin S, a' = 1, 2. \end{aligned} \quad (47)$$

The above bound is given in Eq. (30c) in terms of the quantity $E_{S,k}(\varepsilon)$ (Eq. (28)), where we took $\|\Delta L_S\|_F = 2\sqrt{k}E_{S,k}(\varepsilon)$ and brought back all the necessary subscripts. Again, using the relation in Eq. (55) we can derive bounds on the fidelity between the states ρ_a^x and $\tilde{\rho}_a^x$ for $x \notin S$.

In the fourth, final, part of the proof we derive the bounds for the measurements. We do it by connecting the distance between the experimental measurements and the distance between the corresponding states. Indeed, we can write the following:

$$\left\| U(\tilde{M}_1^y)^{(T)}U^\dagger - M_1^y \right\| \leq \left\| U(\tilde{\rho}_1^y)^{(T)}U^\dagger - \rho_1^y \right\| + \left\| \tilde{\rho}_1^y - \tilde{M}_1^y \right\|, \quad \forall y, \quad (48)$$

where we used the triangle inequality and the invariance of the norm $\left\| \tilde{\rho}_1^y - \tilde{M}_1^y \right\|$ under unitary transformations. Remembering that $\left\| \tilde{\rho}_1^y - \tilde{M}_1^y \right\| \leq \sqrt{\varepsilon}$ (Lemma 1, Eq. (77)) and $\left\| U(\tilde{\rho}_1^y)^{(T)}U^\dagger - \rho_1^y \right\| = d_{\text{tr}}(U(\tilde{\rho}_1^y)^{(T)}U^\dagger, \rho_1^y)$, we produce the bounds in Eqs. (30d,30e). \square

Finally, we proceed to the proof of Theorem 2, which gives a quantitative robustness analysis expressed in terms of average fidelity.

Proof of Theorem 2. Let us first state a useful identity between Frobenius distance and the fidelity, valid for an arbitrary pure states ρ and a Hermitian operator X :

$$\|X - \rho\|_F^2 = 1 + \text{tr}(X^2) - 2\text{tr}(X\rho). \quad (49)$$

We follow exactly the same steps that were given in the proof of Theorem 3. We assume that the reader is familiar with the notation introduced there. First, in order to derive the bound in Eq. (33a) we repeat the reasoning preceding Eq. (38). Analogously to the trace distance discussed there, we use the invariance of the fidelity under unitary transformations and transposition which gives us for all $x \in S = [k]$:

$$\text{tr}(\tilde{\tau}^x \tau^x) = \text{tr}(U(\tilde{\rho}_1^x)^{(T)}U^\dagger \rho_1^x), \quad (50)$$

where $U = V^\dagger \tilde{V}$. Using the above formula and standard algebra involving Pauli matrices we obtain:

$$\frac{1}{k} \sum_{x=1}^k \text{tr}(U(\tilde{\rho}_1^x)^{(T)}U^\dagger \rho_1^x) = \frac{1}{2} + \frac{1}{2k} \sum_{x=1}^k \mathbf{I}^x \cdot \tilde{\mathbf{I}}^x. \quad (51)$$

On the other hand, we have the following identity:

$$\|\Delta L\|_F^2 = \sum_{x=1}^k |\mathbf{I}^x|^2 + \sum_{x=1}^k |\tilde{\mathbf{I}}^x|^2 - 2 \sum_{x=1}^k \mathbf{I}^x \cdot \tilde{\mathbf{I}}^x, \quad (52)$$

where $|\cdot|$ is a standard Euclidean norm in \mathbb{R}^3 . Using the identity $|\tilde{\mathbf{I}}^x|^2 = 2\text{tr}((\tilde{\varrho}_1^x)^2) - 1$ and the fact that target states are pure (which implies $|\mathbf{I}^x| = 1$) we obtain:

$$\|\Delta L\|_F^2 = 2 \sum_{x=1}^k \text{tr}((\tilde{\varrho}_1^x)^2) - 2 \sum_{x=1}^k \mathbf{I}^x \cdot \tilde{\mathbf{I}}^x. \quad (53)$$

Inserting the above into Eq. (51) and using the fact that $\text{tr}(\tilde{\varrho}_1^x)^2 \geq 1 - \frac{2\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2}$ (this follows straightforwardly from Lemma 2), we finally obtain:

$$\frac{1}{k} \sum_{x=1}^k \text{tr}(U(\tilde{\varrho}_1^x)^{(T)}U^\dagger \varrho_1^x) \geq 1 - \frac{\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2} - \frac{\|\Delta L\|_F^2}{4k}. \quad (54)$$

We complete the proof of Eq. (33a) by again employing, exactly as before, Theorem [Sun 1991] (Ref. [38]) in order to upper-bound $\|\Delta L\|_F$.

Now, to derive Eq. (33b), which gives the bounds for the fidelity of states for $x \in [k]$ and $a = 2$, we can use the following inequality which can be derived from Eq. (49) and from the relation $\text{tr}(\tilde{\varrho}_1^x)^2 \geq 1 - \frac{2\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2}$:

$$\text{tr}(U(\tilde{\varrho}_a^x)^{(T)}U^\dagger \varrho_a^x) \geq 1 - \frac{\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2} - (\text{d}_{\text{tr}}(U(\tilde{\varrho}_a^x)^{(T)}U^\dagger, \varrho_a^x))^2. \quad (55)$$

From Eqs. (40,41) it follows that:

$$\begin{aligned} \frac{1}{k} \sum_{x=1}^k (\text{d}_{\text{tr}}(\tilde{\varrho}_2^x)^{(T)}U^\dagger, \varrho_2^x)^2 &\leq \frac{1}{k} \sum_{x=1}^k (\text{d}_{\text{tr}}(\tilde{\varrho}_1^x)^{(T)}U^\dagger, \varrho_1^x) + 2\sqrt{\varepsilon})^2 \\ &\leq \frac{\|\Delta L\|_F^2}{4k} + \frac{2\|\Delta L\|_F}{\sqrt{k}}\sqrt{\varepsilon} + 4\varepsilon = \left(\frac{\|\Delta L\|_F}{2\sqrt{k}} + 2\sqrt{\varepsilon} \right)^2, \end{aligned} \quad (56)$$

which gives the desired bound from Eq. (33b):

$$\frac{1}{k} \sum_{x=1}^k \text{tr}(U(\tilde{\varrho}_2^x)^{(T)}U^\dagger \varrho_2^x) \geq 1 - \frac{\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2} - \left(\frac{\|\Delta L\|_F}{2\sqrt{k}} + 2\sqrt{\varepsilon} \right)^2. \quad (57)$$

The proof of the remaining bounds in Eqs. (33c,33d,33e) is straightforward and follows directly from the formula in Eq. (49). In particular, in order to prove Eq. (33c) we set $X = \tilde{\varrho}_a^x$ and $\varrho = \varrho_a^x$ in Eq. (49), use the inequality $\text{tr}(\tilde{\varrho}_a^x)^2 \geq 1 - \frac{2\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2}$ and the relation $\|\sigma - \varrho\|_F^2 = 2\text{d}_{\text{tr}}(\varrho, \sigma)^2$, with latter being true for arbitrary qubits states ϱ and σ . Derivation of Eq. (33d) is completely analogous to the one of Eq. (33b). Finally, for Eq. (33e), the reasoning is again analogous to Eq. (33c), where, after setting $X = \tilde{M}_1^y$ and $\varrho = \varrho_1^y$, it is additionally necessary to use a simple lower bound: $\text{tr}((M_1^y)^2) \geq (1-\varepsilon)^2$, which follows from the conditions of Theorem 2. \square

C Alternative bounds from Procrustes

Lemma 3 (Robust self-testing for qubits from Procrustes). *Consider pure target qubit preparation states ϱ_a^x and target projective measurements \mathbf{M}^y , where $a = 1, 2$ and $x, y \in [n]$. Assume that $\varrho_a^x = M_a^x$ for all a, x and, furthermore, that experimental states $\tilde{\varrho}_a^x$ and measurements $\tilde{\mathbf{M}}^y$ act on Hilbert space of dimension at most d and generate statistics $\tilde{p}(b|a, x, y) = \text{tr}(\tilde{\varrho}_a^x \tilde{M}_b^y)$ such that $|\tilde{p}(b|a, x, y) - \text{tr}(\varrho_a^x M_b^y)| \leq \varepsilon$, for all a, b, x, y .*

Let $\{\varrho_i\}_{i=1}^m$ be a subset of m considered states among ϱ_a^x . Let L be a matrix whose rows are the Bloch vectors of states ϱ_i , $i \in [m]$, and let $k \in \{2, 3\}$ be its rank ($m \geq k$). Assume, without loss of generality, that for $k = 2$ the third component of the Bloch vectors is 0. In that case, truncate L to the first two columns. Let us define two auxiliary functions:

$$P_m(\varepsilon, L) = \begin{cases} \left\| L^\dagger \right\| F_m(\varepsilon) + \min \left[\frac{\|L^\dagger\| F_m(\varepsilon)}{\sqrt{1 - \|L^\dagger\|^2 F_m(\varepsilon)}}, \sqrt{\sqrt{k} F_m(\varepsilon)} \right], & \text{if } \|L^\dagger\| \sqrt{F_m(\varepsilon)} < 1 \\ \left\| L^\dagger \right\| F_m(\varepsilon) + \sqrt{\sqrt{k} F_m(\varepsilon)} & \text{otherwise,} \end{cases} \quad (58)$$

and

$$F_m(\varepsilon) = \sqrt{4m(m-1)\varepsilon \left(1 + 2\sqrt{\varepsilon} + \frac{m+3}{m-1}\varepsilon \right)}, \quad (59)$$

where $L^\dagger = (LL^T)^{-1}L^T$. There exist a unitary matrix U such that:

$$\frac{1}{m} \sum_{i=1}^m \text{tr}(\varrho_i U \tilde{\varrho}_i U^\dagger) \geq 1 - \frac{\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2} - \frac{1}{4m} P_m^2(\varepsilon, L). \quad (60)$$

In this section we derive alternative bounds for the fidelity between preparation states that follows from the bounds on the so-called orthogonal Procrustes problem [48]. The problem itself can be formulated as follows. Given two sets of vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ and $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$ in \mathbb{R}^d find an orthogonal transformation $O \in O(d)$ in \mathbb{R}^d that minimizes $\sum_{i=1}^m |\mathbf{x}_i - O\mathbf{y}_i|$. This problem has a clear relevance to our task. Indeed, if we take \mathbf{x}_i to be the Bloch vectors of the target qubit preparation states and \mathbf{y}_i the Bloch vectors of the experimental states, then minimization over $O(3)$ is the same as the problem of finding a unitary transformation that connects those qubit states. In Ref. [39] the bounds on Procrustes problem were derived. We give formulation of Theorem 1 from Ref. [39] below, where we change the notation according to our problem.

Theorem (Arias-Castro et.al. 2020 - A perturbation bound for Procrustes). *Given two tall matrices L and \tilde{L} of same size with L having full rank, and set $\delta^2 = \left\| \tilde{L}\tilde{L}^T - LL^T \right\|_F$. Then we have*

$$\min_{O \in O(k)} \left\| L - O\tilde{L} \right\|_F \leq \begin{cases} \left\| L^\dagger \right\| \delta^2 + \min \left[\frac{\|L^\dagger\| \delta^2}{\sqrt{1 - \|L^\dagger\|^2 \delta^2}}, k^{\frac{1}{4}} \delta \right], & \text{if } \|L^\dagger\| \delta < 1 \\ \left\| L^\dagger \right\| \delta^2 + k^{\frac{1}{4}} \delta & \text{otherwise.} \end{cases} \quad (61)$$

In the formulation of the above theorem L^\dagger stands for the Moore–Penrose inverse, which can be defined as $L^\dagger = (LL^T)^{-1}L^T$ for tall matrices of full rank.

Proof of Lemma 3. Unitizing the results of the Theorem [Arias-Castro et.al. 2020] is rather straightforward. Let L be a matrix which rows are the Bloch vectors of all target

preparation states $\{\varrho_a^x\}_{a,x}$. Let \tilde{L} in turn be the matrix of Bloch vectors of the experimental states $\{\tilde{\varrho}_a^x\}_{a,x}$. The matrices LL^T and $\tilde{L}\tilde{L}^T$ are then, of course, the full Gram matrices Γ and $\tilde{\Gamma}$. By "full" we mean that now we do not select a subset S of linearly independent vectors among the Bloch vectors of ϱ_a^x . Instead, Γ and $\tilde{\Gamma}$ are formed by all the considered states, which can still be a subset of the $2n$ states ($x \in [n], a \in [2]$). We will be using m to denote the number of the considered states, and a simple one-indexed set $\{\varrho_i\}_i$ to denote the states themselves. Estimating δ^2 from the formulation of the Theorem [Arias-Castro et.al. 2020] is a direct application of the bound on $\|\Delta\Gamma\|_F$ from Lemma 2, where now instead of k one should put the number m of the considered preparation states.

As for the left-hand side of Eq. (61), we can write the following:

$$\begin{aligned} \|L - O\tilde{L}\|_F^2 &= \text{tr}(LL^T) + \text{tr}(\tilde{L}\tilde{L}^T) - 2\text{tr}(L^T O\tilde{L}) \\ &\geq m + m - m \frac{4\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2} - 4 \sum_{i=1}^m \text{tr}(\varrho_i U \tilde{\varrho}_i U^\dagger) + 2m \end{aligned} \quad (62)$$

where we use the following identity:

$$\text{tr}(L^T O\tilde{L}) = \sum_{i=1}^m \mathbf{n}_i \cdot O\tilde{\mathbf{n}}_i = 2 \sum_{i=1}^m \text{tr}(\varrho_i U \tilde{\varrho}_i U^\dagger) - m, \quad (63)$$

and U is the unitary transformation in $SU(2)$ corresponding to the orthogonal transformation O of the Bloch vectors. The bound on the average fidelity between m target preparation states and the corresponding experimental states is then simply:

$$\frac{1}{m} \sum_{i=1}^m \text{tr}(\varrho_i U \tilde{\varrho}_i U^\dagger) \geq 1 - \frac{\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2} - \frac{1}{4m} P_m^2(\varepsilon, L) \quad (64)$$

where

$$P_m(\varepsilon, L) = \begin{cases} \left\| L^\dagger \right\| F_m(\varepsilon) + \min \left[\frac{\|L^\dagger\| F_m(\varepsilon)}{\sqrt{1 - \|L^\dagger\|^2 F_m(\varepsilon)}}, \sqrt{\sqrt{k} F_m(\varepsilon)} \right], & \text{if } \|L^\dagger\| \sqrt{F_m(\varepsilon)} < 1 \\ \left\| L^\dagger \right\| F_m(\varepsilon) + \sqrt{\sqrt{k} F_m(\varepsilon)} & \text{otherwise,} \end{cases} \quad (65)$$

and

$$F_m(\varepsilon) = \sqrt{4m(m-1)\varepsilon \left(1 + 2\sqrt{\varepsilon} + \frac{m+3}{m-1}\varepsilon \right)}. \quad (66)$$

□

D Examples

Below we provide detailed derivations of the results presented in Table 1.

The first example concerns $n = 2, 3$ MUBs in $d = 2$. Since for MUBs $\text{tr}(\varrho_a^x \varrho_{a'}^{x'}) = \frac{1}{2}$, for $x \neq x', \forall a, a'$, it follows that Γ is an identity matrix in \mathbb{R}^n ($n \in \{2, 3\}$). Hence, in Theorem 2 (see Appendix B) we should take $\|\Gamma_S\| = \|\Gamma_S^{-1}\| = \|L_S\| = 1$, and $\|\Gamma\|_F = \sqrt{n}$. The resulting bound is the average between expressions in Eq. (30a) and Eq. (30b) with the function $E_{S,k}(\varepsilon)$ being simply:

$$E_{S,k}(\varepsilon) = \frac{1}{2\sqrt{2n}} \frac{F_k(\varepsilon)}{\sqrt{1 - O_k(\varepsilon)}}. \quad (67)$$

The leading linear term is given in Table 1 for both $n = 2, 3$.

The second example is a little less straightforward. From the condition $\text{tr}(\varrho_1^1 \varrho_1^2) = \frac{1+\alpha}{2}$, $\alpha \in (-1, 1)$ we obtain that $\Gamma = \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix}$, and hence $\|\Gamma\| = 1 + |\alpha|$, $\|\Gamma^{-1}\| = \frac{1}{1-|\alpha|}$, and $\|\Gamma\|_F = \sqrt{2 + 2\alpha^2}$. The output L of the Cholesky factorization is $L = \begin{pmatrix} 1 & 0 \\ \alpha & \sqrt{1-\alpha^2} \end{pmatrix}$, which leads to $\|L\| = \sqrt{1 + |\alpha|}$. This also determines the minimum in Eq. (28) to be $\frac{\sqrt{1+|\alpha|}}{\sqrt{2}}$. Plugging this values in Eq. (28) gives:

$$E_{\{1,2\},2}(\varepsilon) = \frac{1}{4} \frac{\sqrt{1+|\alpha|} F_k(\varepsilon)}{\sqrt{1-|\alpha|} - O_k(\varepsilon)}. \quad (68)$$

The final bound is again the average of the bounds in Eq. (30a) and Eq. (30b). The first order in ε for this bound is given in Table 1. The applicability of the above bound is determined by the inequality $1 - |\alpha| - O_2(\varepsilon) \geq 0$. Importantly, the latter condition gives nonempty region $\varepsilon \in [0, \varepsilon_0)$ whenever $|\alpha| > 0$.

The third example is a trine ensemble of states $(\varrho_1^1, \varrho_2^2, \varrho_3^3)$, with $\varrho_x^x = \frac{1}{2} + \frac{1}{2} \mathbf{n}_x \cdot \boldsymbol{\sigma}$, $x = 1, 2, 3$, and where $\mathbf{n}_1 = (1, 0, 0)$, $\mathbf{n}_2 = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}, 0\right)$, and $\mathbf{n}_3 = \left(-\frac{1}{2}, -\frac{\sqrt{3}}{2}, 0\right)$. For this configuration of the preparation states the alternative robustness analysis via Procrustes (see Appendix C) gives better bounds. Given the vectors \mathbf{n}_i , $i = 1, 2, 3$ we can directly compute that $\|L^\dagger\| = \sqrt{\frac{2}{3}}$. Inserting this value to Lemma 3 produces the results given in Table 1.

The fourth example is the tetrahedron, with $\mathbf{n}_1 = (0, 0, 1)$, $\mathbf{n}_2 = \left(\sqrt{\frac{8}{9}}, 0, -\frac{1}{3}\right)$, $\mathbf{n}_3 = \left(-\sqrt{\frac{2}{9}}, \sqrt{\frac{2}{3}}, -\frac{1}{3}\right)$, $\mathbf{n}_4 = \left(-\sqrt{\frac{2}{9}}, -\sqrt{\frac{2}{3}}, -\frac{1}{3}\right)$, and $\varrho_x^x = \frac{1}{2} + \frac{1}{2} \mathbf{n}_x \cdot \boldsymbol{\sigma}$, $x = 1, 2, 3, 4$ as before. In this case, we also employ the bounds from Procrustes. For the above configuration of states we have that $\|L^\dagger\| = \frac{\sqrt{3}}{2}$, which leads to the results in Table 1.

E Proofs of auxiliary results

Proof of Lemma 1. First of all, we improve the bound on the norm of each of the experimental states. From $\|\tilde{M}_b^y\| \geq 1 - \varepsilon$, for $d = 2$ it follows immediately that the second (second largest) eigenvalue of each of the effects \tilde{M}_b^y cannot exceed ε . Hence, we can conclude that $\text{tr}(\tilde{M}_b^y) \leq 1 + \varepsilon, \forall y, b$.

Secondly, we can improve the bound on the norm of each of the experimental states. For that let us write the spectral decomposition of each experimental state and POVM effect:

$$\begin{aligned} \tilde{\varrho}_a^x &= \eta(\mathbb{1} - |\psi\rangle\langle\psi|) + (1 - \eta)|\psi\rangle\langle\psi| = \eta\mathbb{1} + (1 - 2\eta)|\psi\rangle\langle\psi| \\ \tilde{M}_a^x &= \lambda_0|\phi\rangle\langle\phi| + \lambda_1(\mathbb{1} - |\phi\rangle\langle\phi|), \end{aligned} \quad (69)$$

where we assume that $\lambda_0 \geq \lambda_1$. We omitted the indices x, a for $\eta, \lambda_0, \lambda_1$ and ψ, ϕ for simplicity. We can assume, without loss of generality, that $\eta \geq \frac{1}{2}$, and let us also assume for now that $|\langle\phi|\psi\rangle|^2 \leq \frac{1}{2}$. From the condition $\text{tr}(\tilde{\varrho}_a^x \tilde{M}_a^x) \geq 1 - \varepsilon$, it then follows that:

$$\eta(\lambda_0 + \lambda_1) + (1 - 2\eta)(\lambda_0|\langle\phi|\psi\rangle|^2 + \lambda_1(1 - |\langle\phi|\psi\rangle|^2)) \geq 1 - \varepsilon, \quad (70)$$

from where we can obtain a lower bound on η , namely:

$$\eta \geq \frac{1}{1 - 2|\langle\phi|\psi\rangle|^2} \left(\frac{1 - \varepsilon - \lambda_1}{\lambda_0 - \lambda_1} - \frac{1}{2} \right) + \frac{1}{2}. \quad (71)$$

The expression on the right-hand side of the above inequality is maximal when $|\langle\phi|\psi\rangle|^2 = 0$ and $\lambda_0 = 1, \lambda_1 = \varepsilon$, which returns the bound $\|\tilde{\varrho}_a^x\| \geq \frac{1-2\varepsilon}{1-\varepsilon}$.

Now let us return to our assumption $|\langle\phi|\psi\rangle|^2 \leq \frac{1}{2}$, for which the above bound is valid. We can upper-bound η by 1 in Eq. (70), which returns nontrivial upper bound on $|\langle\phi|\psi\rangle|^2$ that happens to be $1 - \frac{1-2\varepsilon}{1-\varepsilon}$. This function is below $\frac{1}{2}$ for $\varepsilon \leq \frac{1}{3}$, i.e., for $\varepsilon \leq \frac{1}{3}$ our newly-derived bound on $\|\tilde{\varrho}_a^x\|$ is valid. The region $\varepsilon \in [0, \frac{1}{3}]$ is significantly larger than the resulting region in which our self-testing argument are valid, so this assumption does not affect our final results.

Let us now try to improve the bounds for the overlaps. In the proof of Theorem 1 (see Eq. (7)) we have already established that:

$$|\text{tr}(\tilde{\varrho}_a^x \tilde{\varrho}_{a'}^{x'}) - \text{tr}(\varrho_a^x \varrho_{a'}^{x'})| \leq \varepsilon + \left\| \tilde{\varrho}_a^x - \tilde{M}_a^x \right\|, \quad \forall x \neq x', \forall a, a'. \quad (72)$$

Let us now refine the bound on $\left\| \tilde{\varrho}_a^x - \tilde{M}_a^x \right\|$. For simplicity, we present this result below for a pair of operators $\varrho = (1 - \eta)\mathbb{1} + (2\eta - 1)|\psi\rangle\langle\psi|$ and $M = \lambda_1\mathbb{1} + (\lambda_0 - \lambda_1)|\phi\rangle\langle\phi|$, that satisfy the conditions $\text{tr}(\varrho M) \geq 1 - \varepsilon$, and $1 - \varepsilon \leq \lambda_0 \leq 1, 0 \leq \lambda_1 \leq \varepsilon$. To compute the norm, we look for the eigenvector $|\xi\rangle$ of the operator $(\varrho - M)$, i.e., $(\varrho - M)|\xi\rangle = \Lambda|\xi\rangle$, corresponding to the maximal eigenvalue. We have the following quadratic equation for the eigenvalue Λ :

$$\Lambda^2 - \Lambda(1 - \lambda_0 - \lambda_1) + \eta(1 - \eta) - \lambda_1 + \lambda_0\lambda_1 - \eta(\lambda_0 - \lambda_1) + (2\eta - 1)(\lambda_0 - \lambda_1)|\langle\phi|\psi\rangle|^2 = 0. \quad (73)$$

The sum of the roots of this equation is equal to $1 - \lambda_0 - \lambda_1$. Since we know that $|1 - \lambda_0 - \lambda_1| \leq \varepsilon$, then either both roots are of the same sign, in which case the largest eigenvalue could only be ε , or they are of the opposite sign. In the latter case, it is evident that the absolute values of both roots are maximal whenever the free term in Eq. (73) is minimal. We then can upper-bound the solutions to Eq. (73) by lower-bounding the free term using the condition $\text{tr}(\varrho M) \geq 1 - \varepsilon$. Indeed, from $\text{tr}(\varrho M) \geq 1 - \varepsilon$ we infer immediately that:

$$\text{tr}(\varrho M) = \lambda_0 - \eta(\lambda_0 - \lambda_1) + (2\eta - 1)(\lambda_0 - \lambda_1)|\langle\phi|\psi\rangle|^2 \geq 1 - \varepsilon, \quad (74)$$

and thus we reduce Eq. (73) to:

$$\Lambda^2 - \Lambda(1 - \lambda_0 - \lambda_1) + \eta(1 - \eta) - \lambda_0 - \lambda_1 + \lambda_0\lambda_1 + 1 - \varepsilon = 0. \quad (75)$$

Using the same argument we can set $\eta = 1$, because $\eta(1 - \eta) \geq 0$. The positive root of Eq. (75) is equal to:

$$\Lambda = \frac{1 - \lambda_0 - \lambda_1}{2} + \sqrt{\left(\frac{1 - \lambda_0 - \lambda_1}{2}\right)^2 - (1 - \lambda_0)(1 - \lambda_1) + \varepsilon}. \quad (76)$$

It is easy to check that the above expression does not have any local maxima w.r.t. λ_0, λ_1 on the domain $1 - \varepsilon \leq \lambda_0 \leq 1, 0 \leq \lambda_1 \leq \varepsilon$, whenever $\varepsilon < \frac{1}{2}$, which we assume to be the case. Thus, we conclude that the maximal value of Λ corresponds to the boundary of the region of (λ_0, λ_1) . By considering this boundary we find that this maximal value corresponds to the case of $\lambda_0 = 1$ and $\lambda_1 = 0$ which yields $\Lambda = \sqrt{\varepsilon}$. From the above argument we finally conclude that:

$$\left\| \tilde{\varrho}_a^x - \tilde{M}_a^x \right\| \leq \sqrt{\varepsilon}, \quad \forall a, \quad (77)$$

which completes the proof for the state overlaps. From Eq. (77) and the fact that $\text{tr}(\tilde{M}_b^y) \leq 1 + \varepsilon$, it is easy to obtain the improved bound on the overlaps between measurement effect. \square

Proof of Lemma 2. Let us start by deriving the bound on the Frobenius norm $\|\Delta\Gamma\|_F$:

$$\begin{aligned}\|\Delta\Gamma\|_F^2 &= \sum_{x=1}^k (|\Gamma_{x,x} - \tilde{\Gamma}_{x,x}|^2 + \sum_{x' \neq x} |\Gamma_{x,x'} - \tilde{\Gamma}_{x,x'}|^2) \\ &= 4 \sum_{x=1}^k \left((1 - \text{tr}(\tilde{\varrho}_1^x))^2 + \sum_{x' \neq x} |\text{tr}(\varrho_1^x \varrho_1^{x'}) - \text{tr}(\tilde{\varrho}_1^x \tilde{\varrho}_1^{x'})|^2 \right).\end{aligned}\quad (78)$$

We have already established the bound on $|\text{tr}(\varrho_1^x \varrho_1^{x'}) - \text{tr}(\tilde{\varrho}_1^x \tilde{\varrho}_1^{x'})|$ in Lemma 1. The bound on $(1 - \text{tr}(\tilde{\varrho}_1^x))^2$ can be obtained from the bound on the norm of each $\tilde{\varrho}_1^x$. Namely, from the condition $\|\tilde{\varrho}_a^x\| \geq \frac{1-2\varepsilon}{1-\varepsilon}$ we can immediately conclude that:

$$1 - \text{tr}(\tilde{\varrho}_1^x)^2 \leq \frac{2\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2}.\quad (79)$$

From here, it is easy to get to the final bound on $\|\Delta\Gamma\|_F^2$, which reads:

$$\|\Delta\Gamma\|_F^2 \leq 4k(k-1)(\varepsilon + \sqrt{\varepsilon})^2 + \frac{16k\varepsilon^2(1-2\varepsilon)^2}{(1-\varepsilon)^4} \leq 4k(k-1)\varepsilon \left(1 + 2\sqrt{\varepsilon} + \frac{k+3}{k-1}\varepsilon \right),\quad (80)$$

where we made some approximations to simplify the result.

Now, let us derive the bound on $\|\Delta\Gamma\|$. In principle, we know that $\|\Delta\Gamma\| \leq \|\Delta\Gamma\|_F$, but we can derive a better bound based on the fact that the diagonal entries of $\Delta\Gamma$ are much less than the off-diagonal entries.

First of all, due to the triangle inequality, we can write $\|\Delta\Gamma\| \leq \|\text{diag}(\Delta\Gamma)\| + \|\text{offdiag}(\Delta\Gamma)\|$, where we split $\Delta\Gamma$ on the diagonal and off-diagonal parts. The first term, $\|\text{diag}(\Delta\Gamma)\|$, can be easily bounded as follows:

$$\|\text{diag}(\Delta\Gamma)\| = 2 \max_x (1 - \text{tr}(\tilde{\varrho}_1^x)^2) \leq \frac{4\varepsilon(1-2\varepsilon)}{(1-\varepsilon)^2} \leq 4\varepsilon.\quad (81)$$

As for the off-diagonal part, we give the proof for two cases $k=2$ and $k=3$ separately. For $k=2$, $\|\text{offdiag}(\Delta\Gamma)\| = 2|\text{tr}(\varrho_1^1 \varrho_1^2) - \text{tr}(\tilde{\varrho}_1^1 \tilde{\varrho}_1^2)| \leq 2\sqrt{\varepsilon} + 2\varepsilon$, where we used the results of Lemma 1. As for $k=3$, we will need some intermediate result, namely the following relation:

$$\|A\| \leq \sqrt{\frac{k-1}{k}} \|A\|_F,\quad (82)$$

where k is the size of the matrix A with $\text{tr}(A) = 0$. We give a proof of this below.

Let us assume that $\{\lambda_i\}_{i=1}^k$ are the eigenvalues of matrix A , hence we know that $\sum_{i=1}^k \lambda_i = 0$. Let λ_1 be the largest eigenvalue, i.e., the operator norm of A , if $\lambda_1 \geq 0$. If we wish to maximize the Frobenius norm of A for fixed λ_1 , the following lower-bound has to be satisfied:

$$\|A\|_F^2 = \lambda_1^2 + \sum_{i=2}^k \lambda_i^2 \geq \lambda_1^2 + \frac{1}{k-1} \left(\sum_{i=2}^k |\lambda_i| \right)^2 \geq \lambda_1^2 + \frac{1}{k-1} \left(\sum_{i=2}^k \lambda_i \right)^2 = \lambda_1^2 \frac{k}{k-1},\quad (83)$$

which proves the bound in Eq. (82). Using the above result, we obtain the following bound:

$$\|\text{offdiag}(\Delta\Gamma)\| \leq \sqrt{\frac{2}{3}} \|\text{offdiag}(\Delta\Gamma)\|_F = 2\sqrt{\frac{2}{3}} \sqrt{\sum_{x \neq x'} |\text{tr}(\varrho_1^x \varrho_1^{x'}) - \text{tr}(\tilde{\varrho}_1^x \tilde{\varrho}_1^{x'})|^2} \leq 4(\sqrt{\varepsilon} + \varepsilon), \quad (84)$$

which completes out proof. \square

References

- [1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
- [2] C. L. Degen, F. Reinhard, and P. Cappellaro, Quantum sensing, *Reviews of Modern Physics* **89**, 035002 (2017).
- [3] I. M. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, *Reviews of Modern Physics* **86**, 153 (2014).
- [4] A. Montanaro, Quantum algorithms: an overview, *npj Quantum Information* **2**, 1 (2016).
- [5] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature* **549**, 195 (2017).
- [6] V. Dunjko and H. J. Briegel, Machine learning & artificial intelligence in the quantum domain: a review of recent progress, *Reports on Progress in Physics* **81**, 074001 (2018).
- [7] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, *Nature Reviews Physics* **2**, 382 (2020).
- [8] D. Mayers and A. Yao, Self testing quantum apparatus, *arXiv preprint quant-ph/0307205* (2003).
- [9] I. Šupić and J. Bowles, Self-testing of quantum systems: a review, *Quantum* **4**, 337 (2020).
- [10] J. S. Bell, On the Einstein Podolsky Rosen paradox, *Physics Physique Fizika* **1**, 195 (1964).
- [11] S.-L. Chen, C. Budroni, Y.-C. Liang, and Y.-N. Chen, Natural framework for device-independent quantification of quantum steerability, measurement incompatibility, and self-testing, *Physical Review Letters* **116**, 240401 (2016).
- [12] A. Coladangelo, K. T. Goh, and V. Scarani, All pure bipartite entangled states can be self-tested, *Nature Communications* **8**, 15485 (2017).
- [13] J. Bowles, I. Šupić, D. Cavalcanti, and A. Acín, Device-independent entanglement certification of all entangled states, *Physical Review Letters* **121**, 180503 (2018).

- [14] S. Popescu and D. Rohrlich, Which states violate Bell’s inequality maximally? *Physics Letters A* **169**, 411 (1992).
- [15] J. Ahrens, P. Badziąg, M. Pawłowski, M. Żukowski, and M. Bourennane, Experimental tests of classical and quantum dimensionality, *Physical Review Letters* **112**, 140401 (2014).
- [16] J. B. Brask, A. Martin, W. Esposito, R. Houlmann, J. Bowles, H. Zbinden, and N. Brunner, Megahertz-rate semi-device-independent quantum random number generators based on unambiguous state discrimination, *Physical Review Applied* **7**, 054018 (2017).
- [17] E. A. Aguilar, M. Farkas, D. Martínez, M. Alvarado, J. Cariñe, G. B. Xavier, J. F. Barra, G. Cañas, M. Pawłowski, and G. Lima, Certifying an irreducible 1024-dimensional photonic state using refined dimension witnesses, *Physical Review Letters* **120**, 230503 (2018).
- [18] H. Anwer, S. Muhammad, W. Cherifi, N. Miklin, A. Tavakoli, and M. Bourennane, Experimental characterization of unsharp qubit observables and sequential measurement incompatibility via quantum random access codes, *Physical Review Letters* **125**, 080403 (2020).
- [19] M. Pawłowski and N. Brunner, Semi-device-independent security of one-way quantum key distribution, *Physical Review A* **84**, 010302 (2011).
- [20] T. Van Himbeeck, E. Woodhead, N. J. Cerf, R. García-Patrón, and S. Pironio, Semi-device-independent framework based on natural physical assumptions, *Quantum* **1**, 33 (2017).
- [21] R. Chaves, J. B. Brask, and N. Brunner, Device-independent tests of entropy, *Physical Review Letters* **115**, 110501 (2015).
- [22] T. Fritz, Quantum correlations in the temporal Clauser–Horne–Shimony–Holt (CHSH) scenario, *New Journal of Physics* **12**, 083055 (2010).
- [23] A. Tavakoli, J. Kaniewski, T. Vértesi, D. Rosset, and N. Brunner, Self-testing quantum states and measurements in the prepare-and-measure scenario, *Physical Review A* **98**, 062307 (2018).
- [24] M. Farkas and J. Kaniewski, Self-testing mutually unbiased bases in the prepare-and-measure scenario, *Physical Review A* **99**, 032316 (2019).
- [25] P. Mironowicz and M. Pawłowski, Experimentally feasible semi-device-independent certification of four-outcome positive-operator-valued measurements, *Physical Review A* **100**, 030301 (2019).
- [26] A. Tavakoli, M. Smania, T. Vértesi, N. Brunner, and M. Bourennane, Self-testing non-projective quantum measurements, *Science Advances* **6**, eaaw6664 (2020).
- [27] A. Tavakoli, D. Rosset, and M.-O. Renou, Enabling computation of correlation bounds for finite-dimensional quantum systems via symmetrization, *Physical Review Letters* **122**, 070501 (2019).
- [28] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, *Nature Communications* **8**, 1 (2017).
- [29] C. H. Bennett and G. Brassard, Quantum cryptography: Public-key distribution and coin tossing, *Theoretical Computer Science* **560**, 7 (2014).
- [30] C. H. Bennett, Quantum cryptography using any two nonorthogonal states, *Physical Review Letters* **68**, 3121 (1992).
- [31] E. Woodhead, C. C. W. Lim, and S. Pironio, Semi-device-independent QKD based on BB84 and a CHSH-type estimation, In *Theory of Quantum Computation, Com-*

- munication, and Cryptography*, vol 7582, pages 107–115, Springer, Berlin, Heidelberg (2013).
- [32] J. I. de Vicente, Shared randomness and device-independent dimension witnessing, *Physical Review A* **95**, 012340 (2017).
 - [33] M. Navascués and S. Popescu, How energy conservation limits our measurements, *Physical Review Letters* **112**, 140502 (2014).
 - [34] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, Symmetric informationally complete quantum measurements, *Journal of Mathematical Physics* **45**, 2171 (2004).
 - [35] S. Brierley, S. Weigert, and I. Bengtsson, All mutually unbiased bases in dimensions two to five, *arXiv preprint arXiv:0907.4097* (2009).
 - [36] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, Cambridge University Press, USA, 10th edition (2011).
 - [37] V. Bargmann, Note on Wigner’s theorem on symmetry operations, *Journal of Mathematical Physics* **5**, 862 (1964).
 - [38] J.-G. Sun, Perturbation bounds for the Cholesky and QR factorizations, *BIT Numerical Mathematics* **31**, 341 (1991).
 - [39] E. Arias-Castro, A. Javanmard, and B. Pelletier, Perturbation bounds for Procrustes, classical scaling, and trilateration, with applications to manifold learning, *Journal of Machine Learning Research* **21**, 15 (2020).
 - [40] H. Bechmann-Pasquinucci and N. Gisin, Intermediate states in quantum cryptography and Bell inequalities, *Physical Review A* **67**, 062310 (2003).
 - [41] G. M. D’Ariano, P. L. Presti, and P. Perinotti, Classical randomness in quantum measurements, *Journal of Physics A: Mathematical and General* **38**, 5979 (2005).
 - [42] M. Oszmaniec and T. Biswas, Operational relevance of resource theories of quantum measurements, *Quantum* **3**, 133 (2019).
 - [43] R. Takagi and B. Regula, General resource theories in quantum mechanics and beyond: operational characterization via discrimination tasks, *Physical Review X* **9**, 031053 (2019).
 - [44] R. Uola, T. Kraft, J. Shang, X.-D. Yu, and O. Gühne, Quantifying quantum resources with conic programming, *Physical Review Letters* **122**, 130404 (2019).
 - [45] A. Tavakoli, Semi-device-independent certification of independent quantum state and measurement devices, *Physical Review Letters* **125**, 150503 (2020).
 - [46] N. J. Higham, *Accuracy and stability of numerical algorithms*, Society for Industrial and Applied Mathematics, USA, 2nd edition (2002).
 - [47] X.-W. Chang and D. Stehlé, Rigorous perturbation bounds of some matrix factorizations, *SIAM Journal on Matrix Analysis and Applications* **31**, 2841 (2010).
 - [48] J. R. Hurley and R. B. Cattell, The Procrustes program: Producing direct rotation to test a hypothesized factor structure, *Behavioral Science* **7**, 258 (1962).