Self-testing quantum states and measurements in the prepare-and-measure scenario

Armin Tavakoli,[1] Jędrzej Kaniewski,[2] Tamás Vértesi,[3] Denis Rosset,[4,5] and Nicolas Brunner[1]

[1]*Département de Physique Appliquée, Université de Genève, CH-1211 Genève, Switzerland*

[2]*QMATH, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark*

[3]*Institute for Nuclear Research, Hungarian Academy of Sciences, P.O. Box 51, 4001 Debrecen, Hungary*

[4]*Perimeter Institute for Theoretical Physics, 31 Caroline St. N, Waterloo, Ontario, Canada N2L 2Y5*

[5]*Institute for Quantum Optics and Quantum Information (IQOQI), Boltzmangasse 3, 1090 Vienna, Austria*

(Received 30 January 2018; revised manuscript received 11 September 2018; published 6 December 2018)

The goal of self-testing is to characterize an *a priori* unknown quantum system based solely on measurement statistics, i.e., using an uncharacterized measurement device. Here we develop self-testing methods for quantum prepare-and-measure experiments, thus not necessarily relying on entanglement and/or violation of a Bell inequality. We present noise-robust techniques for self-testing sets of quantum states and measurements, assuming an upper bound on the Hilbert space dimension. We discuss in detail the case of a $2 \to 1$ random access code with qubits, for which we provide analytically optimal self-tests. The simplicity and noise robustness of our methods should make them directly applicable to experiments.

DOI: 10.1103/PhysRevA.98.062307

I. INTRODUCTION

Predicting the results of measurements performed on a given physical system has traditionally been the main concern of physics. However, with the advent of device-independent quantum information processing [1–3], the opposite question has become relevant. More specifically, given an initially unknown system and an uncharacterized measurement device, what can be inferred about the physics of the experiment based solely on the observed measurement statistics? Despite the apparent generality of this question, certain cases do allow for a precise characterization of the system. This is referred to as self-testing [4,5].

The possibility to self-test quantum states and measurements usually relies on quantum nonlocality. Consider two distant observers performing local measurements on a shared quantum state. When the resulting statistics leads to violation of a Bell inequality [6], it is necessarily the case that the shared quantum state is entangled and, moreover, that the local quantum measurements are incompatible; see, e.g., Ref. [7]. Furthermore, for specific Bell inequalities, maximal violation (i.e., the largest possible value in quantum theory) implies that the quantum state and the measurements can be uniquely identified (up to local isometries). For instance, a maximal violation of the Clauser-Horne-Shimony-Holt (CHSH) Bell inequality [8] implies maximally incompatible measurements (two anticommuting Pauli observables) and a shared maximally entangled two-qubit state [9–12]. More recently, it has been demonstrated that all bipartite pure entangled states can be self-tested [13], as well as certain multipartite entangled states [14–16]. Another important progress is the development of self-testing methods robust to noise [17–23]. For instance, given a certain level of violation of a Bell inequality (but not necessarily maximal), the fidelity between the initially unknown state and a given target state can be lower bounded.

Self-testing thus offers promising perspectives for the certification of quantum systems in experiments (see, e.g., Ref. [24]), as well as for device-independent quantum information protocols [25]. It is therefore natural to ask whether the concept of self-testing can be applied to more general quantum experiments, beyond those based on entanglement and nonlocality.

In the present work, we develop self-testing methods tailored to the prepare-and-measure scenario. This covers a broad class of experiments, where quantum communication schemes [e.g., the BB84 quantum key distribution (QKD) protocol] are prominent examples. In this setting, a preparation device initially prepares a quantum system in different possible states. The system is then transmitted to a measurement device, which performs different possible measurements on it. While it is still possible in this case to characterize certain physical properties of the system based only on statistics, this requires in general an assumption on the devices. One possibility, which we will follow here, is to assume that the set of quantum states and measurements admit a full description in a Hilbert space of given dimension [26–28]. Intuitively this means that the amount of information communicated from the preparation device to the measurement device is assumed to be upper bounded. Such a scenario considering quantum systems of fixed dimension, but otherwise uncharacterized, is referred to as semi-device-independent, and opens interesting possibilities for quantum information processing [29–33].

Here we demonstrate techniques for robustly self-testing sets of prepared quantum states, as well as sets of quantum measurements. These methods allow one to (i) assess the compatibility of given sets of preparations and measurements with the observed statistics and (ii) lower bound the average fidelity between the unknown preparations (measurements) and a set of ideal quantum states (measurements). We discuss in detail a simple prepare-and-measure scenario, namely the $2 \to 1$ random access code (RAC). This allows us to provide analytically optimal self-tests for a pair of anticommuting Pauli observables, and for a set of four qubit states corresponding

to the eigenstates of two anticommuting Pauli observables. We then generalize these results to other prepare-and-measure scenarios. The simplicity and robustness of our methods should make them directly applicable to experiments. We conclude with a number of open questions.

II. SCENARIO

We consider a quantum prepare-and-measure experiment. Upon receiving input x, a preparation device emits a physical system in a quantum state ρ_x. The system is then transmitted to a measurement device, which, upon receiving an input y, performs a quantum measurement returning an outcome b. Formally, the measurement is described by a set of positive operators M_y^b, that equal identity when summed over b. Importantly both the specific states ρ_x and measurements M_y^b are *a priori* unknown to the observer. The statistics of the experiment is then given by $P(b|x, y) = \mathrm{tr}(\rho_x M_y^b)$. In this setting, any possible probability distribution can be obtained, given that the prepared states ρ_x can be taken in a sufficiently large Hilbert space. This is however no longer the case when we limit the Hilbert space dimension; specifically we impose that $\rho_x \in \mathcal{L}(\mathbb{C}^d)$ for some given $d < |x|$ (where $|x|$ denotes the number of possible inputs x). In this case, limits on the set of possible distributions can be captured via inequalities of the form

$$\mathcal{A} = \sum_{x,y,b} \alpha_{xyb} P(b|x, y) \leqslant Q_d, \tag{1}$$

where α_{xyb} are real coefficients. These "dimension witnesses" allow one to place device-independent lower bounds on the dimension of the quantum system [26].

Subsequently, one can ask what the limitations are on the set of distributions $P(b|x, y)$ given that the preparations admit a classical d-dimensional representation, i.e., there exists a d-dimensional basis such that all states ρ_x are diagonal in this basis. We denote by C_d the maximal value of the quantity $\mathcal{A}$ in this case. Interestingly, for well-chosen quantities $\mathcal{A}$, one finds that $C_d < Q_d$. Thus, for a given system dimension d, quantum systems outperform classical ones, in the sense that certain quantum distributions cannot be reproduced classically [26]. This quantum advantage can be viewed as the origin for the possibility of developing self-testing methods for the prepare-and-measure scenario, in analogy to Bell inequality violation being the root for self-testing entangled states.

In the following we present robust self-testing techniques based on specific dimension witnesses $\mathcal{A}$. Based only on the value of $\mathcal{A}$, which is directly accessible from the experiment statistics, we characterize the (initially unknown) prepared states and measurements. In particular, when the maximal value of the witness is obtained, i.e., $\mathcal{A} = Q_d$, then a specific set of pure states $\rho_x = |\psi_x\rangle\langle\psi_x|$ and a specific set of projective measurements M_y^b must have been used (up to a unitary). Moreover, when a nonmaximal value $\mathcal{A} < Q_d$ is obtained, the compatibility of given sets of preparations and measurements can be assessed. Finally, one can efficiently lower bound the fidelity between the prepared states and measurements and the ideal (or target) states and measurements leading to $\mathcal{A} = Q_d$.

Note that a recent series of works followed a related though conceptually different approach, based on hypothesis testing [34–36]. This method does however not allow for self-testing.

III. 2 → 1 RANDOM ACCESS CODE

We discuss in detail a simple prepare-and-measure experiment. This involves four possible preparations, denoted by $x = (x_0, x_1)$ (where $x_j \in \{0, 1\}$), and two possible binary measurements, $y \in \{0, 1\}$ and $b \in \{0, 1\}$. The score is given by

$$\mathcal{A}_2 = \frac{1}{8} \sum_{x_0,x_1,y} P(b = x_y|x_0, x_1, y). \tag{2}$$

This means that, upon receiving input y, the measurement device should return the output $b = x_y$, i.e., the yth bit of the input bit-string x received by the preparation device. Hence the name of a $2 \to 1$ RAC [37–39]. Note that all inputs are assumed to be chosen uniformly at random. Indeed, this task is nontrivial only when $d < 4$; here we will consider the case $d = 2$, i.e., qubits. In this case, one finds the tight bounds $C_2 = 3/4$ and $Q_2 = (1 + 1/\sqrt{2})/2 \approx 0.85$ [37]. The classical bound C_2 can be obtained by simply always sending the bit x_0. The quantum bound Q_2 is obtained via the following "ideal" strategy. The four qubit preparations correspond to the pure states

$$\rho_{jj}^{\text{ideal}} = \frac{\mathbb{1} + (-1)^j \sigma_x}{2}, \quad \rho_{j\bar{j}}^{\text{ideal}} = \frac{\mathbb{1} + (-1)^j \sigma_z}{2} \tag{3}$$

for $j \in \{0, 1\}$ and $\bar{j} = 1 - j$. These are simply the eigenstates of the Pauli observables σ_x and σ_z. Next, the measurements are projective and given by two anticommuting Pauli observables

$$M_y^{\text{ideal}} = \left(M_y^0\right)^{\text{ideal}} - \left(M_y^1\right)^{\text{ideal}} = \frac{\sigma_x + (-1)^y \sigma_z}{\sqrt{2}}. \tag{4}$$

These qubit preparations and measurements represent the ideal situation, where the maximal value $\mathcal{A}_2 = Q_2$ is achieved. In the following we will determine what restrictions apply to the possible preparations and measurements, given that a particular value of $\mathcal{A}_2$ is observed. In particular, when the maximal value $\mathcal{A}_2 = Q_2$ is attained, both the states and the measurements must be the ideal ones as given above (up to a unitary).

IV. SELF-TESTING PREPARATIONS

Here we find restrictions on the set of prepared states given an observed value of $\mathcal{A}_2$. For convenience, we write the qubit preparations as $\rho_{x_0x_1} = (\mathbb{1} + \vec{m}_{x_0x_1} \cdot \vec{\sigma})/2$, where $\vec{m}_{x_0x_1}$ denotes the Bloch vector (satisfying $|\vec{m}_{x_0x_1}| \leqslant 1$) and $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ denotes the vector of Pauli matrices.

The first step consists in reexpressing

$$\begin{aligned} \mathcal{A}_2 &= \frac{1}{2} + \frac{1}{8} \sum_y \mathrm{tr}\left(M_y^0 V_y\right) \\ &\leqslant \frac{1}{2} + \frac{1}{8} \sum_y \sqrt{\mathrm{tr}\left(M_y^0 V_y^2\right) \mathrm{tr}\left(M_y^0\right)}, \end{aligned} \tag{5}$$

where $V_y = \sum_{x_0,x_1}(-1)^{x_y}\rho_{x_0x_1}$. In the second step we used that for a positive semidefinite O and a Hermitian operator R, it holds that $|\operatorname{tr}(OR)|^2 \leqslant \operatorname{tr}(OR^2)\operatorname{tr}(O)$ [23]. Without loss of generality, we can restrict ourselves to extremal qubit measurements, which are here projective rank-one operators. Consequently, we have that $\operatorname{tr}(M_y^0) = 1$. Next, we obtain $V_y^2 = \frac{1}{2}[\beta + (-1)^y\alpha]\mathbb{1}$, where $\beta = \frac{1}{2}\sum_{x_0,x_1}|\vec{m}_{x_0x_1}|^2 - \vec{m}_{00}\cdot\vec{m}_{11} - \vec{m}_{01}\cdot\vec{m}_{10}$ and $\alpha = (\vec{m}_{00} - \vec{m}_{11})\cdot(\vec{m}_{01} - \vec{m}_{10})$. Finally, we find that Eq. (5) reduces to

$$\mathcal{A}_2 \leqslant \frac{1}{2} + \frac{1}{8\sqrt{2}}[\sqrt{\beta+\alpha} + \sqrt{\beta-\alpha}]. \tag{6}$$

This provides a tight self-test of the prepared states (in terms of their Bloch vectors), for any given value of $\mathcal{A}_2$. Let us start with the case $\mathcal{A}_2 = Q_2$. Since $\sqrt{\beta+\alpha} + \sqrt{\beta-\alpha} = \sqrt{2\beta + 2\sqrt{\beta^2-\alpha^2}}$, we see that Eq. (6) is maximal iff $\alpha = 0$ and β is maximal. This turns out to be achievable. In order to maximize β, we need (i) $\forall x_0x_1 : |\vec{m}_{x_0x_1}| = 1$, i.e., that all preparations are pure states, and (ii) that $\vec{m}_{00}\cdot\vec{m}_{11} = \vec{m}_{01}\cdot\vec{m}_{10} = -1$, i.e., the states correspond to (pairwise) antipodal Bloch vectors. We define $\vec{r}_0 = \vec{m}_{00} = -\vec{m}_{11}$ and $\vec{r}_1 = \vec{m}_{01} = -\vec{m}_{10}$. Consequently, we find $\alpha = 4\vec{r}_0\cdot\vec{r}_1$. Therefore, in order to have $\alpha = 0$, we must choose $\vec{r}_0\cdot\vec{r}_1 = 0$. This implies that the right-hand side of Eq. (6) is upper bounded by Q_2. Therefore, we conclude that when observing maximal value $\mathcal{A}_2 = Q_2$, the set of four prepared states must be equivalent (up to a unitary rotation) to the set of four ideal states; we note that this was also shown in Ref. [40] in the context of QKD.

More generally, for any value $\mathcal{A}_2$, one can find a set of preparations (and corresponding measurements) such that the inequality (6) is saturated; see Appendix A. For the case of classical preparations (i.e., diagonal in a given basis), the Bloch vectors can simply be replaced by numbers $m_{x_0x_1} \in [-1, 1]$, and we get $\mathcal{A}_2 \leqslant C_2$.

V. SELF-TESTING MEASUREMENTS

Let us now consider self-testing of measurements. Using that $M_y = M_y^0 - M_y^1$, we write

$$\mathcal{A}_2 \leqslant \frac{1}{2} + \frac{1}{16}\sum_{x_0,x_1}\lambda_{\max}[(-1)^{x_0}M_0 + (-1)^{x_1}M_1], \tag{7}$$

where $\lambda_{\max}[X]$ is the largest eigenvalue of the (Hermitian) operator X. Since the upper bound corresponds to choosing the optimal preparations for a fixed pair of observables, it simply quantifies the optimal performance achievable using these observables. If M_0 and M_1 are qubit observables the upper bound can be evaluated exactly (see Appendix A) to give

$$\mathcal{A}_2 \leqslant \tfrac{1}{2} + \tfrac{1}{16}(\sqrt{2\mu + 2\nu - \eta_+^2} + \sqrt{2\mu - 2\nu - \eta_-^2}), \tag{8}$$

where $\mu = \operatorname{tr}(M_0^2 + M_1^2)$, $\nu = \operatorname{tr}\{M_0, M_1\}$, and $\eta_\pm = \operatorname{tr}(M_0 \pm M_1)$. The right-hand side reaches the optimal value Q_2 iff $\mu = 4$, $\eta_\pm = 0$, and $\nu = 0$, which implies anticommuting projective observables (i.e., projective measurement operators). In other words, observing $\mathcal{A}_2 = Q_2$ implies that the measurements are unitarily equivalent to the ideal ones. Moreover, note that inequality (8) is tight; for any value of $\mathcal{A}_2$ one can find measurements (and corresponding states) such that inequality is saturated (see Appendix A). It follows that any pair of projective, rank-one observables that is incompatible ($|\nu| < 4$) can lead to $\mathcal{A}_2 > C_2$.

VI. ROBUST SELF-TESTING OF THE PREPARATIONS

We now discuss the problem of characterizing the fidelity between the realized preparations and the ideal ones. This will allow us to quantify the distance of the prepared states with respect to the ideal ones. Again, we want to develop self-testing methods which are based only on the value of $\mathcal{A}_2$.

More formally, given an arbitrary set of preparations, we define the average fidelity with the ideal preparations to be $S(\{\rho_{x_0x_1}\}) = \max_\Lambda \sum_{x_0,x_1} F(\rho_{x_0x_1}^{\text{ideal}}, \Lambda[\rho_{x_0x_1}])/4$, where Λ is a quantum channel, i.e., a completely positive trace-preserving map. Here the fidelities $F(\rho,\sigma) = \operatorname{tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})$ simplify to $F(\rho_{x_0x_1}^{\text{ideal}}, \Lambda[\rho_{x_0x_1}]) = \operatorname{tr}(\Lambda[\rho_{x_0x_1}]\rho_{x_0x_1}^{\text{ideal}})$, as the $\rho_{x_0x_1}^{\text{ideal}}$ are pure states. We derive lower bounds on the smallest possible value of S given a value of $\mathcal{A}_2$, i.e.,

$$\mathcal{F}(\mathcal{A}_2) = \min_{\{\rho_{x_0x_1}\}\in R(\mathcal{A}_2)} S[\{\rho_{x_0x_1}\}]. \tag{9}$$

Note that this involves a minimization over all sets of four preparations $R(\mathcal{A}_2)$ that are compatible with an observed value $\mathcal{A}_2$.

In order to lower bound $\mathcal{F}$, we use an approach inspired by Ref. [22]. From Eq. (7), we have $\mathcal{A}_2 = \frac{1}{2} + \sum_{x_0,x_1}\operatorname{tr}(W_{x_0x_1}\rho_{x_0x_1})$, where $W_{x_0x_1} = \frac{1}{16}\sum_y(-1)^{x_y}M_y$. We define operators corresponding to some suitably chosen channel acting on the ideal preparations:

$$K_{x_0x_1}(M_0, M_1) = \Lambda^\dagger(M_0, M_1)\big[\rho_{x_0x_1}^{\text{ideal}}\big], \tag{10}$$

where $\Lambda^\dagger$ is the channel dual to Λ. We aim to construct operator inequalities of the form

$$K_{x_0x_1}(M_0, M_1) \geqslant sW_{x_0x_1} + t_{x_0x_1}(M_0, M_1)\mathbb{1}, \tag{11}$$

for all inputs (x_0, x_1), for any given measurements, where s and $t_{x_0x_1}(M_0, M_1)$ are real coefficients. Finding such inequalities, as well as a suitable channel Λ, allows us to lower bound S as follows:

$$\begin{aligned} S &\geqslant \frac{1}{4}\sum_{x_0,x_1}\operatorname{tr}\left(K_{x_0x_1}\rho_{x_0x_1}\right) \geqslant \frac{s}{4}\sum_{x_0,x_1}\operatorname{tr}\left(W_{x_0x_1}\rho_{x_0x_1}\right) \\ &\quad + \frac{1}{4}\sum_{x_0,x_1}t_{x_0x_1} = \frac{s}{4}(\mathcal{A}_2 - 1/2) + \frac{1}{4}\sum_{x_0,x_1}t_{x_0x_1}. \end{aligned} \tag{12}$$

Applying a minimization over M_0 and M_1 to the right-hand side, the above inequality becomes valid for all preparations. Consequently,

$$\mathcal{F}(\mathcal{A}_2) \geqslant \frac{s}{4}(\mathcal{A}_2 - 1/2) + t \equiv L(\mathcal{A}_2), \tag{13}$$

where $t \equiv 1/4\,\min_{M_0,M_1}\sum_{x_0,x_1}t_{x_0x_1}(M_0, M_1)$. In Appendix B, we construct explicitly the channel and derive an operator inequality leading to a lower bound, given by $s = 4(1+\sqrt{2})$ and $t = (2-\sqrt{2})/4$.

This provides a robust self-testing for the preparations. A maximal value $\mathcal{A}_2 = Q_2$ implies $\mathcal{F} = 1$, i.e., the preparations must be the ideal ones (up to a unitary). For $\mathcal{A}_2 = C_2$, i.e.,

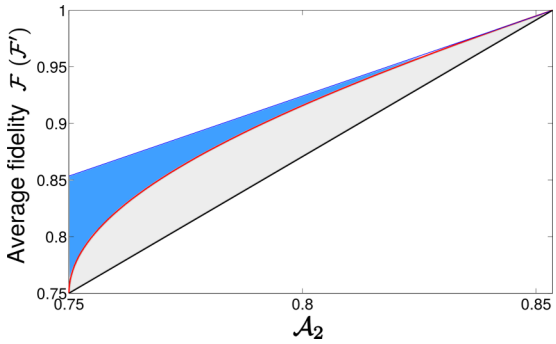

FIG. 1. Average fidelity $\mathcal{F}$ ($\mathcal{F}'$) for prepared states (measurements), as a function of the observed value of $\mathcal{A}_2$. The black line is our analytical lower bound of Eq. (13). The blue region is accessible via single qubit strategies without shared randomness, as confirmed by strong numerical evidence (see Appendixes). When allowing for shared randomness between the devices, the accessible region (obtained by taking the convex hull of the blue region) now also includes the gray area, and our analytic lower bound is tight in general.

a maximal value given a set of classical states, we get that $\mathcal{F} \geqslant 3/4$. This bound can be attained via the set of pure states $\rho_{x_0x_1} = [\mathbb{1} + (-1)^{x_0x_1}\sigma_z]/2$ (diagonal in the same basis, hence classical), combined with the measurements $M_0 = M_1 = \sigma_z$. Therefore, we see that our bound $\mathcal{F}(\mathcal{A}_2) \geqslant L(\mathcal{A}_2)$ is optimal, as far as linear inequalities are concerned (see Fig. 1). It is then interesting to consider the intermediate region $C_2 < \mathcal{A}_2 < Q_2$. First, focusing on strategies involving a single set of states and measurements, we observe numerically that the linear bound $\mathcal{F}(\mathcal{A}_2) \geqslant L(\mathcal{A}_2)$ cannot be saturated anymore, and conjecture the form of optimal states and measurements; see red curve in Fig. 1 and Appendix C for details. Second, allowing for shared randomness between the preparation and measurement device (such that convex combinations of qubit strategies are now possible), the linear bound becomes tight, a direct consequence of the linearity of $\mathcal{F}$ and $\mathcal{A}_2$ in terms of the states and measurements.

VII. ROBUST SELF-TESTING OF THE MEASUREMENTS

Similarly, we can quantify the average fidelity of the measurements with respect to the ideal ones: $S'(\{M_y^b\}) = \max_\Lambda \sum_{y,b} F((M_y^b)^{\text{ideal}}, \Lambda[M_y^b])/4$, where Λ must be a unital channel (i.e., mapping the identity to itself), in order to ensure that measurements are mapped to measurements. In analogy with the case of preparations, our goal is to lower bound the following quantity:

$$\mathcal{F}'(\mathcal{A}_2) = \min_{\{M_y^b\} \in R'(\mathcal{A}_2)} S'\left(\{M_y^b\}\right), \tag{14}$$

where $R'(\mathcal{A}_2)$ represents all sets of measurements compatible with a certain value of $\mathcal{A}_2$.

We first rewrite $\mathcal{A}_2 = \sum_{y,b} \text{tr}(M_y^b Z_{yb})$, where $Z_{yb} = \frac{1}{8}\sum_{x_0,x_1} \rho_{x_0x_1}\delta_{b,x_y}$. Next, we construct operator inequalities

$$K_{yb}(\{\rho_{x_0x_1}\}) \geqslant sZ_{yb} + t_y(\{\rho_{x_0x_1}\})\mathbb{1}, \tag{15}$$

given the unital channel $K_{yb} = \Lambda^\dagger[(M_y^b)^{\text{ideal}}]$. Similar to the case of preparations, strong operator inequalities can be derived by choosing carefully the channel; all details are given in Appendix D. Finally, this leads to a lower bound on the average fidelity

$$\mathcal{F}'(\mathcal{A}_2) \geqslant \min_{\{\rho_{x_0x_1}\}} \frac{1}{4}\sum_{y,b} \text{tr}\left(K_{yb}M_y^b\right) \geqslant L(\mathcal{A}_2). \tag{16}$$

That is, we find that $\mathcal{F}'$ can be lower bounded by a linear expression in terms of $\mathcal{A}_2$, which turns out to be the same as for the case of preparations.

This provides a robust self-test for the measurements. Observing $\mathcal{A}_2 = Q_2$ implies that $\mathcal{F}' = 1$; hence the measurements are equivalent to the ideal ones (up to a unitary). For $\mathcal{A}_2 = C_2$, we have that $\mathcal{F} \geqslant 3/4$. This lower bound can be attained by choosing $M_0 = \sigma_z$ and $M_1 = \mathbb{1}$, with the states $\rho_{00} = \rho_{01} = (\mathbb{1} + \sigma_z)/2$ and $\rho_{10} = \rho_{11} = (\mathbb{1} - \sigma_z)/2$. For $C_2 < \mathcal{A}_2 < Q_2$, we find numerically that the inequality (16) cannot be saturated using a single set of measurements and states (see Fig. 1). Details, in particular a conjecture for the form of the optimal measurements, are given in Appendix C. Similarly as for the case of states, when allowing for convex combinations of qubit strategies, our linear bound is tight.

VIII. GENERALIZATIONS

The above results can be generalized in several directions. First, a generalization of the $2 \to 1$ RAC enables self-testing of any pair of incompatible Pauli observables (see Appendix E). Secondly, we consider the $N \to 1$ RAC, where the preparation device receives as input an N-bit string $x = (x_1, \ldots, x_N)$ and the measurement device gets input $y \in \{1, \ldots, N\}$. The average score is then given by

$$\mathcal{A}_N = \frac{1}{N2^N}\sum_{x,y} P(b = x_y|x, y). \tag{17}$$

The methods discussed above (for $N = 2$) can be generalized and lead to self-testing conditions for states and measurements; details are given in Appendix F. The case of $N = 3$ is of particular interest. Here, the best possible score with qubits is $\mathcal{A}_3 = (1 + 1/\sqrt{3})/2$; see, e.g., Ref. [39]. In this case, our self-testing conditions can certify that (i) the eight prepared states correspond to Bloch vectors forming a cube on the Bloch sphere and (ii) the measurements correspond to three mutually unbiased bases (i.e., three pairwise anticommuting Pauli observables). Thirdly, we self-test qutrit preparations and projective measurements in the $2 \to 1$ RAC (see Appendix G).

Finally, we present a numerical method for robust self-testing of preparations applicable in scenarios beyond RACs. The method is based on semidefinite programing and combines (i) the swap method [21] used for self-testing in Bell scenarios with (ii) the hierarchy of finite-dimensional quantum correlations [41–43]. The idea is to first construct a swap operator, based on the measurement operators, which maps the state of the preparation onto an ancilla. The average fidelity between the ancilla and the ideal states can then be expressed in terms of strings of products of measurement operators and the extracted states. The last step is to miminize this average fidelity over all quantum realizations that are compatible with a given witness value, using the hierarchy of Refs. [41–43].

Although typically returning suboptimal bounds on $\mathcal{F}$, this method is widely applicable. In Appendix H, we describe in detail the methodology and apply to two examples, including the $2 \to 1$ RAC.

IX. OUTLOOK

We presented methods for self-testing quantum states and measurements in the prepare-and-measure scenario. These techniques demonstrate strong robustness to noise, and should therefore be directly amenable to experiments, providing useful certification techniques in a semi-device-independent setting. Moreover, these ideas should find applications in quantum communications. Our methods apply to the states and measurements used in QKD (e.g., in BB84), as well as in semi-device-independent QKD and randomness generation protocols [29–33].

It would be interesting to develop robust self-testing techniques for more general scenarios, e.g., for higher-dimensional quantum systems. Another direction would be to consider scenarios beyond prepare and measure, for instance, adding between the preparation and measurement devices a transformation device [44,45] and self-testing the latter.

Finally, while we have focused here on self-testing based on an assumption on the dimension, one could develop methods based on different assumptions, such as a bound on the mean energy [46], the overlap [47], or the entropy [48].

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation (Starting grant DIAQ, QSIT, and Early Postdoc Mobility fellowship No. P2GEP2_162060), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Action ROSETTA (Grant No. 749316), the European Research Council (Grant No. 337603), the Danish Council for Independent Research (Sapere Aude), VILLUM FONDEN via the QMATH Centre of Excellence (Grant No. 10059), and the National Research, Development and Innovation Office NKFIH (Grants No. K111734 and No. KH125096).

APPENDIX A: SELF-TESTING RELATIONS FOR PREPARATIONS AND MEASUREMENTS

In this section we provide a simple example of preparations that saturate the compatibility bound for $\mathcal{A}_2$ given in the main text. Moreover, we derive the upper bound for compatibility of measurements given in the main text.

First, we consider the case of preparations. Consider preparations such that ρ_{00} and ρ_{11}, and ρ_{01} and ρ_{10} correspond to antipodal Bloch vectors with a relative angle θ, the maximal quantum value of $\mathcal{A}_2$, is obtained from

$$\mathcal{A}_2 = \frac{1}{2} + \frac{1}{8}\sum_y \lambda_{\max}[V_y], \tag{A1}$$

where $V_y = \sum_{x_0,x_1}(-1)^{x_y}\rho_{x_0x_1}$. We represent the preparations on the Bloch sphere as $\rho_{x_0x_1} = 1/2(\mathbb{1} + \vec{m}_{x_0x_1}\cdot\vec{\sigma})$, where $\vec{m}_{00} = [\cos(\theta/2), 0, \sin(\theta/2)]$ and $\vec{m}_{01} = [\cos(\theta/2), 0, -\sin(\theta/2)]$, with $\vec{m}_{11} = -\vec{m}_{00}$ and $\vec{m}_{10} = -\vec{m}_{01}$. This gives $V_0 = 2\cos(\theta/2)\sigma_x$ and $V_1 = 2\sin(\theta/2)\sigma_z$. The respective largest eigenvalues are $\lambda_{\max}[V_0] = 2\cos(\theta/2)$ and $\lambda_{\max}[V_1] = 2\sin(\theta/2)$, leading to

$$\mathcal{A}_2 = \frac{1}{2} + \frac{1}{4\sqrt{2}}[\sqrt{1+\cos\theta} + \sqrt{1-\cos\theta}]. \tag{A2}$$

It is straightforward to see that this achieves the upper bound in the main text; indeed the above choice of preparations leads to $\beta = 4$ and $\alpha = 4\cos\theta$.

In order to derive the upper bound on $\mathcal{A}_2$ for compatibility of measurements in the main text we evaluate

$$\sum_{x_0,x_1}\lambda_{\max}[(-1)^{x_0}M_0 + (-1)^{x_1}M_1] \tag{A3}$$

for arbitrary qubit observables M_0, M_1. We take advantage of the fact that

$$\lambda_{\max}[T] + \lambda_{\max}[-T] = \lambda_{\max}[T] - \lambda_{\min}[T], \tag{A4}$$

which for a 2×2 matrix can be evaluated analytically. More specifically, if T is a 2×2 Hermitian matrix with eigenvalues $\lambda_0 \geqslant \lambda_1$, let

$$\chi := \operatorname{tr} T = \lambda_0 + \lambda_1,$$
$$\zeta := \operatorname{tr} T^2 = \lambda_0^2 + \lambda_1^2$$

and then

$$\lambda_0 - \lambda_1 = \sqrt{2\zeta - \chi^2}. \tag{A5}$$

Evaluating this expression for $T = M_0 \pm M_1$ gives the desired upper bound.

APPENDIX B: OPERATOR INEQUALITIES FOR ROBUST SELF-TESTING OF PREPARATIONS

In this section we provide a detailed derivation of the lower bound on the average fidelity $\mathcal{F}(\mathcal{A}_2)$. For a real constant $s > 0$, to be chosen later, consider for each pair (x_0, x_1) the operator $K_{x_0x_1} - sW_{x_0x_1}$, where $W_{x_0x_1} = \frac{1}{16}\sum_y(-1)^{x_y}M_y$ and $K_{x_0x_1} = \Lambda^\dagger[\rho^{\text{ideal}}_{x_0x_1}]$, for some channel Λ. Suppose now that $t_{x_0x_1} \in \mathbb{R}$ is a lower bound on its eigenvalues, or, equivalently, that the operator inequality

$$K_{x_0x_1} \geqslant s\, W_{x_0x_1} + t_{x_0x_1}\,\mathbb{1} \tag{B1}$$

holds. Then, computing the trace of this inequality with $\rho_{x_0x_1}$ and averaging over inputs leads to

$$S \geqslant \frac{1}{4}\sum_{x_0,x_1} F\left(\rho^{\text{ideal}}_{x_0x_1}, \Lambda[\rho_{x_0x_1}]\right) \geqslant \frac{s}{4}\left(\mathcal{A}_2 - \frac{1}{2}\right) + t,$$
$$t \equiv \frac{1}{4}\sum_{x_0,x_1} t_{x_0x_1}, \tag{B2}$$

where the first inequality holds because S is defined as maximization over all possible channels, and the Λ used there is one possible choice. In turn, if (B1) holds *as an operator inequality*, it is valid for any set of preparations $\{\rho_{x_0x_1}\}$, and thus $\mathcal{F}(\mathcal{A}_2) \geqslant \frac{s}{4}(\mathcal{A}_2 - \frac{1}{2}) + t$. Note that (B1) has a dependence on M_0, M_1 through $W_{x_0x_1}$. If (B1) holds for a particular choice of measurement operators M_0, M_1, then the bound on $\mathcal{F}(\mathcal{A}_2)$ holds for all preparations, for that particular choice of M_0, M_1. However, if (B1) holds for *all* possible

M_0, M_1, then the bound on $\mathcal{F}(\mathcal{A}_2)$ is valid for all quantum setups and is thus a robust self-testing inequality. To derive the appropriate constants s and $t_{x_0x_1}$, we first allow $t_{x_0x_1}$ and Λ to have a dependence on M_0 and M_1. We then minimize over M_0 and M_1 the constants $t_{x_0x_1}$, for a suitable choice of s, such that, at the end, Eq. (B1) holds regardless of the choice of measurement operators.

We choose a dephasing channel of the form

$$\Lambda_\theta(\rho) = \frac{1+c(\theta)}{2}\rho + \frac{1-c(\theta)}{2}\Gamma(\theta)\rho\Gamma(\theta), \tag{B3}$$

where for $0 \leqslant \theta \leqslant \pi/4$ we use $\Gamma = \sigma_x$, while for $\pi/4 < \theta \leqslant \pi/2$ we use $\Gamma = \sigma_z$. The function $c(\theta) \in [-1, 1]$ will be specified later.

In the interval $0 \leqslant \theta \leqslant \pi/4$, the action of the channel leads to

$$\begin{aligned} K_{00} &= \frac{\mathbb{1}+\sigma_x}{2}, \quad K_{01} = \frac{\mathbb{1}+c(\theta)\sigma_z}{2}, \\ K_{10} &= \frac{\mathbb{1}-c(\theta)\sigma_z}{2}, \quad K_{11} = \frac{\mathbb{1}-\sigma_x}{2}, \end{aligned} \tag{B4}$$

whereas in the interval $\pi/4 < \theta \leqslant \pi/2$, we have

$$\begin{aligned} K_{00} &= \frac{\mathbb{1}+c(\theta)\sigma_x}{2}, \quad K_{01} = \frac{\mathbb{1}+\sigma_z}{2}, \\ K_{10} &= \frac{\mathbb{1}-\sigma_z}{2}, \quad K_{11} = \frac{\mathbb{1}-c(\theta)\sigma_x}{2}. \end{aligned} \tag{B5}$$

As discussed in the main text, for any given set of preparations, the optimal measurements are projective and rank-one. Furthermore, any two such measurements can be represented on an equator of the Bloch sphere. Due to the freedom of setting the reference frame, we can without loss of generality represent the two measurements in the xz plane, i.e.,

$$\begin{aligned} M_0 &= \cos\theta\ \sigma_x + \sin\theta\ \sigma_z, \\ M_1 &= \cos\theta\ \sigma_x - \sin\theta\ \sigma_z. \end{aligned}$$

We can therefore write $W_{x_0x_1}$ as

$$\begin{aligned} W_{00} &= \tfrac{1}{8}\cos\theta\sigma_x, \quad W_{01} = \tfrac{1}{8}\sin\theta\sigma_z, \\ W_{10} &= -\tfrac{1}{8}\sin\theta\sigma_z, \quad W_{11} = -\tfrac{1}{8}\cos\theta\sigma_x. \end{aligned} \tag{B6}$$

We can reduce the number of operator inequalities (B1) by exploiting the apparent symmetries in the expressions for $W_{x_0x_1}$ and $K_{x_0x_1}$: we restrict ourselves so that $t_o \equiv t_{01} = t_{10}$ and $t_e \equiv t_{00} = t_{11}$. Thus we have to consider two operator inequalities in each interval $\theta \in [0, \pi/4]$ and $\theta \in (\pi/4, \pi/2]$. In the first interval, the two operator inequalities are

$$\begin{aligned} \frac{1+\sigma_x}{2} - \frac{s}{8}\cos\theta\sigma_x - t_e\mathbb{1} &\geqslant 0, \\ \frac{1+c(\theta)\sigma_z}{2} - \frac{s}{8}\sin\theta\sigma_z - t_o\mathbb{1} &\geqslant 0. \end{aligned} \tag{B7}$$

In the second interval, the two operator inequalities are

$$\begin{aligned} \frac{1+c(\theta)\sigma_x}{2} - \frac{s}{8}\cos\theta\sigma_x - t_e\mathbb{1} &\geqslant 0, \\ \frac{1+\sigma_z}{2} - \frac{s}{8}\sin\theta\sigma_z - t_o\mathbb{1} &\geqslant 0. \end{aligned} \tag{B8}$$

We now focus on the former interval. Solving the two inequalities for t_o and t_e we obtain

$$t_e \leqslant 1 - \frac{s}{8}\cos\theta, \quad t_o \leqslant \frac{1}{8}[4 + 4c(\theta) - s\ \sin\theta], \tag{B9}$$

$$t_e \leqslant \frac{s}{8}\cos\theta, \quad t_o \leqslant \frac{1}{8}[4 - 4c(\theta) + s\ \sin\theta]. \tag{B10}$$

Any choice of t_o and t_e satisfying these constraints gives rise to valid operator inequalities. In order to obtain the strongest bound, we choose the largest values of t_o and t_e consistent with their respective constraints, i.e.,

$$\begin{aligned} t_e &= \min\left\{1 - \frac{s}{8}\cos\theta, \frac{s}{8}\cos\theta\right\}, \\ t_o &= \min\left\{\frac{1}{8}[4 + 4c(\theta) - s\sin\theta], \frac{1}{8}[4 - 4c(\theta) + s\sin\theta]\right\}. \end{aligned} \tag{B11}$$

A similar procedure for the interval $\theta \in (\pi/4, \pi/2]$ leads to

$$\begin{aligned} t_e &= \min\left\{\frac{1}{8}[4 + 4c(\theta) - s\ \cos\theta], \frac{1}{8}[4 - 4c(\theta) + s\ \cos\theta]\right\}, \\ t_o &= \min\left\{1 - \frac{s}{8}\sin\theta, \frac{s}{8}\sin\theta\right\}. \end{aligned} \tag{B12}$$

It is worth pointing out that the two intervals only differ by exchanging $t_e \leftrightarrow t_o$ and $\sin\theta \leftrightarrow \cos\theta$. Hence, for any given θ, we have constructed operator inequalities of the form (B1).

As shown in the main text, we obtain our lower bound on the average fidelity from

$$\mathcal{F}(\mathcal{A}_2) \geqslant \frac{s}{4}(\mathcal{A}_2 - 1/2) + \min_{M_0, M_1} t(M_0, M_1) \equiv L(\mathcal{A}_2), \tag{B13}$$

where $t(M_0, M_1) = (t_e + t_o)/2$. To compute this quantity we fix the value of s to be

$$s = 4(1 + \sqrt{2}) \tag{B14}$$

and choose the dephasing function as $c(\theta) = \min\{1, \frac{s}{4}\sin\theta\}$ whenever $\theta \in [0, \pi/4]$ and $c(\theta) = \min\{1, \frac{s}{4}\cos\theta\}$ whenever $\theta \in (\pi/4, \pi/2]$. It is easy to see that $c(\theta) \in [0, 1]$, which ensures that Λ_θ is a valid quantum channel, and that $c(\theta)$ is continuous at $\theta = \pi/4$. A simple calculation shows that in this case

$$t = \frac{2 - \sqrt{2}}{4}, \tag{B15}$$

which gives the lower bound

$$\mathcal{F}(\mathcal{A}_2) \geqslant (1 + \sqrt{2})\mathcal{A}_2 - \frac{3}{2\sqrt{2}} \equiv L(\mathcal{A}_2). \tag{B16}$$

One can check that choosing distinct values of s will not lead to improved lower bounds.

APPENDIX C: TIGHTNESS OF FIDELITY BOUNDS

In the main text, we have derived fidelity bounds for both the preparations and the measurements, based on operator

inequalities. Specifically, we obtain a lower bound on the average fidelity $\mathcal{F}$ of the prepared states (with respect to the ideal ones) given by the linear expression

$$\mathcal{F}(\mathcal{A}_2) \geqslant (1+\sqrt{2})\mathcal{A}_2 - \frac{3}{2\sqrt{2}} \equiv L(\mathcal{A}_2). \tag{C1}$$

For measurements, a similar bound is obtained on the average fidelity $\mathcal{F}'$ with respect to the ideal ones. In the present appendix, we discuss the tightness of these bounds.

We start with our bound on the fidelity of the states. As discussed in the main text, obtaining $\mathcal{A}_2 = Q_2$ implies $\mathcal{F} = 1$, i.e., the states are the ideal ones (up to a unitary). Let us refer to the optimal strategy (with the ideal states) as strategy S_1. Then, for $\mathcal{A}_2 = C_2$, our bound gives $\mathcal{F} \geqslant 3/4$. This bound is tight and can be obtained via the set of pure states $\rho_{x_0x_1} = [\mathbb{1} + (-1)^{x_0x_1}\sigma_z]/2$ (diagonal in the same basis, hence classical), combined with the measurements $M_0 = M_1 = \sigma_z$. Let us refer to this strategy as S_2.

The above shows that our bound (C1) is tight as far as linear inequalities are concerned. More generally, the bound is in fact tight in general, when shared randomness between the preparation and measurement devices is taken into account. In this case, taking a convex combination between strategies S_1 and S_2 allows us to get any point on the line (i.e., pair of values $\mathcal{F}$ and $\mathcal{A}_2$) between S_1 and S_2.

It is also interesting to understand what happens when shared randomness between the devices is not taken into account. In this case, the end points ($\mathcal{A}_2 = Q_2, \mathcal{F} = 1$) and ($\mathcal{A}_2 = C_2, \mathcal{F} = 3/4$) can still be obtained. To understand what happens in the intermediate region $C_2 < \mathcal{A}_2 < Q_2$, we first performed a numerical analysis. Specifically, we choose randomly four qubit states, and compute (i) the maximal value of $\mathcal{A}_2$ (optimizing over the measurements) and (ii) the average fidelity $\mathcal{F}$ (where the optimization over channels is restricted here to unitaries). The resulting points are shown on Fig. 2 (blue circles). This indicates that, for $C_2 < \mathcal{A}_2 < Q_2$, the bound (C1) cannot be saturated anymore. Moreover, we conjecture that an optimal class of strategies is given by the pure states

$$|\psi_{00}\rangle = |0\rangle, \quad |\psi_{11}\rangle = |1\rangle, \quad |\psi_{01}\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle,$$
$$|\psi_{10}\rangle = \cos\theta|0\rangle - \sin\theta|1\rangle \tag{C2}$$

and the measurements $M_y = \cos(\varphi)\sigma_z + (-1)^y \sin(\varphi)\sigma_x$. Straightforward calculations show that taking $\tan\varphi = \sin 2\theta$ leads to

$$\mathcal{A}_2 = \tfrac{1}{2} + \tfrac{1}{4}\sqrt{1+\tan^2(\varphi)}, \quad \mathcal{F} = \tfrac{1}{4}(3+\tan\varphi). \tag{C3}$$

This gives a parametric curve, as a function of $\varphi \in [0, \pi/4]$, given by the red curve in Fig. 2. This curve is in excellent agreement with the numerical results obtained before. Note that this class of strategies interpolates between the strategies S_1 (setting $\varphi = 0$) and S_2 (setting $\varphi = \pi/4$).

Next we discuss the bound on the average fidelity of measurements. As discussed in the main text, the linear bound $\mathcal{F}'(\mathcal{A}_2) \geqslant L(\mathcal{A}_2)$ is optimal as far as linear inequalities are concerned. Moreover, when allowing for shared randomness the bound is tight in general for $C_2 \leqslant \mathcal{A}_2 \leqslant Q_2$. This is obtained by considering convex combinations of strategy S_1' (defined as the optimal strategy S_1, up to a rotation of $\pi/8$ around the y axis; see below), and the following strategy (referred to as S_3): take $M_0 = \sigma_z$ and $M_1 = \mathbb{1}$, with the states $\rho_{00} = \rho_{01} = (\mathbb{1} + \sigma_z)/2$ and $\rho_{10} = \rho_{11} = (\mathbb{1} - \sigma_z)/2$.

Similar to the case of states, we now consider the situation where shared randomness between the devices is not allowed. Performing a numerical analysis similar to the one described above (except that measurements are now generated randomly), we observe that the accessible region (in terms of $\mathcal{F}'$ vs $\mathcal{A}_2$) appears to be exactly the same as for the case of states (i.e., the blue region in Fig. 2). We conjecture that the lower bound is given by the following class of optimal strategies: take the measurements

$$M_0 = \sigma_z, \quad M_1 = \eta\sigma_x + (1-\eta)\mathbb{1}, \tag{C4}$$

with the states $|\psi_{00}\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle$, $|\psi_{01}\rangle = \cos\theta|0\rangle - \sin\theta|1\rangle$, $|\psi_{10}\rangle = \cos\theta|1\rangle + \sin\theta|0\rangle$, and $|\psi_{11}\rangle = \cos\theta|1\rangle - \sin\theta|0\rangle$. Setting $\eta = \tan 2\theta$, we get

$$\mathcal{A}_2 = \frac{\cos^2(\theta)}{2} + \frac{1}{4} + \frac{\sin^2(2\theta)}{\cos(2\theta)}, \quad \mathcal{F} = \frac{1}{4}[3 + \tan(2\theta)]. \tag{C5}$$

This gives a parametric curve, as a function of $\theta \in [0, \pi/8]$, given by the red curve in Fig. 2. This curve is in excellent agreement with the numerical results obtained before. Also, this curve turns out to be exactly the same as the curve we obtained above for the case of states. Note that this class of strategies interpolates between the strategies S_1' (setting $\theta = \pi/8$) and S_3 (setting $\theta = 0$).

Finally, note that the numerics also suggests that there is a linear upper bound on the average fidelities $\mathcal{F}$ ($\mathcal{F}'$) as a function of $\mathcal{A}_2$ (see Fig. 2); specifically $\mathcal{F} \leqslant \frac{1-Q_2}{Q_2-3/4}\mathcal{A}_2 +$

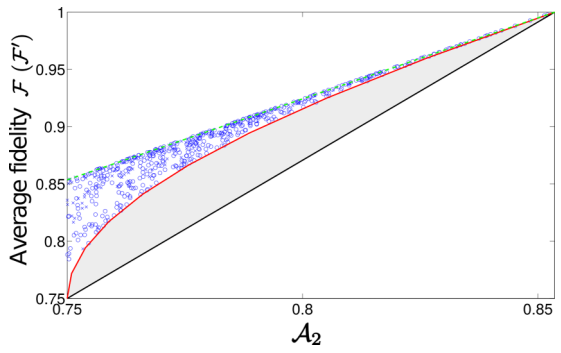

FIG. 2. Black line is the analytic lower bound on the average fidelity $\mathcal{F}$ ($\mathcal{F}'$) for prepared states (measurements), as a function of the observed value of $\mathcal{A}_2$. To characterize the region accessible via pure qubit strategies (i.e., without shared randomness), we perform numerics generating randomly sets of qubit preparations (blue circles and crosses); here we show the numerical results for the case of states, but similar results are obtained for the case of measurements. In the region $C_2 < \mathcal{A}_2 < Q_2$, we conjecture that the class of strategies given in the text (corresponding to the red curve) are optimal, both for $\mathcal{F}$ and $\mathcal{F}'$. Finally, the green dashed line is our conjectured upper bound on the average fidelity.

$\frac{Q_2^2-3/4}{Q_2-3/4}$ and similarly for $\mathcal{F}'$. It would be interesting to provide a proof of these upper bounds.

APPENDIX D: OPERATOR INEQUALITIES FOR ROBUST SELF-TESTING OF MEASUREMENTS

In this section, we account for the detailed derivation of the lower bound on the average fidelity of the measurements $\mathcal{F}'(\mathcal{A}_2)$. The approach bears significant resemblance to the case of robustly self-testing preparations, as outlined in Appendix B.

We aim to derive operator inequalities of the form

$$K_{yb}(\{\rho_{x_0x_1}\}) \geqslant s Z_{yb} + t_y(\{\rho_{x_0x_1}\})\mathbb{1}, \tag{D1}$$

where $Z_{yb} = \frac{1}{8}\sum_{x_0,x_1}\rho_{x_0x_1}\delta_{b,x_y}$ and $K_{yb}(\{\rho_{x_0x_1}\}) = \Lambda^\dagger[(M_y^b)^{\text{ideal}}]$. For the sake of simplicity, we first apply a unitary channel to $(M_y^b)^{\text{ideal}}$ to align these operators with the eigenstates of σ_x and σ_z. Then, we adopt the same (unital, trace-preserving) channel Λ as specified in the main text, with the same coefficients as used to robustly self-test the preparations: $c(\theta) = \min\{1, \frac{s}{4}\sin\theta\}$ when $\theta \in [0, \pi/4]$ and $c(\theta) = \min\{1, \frac{s}{4}\cos\theta\}$ when $\theta \in (\pi/4, \pi/2]$.

It is straightforward to see that, for any given pair of measurements, the optimal choice of preparations are four pure qubit states, such that ρ_{00} and ρ_{11}, and ρ_{01} and ρ_{10}, respectively, correspond to antipodal vectors on the Bloch sphere. Therefore, we can without loss of generality restrict to such preparations since these impose the weakest constraints on the measurements of our interest. We can therefore parametrize the preparations $\rho_{x_0x_1} = 1/2(\mathbb{1} + \vec{m}_{x_0x_1}\cdot\vec{\sigma})$ by Bloch vectors

$$\begin{aligned}\vec{m}_{00} &= [\cos\theta, 0, \sin\theta], \quad \vec{m}_{11} = -[\cos\theta, 0, \sin\theta],\\ \vec{m}_{01} &= [\cos\theta, 0, -\sin\theta], \quad \vec{m}_{10} = [-\cos\theta, 0, \sin\theta].\end{aligned} \tag{D2}$$

Expressing Z_{yb} in terms of these preparations gives

$$\begin{aligned}Z_{00} &= \tfrac{1}{8}(\mathbb{1} + \cos\theta\sigma_x), \quad Z_{01} = \tfrac{1}{8}(\mathbb{1} - \cos\theta\sigma_x),\\ Z_{10} &= \tfrac{1}{8}(\mathbb{1} + \sin\theta\sigma_z)\sigma_z, \quad Z_{11} = \tfrac{1}{8}(\mathbb{1} - \sin\theta\sigma_z).\end{aligned} \tag{D3}$$

Due to symmetries, we restrict ourselves so that $t_o \equiv t_{01} = t_{10}$ and $t_e \equiv t_{00} = t_{11}$. Thus we have to consider two operator inequalities in each interval $\theta \in [0, \pi/4]$ and $\theta \in (\pi/4, \pi/2]$. In the first interval, the two operator inequalities are

$$\begin{aligned}\frac{1+\sigma_x}{2} - \frac{s}{8}(\mathbb{1} + \cos\theta\sigma_x) - t_e\mathbb{1} &\geqslant 0,\\ \frac{1+c(\theta)\sigma_z}{2} - \frac{s}{8}(\mathbb{1} + \sin\theta\sigma_z) - t_o\mathbb{1} &\geqslant 0.\end{aligned} \tag{D4}$$

In the second interval, the two operator inequalities are

$$\begin{aligned}\frac{1+c(\theta)\sigma_x}{2} - \frac{s}{8}(\mathbb{1} + \cos\theta\sigma_x) - t_e\mathbb{1} &\geqslant 0,\\ \frac{1+\sigma_z}{2} - \frac{s}{8}(\mathbb{1} + \sin\theta\sigma_z) - t_o\mathbb{1} &\geqslant 0.\end{aligned} \tag{D5}$$

Just as in Appendix B, we solve these inequalities for t_e and t_o, and choose the largest value compatible with the solutions. In the first interval, this gives

$$\begin{aligned}t_e &= \min\left\{\frac{1}{8}(8 - s - s\cos\theta), \frac{s}{8}(\cos\theta - 1)\right\},\\ t_o &= \min\left\{\frac{1}{8}[4c(\theta) - s\sin\theta - s + 4],\right.\\ &\qquad \left.\frac{1}{8}[-4c(\theta) + s\sin\theta - s + 4]\right\}.\end{aligned} \tag{D6}$$

A similar procedure for the interval $\theta \in (\pi/4, \pi/2]$ leads to

$$\begin{aligned}t_e &= \min\left\{\frac{1}{8}[4c(\theta) - s\cos\theta - s + 4],\right.\\ &\qquad \left.\frac{1}{8}[-4c(\theta) + s\cos\theta - s + 4]\right\},\\ t_o &= \min\left\{\frac{s}{8}(\sin\theta - 1), \frac{1}{8}(8 - s - s\sin\theta)\right\}.\end{aligned} \tag{D7}$$

For any choice of θ, we have constructed operator inequalities of the form (D1).

In order to obtain our lower bound on $\mathcal{F}'$, we must minimise the quantity $t(\theta) = (t_e + t_o)/2$ for a specific choice of s. In analogy with the procedure in Appendix D, we choose $s = 4(1+\sqrt{2})$, which returns $\min_\theta t(\theta) = -3/(2\sqrt{2})$. Hence we have obtained the lower bound

$$\mathcal{F}'(\mathcal{A}_2) \geqslant (1+\sqrt{2})\mathcal{A}_2 - \frac{3}{2\sqrt{2}} = L(\mathcal{A}_2). \tag{D8}$$

APPENDIX E: SELF-TESTING ALL PAIRS OF INCOMPATIBLE PAULI OBSERVABLES

Consider a generalization of the $2 \to 1$ RAC, in which we introduce a bias on the score associated to certain inputs. Specifically, whenever the game is successful, i.e., $b = x_y$, the awarded score is $q/2$ if $x_0 \oplus x_1 = 0$, and $(1-q)/2$ if $x_0 \oplus x_1 = 1$, for some $q \in [0, 1]$. The average score reads

$$\mathcal{A}_2^q = \frac{1}{2}\sum_{x_0,x_1,y} r(x_0, x_1)P(b = x_y|x_0, x_1, y), \tag{E1}$$

where $r(x_0, x_1) = q/2$ if $x_0 \oplus x_1 = 0$ and $r(x_0, x_1) = (1-q)/2$ if $x_0 \oplus x_1 = 1$. Note that, for $q = 1/2$, we recover the standard $2 \to 1$ RAC. Based on the quantity $\mathcal{A}_2^q$, we will now see how to derive a self-testing condition for any pair of incompatible Pauli observables, i.e., any pair of noncommuting projective rank-one qubit measurements.

We start by expressing $\mathcal{A}_2^q$ for a quantum strategy:

$$\begin{aligned}\mathcal{A}_2^q &= \frac{1}{2} + \frac{1}{4}\sum_{x_0,x_1} r(x_0, x_1)\,\text{tr}\{\rho_{x_0x_1}[(-1)^{x_0}M_0 + (-1)^{x_1}M_1]\}\\ &\leqslant \frac{1}{2} + \frac{1}{4}\sum_{x_0,x_1} r(x_0, x_1)\lambda_{\max}[(-1)^{x_0}M_0 + (-1)^{x_1}M_1].\end{aligned} \tag{E2}$$

Denoting $\mu_k = \lambda_{\min}[M_0 + (-1)^k M_1]$ and $\nu_k = \lambda_{\max}[M_0 + (-1)^k M_1]$, for $k = 0, 1$, we obtain

$$\mathcal{A}_2^q \leqslant \tfrac{1}{2} + \tfrac{1}{8}[q(\mu_0 - \nu_0) + (1-q)(\mu_1 - \nu_1)]. \tag{E3}$$

Following a derivation analogous to that appearing in Appendix A to obtain, we obtain

$$\mathcal{A}_2^q \leqslant \tfrac{1}{2} + \tfrac{1}{8}[q\sqrt{\beta+\alpha} + (1-q)\sqrt{\beta-\alpha}], \tag{E4}$$

where $\beta = 2\,\mathrm{tr}\,(M_0^2 + M_1^2) - \mathrm{tr}\,(M_0)^2 - \mathrm{tr}\,(M_1)^2$ and $\alpha = 2\,\mathrm{tr}\,(\{M_0, M_1\}) - 2\,\mathrm{tr}\,(M_0)\,\mathrm{tr}\,(M_1)$. Treating α and β as independent variables, we obtain the largest value of the right-hand side of Eq. (E4) by demanding that the derivative with respect to α equals zero, and checking that the second derivative is negative at this point. We obtain the optimality constraint

$$\alpha = \frac{2q-1}{1-2q+2q^2}\beta. \tag{E5}$$

Inserting this value back into Eq. (E4), we find an upper bound on $\mathcal{A}_2^q$ as obtained by independent variables α and β. It turns out that this bound can be saturated by the de facto coupled variables α and β. From Eq. (E4), it is clear that a necessary condition for optimality is to maximize β. This amounts to the observables M_0 and M_1 being traceless and such that $M_0^2 = M_1^2 = \mathbb{1}$, leading to $\beta = 8$. This implies that the observables represent projective rank-one measurements. Hence we can write $M_y = \vec{n}_y \cdot \vec{\sigma}$ where the Bloch vector satisfies $|\vec{n}_y| = 1$. Hence we have $\alpha = 8\vec{n}_0 \cdot \vec{n}_1$. Thus Eq. (E5) becomes

$$\vec{n}_0 \cdot \vec{n}_1 = \frac{2q-1}{1-2q+2q^2}, \tag{E6}$$

which has a solution for any choice of q. Note that setting $q = 1/2$ reduces the above to $\vec{n}_0 \cdot \vec{n}_1 = 0$, which we recognize as the optimality constraint for the standard $2 \to 1$ random access code. In conclusion, for any pair of incompatible Pauli observables (characterized by the scalar product $\vec{n}_0 \cdot \vec{n}_1$), we have a game $\mathcal{A}_2^q$ (where q is chosen in order to satisfy the above equation), such that the maximal score can only be attained by using that specific pair of Pauli observables. We thus obtain a general class of self-tests for any pair of Pauli observables, corresponding to saturating the maximal quantum value of $\mathcal{A}_2^q$ for a given value of q:

$$\mathcal{A}_2^q \leqslant \tfrac{1}{2}(1 + \sqrt{1-2q+2q^2}). \tag{E7}$$

APPENDIX F: SELF-TESTING FOR THE $N \to 1$ RANDOM ACCESS CODE

In this appendix, we extend the results presented in the main text to self-test the preparations and measurements in an $N \to 1$ RAC. The latter is a straightforward generalization of the $2 \to 1$ RAC considered in the main text. The input of the preparation device is a random N-bit string $x \equiv (x_1, \ldots, x_N)$, while the input of the measurement device is $y \in \{1, \ldots, N\}$. The average score is

$$\mathcal{A}_N = \frac{1}{N2^N} \sum_{x,y} P(b = x_y | x, y). \tag{F1}$$

Considering qubit states ρ_x, and measurement observables M_y, we get

$$\mathcal{A}_N = \frac{1}{2} + \frac{1}{N2^{N+1}} \sum_{x,y} (-1)^{x_y}\,\mathrm{tr}(\rho_x M_y). \tag{F2}$$

1. Compatibility of measurements

We determine whether a set of measurements can explain (i.e., are compatible with) a given value of $\mathcal{A}_N$. Since rank-one projective measurements are optimal for any set of preparations, we choose for simplicity to restrict our consideration to such measurements. However, it is straightforward to consider general measurements using the method outlined in the main text and Appendix A.

Specifically, we first write

$$\begin{aligned}\mathcal{A}_N &= \frac{1}{2} + \frac{1}{N2^{N+1}} \sum_x \mathrm{tr}\,(\rho_x W_x) \\ &\leqslant \frac{1}{2} + \frac{1}{N2^{N+1}} \sum_x \lambda_{\max}[W_x],\end{aligned} \tag{F3}$$

where $W_x = \sum_y (-1)^{x_y} M_y$.

Note $\lambda_{\max}[W_x] = \lambda_{\min}[W_{\bar{x}}]$, where $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_N)$ is the bit string obtained from x by flipping all bits. Thus it is sufficient to only calculate eigenvalues for the strings not obtainable from each other under a full bit-flip operation. To this end let $z = x_1 \ldots x_{N-1}, 0$ and $\lambda_{z,0}$ ($\lambda_{z,1}$) be the largest (smallest) eigenvalue of W_z. Thus we write

$$\mathcal{A}_N \leqslant \frac{1}{2} + \frac{1}{N2^{N+1}} \sum_z [\lambda_{z,0} - \lambda_{z,1}]. \tag{F4}$$

Since $\lambda_{z,0}^2$ and $\lambda_{z,1}^2$ are eigenvalues of W_z^2, we have $\lambda_{z,0}^2 + \lambda_{z,1}^2 = \mathrm{tr}\,(W_z^2)$, which is equivalent to

$$\lambda_{z,0}^2 + \lambda_{z,1}^2 = \sum_{y=1}^N \mathrm{tr}\left(M_y^2\right) + \sum_{k<l} (-1)^{z_k + z_l}\,\mathrm{tr}\,(\{M_k, M_l\}). \tag{F5}$$

This equation, together with the relation $(\lambda_{z,0} - \lambda_{z,1})^2 \leqslant 2(\lambda_{z,0}^2 + \lambda_{z,1}^2)$, imply that Eq. (F4) becomes

$$\begin{aligned}\mathcal{A}_N \leqslant \frac{1}{2} + \frac{\sqrt{2}}{N2^{N+1}} \sum_z \Bigg[&\sum_{y=1}^N \mathrm{tr}\left(M_y^2\right) \\ &+ \sum_{k<l} (-1)^{z_k + z_l}\,\mathrm{tr}\,(\{M_k, M_l\}) \Bigg]^{1/2}.\end{aligned} \tag{F6}$$

This provides a robust self-testing condition, allowing one to determine whether a given set of measurements is compatible with the observed value of $\mathcal{A}_N$. Furthermore, we can derive an upper bound on the maximal value of $\mathcal{A}_N$ by assuming (incorrectly for $N > 3$) that there exists N mutually unbiased bases in $\mathbb{C}^2$. This means that all measurements are maximally incompatible, i.e., that $\mathrm{tr}\,(\{M_k, M_l\}) = 0$ for $k \neq l$. Consequently, Eq. (F6) reduces to

$$\mathcal{A}_N \leqslant \frac{1}{2}\left(1 + \frac{1}{\sqrt{N}}\right). \tag{F7}$$

We emphasize that only three mutually unbiased bases exist in $\mathbb{C}^2$ and hence this bound is only tight for $N = 2, 3$. For $N = 2$, we recover the result presented in the main text. For $N = 3$, this implies that a maximal value of $\mathcal{A}_3$ (i.e., achieving the right-hand side of the above inequality) ensures that the

measurements are three mutually unbiased qubit observables, such as the three Pauli observables σ_x, σ_y, and σ_z.

Going one step further, we can then also self-test the preparations (still assuming maximal value of $\mathcal{A}_3$). Indeed, each preparation ρ_x must be pure, and correspond to the eigenvector of W_x associated to its largest eigenvalue. Such a set of preparations corresponds to a set of Bloch vectors forming a cube on the surface of the Bloch sphere.

2. Compatibility of preparations

We ask whether a given value of $\mathcal{A}_N$ can be explained by a particular set of preparations. We suitably express (F2) in a quantum model and subsequently apply the Cauchy-Schwarz inequality for operators to obtain

$$\mathcal{A}_N = \frac{1}{2} + \frac{1}{N2^N}\sum_{y=1}^{N} \mathrm{tr}\left[M_y^0 \sum_x (-1)^{x_y}\rho_x\right] \leqslant \frac{1}{2} + \frac{1}{N2^N}\sum_{y=1}^{N}\sqrt{\mathrm{tr}\left[M_y^0\left(\sum_x (-1)^{x_y}\rho_x\right)^2\right]}. \quad \text{(F8)}$$

In the last expression, the squared operator is evaluated to

$$\left(\sum_x (-1)^{x_y}\rho_x\right)^2 = \sum_x \rho_x^2 + \sum_{k<l}(-1)^{k_y+l_y}\{\rho_k, \rho_l\}. \quad \text{(F9)}$$

If necessary, the anticommutators can be evaluated using Bloch sphere representation with the relation $\{\rho_k, \rho_l\} = 1/2[(1+\vec{m}_k\cdot\vec{m}_l)\mathbb{1} + (\vec{m}_k+\vec{m}_l)\cdot\vec{\sigma}]$. However, it is more convenient to consider a basis-independent representation. Importantly, note that since an equal number of positive and negative terms appear inside the square, the operator $\sum_x (-1)^{x_y}\rho_x$ is a linear combination of $\{\sigma_x, \sigma_y, \sigma_z\}$ and hence its square is proportional to the identity operator. Therefore, when reinserting Eq. (F9) into Eq. (F8), we find

$$\mathcal{A}_N \leqslant \frac{1}{2} + \frac{1}{N2^N}\sum_{y=1}^{N}\left[\sum_x \mathrm{tr}\left(\rho_x^2\right) + \sum_{k<l}(-1)^{k_y+l_y}\,\mathrm{tr}\left(\{\rho_k, \rho_l\}\right)\right]^{1/2}. \quad \text{(F10)}$$

This is a self-testing condition for preparations, assessing whether a given set of preparations is compatible with a given value of $\mathcal{A}_N$. In particular, a classical strategy in which the preparations are binary messages corresponds to $\forall x$: $\mathrm{tr}\,(\rho_x^2) = 1$ and $\mathrm{tr}\,(\{\rho_k, \rho_l\}) = 2\delta_{E(k),E(l)}$, where E is the specific classical encoding strategy, i.e., a function $E : \{0,1\}^N \to \{0,1\}$.

APPENDIX G: SELF-TESTING WITH THREE-LEVEL SYSTEMS

In the main text, we have considered self-testing in the $2 \to 1$ random access code when the physical system transmitted from Alice to Bob is a qubit. Clearly, if that system is allowed to carry two bits of information, the task is trivial since Alice can send both her inputs to Bob. Here, we consider the remaining nontrivial case of Alice communicating a three-level quantum system. To simplify the analysis we restrict ourselves to projective measurements for which all possible arrangements admit a compact characterization. We show that the optimal quantum value equals $\mathcal{A}_2 = (5+\sqrt{5})/8 \approx 0.9045$ and find all the optimal arrangements of observables (we argue that the optimal value is achieved only if both measurements are projective). Our argument is robust in the sense that we are able to certify incompatibility of M_0 and M_1 whenever the success probability exceeds the classical bound for three-level systems, which turns out to be $\mathcal{A}_2 \leqslant 7/8$.

To obtain a statement which only depends on the observables we follow the main text and evaluate the sum

$$\sum_{x_0,x_1}\lambda_{\max}[(-1)^{x_0}M_0 + (-1)^{x_1}M_1]. \quad \text{(G1)}$$

Jordan's lemma states that any two projective observables can be simultaneously diagonalized such that the resulting blocks are 1×1 or 2×2. For observables acting on a qutrit, we only need to consider two cases: (a) three one-dimensional subspaces or (b) one subspace of each type. Case (a) corresponds to classical strategies and it is easy to check that these satisfy $\mathcal{A}_2 \leqslant 7/8$. In case (b) the observables (up to a unitary) can be written as

$$M_0 = \begin{pmatrix}\cos\alpha\,\sigma_x + \sin\alpha\,\sigma_z & \\ & r\end{pmatrix}, \quad M_1 = \begin{pmatrix}\cos\alpha\,\sigma_x - \sin\alpha\,\sigma_z & \\ & s\end{pmatrix} \quad \text{(G2)}$$

for some angle $\alpha \in [0, 2\pi]$ and $r, s \in \{\pm 1\}$. A simple calculation yields

$$\begin{aligned}\lambda_{\max}[M_0+M_1] &= \max\{2|\cos\alpha|, r+s\},\\ \lambda_{\max}[M_0-M_1] &= \max\{2|\sin\alpha|, r-s\},\\ \lambda_{\max}[-M_0+M_1] &= \max\{2|\sin\alpha|, -r+s\},\\ \lambda_{\max}[-M_0-M_1] &= \max\{2|\cos\alpha|, -r-s\}\end{aligned}$$

and, therefore,

$$\sum_{x_0,x_1}\lambda_{\max}[(-1)^{x_0}M_0 + (-1)^{x_1}M_1] = \begin{cases}2+4|\sin\alpha|+2|\cos\alpha| & \text{if } r=s,\\ 2+2|\sin\alpha|+4|\cos\alpha| & \text{if } r\neq s.\end{cases} \quad \text{(G3)}$$

For $r=s$ the right-hand side is maximized for $\alpha \in \{c_1, c_1+\pi, -c_1+\pi, -c_1+2\pi\}$, where c_1 is the unique solution to $\tan c_1 = 2$ in the interval $[0, \pi/2]$. Similarly, for $r\neq s$ the right-hand side is maximized for $\alpha \in \{c_2, c_2+\pi, -c_2+\pi, -c_2+2\pi\}$, where c_2 is the unique solution to $\tan c_2 = 1/2$ in the interval $[0, \pi/2]$.

While the different optimal arrangements are not unitarily equivalent, they are of similar form. The optimal arrangement characterized by $r=s=1$ and $\alpha = c_1$ yields the following

optimal preparations:

$$\rho_{00} = \begin{pmatrix} 0 & \\ & 1 \end{pmatrix}, \quad \rho_{01} = \begin{pmatrix} (\mathbb{1}+\sigma_z)/2 & \\ & 0 \end{pmatrix},$$
$$\rho_{10} = \begin{pmatrix} (\mathbb{1}-\sigma_z)/2 & \\ & 0 \end{pmatrix}, \quad \rho_{11} = \begin{pmatrix} (\mathbb{1}-\sigma_x)/2 & \\ & 0 \end{pmatrix}. \tag{G4}$$

Indeed, it is always the case that one preparation lives in the 1×1 subspace, whereas the other three occupy the 2×2 subspace (two of them form a basis to which the last one is unbiased). To see that the optimal winning probability requires projective measurements, note that for every set of preparations the optimal observables can be chosen projective. However, all sets of preparations optimal for projective observables are of the form given above and one can check that for these preparations the optimal measurements must be projective (a direct consequence of the fact that the operators $\rho_{00} + \rho_{01} - \rho_{10} - \rho_{11}$ and $\rho_{00} - \rho_{01} + \rho_{10} - \rho_{11}$ are full rank).

It is the presence of multiple inequivalent maximizers that prevents us from writing down a simple self-testing statement. However, Eq. (G3) allows us to deduce the range of α compatible with the observed value of $\mathcal{A}_2$ (note that the conclusion will be stronger if we know whether $r = s$ or $r \neq s$). In particular, any value exceeding the classical bound of 7/8 implies a lower bound on the incompatibility between M_0 and M_1 on the 2×2 subspace.

APPENDIX H: NUMERICAL METHOD FOR ROBUST SELF-TESTING

In the main text, we focused on the RAC and derived an optimal robust self-test. However, robust self-testing is relevant also for many other tasks that are not RACs. Here, we outline a numerical method based on semidefinite programming for inferring lower bounds on the worst-case average fidelity of preparations $\mathcal{F}$ in more general tasks. Specifically, we adapt the so-called swap method of [21] (constructed for Bell scenarios) to prepare-and-measure scenarios by combining it with the hierarchy of dimensionally bounded quantum correlations [41]. For sake of instruction, we first present the method by applying it to the RAC, and then use it to robustly self-test preparations in another prepare-and-measure scenario.

The preparations in the random access code are self-tested up to a collective unitary transformation. A robust self-test must therefore be valid under this degree of freedom. However, one can only consider the fidelity of the unknown preparations with respect to the optimal states in some chosen basis. Therefore, in order to achieve a robust self-test, one needs to find a way to avoid the possibility of a collective unitary misaligning the bases. This can be done by supplying Bob's measurement device with an auxillary system, say it is initialized in the state $|0\rangle_{\mathrm{A}}$, into which the unknown received preparations can be swapped [21]. In the RAC, the optimal measurements are anticommuting Pauli measurements. Therefore, with inspiration from this ideal case, Bob's swap operator S can be composed as follows: $S = UVU$, where

$$\begin{aligned} U &= \mathbb{1} \otimes |0\rangle\langle 0| + B_1 \otimes |1\rangle\langle 1|, \\ V &= \frac{1+B_0}{2} \otimes \mathbb{1} + \frac{1-B_0}{2} \otimes \sigma_x, \end{aligned} \tag{H1}$$

where B_0 and B_1 denote the observables of Bob. If B_0 and B_1 correspond to σ_z and σ_x, respectively, the above returns the two-qubit swap operator. Bob applies S to the joint system of received preparation (labeled B) and ancilla (labeled A). The state swapped into Bob's ancilla reads

$$\rho^{\mathrm{SWAP}}_{x_0 x_1} = \mathrm{tr}_{\mathrm{B}}[S(\rho_{x_0 x_1} \otimes |0\rangle_{\mathrm{AA}}\langle 0|)S^\dagger]. \tag{H2}$$

Consequently, the worst-case average fidelity of Alice's preparations with her optimal preparations is

$$\begin{aligned} \mathcal{F}(\mathcal{A}_2^*) &= \min_{\rho \in R(\mathcal{A}_2^*)} \max_{\Lambda} \frac{1}{4} \sum_{x_0 x_1} \mathrm{tr}\left[\Lambda\left[\rho^{\mathrm{ideal}}_{x_0 x_1}\right] \rho^{\mathrm{SWAP}}_{x_0 x_1}\right] \\ &= \min_{\rho \in R(\mathcal{A}_2^*)} \max_{\Lambda} \frac{1}{4} \sum_{x_0 x_1} \mathrm{tr}\left[S(\Lambda[\rho_{x_0 x_1}] \otimes |0\rangle_{\mathrm{AA}}\langle 0|)\right. \\ &\qquad \left. S^\dagger\left(\mathbb{1} \otimes \rho^{\mathrm{ideal}}_{x_0 x_1}\right)\right], \end{aligned} \tag{H3}$$

where $R(\mathcal{A}_2^*)$ is the set of all preparations that are compatible with the value $\mathcal{A}_2^*$ and Λ is the extraction channel, the duality of which is used above.

We may write the operator S in terms Bob's observables as follows:

$$S = \frac{1}{2} \sum_{ij} s_{ij} \otimes |i\rangle_{\mathrm{AA}}\langle j|, \tag{H4}$$

where

$$\begin{aligned} s_{00} &= \mathbb{1} + B_0, \quad s_{01} = B_1 - B_0 B_1, \\ s_{10} &= B_1 - B_1 B_0, \quad s_{11} = \mathbb{1} + B_1 B_0 B_1. \end{aligned} \tag{H5}$$

Inserting this into (H3) we find

$$\begin{aligned} \mathcal{F}(\mathcal{A}_2^*) &= \min_{\rho \in R(\mathcal{A}^*)} \max_{\Lambda} \frac{1}{16} \sum_{x_0 x_1} \sum_{ijkl} \mathrm{tr}\left[(s_{ij} \otimes |i\rangle_{\mathrm{AA}}\langle j|)(\Lambda[\rho_{x_0 x_1}] \otimes |0\rangle_{\mathrm{AA}}\langle 0|)(s_{kl} \otimes |k\rangle_{\mathrm{AA}}\langle l|)^\dagger\left(\mathbb{1} \otimes \rho^{\mathrm{ideal}}_{x_0 x_1}\right)\right] \\ &= \min_{\rho \in R(\mathcal{A}_2^*)} \max_{\Lambda} \frac{1}{16} \sum_{x_0 x_1} \sum_{ijkl} \mathrm{tr}[s_{ij}\Lambda[\rho_{x_0 x_1}]s_{kl}^\dagger]\,\mathrm{tr}\left[|i\rangle\langle j|0\rangle\langle 0|l\rangle\langle k|\rho^{\mathrm{ideal}}_{x_0 x_1}\right] \\ &= \min_{\rho \in R(\mathcal{A}_2^*)} \max_{\Lambda} \frac{1}{16} \sum_{x_0 x_1} \sum_{ik} \mathrm{tr}[s_{k0}^\dagger s_{i0}\Lambda[\rho_{x_0 x_1}]]\langle k|\rho^{\mathrm{ideal}}_{x_0 x_1}|i\rangle \\ &= \min_{\rho \in R(\mathcal{A}_2^*)} \max_{\Lambda} \frac{1}{16} \sum_{x_0 x_1} \sum_{ik} \mathrm{tr}[T_{ik}\Lambda[\rho_{x_0 x_1}]]\langle k|\rho^{\mathrm{ideal}}_{x_0 x_1}|i\rangle, \end{aligned} \tag{H6}$$

where we defined $T_{ik} = s_{k0}^{\dagger} s_{i0}$. The four elements of T are straightforwardly computed to

$$T_{00} = 2(\mathbb{1} + B_0), \quad T_{01} = B_1(\mathbb{1} - B_0) + B_0 B_1(\mathbb{1} - B_0), \tag{H7}$$

$$T_{11} = 2(\mathbb{1} - B_0), \quad T_{10} = B_1(\mathbb{1} + B_0) - B_0 B_1(\mathbb{1} + B_0). \tag{H8}$$

In the calculation of the fidelity, the same channel is applied to all Alice's preparations. We may simply consider that as four other valid preparations $\bar{\rho}_{x_0 x_1} = \Lambda[\rho_{x_0 x_1}]$. The fidelity in (H6) is then a linear combination of variables $\{\mathrm{tr}\,(\bar{\rho}_{x_0 x_1}\mathbb{1}), \mathrm{tr}\,(\bar{\rho}_{x_0 x_1} B_0), \ldots, \mathrm{tr}\,(\bar{\rho}_{x_0 x_1} B_0 B_1 B_0)\}$. Therefore, we may establish a lower bound on (H6) using the dimensionally bounded hierarchy of quantum correlations [41]. The accuracy of this bound depends on the level of the hierarchy employed. We choose to consider the following level: we define a moment matrix

$$\begin{aligned} \chi_{ijkl} &= \mathrm{tr}[R_j^{\dagger} Q_i^{\dagger} Q_k R_l], \quad \text{where} \\ Q &= (\mathbb{1}, B_0, B_1, B_0 B_1, B_1 B_0), \\ R &= (\mathbb{1}, \bar{\rho}_{00}, \bar{\rho}_{01}, \bar{\rho}_{10}, \bar{\rho}_{11}), \end{aligned} \tag{H9}$$

for $i, j, k, l = 1, \ldots, 5$. From the moment matrix we calculate all terms needed to evaluate the average fidelity (H6), using the labels $x = 2x_0 + x_1 + 2$,

$$\begin{aligned} \mathrm{tr}\left[T_{00}\bar{\rho}_{x_0 x_1}\right] &= 2\chi_{111x} + 2\chi_{112x}, \\ \mathrm{tr}\left[T_{11}\bar{\rho}_{x_0 x_1}\right] &= 2\chi_{111x} - 2\chi_{112x}, \end{aligned} \tag{H10}$$

$$\begin{aligned} \mathrm{tr}\left[T_{01}\bar{\rho}_{x_0 x_1}\right] &= \chi_{113x} + \chi_{114x} - \chi_{115x} - \chi_{215x}, \\ \mathrm{tr}\left[T_{10}\bar{\rho}_{x_0 x_1}\right] &= \chi_{113x} - \chi_{114x} + \chi_{115x} - \chi_{215x}. \end{aligned} \tag{H11}$$

In order to enforce that the average fidelity is extremized for a particular value $\mathcal{A}_2^*$ of the random access code, we write the probability distribution of Bob's outcomes in terms of the moment matrix as

$$P(b|x_0, x_1, y) = \frac{1 + (-1)^b \chi_{1,1,y+2,x}}{2}. \tag{H12}$$

Thus we can evaluate $\mathcal{A}_2$ as a linear combination of moment matrix elements. Fixing the value of $\mathcal{A}_2$ corresponds to introducing an affine constraint on the moment matrix. Therefore, the following semidefinite program establishes a lower bound on $\mathcal{F}(\mathcal{A}_2)$:

$$\mathcal{F}(\mathcal{A}_2^*) \geqslant \min_{\chi} \frac{1}{16} \sum_{x_0 x_1} \sum_{i,k=0}^{1} \mathrm{tr}\left(T_{ik}\bar{\rho}_{x_0 x_1}\right)\langle k|\rho_{x_0 x_1}^{\mathrm{ideal}}|i\rangle \tag{H13}$$

$$\text{such that } \chi \geqslant 0, \quad \mathcal{A}_2 \geqslant \mathcal{A}_2^*.$$

We have implemented the semidefinite program and the results are presented in Fig. 3, together with the lower bound on $\mathcal{F}(\mathcal{A}_2)$ obtained from the analytical method presented in the main text. Evidently, the swap method returns a suboptimal but still nontrivial result. Using the swap method, we find a higher-than-classical value of $\mathcal{F}(\mathcal{A}_2)$, i.e., $\mathcal{F}(\mathcal{A}_2) > 3/4$, whenever $\mathcal{A}_2 > 0.802$.

The advantage of the swap method is that it applies also to other prepare-and-measure scenarios beyond RACs. The drawback of the method is that the self-tests are typically not optimal, and that the complexity of evaluating the dimensionally bounded hierarchy of quantum correlations increases exponentially with the number of preparations and measurements, thus making more complicated scenarios infeasible to study.

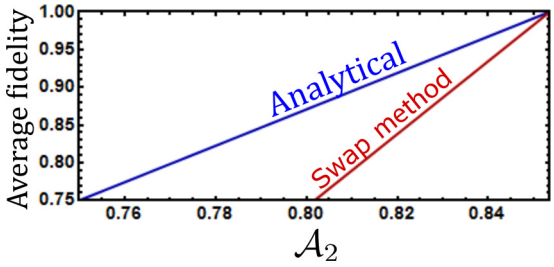

FIG. 3. Lower bound on $\mathcal{F}(\mathcal{A}_2)$ as obtained by the swap method and by analytical technique.

To exemplify the usefulness of this method also for other prepare-and-measure scenarios, we present a second example. Consider a prepare-and-measure scenario in which Alice has a random input $x \in \{0, 1, 2\}$ and Bob has a random input $y \in \{0, 1\}$. Alice may only communicate a qubit to Bob. The objective of the scenario reads

$$\mathcal{A} = \sum_{x,y} c_{x,y} E(x, y), \tag{H14}$$

where $E(x, y) = p(b = 0|x, y) - p(b = 1|x, y)$ and $c_{x,0} = [1, 1, -1]$ and $c_{x,1} = [\sqrt{3}, -\sqrt{3}, 0]$. One straightforwardly finds that the maximal classical value is $\mathcal{A} = 1 + 2\sqrt{3}$. We wish to robustly self-test Alice's preparations solely based on the value of $\mathcal{A}$. From numerical brute-force maximizations of $\mathcal{A}$, we find that its maximal value is $\mathcal{A} = 5$ and that this value is saturated using anticommuting Pauli measurements and preparations forming an equilateral triangle in a disk of the Bloch sphere. Such preparations can up to a unitary be written

$$\begin{aligned} \rho_0^{\mathrm{ideal}} &= \frac{1}{2}(\mathbb{1} + \sigma_x), \quad \rho_1^{\mathrm{ideal}} = \frac{1}{2}\left(\mathbb{1} + \frac{\sqrt{3}}{2}\sigma_z - \frac{1}{2}\sigma_x\right), \\ \rho_2^{\mathrm{ideal}} &= \frac{1}{2}\left(\mathbb{1} - \frac{\sqrt{3}}{2}\sigma_z - \frac{1}{2}\sigma_x\right). \end{aligned} \tag{H15}$$

We make the ansatz that this constitutes a self-test of the preparations. We supply Bob with an ancilla state and define the swap operator as done in the RAC. Performing calculations fully analogous to the case of the RAC, we obtain a semidefinite program that gives a lower bound on the

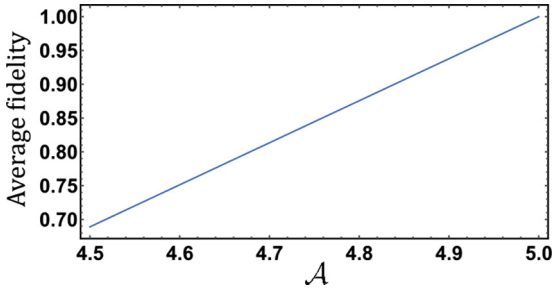

FIG. 4. Lower bound on $\mathcal{F}(\mathcal{A})$ as obtained by the swap method.

worst-case average fidelity

$$\mathcal{F}(\mathcal{A}) = \min_{\rho\in R(\mathcal{A})} \max_{\Lambda} \frac{1}{3}\sum_{x} \mathrm{tr}\left[\Lambda\left[\rho_x^{\mathrm{ideal}}\right]\rho_x\right], \tag{H16}$$

where $\mathcal{R}(\mathcal{A})$ is the set of preparations compatible with the value $\mathcal{A}$ and Λ is the extraction channel. We have used an intermediate level of the hierarchy of dimensionally bounded quantum correlations (sometimes referred to as 1+AB+BB+BBA) corresponding to an SDP matrix of size 20. The corresponding lower bound on $\mathcal{F}(\mathcal{A})$ is presented in Fig. 4. We first see that the maximal value $\mathcal{A} = 5$ indeed self-tests (up to numerical precision) the preparations of Alice to form an equilateral triangle on the Bloch sphere (the fidelity is one). For nonmaximal values of $\mathcal{A}$, we still obtain a nontrivial bound on the average fidelity of Alice's preparations with the optimal ones.

[1] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).

[2] R. Colbeck, Ph.D. thesis, University of Cambridge, 2007; arXiv:0911.3814.

[3] S. Pironio, A. Acín, S. Massar, A. Boyer De La Giroday, N. D. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, Nature (London) **464**, 1021 (2010).

[4] D. Mayers and A. Yao, *Proceedings of the 39th FOCS* (IEEE Computer Society, Washington, DC, 1998), p. 503.

[5] D. Mayers and A. Yao, Quantum Inf. Comput. **4**, 273 (2004).

[6] J. S. Bell, Physics **1**, 195 (1964).

[7] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).

[8] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).

[9] S. J. Summers and R. Werner, J. Math. Phys. **28**, 2440 (1987).

[10] S. Popescu and D. Rohrlich, Phys. Lett. A **169**, 411 (1992).

[11] B. S. Tsirelson, Hadron. J. Suppl. **8**, 329 (1993).

[12] B. W. Reichardt, F. Unger, and U. Vazirani, Nature (London) **496**, 456 (2013).

[13] A. Coladangelo, K. T. Goh, and V. Scarani, Nat. Commun. **8**, 15485 (2017).

[14] M. McKague, in *Theory of Quantum Computation, Communication, and Cryptography TQC 2011*, edited by D. Bacon, M. Martin-Delgado, and M. Roetteler, Lecture Notes in Computer Science Vol. 6745 (Springer, Berlin, Heidelberg, 2014), pp. 104–120.

[15] K. F. Pál, T. Vértesi, and M. Navascués, Phys. Rev. A **90**, 042340 (2014).

[16] X. Wu, Y. Cai, T. H. Yang, H. N. Le, J.-D. Bancal, and V. Scarani, Phys. Rev. A **90**, 042339 (2014).

[17] C.-E. Bardyn, T. C. H. Liew, S. Massar, M. McKague, and V. Scarani, Phys. Rev. A **80**, 062327 (2009).

[18] M. McKague, T. H. Yang, and V. Scarani, J. Phys. A: Math. Theor. **45**, 455304 (2012).

[19] T. H. Yang and M. Navascués, Phys. Rev. A **87**, 050102(R) (2013).

[20] C. Bamps and S. Pironio, Phys. Rev. A **91**, 052111 (2015).

[21] T. H. Yang, T. Vértesi, J.-D. Bancal, V. Scarani, and M. Navascués, Phys. Rev. Lett. **113**, 040401 (2014).

[22] J. Kaniewski, Phys. Rev. Lett. **117**, 070402 (2016).

[23] J. Kaniewski, Phys. Rev. A **95**, 062323 (2017).

[24] T. R. Tan, Y. Wan, S. Erickson, P. Bierhorst, D. Kienzler, S. Glancy, E. Knill, D. Leibfried, and D. J. Wineland, Phys. Rev. Lett. **118**, 130403 (2017).

[25] I. Supic, R. Augusiak, A. Salavrakos, and A. Acin, New J. Phys. **18**, 035013 (2016).

[26] R. Gallego, N. Brunner, C. Hadley, and A. Acin, Phys. Rev. Lett. **105**, 230501 (2010).

[27] M. Hendrych, R. Gallego, M. Micuda, N. Brunner, A. Acin, and J. P. Torres, Nat. Phys. **8**, 588 (2012).

[28] J. Ahrens, P. Badziag, A. Cabello, and M. Bourennane, Nat. Phys. **8**, 592 (2012).

[29] M. Pawłowski and N. Brunner, Phys. Rev. A **84**, 010302(R) (2011).

[30] H.-W. Li, Z.-Q. Yin, Y.-C. Wu, X.-B. Zou, S. Wang, W. Chen, G.-C. Guo,and Z.-F. Han, Phys. Rev. A **84**, 034301 (2011).

[31] E. Woodhead and S. Pironio, Phys. Rev. Lett. **115**, 150501 (2015).

[32] T. Lunghi, J. B. Brask, C. C. W. Lim, Q. Lavigne, J. Bowles, A. Martin, H. Zbinden, and N. Brunner, Phys. Rev. Lett. **114**, 150501 (2015).

[33] P. Mironowicz, A. Tavakoli, A. Hameedi, B. Marques, M. Pawłowski, and M. Bourennane, New J. Phys. **18**, 065004 (2016).

[34] M. Dall'Arno, S. Brandsen, F. Buscemi, and V. Vedral, Phys. Rev. Lett. **118**, 250501 (2017).

[35] M. Dall'Arno, S. Brandsen, and F. Buscemi, Proc. R. Soc. A **473**, 20160721 (2017).

[36] M. Dall'Arno, arXiv:1702.00575.

[37] A. Ambainis, A. Nayak, A. Ta-Shama, and U. Vazirani, *Proceedings of 31st ACM Symposium on Theory of Computing* (ACM, Atlanta, 1999), pp. 376–383.

[38] A. Nayak, *Proceedings of the 40th IEEE Symposium on Foundations of Computer Science (FOCS'99)* (IEEE, New York, 1999), pp. 369–376.

[39] A. Tavakoli, A. Hameedi, B. Marques, and M. Bourennane, Phys. Rev. Lett. **114**, 170502 (2015).

[40] E. Woodhead, C. C. W. Lim, and S. Pironio, in *Theory of Quantum Computation, Communication, and Cryptography, TQC 2012*, edited by K. Iwama, Y. Kawano, and M. Murao, Lecture Notes in Computer Science Vol. 7582 (Springer, Berlin, Heidelberg, 2013), pp. 107–115

[41] M. Navascués and T. Vértesi, Phys. Rev. Lett. **115**, 020501 (2015).

[42] M. Navascués, A. Feix, M. Araújo, and T. Vértesi, Phys. Rev. A **92**, 042117 (2015).

[43] A. Tavakoli, D. Rosset, and M. O. Renou, arXiv:1808.02412.

[44] P. Trojek, C. Schmid, M. Bourennane, C. Brukner, M. Żukowski, and H. Weinfurter, Phys. Rev. A **72**, 050305(R) (2005).

[45] J. Bowles, N. Brunner, and M. Pawłowski, Phys. Rev. A **92**, 022351 (2015).

[46] T. Van Himbeeck, E. Woodhead, N. J. Cerf, R. García-Patron, and S. Pironio, Quantum **1**, 33 (2017).

[47] J. B. Brask, A. Martin, W. Esposito, R. Houlmann, J. Bowles, H. Zbinden, and N. Brunner, Phys. Rev. Appl. **7**, 054018 (2017).

[48] R. Chaves, J. B. Brask, and N. Brunner, Phys. Rev. Lett. **115**, 110501 (2015).