# Robust self-testing of two-qubit states

Coopmans, Tim; Kaniewski, Jdrzej; Schaffner, Christian

**DOI**
10.1103/PhysRevA.99.052123

**Publication date**
2019

**Document Version**
Final published version

**Published in**
Physical Review A

**Citation (APA)**
Coopmans, T., Kaniewski, J., & Schaffner, C. (2019). Robust self-testing of two-qubit states. *Physical Review A*, *99*(5), Article 052123. https://doi.org/10.1103/PhysRevA.99.052123

# Robust self-testing of two-qubit states

Tim Coopmans,[1] Jędrzej Kaniewski,[2] and Christian Schaffner[3]
[1] *QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, Netherlands*
[2] *Center for Theoretical Physics, Polish Academy of Sciences, Aleja Lotników 32/46, 02-668 Warsaw, Poland*
[3] *QuSoft, University of Amsterdam, Science Park 123, 1098 XG Amsterdam, Netherlands*

(Received 15 February 2019; published 24 May 2019)

It is well known that observing nonlocal correlations allows us to draw conclusions about the quantum systems under consideration. In some cases this yields a characterisation which is essentially complete, a phenomenon known as self-testing. Self-testing becomes particularly interesting if we can make the statement robust, so that it can be applied to a real experimental setup. For the simplest self-testing scenarios the most robust bounds come from the method based on operator inequalities. In this work we elaborate on this idea and apply it to the family of tilted Clauser-Horne-Shimony-Holt (CHSH) inequalities. These inequalities are maximally violated by partially entangled two-qubit states and our goal is to estimate the quality of the state based only on the observed violation. For these inequalities we have reached a candidate bound and while we have not been able to prove it analytically, we have gathered convincing numerical evidence that it holds. Our final contribution is a proof that in the usual formulation, the CHSH inequality only becomes a self-test when the violation exceeds a certain threshold. This shows that self-testing scenarios fall into two distinct classes depending on whether they exhibit such a threshold or not.

DOI: 10.1103/PhysRevA.99.052123

## I. INTRODUCTION

Among the many sins of quantum mechanics, correlations between spacelike separated systems occupy a rather special place. Stronger-than-classical correlations [1,2] were initially seen as a problem, but have now become an inherent (and useful) feature of the quantum world. Investigating the difference between correlations achievable in quantum mechanics and in classical (local-realistic) theories goes under the name of Bell nonlocality [3], and one of the great achievements of this field is the ability to rule out any classical description of the system under consideration based only on the observed statistics. While clearly of fundamental importance, it turns out that this argument can be pushed one step further.

If we can rule out a classical description, our next guess is that the system is governed by quantum mechanics. Under this assumption, it makes sense to ask which features of the quantum system give rise to such strikingly nonclassical behavior. Can we, for instance, deduce something about the quantum state or the measurements performed?

While it is clear that in order to observe nonlocal correlations one must perform incompatible measurements on entangled quantum systems, it is not clear which meaningful quantitative statements can be made. It might therefore come as a surprise that certain nonlocal correlations can be realized in an essentially unique manner. While this observation can be found in the early works of Tsirelson [4,5], Summers and Werner [6], and Popescu and Rohrlich [7], it did not attract much attention until the seminal work of Mayers and Yao [8,9]. Mayers and Yao realized that this effect can be used to certify quantum devices under minimal assumptions and they called this phenomenon self-testing.

The goal of self-testing is to make quantitative statements about the quantum realization, e.g., about the entanglement present in a quantum state or about the incompatibility of the measurements performed. Self-testing is closely related to the field of device-independent cryptography whose goal is to certify properties of the classical output produced by quantum devices. Device-independent cryptography is a promising solution for randomness generation [10–15], quantum key distribution [16–21], and several other tasks [22–26]. For a brief overview of device-independent cryptography, we recommend Ref. [27] (focus on quantum key distribution) and Ref. [28] (focus on randomness generation). For a comprehensive review of both philosophical and technological aspects of randomness in quantum physics, we refer the reader to Ref. [29].

In this work we focus solely on the task of self-testing in its most common formulation, i.e., when the goal is to certify the state and the measurements performed on it.[1] While there is a large class of scenarios in which self-testing statements have been proven, most results only apply if the observed statistics are (almost) perfect [34–47]. While such results are robust in the sense that they are stable under sufficiently small perturbations, the obtained noise tolerance is not relevant from the experimental point of view. Our goal, on the other hand, is to derive self-testing statements which can be applied to real statistics collected in real experiments.[2] Such results are

[1] Note that other quantum objects such as quantum channels [30], entangled measurements [31,32], or weak measurements [33] can be self-tested in more complex scenarios.

[2] For an intuitive explanation of the difference between robustness and experimentally-relevant robustness see Sec. I of Ref. [48].

of interest to both experimentalists [49–51] and theoreticians investigating specific physical setups [52], but deriving them turns out to be significantly more challenging.

The first result of this type is due to Bardyn *et al.* [34] and there are currently two methods of deriving such results: the swap method [53–55] and the self-testing from operator inequalities (STOPI) method [56]. While the swap method is extremely versatile and can be (at least in principle) immediately applied to any Bell scenario, it has two weaknesses. First of all, it is a numerical method which scales unfavorably with the dimension of the system we wish to certify: The largest states certified using this method until today consist of two ququarts [57] or four qubits [55]. The second, and more severe, disadvantage of the swap method is that the output of the computation is just a number, which gives little intuition about the underlying physics.

The STOPI method, on the other hand, is more time consuming, as it requires a more thorough understanding of the particular self-testing scenario, but the resulting bounds are significantly stronger (in some cases even tight). In Ref. [56] the STOPI method was used to derive analytic self-testing bounds for the Clauser-Horne-Shimony-Holt (CHSH) [58] and Mermin [59] inequalities. In this work we applied this method to self-test partially entangled pure two-qubit states using the family of tilted CHSH inequalities. Investigating some special cases led us to conjecture a particular form of the self-testing statement. While we were not able to prove it analytically, we have gathered strongly convincing numerical evidence that it holds. The conjectured statement improves on the bounds obtained from the swap method [53].

Our second contribution is a proof that the CHSH inequality becomes a self-test only above a certain violation. More specifically, we have constructed a state which violates the CHSH inequality, but does not satisfy the usual self-testing criteria. This is in contrast with the Mermin inequality in which the value of the self-testing threshold coincides with the maximum value achievable if only two out of three parties are entangled.

In Sec. II we formalize the problem of self-testing, while in Sec. III we explain the STOPI method. In Sec. IV we present the conjectured robust self-testing bounds for partially entangled two-qubit states. In Sec. V we explain the construction of the state that violates the CHSH inequality but for which none of the usual self-testing statements can be made. In Sec. VI we summarize our results and discuss some open problems.

## II. PRELIMINARIES

In this section we establish the basic notation and formalize the problem of self-testing.

### A. Notation

We denote the identity matrix by $\mathbb{1}$ and the Pauli matrices by $\mathsf{X}$, $\mathsf{Y}$, and $\mathsf{Z}$. For a Hermitian matrix $X$ we use $\lambda_{\max}(X)$ and $\lambda_{\min}(X)$ to denote its largest and smallest eigenvalue, respectively.

For arbitrary linear operators $X$ and $Y$ we use $\langle X, Y\rangle := \operatorname{tr}(X^\dagger Y)$ to denote the Hilbert-Schmidt inner product and $\|X\|_p$ to denote the Schatten $p$-norm. For a positive-semidefinite operator $A$, $B := \sqrt{A}$ is the unique positive-semidefinite operator satisfying $B^2 = A$. The fidelity of two positive-semidefinite operators $A$ and $B$ is defined as $F(A, B) = \|\sqrt{A}\sqrt{B}\|_1^2$.

The Hilbert space corresponding to register $A$ is denoted by $\mathcal{H}_A$ and in this work we assume all the Hilbert spaces to be finite dimensional. The set of linear operators acting on $\mathcal{H}$ is denoted by $\mathcal{L}(\mathcal{H})$.

For a completely positive map $\Lambda : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$, the dual map $\Lambda^\dagger : \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_A)$ is the unique linear map which satisfies $\langle \Lambda(X), Y\rangle = \langle X, \Lambda^\dagger(Y)\rangle$ for all $X \in \mathcal{L}(\mathcal{H}_A)$ and $Y \in \mathcal{L}(\mathcal{H}_B)$. The map $\Lambda$ is a quantum channel if it is trace preserving, which is equivalent to the dual map $\Lambda^\dagger$ being unital, i.e., $\Lambda^\dagger(\mathbb{1}_B) = \mathbb{1}_A$.

The Choi-Jamiołkowski isomorphism states that completely positive maps $\Lambda : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ are in 1:1 correspondence with positive semidefinite operators acting on $\mathcal{H}_A \otimes \mathcal{H}_B$. Let $\{|j\rangle\}_{j=1}^d$ be the standard basis on $\mathcal{H}_A$, let $\mathcal{H}_{A'} \cong \mathcal{H}_A$, and let

$$\Omega_{AA'} = |\Omega\rangle\langle\Omega|_{AA'} \quad \text{for } |\Omega\rangle_{AA'} = \sum_{j=1}^{d} |j\rangle_A |j\rangle_{A'}$$

be an unnormalized maximally entangled state. The (unnormalized) Choi state of $\Lambda$, denoted by $C_{AB}$, is defined as

$$C_{AB} := (\mathrm{id}_A \otimes \Lambda_{A'})(\Omega_{AA'})$$

and it is well known that for any $X \in \mathcal{L}(\mathcal{H}_A)$,

$$\Lambda(X) = \operatorname{tr}_A\left[C_{AB}\left(X_A^{\mathsf{T}} \otimes \mathbb{1}_B\right)\right],$$

where $\mathsf{T}$ denotes the transpose in the standard basis. If $\Lambda$ is trace preserving, then $C_A = \mathbb{1}$, whereas if $\Lambda$ is unital, then $C_B = \mathbb{1}$.

### B. Self-testing of quantum states

Consider the usual Bell scenario in which two spacelike separated parties, Alice and Bob, perform local measurements on a shared quantum state. Alice and Bob would like to certify that the state they share is entangled, but as they do not trust their measurement devices, they are unable to perform full state tomography. Their only option is to choose measurement settings, observe the outcomes, and collect statistics. To simplify the problem we assume that their devices behave in the same way every time they are used, i.e., that they give rise to a well-defined conditional probability distribution $\Pr(a, b|x, y)$, where $a$ and $b$ are outputs and $x$ and $y$ are inputs of Alice and Bob, respectively. Since the probability vector $\mathrm{P} = [\Pr(a, b|x, y)]_{abxy}$ can be estimated to arbitrary precision and we are interested in the fundamental aspects of self-testing, we assume to have access directly to the exact probability distribution P.[3]

From a mathematical point of view, self-testing of quantum states is in essence a matter of the following question: Given a conditional probability distribution

$$\mathrm{P} = [\Pr(a, b|x, y)]_{abxy},$$

[3] Not surprisingly, drawing conclusions from a finite set of data is significantly harder (see Refs. [60–62]).

which comes from measuring a quantum system, i.e.,

$$\Pr(a, b|x, y) = \operatorname{tr}\left[\left(P_a^x \otimes Q_b^y\right)\rho_{AB}\right],$$

what can we deduce about the unknown state $\rho_{AB}$? We intentionally denote the unknown state by $\rho_{AB}$, as we do not want to assume its purity.[4] Let us also emphasize here that no knowledge of the observables is assumed, which makes self-testing a significantly different problem from quantum state tomography.

It is important to realize that the observed statistics $\Pr(a, b|x, y)$ can never uniquely determine the state. Indeed, the two equivalences we must always allow for are (i) local isometries and (ii) the presence of additional degrees of freedom. Motivated by these limitations, we say that $\rho_{AB}$ contains $\sigma_{A'B'}$ if there exist local quantum channels $\Lambda_A : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_{A'})$ and $\Lambda_B : \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_{B'})$ that extract a perfect copy of $\sigma_{A'B'}$ from $\rho_{AB}$, i.e.,

$$(\Lambda_A \otimes \Lambda_B)(\rho_{AB}) = \sigma_{A'B'}. \tag{1}$$

It is intuitively clear that this formulation is equivalent to the usual formulation using isometries and an auxiliary state, but for completeness we provide a proof in Appendix A.[5]

The concept of local extraction channels is well aligned with the conditions of a Bell test in which Alice and Bob are only allowed local measurements (no communication) and they must always produce an outcome (from a fixed alphabet). Similarly, we require the extraction channels to act locally and deterministically produce a state (of the correct dimension).

Replacing local extraction channels by a distillation procedure, i.e., allowing for classical communication, completely changes the problem. Note that the same phenomenon occurs in Bell nonlocality, where states can be preprocessed to enhance their nonlocal properties [64].

A self-testing statement consists of two components: (i) a quantum-realizable probability distribution P* and (ii) a pure bipartite state $\Phi_{A'B'}$. The statement asserts that if an unknown state $\rho_{AB}$ is capable of producing the probability distribution P* (under some local measurements), then $\rho_{AB}$ must contain $\Phi_{A'B'}$.

Of course, in a real experiment one never actually observes the exact probability distribution P*,[6] which means that an improved, robust version of Eq. (1) is needed. For exactly that purpose the channel formulation is particularly convenient, as it is immediately clear how to turn the original requirement into an approximate statement. We define the extractability of $\Phi_{A'B'}$ from $\rho_{AB}$ as [34,56]

$$\Xi(\rho_{AB} \to \Phi_{A'B'}) := \max_{\Lambda_A, \Lambda_B} F((\Lambda_A \otimes \Lambda_B)(\rho_{AB}), \Phi_{A'B'}), \tag{2}$$

where the maximization is taken over all quantum channels from $A$ to $A'$ and $B$ to $B'$, respectively. It is clear that extractability is invariant under local unitaries applied to $\Phi_{A'B'}$, i.e., it depends only on the Schmidt coefficients of the target state. The maximal value of extractability equals 1 and implies that $\rho_{AB}$ contains $\Phi_{A'B'}$. The lowest value, on the other hand, equals $\lambda_0^2$, where $\lambda_0$ is the largest Schmidt coefficient of $\Phi_{A'B'}$, because Alice and Bob can always replace $\rho_{AB}$ with a pure product state. Moreover, extractability is convex in the input state, which implies that $\Xi(\rho_{AB} \to \Phi_{A'B'}) = \lambda_0^2$ whenever $\rho_{AB}$ is separable. Note that there exist other measures for robust self-testing, but extractability is the only one for which experimentally-relevant robustness has been proven (see Appendix A 2 for details).

In this language a self-testing statement says that if $\rho_{AB}$ is capable of producing P*, then $\Xi(\rho_{AB} \to \Phi_{A'B'}) = 1$. A robust version states that observing statistics close to P* implies that the extractability is close to 1. More generally, we are interested in deriving a nontrivial lower bound on $\Xi(\rho_{AB} \to \Phi_{A'B'})$ as a function of the observed statistics.

In this work, instead of looking at the entire probability distribution P, we focus on some suitably chosen Bell function. A Bell function is specified by a vector of real coefficients $(c_{abxy})_{abxy}$, and its value evaluated on the probability distribution P equals

$$\beta := \sum_{abxy} c_{abxy} \Pr(a, b|x, y).$$

If $\beta_C$ and $\beta_Q$ are the maximal classical and quantum values, respectively, then our goal is to prove

$$\Xi(\rho_{AB} \to \Phi_{A'B'}) \geqslant f(\beta) \tag{3}$$

for some explicit function $f : [\beta_C, \beta_Q] \to [0, 1]$. While in principle $f$ could be an arbitrary function, we can without loss of generality assume that it is nondecreasing. Since any state capable of producing the Bell violation of $\beta$ is also capable of producing any violation in the interval $[\beta_C, \beta]$, we can define

$$f_{\text{nd}}(\beta) := \sup_{x \in [\beta_C, \beta]} f(x),$$

where the subscript in $f_{\text{nd}}$ stands for nondecreasing, and we immediately see that

$$\Xi(\rho_{AB} \to \Phi_{A'B'}) \geqslant f_{\text{nd}}(\beta).$$

While such trade-offs could be investigated for arbitrary combinations of target state and Bell function, the term self-testing is only used if the maximal violation of the Bell function certifies the presence of the target state, i.e., $f(\beta_Q) = 1$. A self-testing statement is called robust if $f(\beta) \to 1$ as $\beta \to \beta_Q$.

An important advantage of self-testing statements based only on the Bell value is the fact that we can assess their tightness by deriving an explicit upper bound on $f(\beta)$. If the Bell inequality is not violated, we cannot improve over the trivial bound of $\lambda_0^2$, i.e., $f(\beta_C) = \lambda_0^2$. On the other extreme, by assumption we have $f(\beta_Q) = 1$. Since every intermediate violation can be achieved as a mixture of these two points, we cannot hope to certify extractability larger than the value corresponding to such a mixture. This leads to an upper bound of the form

$$f(\beta) \leqslant \lambda_0^2 + \left(1 - \lambda_0^2\right)\frac{\beta - \beta_C}{\beta_Q - \beta_C}. \tag{4}$$

[4] Under the purity assumption, even classical correlations are sufficient to certify entanglement [63].

[5] At the end of Appendix A, we also point out that the formulation with unitaries instead of isometries is not quite correct.

[6] The two most obvious obstacles are experimental noise and finite statistics.

This upper bound tells us how much room for improvement there potentially is and it is worth mentioning that in some scenarios, one can prove self-testing statements matching this upper bound [56]. A good indication of the strength of a self-testing bound is the critical Bell value above which the statement becomes nontrivial, i.e.,

$$\beta_f^* := \inf_\beta \{f(\beta) > \lambda_0^2\}.$$

Clearly, $\beta_f^*$ is computed for a specific self-testing bound (i.e., a particular function $f$) and is not a fundamental property of the Bell inequality under consideration.

## III. SELF-TESTING FROM OPERATOR INEQUALITIES

The STOPI method was introduced and applied to two specific examples in Ref. [56]. Here we provide a more detailed discussion of the underlying idea.

Our goal is to prove a lower bound on the extractability as a function of the observed Bell violation $\beta$. The STOPI method is constructive: Given a quantum realization, which consists of the shared state $\rho_{AB}$, the measurements of Alice $\{P_a^x\}$, and the measurements of Bob $\{Q_b^y\}$, we explicitly construct the local extraction channels $\Lambda_A$ and $\Lambda_B$ and we provide a lower bound on their performance as a function of $\beta$. The extraction channel of Alice $\Lambda_A : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_{A'})$ is built out of her measurement operators $\{P_a^x\}$ and similarly the extraction channel of Bob $\Lambda_B : \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_{B'})$ depends only on $\{Q_b^y\}$. We are interested in the fidelity

$$F((\Lambda_A \otimes \Lambda_B)(\rho_{AB}), \Phi_{A'B'}),$$

but since $\Phi_{A'B'}$ is a pure state, we can replace the fidelity by the inner product, which allows us to replace the channels by their duals

$$\begin{aligned} F((\Lambda_A \otimes \Lambda_B)(\rho_{AB}), \Phi_{A'B'}) &= \langle(\Lambda_A \otimes \Lambda_B)(\rho_{AB}), \Phi_{A'B'}\rangle \\ &= \langle\rho_{AB}, (\Lambda_A^\dagger \otimes \Lambda_B^\dagger)(\Phi_{A'B'})\rangle. \end{aligned}$$

Define

$$K := (\Lambda_A^\dagger \otimes \Lambda_B^\dagger)(\Phi_{A'B'}) \tag{5}$$

and note that this operator depends only on the measurement operators (and not on the input state $\rho_{AB}$). Another operator that depends only on the measurement operators is the Bell operator defined as

$$W := \sum_{abxy} c_{abxy} P_a^x \otimes Q_b^y,$$

which by construction satisfies $\mathrm{tr}(W\rho_{AB}) = \beta$. We might therefore hope to prove an operator inequality of the form

$$K \geqslant sW + \mu\mathbb{1} \tag{6}$$

for suitably chosen (real) constants $s$ and $\mu$. If we prove this operator inequality for all choices of local measurements on Alice and Bob, it immediately implies that for any input state $\rho_{AB}$ we have

$$\begin{aligned} \Xi(\rho_{AB} \to \Phi_{A'B'}) &\geqslant F((\Lambda_A \otimes \Lambda_B)(\rho_{AB}), \Phi_{A'B'}) \\ &= \langle\rho_{AB}, K\rangle \geqslant \langle\rho_{AB}, sW + \mu\mathbb{1}\rangle \\ &= s\beta + \mu. \end{aligned}$$

Therefore, we obtain precisely a self-testing statement of the form given in Eq. (3) for

$$f(\beta) = s\beta + \mu.$$

This approach reduces the problem of self-testing to three steps: (i) constructing suitable extraction channels, (ii) choosing the right constants $s$ and $\mu$, and (iii) proving the resulting operator inequality.

### A. Extraction channels from measurement operators

Given a set of measurements operators $\{P_a^x\}$ acting on $\mathcal{H}_A$, we want to construct an extraction channel $\Lambda_A : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_{A'})$, where the Hilbert space $\mathcal{H}_{A'}$ is determined by the target state. Let us first point out that for the purpose of deriving self-testing statements it suffices to construct channels for projective measurement operators. In the case of nonprojective measurement operators, Alice starts her extraction procedure by enlarging her Hilbert space until she can find projective measurements reproducing precisely the same statistics. She would then construct an extraction channel using the new, projective measurement operators.

Instead of first constructing the channel and then taking its dual, it is easier to construct the dual channel $\Lambda^\dagger : \mathcal{L}(\mathcal{H}_{A'}) \to \mathcal{L}(\mathcal{H}_A)$ directly and it is convenient to specify it through its Choi state. The dual channel must be unital, so the Choi state $C_{A'A}$ must satisfy $C_A = \mathbb{1}$. If $\{O_j\}_j \in \mathcal{L}(\mathcal{H}_{A'})$ is an operator basis on $\mathcal{L}(\mathcal{H}_{A'})$, the Choi state can be written as

$$C_{A'A} := \sum_j O_j \otimes F_j(\{P_a^x\})$$

for some collection of functions $\{F_j\}$ such that $F_j : [\mathcal{L}(\mathcal{H}_A)]^{\times k} \to \mathcal{L}(\mathcal{H}_A)$, where $k$ is the product of the number of inputs and outputs. In principle, the only restriction on $\{F_j\}$ is that the resulting operator must be a valid Choi operator for all sets of valid measurement operators $\{P_a^x\}$, but it is natural to choose extraction channels satisfying certain conditions.

First of all, sets of measurement operators which are related by a unitary should be treated in an equivalent manner, i.e.,

$$F_j(\{UP_a^xU^\dagger\}) = UF_j(\{P_a^x\})U^\dagger$$

for all unitaries $U$ and all $j$. We call such extraction channels covariant with respect to the unitary group.

Moreover, whenever the measurement operators exhibit a certain direct-sum structure, the extraction channels should preserve it. Given one set of measurements $\{P_a^{x,0}\}$ acting on $\mathcal{H}_{A_0}$ and another set of measurements $\{P_a^{x,1}\}$ acting on $\mathcal{H}_{A_1}$, we should have

$$F_j(\{P_a^{x,0} \oplus P_a^{x,1}\}) = F_j(\{P_a^{x,0}\}) \oplus F_j(\{P_a^{x,1}\}).$$

Restricting ourselves to extraction channels satisfying these two criteria makes it easier to analyze the resulting operator inequalities. As explained in the next section, these restrictions do not affect the obtained bounds.

Since the target state is pure, we can assume that $\mathcal{H}_{B'} \cong \mathcal{H}_{A'}$ and we can choose the same operator basis for $\mathcal{H}_{B'}$. Analogously to $C_{A'A}$, the Choi state describing $\Lambda_B^\dagger$ reads

$$C_{B'B} := \sum_j O_j \otimes G_j(\{Q_b^y\}).$$

Computing the $K$ operator gives

$$\begin{aligned} K &= (\Lambda_A^\dagger \otimes \Lambda_B^\dagger)(\Phi_{A'B'}) \\ &= \operatorname{tr}_{A'B'}\left[(C_{A'A} \otimes C_{B'B})\big(\Phi_{A'B'}^{\mathsf{T}} \otimes \mathbb{1}_{AB}\big)\right] \\ &= \sum_{jk} \alpha_{jk} F_j\big(\{P_a^x\}\big) \otimes G_k\big(\{Q_b^y\}\big), \end{aligned}$$

where $\alpha_{jk} := \operatorname{tr}[(O_j \otimes O_k)\Phi_{A'B'}^{\mathsf{T}}]$.

### B. Choosing the constants

Since we are interested in nondecreasing functions of $\beta$, we restrict ourselves to the case $s > 0$, but otherwise all values of $s$ are in principle worth considering. For a particular choice of extraction channels and $s$, we define

$$\mu(s) := \inf \lambda_{\min}(K - sW), \tag{7}$$

where the infimum is taken over all possible measurements of Alice and Bob (in all finite dimensions). Clearly, this is simply the largest value of $\mu$ for which the operator inequality (6) holds for all possible measurements. To see that $\mu(s)$ does not diverge to $-\infty$, note that

$$\mu(s) \geqslant \inf \lambda_{\min}(-sW) \geqslant -\sup \|sW\|_\infty \geqslant -s\sum_{abxy} |c_{abxy}|.$$

It should now be clear why the restrictions discussed in the preceding section simplify the computation of $\mu(s)$. Requiring the extraction channels to be covariant ensures that the spectrum of $K - sW$ is not affected by applying local unitaries to the measurement operators of Alice and Bob, which significantly reduces the parameter space. Requiring the channels to preserve the direct-sum structure ensures that the same direct-sum structure is inherited by the operator $K - sW$ which facilitates bounding its spectrum.

The quantity $\mu(s)$ is in general hard to compute, but if we were able to do so for a fixed choice of extraction channels, then we would obtain a family of lower bounds of the form

$$f_s(\beta) = s\beta + \mu(s)$$

parametrized by $s > 0$.[7] All these bounds could be collected in a single function defined as

$$\sup_{s>0}\,(s\beta + \mu(s)).$$

In fact, we could also optimize over the choice of extraction channels. Such an optimization might seem particularly advantageous as we would expect that extraction channels in the regime $\beta \approx \beta_Q$ should be rather different from those in the regime $\beta \approx \beta_C$. It is therefore rather surprising that in all the examples considered in Ref. [56] and in this work, the best lower bounds come from a single choice of extraction channels and a single value of $s$. This situation stands in contrast to the swap method in which it is beneficial to tailor the extraction channels to the observed violation [see Eqs. (33) and (34) of Ref. [53]]. In this work we focus on the case where all the systems are finite dimensional, but the method can be equally well applied to infinite-dimensional systems as long as the construction of extraction channels from measurement operators and the proof of the relevant operator inequality carry over to the infinite-dimensional case.

### C. Extracting a qubit from two binary observables

A binary measurement $\{P_0, P_1\}$ can be conveniently represented as an observable $A := P_0 - P_1$ (and since $P_0 + P_1 = \mathbb{1}$ this mapping is a bijection). An observable is a Hermitian operator $A = A^\dagger$ satisfying $-\mathbb{1} \leqslant A \leqslant \mathbb{1}$, whereas projective measurements give rise to observables satisfying $A^2 = \mathbb{1}$.

The case of two binary observables is particularly simple due to Jordan's lemma, which completely characterizes the interaction between two projective observables. More specifically, it states that given two projective observables $A_0$ and $A_1$, one can find a unitary which simultaneously block diagonalizes $A_0$ and $A_1$ into blocks of size 1 × 1 or 2 × 2. There are four distinct types of 1 × 1 blocks corresponding to $A_0 = \pm 1, A_1 = \pm 1$, whereas the 2 × 2 blocks form a one-parameter family given by

$$A_0 := \cos(a)\mathsf{X} + \sin(a)\mathsf{Z}, \tag{8}$$

$$A_1 := \cos(a)\mathsf{X} - \sin(a)\mathsf{Z} \tag{9}$$

for $a \in (0, \pi/2)$. In Sec. III A we have argued that by enlarging the Hilbert space we can focus solely on projective measurements. Similarly, in this case we could enlarge the Hilbert space to ensure that every 1 × 1 block is paired up with another suitably chosen 1 × 1 block such that the two together are unitarily equivalent to a 2 × 2 block corresponding to $a = 0$ or $a = \pi/2$. As before, this grouping operation would be the first step of the extraction channel. It is not strictly necessary, but it makes the analysis easier, since it ensures that the observables are just a direct sum of 2 × 2 blocks parametrized by $a \in [0, \pi/2]$.

Since we restrict ourselves to covariant extraction channels, we can assume that the observables are already in block-diagonal form. Moreover, the channels respect the direct-sum structure, which implies that we only need to propose a one-parameter family of qubit channels corresponding to the 2 × 2 blocks. If the extraction channels for Alice and Bob are denoted by $\Lambda_A(a)$ and $\Lambda_B(b)$, respectively, then

$$K(a, b) := [\Lambda_A^\dagger(a) \otimes \Lambda_B^\dagger(b)](\Phi_{A'B'})$$

is a 4 × 4 operator. Similarly, let $W(a, b)$ be the 4 × 4 Bell operator constructed from local qubit observables corresponding to angles $a$ and $b$ for Alice and Bob, respectively. Due to the block structure, computing the lowest eigenvalue of $K - sW$ simplifies to

$$\mu(s) := \inf \lambda_{\min}(K - sW) = \min_{a,b} \lambda_{\min}(K(a, b) - sW(a, b)),$$

where the minimization is performed over the square $(a, b) \in [0, \pi/2] \times [0, \pi/2]$. This procedure is precisely the approach used to derive robust self-testing statements in Ref. [56]. In the following section, we apply it to the case of the tilted CHSH inequality.

[7] Every $s > 0$ gives a valid bound, but for poor choices of extraction channels and/or the parameter $s$, the bound might be trivial for the entire range of $\beta \in [\beta_C, \beta_Q]$.

## IV. ROBUST SELF-TESTING OF ALL ENTANGLED TWO-QUBIT STATES

Acín *et al.* introduced a family of Bell functions which are now commonly referred to as the tilted CHSH family [65]. The corresponding Bell operator reads

$$W_\alpha := \alpha A_0 \otimes \mathbb{1} + A_0 \otimes (B_0 + B_1) + A_1 \otimes (B_0 - B_1), \quad (10)$$

where $\alpha \in [0, 2)$ is a parameter. The classical and quantum values of this Bell function equal $\beta_C = 2 + \alpha$ and $\beta_Q = \sqrt{8 + 2\alpha^2}$, respectively. Clearly, for all values of $\alpha$ we have $\beta_Q > \beta_C$, although the gap vanishes as $\alpha \to 2$. The quantum value can be achieved using a pure state of two qubits $\Phi^\alpha_{A'B'} = |\Phi^\alpha\rangle\langle\Phi^\alpha|_{A'B'}$ for

$$|\Phi^\alpha\rangle_{A'B'} := \cos\theta_\alpha |u_0\rangle_{A'}|v_0\rangle_{B'} + \sin\theta_\alpha |u_1\rangle_{A'}|v_1\rangle_{B'},$$

where $\{|u_0\rangle, |u_1\rangle\}$ and $\{|v_0\rangle, |v_1\rangle\}$ are orthonormal bases on a qubit and

$$\theta_\alpha := \frac{1}{2}\arcsin\left(\sqrt{\frac{4-\alpha^2}{4+\alpha^2}}\right). \quad (11)$$

While the optimal observables of Alice are always maximally incompatible, which corresponds to setting $a = \pi/4$ in Eqs. (8) and (9), the optimal angle on Bob's side changes with $\alpha$ according to

$$b^*_\alpha := \arcsin\left(\sqrt{\frac{4-\alpha^2}{8}}\right).$$

Performing these measurements on this particular state turns out to be essentially the only manner of achieving the maximal violation, i.e., this Bell inequality is a self-test [37,38]. Since the range $\alpha \in [0, 2)$ is mapped onto $\theta_\alpha \in (0, \pi/4]$, it allows us to self-test every pure entangled state of two qubits. Clearly, setting $\alpha = 0$ yields the CHSH inequality for which the STOPI method gives strong self-testing bounds [56] and in this work we apply this approach to the entire range $\alpha \in [0, 2)$.

Before stating the conjectured bound, let us briefly explain the construction of extraction channels and the choice of constants $s_\alpha$ and $\mu_\alpha$. The optimal channels for the CHSH case correspond to full dephasing in X for $a = 0$, full dephasing in Z for $a = \pi/2$, and an identity channel for $a = \pi/4$. This choice is correct for Alice, because her optimal angle is always $\pi/4$, but for Bob we must introduce a modification which shifts the occurrence of the identity channel to his optimal angle $b^*_\alpha$. This modification can be achieved by defining an effective angle which uniformly extends the interval $[0, b^*_\alpha]$ to $[0, \pi/4]$ and simultaneously shrinks the interval $[b^*_\alpha, \pi/2]$ to $[\pi/4, \pi/2]$. After this modification one can check that this choice of channels performs well on the vertices of the square $(a, b) \in \{(0, 0), (0, \pi/2), (\pi/2, 0), (\pi/2, \pi/2)\}$ and the point of maximal violation $(a, b) = (\pi/4, b^*_\alpha)$. We choose the constant $s_\alpha$ so that the smallest eigenvalue of the operator $K - s_\alpha W$ occurs at multiple points $(a, b)$. In the case of CHSH, i.e., for $\alpha = 0$, we can obtain the same smallest eigenvalue on all the vertices and the point of maximal violation. However, the case of $\alpha > 0$ is less symmetric and the optimal choice of $s_\alpha$ only equalizes the smallest eigenvalue at two vertices and the point of maximal violation. Since the operators corresponding to the five special points (the vertices and the point of maximal violation) are easy to analyze (the operators $K$ and $W$ are diagonal in the same basis), our choice of $s_\alpha$ and $\mu_\alpha$ is given by analytic expressions. One can then check that the resulting operator inequality holds at these points for the entire range of $\alpha \in [0, 2)$. Unfortunately, verifying the operator inequality on the rest of the square turns out to be much harder and we were not able to do it analytically. However, since the parameter space is bounded ($\alpha \in [0, 2)$, $a, b \in [0, \pi/2]$), one can generate a grid over this space and check the operator inequality at those points numerically. We have found that the operator inequality holds up to numerical accuracy (see Appendix B for details), which lends support to the following conjecture.

*Conjecture 1.* Let $\alpha \in [0, 2)$ and let $\rho_{AB}$ be a bipartite quantum state which achieves the tilted CHSH violation of $\beta_\alpha := \mathrm{tr}(W_\alpha \rho_{AB})$, where $W_\alpha$ is the Bell operator defined in Eq. (10). Then the extractability of $\Phi^\alpha_{A'B'}$ from $\rho_{AB}$ satisfies

$$\Xi\left(\rho_{AB} \to \Phi^\alpha_{A'B'}\right) \geqslant s_\alpha \beta_\alpha + \mu_\alpha$$

for

$$s_\alpha := \frac{(\sqrt{8+2\alpha^2} + 2 + \alpha)(3\sqrt{8+2\alpha^2} - \sqrt{4-\alpha^2} - \alpha\sqrt{2})}{4(2-\alpha)^2\sqrt{8+2\alpha^2}},$$

$$\mu_\alpha := 1 - s_\alpha\sqrt{8+2\alpha^2}.$$

In Fig. 1 we compare the conjectured bounds with the results obtained by Bancal *et al.* using the swap method [53].[8]

Note that if we trust the numerical package used to verify the operator inequality, this conjecture could be made into a rigorous bound by explicitly calculating the error term. The error term would consist of two components: the error observed numerically on the grid (for our grid this value is of the order of $10^{-9}$) and the discretization error. Unfortunately, since both $s_\alpha$ and $\mu_\alpha$ diverge as $\alpha \to 2$, the discretization error would necessarily diverge in this limit. Therefore, no finite grid enables us to obtain certified bounds for $\alpha$ arbitrarily close to 2.

[8] The formulation used in the swap method involves isometries rather than channels, but the auxiliary registers are traced out before computing fidelity with the target state [see Eqs. (10) and (11) in Ref. [53]]. Therefore, in both cases we obtain lower bounds on precisely the same quantity (see Appendix A for more details).

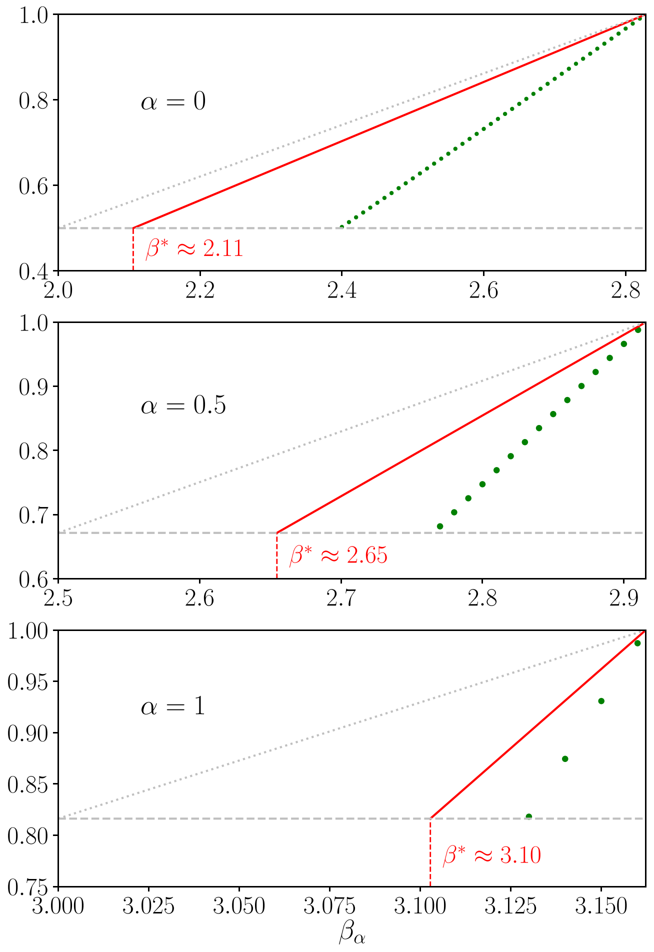

FIG. 1. Comparison of the conjectured lower bounds (solid line) with the previous results of Bancal *et al.* (green points) [53]. The range of $\beta_\alpha$ is chosen to cover the entire range between the classical and the quantum values. The dashed horizontal line indicates the trivial lower bound, whereas the dotted line corresponds to the upper bound given in Eq. (4). Vertical dashed lines mark the threshold violation $\beta^*$ above which the statement becomes nontrivial. The case of $\alpha = 0$ corresponds to the self-testing bound for the CHSH inequality derived in Ref. [56].

## V. NONTRIVIAL THRESHOLD VIOLATION FOR THE CHSH INEQUALITY

In Ref. [56] the STOPI method was used to derive robust bounds on self-testing the singlet[9] using the CHSH inequality. The resulting statement is nontrivial for any violation exceeding the threshold value of $\beta^*_{\text{CHSH}} := (16 + 14\sqrt{2})/17 \approx 2.11$ (recall that for the CHSH inequality we have $\beta_C = 2$ and $\beta_Q = 2\sqrt{2}$). We have tried to improve on this result, but we have not succeeded. In fact, the dephasing channels specified in the original paper seem to be by far the best choice.

This phenomenon made us wonder whether the existence of a threshold is an inherent feature of quantum mechanics, independent of the proof technique. In other words, maybe one can only make a self-testing statement for sufficiently large violations. The example below shows that this is indeed the case. More specifically, we have constructed a bipartite state which violates the CHSH inequality, but whose singlet extractability does not exceed the separable threshold of $\frac{1}{2}$. In this section we explain the construction of the state, briefly outline the idea of the proof, and discuss the implications of this result, while the technical details can be found in Appendix C.

Suppose that the system of Alice (Bob) consists of two subsystems: a three-dimensional classical register denoted by $X$ ($Y$) and a qubit denoted by $A$ ($B$). Consider the joint state

$$\rho_{XYAB} = \sum_{x,y=0}^{2} p_{xy} |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \rho_{AB}^{xy},$$

where $\{p_{xy}\}$ is a normalized probability distribution over $x, y \in \{0, 1, 2\}$ and $\rho_{AB}^{xy}$ are normalized two-qubit states to be specified later. The observables of Alice are given by

$$\begin{aligned} A_0 &= |0\rangle\langle 0|_X \otimes \mathsf{Z}_A + |1\rangle\langle 1|_X \otimes \mathsf{Z}_A + |2\rangle\langle 2|_X \otimes \mathsf{Z}_A, \\ A_1 &= |0\rangle\langle 0|_X \otimes \mathsf{Z}_A + |1\rangle\langle 1|_X \otimes \mathsf{X}_A + |2\rangle\langle 2|_X \otimes (-\mathsf{Z})_A. \end{aligned} \tag{12}$$

The observables of Bob are precisely the same, but act on subsystems $Y$ and $B$ instead of $X$ and $A$. Computing the CHSH operator[10] gives

$$W = \sum_{x,y=0}^{2} |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes W_{AB}^{xy},$$

where $W_{AB}^{xy}$ are the resulting two-qubit operators. Let us arrange the nine possible combination of $(x, y)$ on a $3 \times 3$ grid, where one axis corresponds to $x$ and the other axis corresponds to $y$. We will refer to the point $(x, y) = (1, 1)$ as the center, while the remaining eight points constitute the frame. The center allows for the optimal CHSH violation and we choose $\rho_{AB}^{11}$ to be the corresponding eigenstate of $W_{AB}^{11}$, i.e.,

$$\operatorname{tr}(W_{AB}^{11} \rho_{AB}^{11}) = 2\sqrt{2}.$$

For all the points on the frame the two-qubit operator is a product operator whose eigenvalues are $\{-2, 2\}$. We choose the states $\rho_{AB}^{xy}$ to be classically correlated and satisfy

$$\operatorname{tr}\left(W_{AB}^{xy} \rho_{AB}^{xy}\right) = 2.$$

Clearly, this setup violates the CHSH inequality as long as $p_{11} > 0$.

Now we would like to show that there exists a probability distribution satisfying $p_{11} > 0$ such that the resulting state $\rho_{XYAB}$ has singlet extractability of $\frac{1}{2}$. In general this is a hard task, as we must show that this value cannot be exceeded regardless of the choice of the extraction channels. Fortunately, the presence of classical registers significantly simplifies the problem due the following observation: Any quantum channel that acts simultaneously on classical and quantum registers can be simulated by first reading off the value of the classical register and then applying a particular quantum channel to

[9] As explained in Sec. II B in the context of self-testing, it is only the Schmidt coefficients that matter, so we use the term singlet to mean any (fixed) maximally entangled state of two qubits.

[10] The CHSH operator is obtained by setting $\alpha = 0$ in Eq. (10).

the quantum register (for completeness we provide a proof; see Lemma 2 in Appendix C). This observation implies that instead of considering channels from $\mathcal{L}(\mathbb{C}^3 \otimes \mathbb{C}^2)$ to $\mathcal{L}(\mathbb{C}^2)$, it suffices to consider triples (one corresponding to each value of the classical register) of qubit $[\mathcal{L}(\mathbb{C}^2) \rightarrow \mathcal{L}(\mathbb{C}^2)]$ channels.

All the states on the frame are classically correlated, but the local bases are different for different points. In fact, one can show that the only strategy that achieves optimal extraction (i.e., fidelity of $\frac{1}{2}$) on all the frame points corresponds to erasing the initial state and replacing it with a fixed product state. This operation is achieved precisely by the full amplitude-damping channel. On the other hand, in order to preserve entanglement of the state in the center, one should apply some nondestructive channels, e.g., unitaries. These two requirements are highly incompatible and this incompatibility is precisely what our proof hinges on. We choose a probability distribution concentrated on the frame, which forces Alice and Bob to perform channels close to full amplitude damping, and we show that such channels necessarily destroy the entanglement present in the center. The proof, which consists of a long sequence of elementary inequalities, can be found in Appendix C.

*Proposition 1.* There exists a bipartite state $\rho_{XYAB}$ which produces a CHSH violation of $\beta \approx 2.0014$ but nevertheless exhibits a singlet extractability of $\frac{1}{2}$.

This result can be interpreted in several ways. First of all, it implies that self-testing of the singlet using the CHSH inequality is only possible above some threshold. We find this insight rather surprising, since it shows that self-testing scenarios can be split up into two classes depending on whether they exhibit a threshold (like the CHSH inequality) or not (like the Mermin inequality [56]). Intuitively, one would conjecture that the presence of a threshold is generic and only in some special circumstances can we make self-testing statements arbitrarily close to the classical value $\beta_C$. Note that the Mermin inequality is frustration-free in the sense that the optimal quantum realisation simultaneously saturates every term of the Bell operator (contrary to the CHSH inequality). We conjecture that frustration-freeness is the source of strong self-testing properties.

We do not know what the exact threshold for the CHSH inequality is, but it must lie in between 2.0014 and $\beta^*_{\mathrm{CHSH}} \approx 2.11$. The analysis we perform could certainly be tightened to improve the lower limit of this interval, but one cannot hope for a significant improvement using our method.

It is important to realize that our result crucially relies on choosing the extractability as the quantity relevant for the task of self-testing and one can ask whether the same threshold phenomenon appears if we replace the fidelity with some other distance measure such as the trace distance. While we do not have a definite answer to this question, we would like to point out that extractability is the only quantity for which robust self-testing statements have been proven, i.e., it seems to be the most "forgiving" one. We therefore conjecture that if a threshold occurs for the extractability, it will also appear for any other quantity that accurately captures the task of self-testing (although the actual threshold values will of course be different).

We have shown that from the extractability point of view the state $\rho_{XYAB}$ is as uninteresting as any separable state, but it is clear that the entanglement becomes accessible when more general transformations are allowed. If we allow for nondeterministic entanglement extraction (Alice and Bob apply a local extraction map which either succeeds or fails and we only care about the performance if they both succeed), all the entanglement can be extracted. In a similar fashion the entanglement becomes accessible if we allow classical communication between Alice and Bob, i.e., we perform entanglement distillation. One could therefore ask whether a stronger counterexample could be found, in which we find a state which is not only nonextractable but also nondistillable. Such a counterexample is however not possible, because every state that violates the CHSH inequality is necessarily distillable [66].

At first glance our result seems related to the celebrated conjecture of Peres stating that undistillable states do not violate Bell inequalities [67] (recently disproved by Vértesi and Brunner [68]), but this similarity is rather superficial. Distillability is a fundamental property of entanglement and does not require any particular reference state. Singlet extractability, on the other hand, is defined with respect to a specific target state and is tailored specifically to the task of self-testing.

## VI. CONCLUSION AND OPEN QUESTIONS

In this work we have focused on the problem of self-testing in the channel formulation as proposed by Bardyn *et al.* [34]. We have discussed the recently proposed STOPI method and applied it to the tilted CHSH inequality. Moreover, we have shown that self-testing using the CHSH inequality is only possible above some threshold, which implies the existence of two fundamentally different classes of self-testing scenarios.

Let us conclude by presenting a couple of directions for future research. The first natural extension would be to look at scenarios with more than two parties, but still only two inputs and two outputs per party. The family of Mermin-Ardehali-Belinskii-Klyshko inequalities [59,69,70] is a promising candidate because it is permutation symmetric and the optimal observables are precisely the same as for the CHSH and Mermin inequalities. We therefore expect that applying the same channels could already give satisfactory results. A more challenging goal is to apply the STOPI method to scenarios going beyond Jordan's lemma, i.e., where the number of inputs or outputs is higher than 2. As this is not an easy task, it might be more tractable in a more restrictive setup, e.g., in a semi-device-independent scenario where one of the parties is trusted (equivalent to steering [71–73]). The STOPI method has been successfully applied to prepare-and-measure scenarios in which the transmitted system is a qubit [74] and one might also try to apply it to higher-dimensional cases (although one should remember that they are self-tests in a weaker sense [75]).

Another important concept that arises from this work is the threshold violation. We have shown that the CHSH inequality exhibits a threshold violation, but we have not pinned down the number. Computing the exact number is likely to be hard and moreover the actual value might depend on the specific formulation of self-testing, which makes it less interesting from a fundamental point of view. However, we see the sheer existence of a threshold as something that deserves a better understanding. We would first like to know whether there exists an alternative natural formulation of the self-testing problem for which the threshold does not appear. If that is not the case, it would be interesting to find out which features of the Bell inequality determine whether it exhibits a threshold or not and which of the two behaviors is generic. We would also like to have an example of a bipartite inequality without a threshold.

Let us finish by pointing out that while the current formulation of self-testing works well in some scenarios, there is some recent evidence that the problem of deducing properties of quantum systems from statistics alone is generically much harder, particularly in multipartite scenarios [76]. This evidence motivates more relaxed formulations of the problem, where instead of pinning down the exact state, we are happy to obtain a lower bound on some entanglement measure [77–81].

## ACKNOWLEDGMENTS

We thank Jean-Daniel Bancal for sharing numerical data with us. T.C. would like to thank the QMATH Centre and the Erasmus+ Programme for financial support during his research stay. J.K. acknowledges support under the POLONEZ program, which has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 665778. C.S. was supported by a NWO VIDI grant (Project No. 639.022.519).

## APPENDIX A: FORMULATIONS OF THE SELF-TESTING PROBLEM

In this Appendix we discuss possible formulations of the self-testing problem. In the first section we show that the three commonly used formulations are equivalent. In the second section we explain how to make these formulations robust and discuss the relations between the resulting inequivalent measures for robust self-testing.

### 1. Exact self-testing definitions

A linear map $V : \mathcal{H}_A \to \mathcal{H}_B$ is called an isometry if it satisfies $V^\dagger V = \mathbb{1}_A$. For a Hilbert space $\mathcal{H}$ let $\mathcal{S}(\mathcal{H})$ be the set of density operators acting on $\mathcal{H}$. Let $\mathcal{H}_X$ and $\mathcal{H}_{X'}$ for $X \in \{A, B\}$ be finite-dimensional Hilbert spaces. The target state $\Phi_{A'B'} \in \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$ is pure ($\Phi_{A'B'}^2 = \Phi_{A'B'}$) and its marginals ($\Phi_{A'} := \mathrm{tr}_{B'}\, \Phi_{A'B'}$ and $\Phi_{B'} := \mathrm{tr}_{A'}\, \Phi_{A'B'}$) are full rank [$\mathrm{rank}(\Phi_{X'}) = \dim(\mathcal{H}_{X'})$ for $X \in \{A, B\}$, which immediately implies $\dim(\mathcal{H}_{A'}) = \dim(\mathcal{H}_{B'})$]. The input state $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is arbitrary.

*Proposition 2.* The following three statements are equivalent.

(i) There exist completely positive trace-preserving maps $\Lambda_X : \mathcal{L}(\mathcal{H}_X) \to \mathcal{L}(\mathcal{H}_{X'})$ such that

$$(\Lambda_A \otimes \Lambda_B)(\rho_{AB}) = \Phi_{A'B'}. \tag{A1}$$

(ii) There exist Hilbert spaces $\mathcal{H}_{X''}$, isometries $V_X : \mathcal{H}_X \to \mathcal{H}_{X'} \otimes \mathcal{H}_{X''}$, and an auxiliary state $\sigma_{A''B''} \in \mathcal{S}(\mathcal{H}_{A''} \otimes \mathcal{H}_{B''})$ such that

$$V \rho_{AB} V^\dagger = \Phi_{A'B'} \otimes \sigma_{A''B''}, \tag{A2}$$

where $V = V_A \otimes V_B$ is the combined isometry.

(iii) There exist Hilbert spaces $\mathcal{H}_{X'''}$, isometries $W_X : \mathcal{H}_{X'} \otimes \mathcal{H}_{X'''} \to \mathcal{H}_X$, and an auxiliary state $\tau_{A'''B'''} \in \mathcal{S}(\mathcal{H}_{A'''} \otimes \mathcal{H}_{B'''})$ such that

$$\rho_{AB} = W(\Phi_{A'B'} \otimes \tau_{A'''B'''})W^\dagger, \tag{A3}$$

where $W = W_A \otimes W_B$ is the combined isometry.

Before proceeding to the proof, let us sketch how the three formulations are connected. The equivalence between (i) and (ii) is a direct consequence of Naimark's dilation theorem. The relation between (ii) and (iii), on the other hand, is more subtle and deserves a brief discussion. If the isometry $V$ in Eq. (A2) happens to be a unitary, we can just move it to the other side to obtain Eq. (A3) and the equivalence is trivial. However, if the dimensions do not match, i.e., when $\dim(\mathcal{H}_X)$ is not a multiple of $\dim(\mathcal{H}_{X'})$, the isometry $V_X$ cannot be a unitary and cannot be inverted. Then the solution is to invert it only on the support of the state $\Phi_{X'} \otimes \sigma_{X''}$ and the construction proving that (ii) implies (iii) does precisely that. The proof of (iii) implies (ii) proceeds analogously.

*Proof.* To see that (i) implies (ii) we construct Naimark's dilation of the extraction channels. This gives us Hilbert spaces $\mathcal{H}_{A''}$ and $\mathcal{H}_{B''}$ and local isometries $V_A$ and $V_B$ such that

$$V \rho_{AB} V^\dagger = \eta_{A'B'A''B''}$$

and $\mathrm{tr}_{A''B''}\, \eta_{A'B'A''B''} = \Phi_{A'B'}$. Since the reduced state on $A'B'$ is pure, it must be uncorrelated from the state on $A''B''$, which concludes the proof. The opposite direction is easy: The extraction channel corresponds to applying the isometry and tracing out the auxiliary system.

To prove that (ii) implies (iii) we explicitly construct a new Hilbert space, isometries, and an auxiliary state. Let us start by showing a simple implication of Eq. (A2). Tracing out one of the systems gives

$$V_X \rho_X V_X^\dagger = \Phi_{X'} \otimes \sigma_{X''}.$$

If two operators are equal, their supports must be equal too. Moreover, the support of a tensor product is the tensor product of the supports. Let $\Pi_X$ and $\Pi_{X''}$ be the projectors on the supports of $\rho_X$ and $\sigma_{X''}$, respectively. Since $\Phi_{X'}$ is full rank, we obtain

$$V_X \Pi_X V_X^\dagger = \mathbb{1}_{X'} \otimes \Pi_{X''}. \tag{A4}$$

We can now proceed to the construction. Consider a Hilbert space $\mathcal{H}_{X'''}$ such that $\dim(\mathcal{H}_{X'''}) = \mathrm{tr}(\Pi_{X''})$ equipped with an isometry $T_X : \mathcal{H}_{X'''} \to \mathcal{H}_{X''}$ satisfying

$$T_X T_X^\dagger = \Pi_{X''}. \tag{A5}$$

Define

$$\tau_{A'''B'''} := (T_A^\dagger \otimes T_B^\dagger)\sigma_{A''B''}(T_A \otimes T_B). \tag{A6}$$

To see that $\tau_{A'''B'''}$ is a valid state we need to check that it is positive semidefinite and of unit trace. The first property is clear (if $A \geqslant 0$, then $X^\dagger A X \geqslant 0$ for any $X$), while for the second property we first observe that

$$\mathrm{tr}(\tau_{A'''B'''}) = \mathrm{tr}[(\Pi_{A''} \otimes \Pi_{B''})\sigma_{A''B''}]$$

and then recall that projecting on the local supports does not affect the state, i.e.,

$$(\Pi_{A''} \otimes \Pi_{B''})\sigma_{A''B''} = \sigma_{A''B''}.$$

Define $W_X : \mathcal{H}_{X'} \otimes \mathcal{H}_{X'''} \to \mathcal{H}_X$ as

$$W_X := \Pi_X V_X^\dagger (\mathbb{1}_{X'} \otimes T_X).$$

To see that $W_X$ is an isometry compute

$$\begin{aligned} W_X^\dagger W_X &= (\mathbb{1}_{X'} \otimes T_X^\dagger) V_X \Pi_X V_X^\dagger (\mathbb{1}_{X'} \otimes T_X) \\ &= (\mathbb{1}_{X'} \otimes T_X^\dagger)(\mathbb{1}_{X'} \otimes \Pi_{X''})(\mathbb{1}_{X'} \otimes T_X) \\ &= \mathbb{1}_{X'} \otimes (T_X^\dagger \Pi_{X''} T_X) = \mathbb{1}_{X'} \otimes \mathbb{1}_{X'''}, \end{aligned}$$

where in the first line we have used Eq. (A4), while the last step relies on Eq. (A5). Finally, we must verify that Eq. (A3) holds. Writing out the right-hand side gives

$$\begin{aligned} W(\Phi_{A'B'} \otimes \tau_{A'''B'''})W^\dagger &= (\Pi_A \otimes \Pi_B)(V_A^\dagger \otimes V_B^\dagger)(\mathbb{1}_{A'B'} \otimes T_A \otimes T_B)(\Phi_{A'B'} \otimes \tau_{A'''B'''}) \\ &\quad \times (\mathbb{1}_{A'B'} \otimes T_A^\dagger \otimes T_B^\dagger)(V_A \otimes V_B)(\Pi_A \otimes \Pi_B) \\ &= (\Pi_A \otimes \Pi_B)(V_A^\dagger \otimes V_B^\dagger)[\Phi_{A'B'} \otimes (T_A \otimes T_B)\tau_{A'''B'''}(T_A^\dagger \otimes T_B^\dagger)](V_A \otimes V_B)(\Pi_A \otimes \Pi_B). \end{aligned}$$

We simplify the middle term using Eq. (A6),

$$(T_A \otimes T_B)\tau_{A'''B'''}(T_A^\dagger \otimes T_B^\dagger) = (\Pi_{A''} \otimes \Pi_{B''})\sigma_{A''B''}(\Pi_{A''} \otimes \Pi_{B''}) = \sigma_{A''B''}.$$

Therefore,

$$\begin{aligned} W(\Phi_{A'B'} \otimes \tau_{A'''B'''})W^\dagger &= (\Pi_A \otimes \Pi_B)(V_A^\dagger \otimes V_B^\dagger)(\Phi_{A'B'} \otimes \sigma_{A''B''})(V_A \otimes V_B)(\Pi_A \otimes \Pi_B) \\ &= (\Pi_A \otimes \Pi_B)\rho_{AB}(\Pi_A \otimes \Pi_B) = \rho_{AB}, \end{aligned}$$

where the middle step is a direct consequence of Eq. (A2).

The proof of (iii) implies (ii) is again a construction. Analogously to the previous argument, we find that the projectors on the supports $\Pi_X$ and $\Pi_{X'''}$ satisfy

$$\Pi_X = W_X(\mathbb{1}_{X'} \otimes \Pi_{X'''})W_X^\dagger.$$

Consider a Hilbert space $\mathcal{H}_{X''}$ (dimension to be specified later) and a linear map $L_X : \mathcal{H}_{X''} \to \mathcal{H}_{X'''}$ satisfying

$$L_X L_X^\dagger = \Pi_{X'''}.$$

Let

$$\sigma_{A''B''} := (L_A^\dagger \otimes L_B^\dagger)\tau_{A'''B'''}(L_A \otimes L_B)$$

and it is easy to check that $\sigma_{A''B''}$ is a valid state. Finally, we need an isometry $R_X : \mathcal{H}_X \to \mathcal{H}_{X'} \otimes \mathcal{H}_{X''}$ such that the projectors $R_X(\mathbb{1}_X - \Pi_X)R_X^\dagger$ and $\mathbb{1}_{X'} \otimes L_X^\dagger L_X$ are orthogonal. Finding such an isometry is possible if the Hilbert space $\mathcal{H}_{X''}$ is of sufficiently high dimension. A simple dimension counting argument implies that we must choose $\dim(\mathcal{H}_{X''})$ to satisfy $\dim(\mathcal{H}_{X'}) \times \dim(\mathcal{H}_{X''}) \geqslant \dim(\mathcal{H}_X)$. Define $V_X : \mathcal{H}_X \to \mathcal{H}_{X'} \otimes \mathcal{H}_{X''}$ as

$$V_X = (\mathbb{1}_{X'} \otimes L_X^\dagger)W_X^\dagger + R_X(\mathbb{1}_X - \Pi_X).$$

It is easy to verify that $V_X$ is an isometry and that the combined isometry $V := V_A \otimes V_B$ satisfies Eq. (A2). ■

### 2. Robust self-testing measures

The conditions discussed in the preceding section capture the idea that a perfect copy of the target state can be extracted from the real state. If we want to use these quantities in any real-world situation, we need to introduce their approximate versions. In the ideal case we require the existence of some objects (e.g., channels or isometries) which render the equalities (A1)–(A3) true. In the approximate case we will quantify approximate satisfaction of these equalities by computing the fidelity between the left- and right-hand sides and we will maximize this value over all valid objects. Note that instead of using the fidelity, we could use the trace norm as a measure of distance, but since we are not aware of any robust results involving the trace distance, we do not discuss it here.

The approximate satisfaction of the condition (A1) is quantified by the extractability defined as

$$\Xi(\rho_{AB} \to \Phi_{A'B'}) := \max_{\Lambda_A, \Lambda_B} F((\Lambda_A \otimes \Lambda_B)(\rho_{AB}), \Phi_{A'B'}),$$

where the maximization is taken over all quantum channels from $A$ to $A'$ and $B$ to $B'$, respectively. Basic properties of extractability are discussed in Sec. II B.

The condition (A2) gives rise to a measure which we call isometric fidelity, defined as

$$F_{\mathrm{iso}}(\rho_{AB} \to \Phi_{A'B'}) := \sup_{\sigma_{A''B''}} \sup_{V} F(V\rho_{AB}V^\dagger, \Phi_{A'B'} \otimes \sigma_{A''B''}), \tag{A7}$$

where the supremum is taken over product isometries $V = V_A \otimes V_B$, where $V_X : \mathcal{H}_X \to \mathcal{H}_{X'} \otimes \mathcal{H}_{X''}$, and auxiliary states $\sigma_{A''B''} \in \mathcal{S}(\mathcal{H}_{A''} \otimes \mathcal{H}_{B''})$. Perhaps surprisingly, this quantity turns out to be equal to the extractability as long as the target state is pure [82].

*Proposition 3.* Let $\rho_{AB}$ be an arbitrary input state and $\Phi_{A'B'}$ be an arbitrary pure target state. Then

$$\Xi(\rho_{AB} \to \Phi_{A'B'}) = F_{\mathrm{iso}}(\rho_{AB} \to \Phi_{A'B'}).$$

*Proof.* To see that the extractability is never smaller than the isometric fidelity it suffices to realize that every local isometry can be turned into an extraction channel by performing a partial trace. Since the fidelity is nondecreasing under tracing out, we immediately conclude that

$$\Xi(\rho_{AB} \to \Phi_{A'B'}) \geqslant F_{\text{iso}}(\rho_{AB} \to \Phi_{A'B'}).$$

To show that this inequality holds as an equality we use Uhlmann's theorem. Let $\Lambda_A$ and $\Lambda_B$ be a pair of extraction channels that achieves optimal fidelity in the definition of extractability, i.e., if

$$\zeta_{A'B'} := (\Lambda_A \otimes \Lambda_B)(\rho_{AB}),$$

then

$$\Xi(\rho_{AB} \to \Phi_{A'B'}) = F(\zeta_{A'B'}, \Phi_{A'B'}).$$

Uhlmann's theorem implies that the fidelity between two mixed states equals the highest achievable fidelity between their purifications and moreover that one of the purifications can be fixed. In our case we pick a specific purification of $\zeta_{A'B'}$. Let $\rho_{ABE}$ be a purification of $\rho_{AB}$, for $X \in \{A, B\}$ let $V_X : \mathcal{H}_X \to \mathcal{H}_{X'} \otimes \mathcal{H}_{X''}$ be Naimark's dilation of the extraction channel $\Lambda_X$, and finally let $V_{AB} := V_A \otimes V_B$. Then the state

$$\zeta_{A'B'A''B''E} := (V_{AB} \otimes \mathbb{1}_E)\rho_{ABE}(V_{AB}^\dagger \otimes \mathbb{1}_E)$$

is a purification of $\zeta_{A'B'}$. By Uhlmann's theorem there exists a purification of $\Phi_{A'B'}$, which we denote by $\gamma_{A'B'A''B''E}$, such that

$$F(\zeta_{A'B'}, \Phi_{A'B'}) = F(\zeta_{A'B'A''B''E}, \gamma_{A'B'A''B''E}). \tag{A8}$$

However, since $\Phi_{A'B'}$ is already pure, all its purifications are of the form

$$\gamma_{A'B'A''B''E} = \Phi_{A'B'} \otimes \gamma_{A''B''E}$$

for some pure state $\gamma_{A''B''E}$. Since the fidelity is nondecreasing under tracing out, we have

$$\begin{aligned} F(\zeta_{A'B'A''B''E}, \Phi_{A'B'} \otimes \gamma_{A''B''E}) &\leqslant F(\zeta_{A'B'A''B''}, \Phi_{A'B'} \otimes \gamma_{A''B''}) \\ &\leqslant F(\zeta_{A'B'}, \Phi_{A'B'}), \end{aligned}$$

which together with Eq. (A8) implies that

$$F(\zeta_{A'B'A''B''}, \Phi_{A'B'} \otimes \gamma_{A''B''}) = F(\zeta_{A'B'}, \Phi_{A'B'}).$$

The left-hand side is a lower bound on the isometric fidelity, whereas the right-hand side by construction equals the extractability, which concludes the proof. ■

To finish our discussion of the isometric fidelity, let us point out that in the literature one sometimes sees the isometries in Eqs. (A2) and (A7) replaced by unitaries, but using unitaries is strictly speaking not correct. For instance, the unitary version of isometric fidelity has the unpleasant feature that it is not defined for all input states. The existence of a unitary $U_A : \mathcal{H}_A \to \mathcal{H}_{A'} \otimes \mathcal{H}_{A''}$ implies that $\dim(\mathcal{H}_A) = \dim(\mathcal{H}_{A'})\dim(\mathcal{H}_{A''})$. Since the dimension of the auxiliary Hilbert space $\mathcal{H}_{A''}$ must be an integer, unitarity requires that the dimension of the Hilbert space $\mathcal{H}_A$ is a multiple of the dimension of the target Hilbert space $\mathcal{H}_{A'}$, which does not have to be the case. Clearly, a measure which is not defined for all states is not suitable for the purpose of making self-testing statements.

Finally, the condition (A3) gives rise to the Mayers-Yao fidelity defined as

$$F_{\text{MY}}(\rho_{AB} \to \Phi_{A'B'}) := \sup_{\sigma_{A''B''}} \sup_{W} F(\rho_{AB}, W(\Phi_{A'B'} \otimes \sigma_{A''B''})W^\dagger),$$

where the supremum is taken over product isometries $W = W_A \otimes W_B$ for $W_X : \mathcal{H}_{X'} \otimes \mathcal{H}_{X''} \to \mathcal{H}_X$ and auxiliary states $\sigma_{A''B''} \in \mathcal{S}(\mathcal{H}_{A''} \otimes \mathcal{H}_{B''})$. However, this quantity suffers from the same problem: It is not defined for all states, e.g., when $\dim(\mathcal{H}_A) < \dim(\mathcal{H}_{A'})$.

## APPENDIX B: ROBUST SELF-TESTING OF TWO-QUBIT STATES

In this Appendix we provide the details of the argument discussed in Sec. IV. In the first section we give the definitions of the extraction channels and compute all the operators appearing in the operator inequality. In the second section we discuss the numerical evidence supporting the conjecture.

### 1. Operator inequality

Let us start by writing down the Bell operator. Recall that the observables of Alice and Bob are parametrized by

$$\begin{aligned} A_r &:= \cos(a)\mathsf{X} + (-1)^r \sin(a)\mathsf{Z}, \\ B_r &:= \cos(b)\mathsf{X} + (-1)^r \sin(b)\mathsf{Z} \end{aligned}$$

for $r \in \{0, 1\}$. For these observables the tilted CHSH operator defined in Eq. (10) reads

$$\begin{aligned} W_\alpha(a, b) = {}& \alpha[\cos(a)\mathsf{X} + \sin(a)\mathsf{Z}] \otimes \mathbb{1} + 2\cos a \cos(b)\mathsf{X} \otimes \mathsf{X} \\ &+ 2\cos a \sin(b)\mathsf{X} \otimes \mathsf{Z} \\ &+ 2\sin a \cos(b)\mathsf{Z} \otimes \mathsf{X} - 2\sin a \sin(b)\mathsf{Z} \otimes \mathsf{Z}. \end{aligned}$$

The optimal violation is achieved for $a^* := \pi/4$ and

$$b^*_\alpha := \arcsin\left(\sqrt{\frac{4-\alpha^2}{8}}\right). \tag{B1}$$

The corresponding optimal state is given by

$$\Phi_\alpha := \frac{1}{4}\left(\mathbb{1} \otimes \mathbb{1} + \sqrt{\frac{2\alpha^2}{4+\alpha^2}}\left[\frac{\mathsf{X}+\mathsf{Z}}{\sqrt{2}} \otimes \mathbb{1} + \mathbb{1} \otimes \mathsf{X}\right] + \frac{\mathsf{X}+\mathsf{Z}}{\sqrt{2}} \otimes \mathsf{X} + \sqrt{\frac{4-\alpha^2}{4+\alpha^2}}\left[\mathsf{Y} \otimes \mathsf{Y} + \frac{\mathsf{X}-\mathsf{Z}}{\sqrt{2}} \otimes \mathsf{Z}\right]\right). \tag{B2}$$

To see that this state is unitarily equivalent to $\cos\theta|00\rangle + \sin\theta|11\rangle$ for $\theta$ specified in Eq. (11) note that

$$\sin 2\theta = \sqrt{\frac{4-\alpha^2}{4+\alpha^2}}, \quad \cos 2\theta = \sqrt{\frac{2\alpha^2}{4+\alpha^2}}.$$

The extraction channel for Alice is precisely the channel used in Ref. [56],

$$[\Lambda_A(x)](\rho) := \frac{1+g(x)}{2}\rho + \frac{1-g(x)}{2}\Gamma(x)\rho\Gamma(x),$$

where

$$\Gamma(x) := \begin{cases} \mathsf{X} & \text{if } x \in [0, \pi/4] \\ \mathsf{Z} & \text{if } x \in (\pi/4, \pi/2] \end{cases}$$

and

$$g(x) := (1 + \sqrt{2})(\sin x + \cos x - 1).$$

It is easy to check that $x = \pi/4$ gives the identity channel, whereas $x = 0$ and $x = \pi/2$ correspond to full dephasing. The channel of Bob has the same form except that the identity channel should arise for the angle $b_\alpha^*$ defined in Eq. (B1). Let us define the effective angle $h_\alpha(x)$ as a piecewise linear function which maps the interval $[0, b_\alpha^*]$ onto $[0, \pi/4]$ and $[b_\alpha^*, \pi/2]$ onto $[\pi/4, \pi/2]$:

$$h_\alpha(x) := \begin{cases} \frac{\pi}{4}\frac{x}{b_\alpha^*} & \text{if } x \in [0, b_\alpha^*] \\ \frac{\pi}{2} - \frac{\pi}{4}\frac{\pi - 2x}{\pi - 2b_\alpha^*} & \text{if } x \in (b_\alpha^*, \pi/2]. \end{cases}$$

These definitions allow us to write the extraction channel of Bob as

$$\Lambda_B(x) := \Lambda_A(h_\alpha(x)).$$

The operator $K_\alpha(a, b)$ is obtained by applying the dual channels to the ideal state given in Eq. (B2). Since the dephasing channels are self-dual, we have

$$K_\alpha(a, b) := [\Lambda_A(a) \otimes \Lambda_B(b)](\Phi_\alpha).$$

The operator inequality (6) is equivalent to the operator

$$T_\alpha(a, b) := K_\alpha(a, b) - s_\alpha W_\alpha(a, b) - \mu_\alpha \mathbb{1}$$

being positive semidefinite for

$$s_\alpha := \frac{(\sqrt{8 + 2\alpha^2} + 2 + \alpha)(3\sqrt{8 + 2\alpha^2} - \sqrt{4 - \alpha^2} - \alpha\sqrt{2})}{4(2 - \alpha)^2\sqrt{8 + 2\alpha^2}},$$

$$\mu_\alpha := 1 - s_\alpha\sqrt{8 + 2\alpha^2}.$$

Since the dephasing basis changes at $a = \pi/4$ and $b = b_\alpha^*$, there are in principle four distinct cases that need to be considered. In the case of CHSH the presence of symmetries allows one to reduce the analysis of the entire square ($[0, \pi/2] \times [0, \pi/2]$) to a single quarter ($[0, \pi/4] \times [0, \pi/4]$). In the tilted case this symmetry is partially broken, but we still have

$$T_\alpha(a, b) = U T_\alpha(\pi/2 - a,\ b) U^\dagger, \tag{B3}$$

where

$$U := \frac{X + Z}{\sqrt{2}} \otimes X. \tag{B4}$$

This observation implies that it suffices to analyze the half of the square corresponding to $a \in [0, \pi/4]$.

### 2. Numerical evidence

Our goal is to gather evidence that the operator $T_\alpha(a, b)$ is positive semidefinite for $\alpha \in [0, 2)$, $a \in [0, \pi/4]$, and $b \in [0, \pi/2]$. For this purpose, we have generated a grid over the parameter space in the following manner.

(a) We have chosen $\alpha$ in the range $[0, 1.999]$ with a step size of 0.001.

(b) We have discretized the angle of Alice by splitting the interval $[0, \pi/4]$ into 99 equally spaced intervals $[a_k, a_{k+1}]$, where $a_1 = 0$, $a_{100} = \pi/4$, and $1 \leqslant k \leqslant 100$. Similarly, for the angle of Bob we have discretized $[0, \pi/2]$ as intervals $[b_m, b_{m+1}]$ of equal length, with $b_1 = 0$, $b_{200} = \pi/2$, and $0 \leqslant m \leqslant 200$. For fixed $\alpha$, we thus obtain the grid $\{(a_k, b_m) \mid 1 \leqslant k \leqslant 100, 1 \leqslant m \leqslant 200\}$.

Using the LINALG library from NUMPY (a scientific computing package for PYTHON), we have computed the eigenvalues of $T_\alpha(a, b)$ at every point of the grid. We have found that the smallest value equals $-1.317 \times 10^{-9}$ and occurs for $\alpha = 1.998$. Our code can be freely accessed online [83].

## APPENDIX C: CHSH VIOLATION DOES NOT IMPLY NONTRIVIAL EXTRACTABILITY

In this Appendix we construct a state which violates the CHSH inequality but whose singlet extractability does not exceed the trivial value of $\frac{1}{2}$. The proof hinges on two technical propositions and since proving them within the main argument would be rather distracting, let us use them without proofs. Complete proofs can be found in Appendix C 2.

### 1. Argument

Consider a state $\rho_{XYAB}$ acting on $\mathcal{H}_X \otimes \mathcal{H}_Y \otimes \mathcal{H}_A \otimes \mathcal{H}_B$ for $\mathcal{H}_X, \mathcal{H}_Y \equiv \mathbb{C}^3$ and $\mathcal{H}_A, \mathcal{H}_B \equiv \mathbb{C}^2$, where subsystems $X$ and $A$ belong to Alice and subsystems $Y$ and $B$ belong to Bob. The state is defined with respect to the CHSH operator corresponding to the observables given in Eq. (12) which reads

$$W = \sum_{x,y=0}^{2} |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes W_{AB}^{xy}$$

for the two-qubit operators $W_{AB}^{xy}$ given by

$$\begin{array}{c|ccc} x\backslash y & 0 & 1 & 2 \\ \hline 0 & 2\mathsf{Z} \otimes \mathsf{Z} & 2\mathsf{Z} \otimes \mathsf{Z} & 2\mathsf{Z} \otimes \mathsf{Z} \\ 1 & 2\mathsf{Z} \otimes \mathsf{Z} & \mathsf{X} \otimes (-\mathsf{X} + \mathsf{Z}) + \mathsf{Z} \otimes (\mathsf{X} + \mathsf{Z}) & 2\mathsf{X} \otimes \mathsf{Z} \\ 2 & 2\mathsf{Z} \otimes \mathsf{Z} & 2\mathsf{Z} \otimes \mathsf{X} & -2\mathsf{Z} \otimes \mathsf{Z} \end{array} \tag{C1}$$

We choose the state $\rho_{XYAB}$ to be of the form

$$\rho_{XYAB} = \sum_{x,y=0}^{2} p_{xy} |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \rho_{AB}^{xy}$$

for some probability distribution $\{p_{xy}\}_{x,y=0}^{2}$ and two-qubit states $\rho_{AB}^{xy}$ chosen to satisfy

$$\langle W_{AB}^{xy}, \rho_{AB}^{xy} \rangle = \begin{cases} 2\sqrt{2} & \text{if } x = y = 1 \\ 2 & \text{otherwise.} \end{cases} \tag{C2}$$

The precise form of the states $\rho_{AB}^{xy}$ will be specified later. Recall that we refer to the point $x = y = 1$ as the center and the remaining eight points as the frame. A simple calculation shows that

$$\beta = \langle W, \rho_{XYAB} \rangle = 2 + (2\sqrt{2} - 2)p_{11}, \tag{C3}$$

i.e., the CHSH inequality is violated as long as $p_{11} > 0$. Our goal is to prove that there exists a probability distribution satisfying $p_{11} > 0$ and two-qubit states satisfying Eq. (C2) such that the resulting state $\rho_{XYAB}$ satisfies

$$\Xi(\rho_{XYAB} \to \Phi_{A'B'}^{+}) = \tfrac{1}{2},$$

where $\Phi_{A'B'}^{+}$ is a maximally entangled state of two qubits. The quantity does not depend on which maximally entangled state we choose and for this proof it is convenient to assume that $|\Phi^{+}\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. By definition of extractability, showing the existence of a suitable probability distribution and two-qubit states is equivalent to showing that for all local extraction channels $\Lambda_A, \Lambda_B : \mathcal{L}(\mathbb{C}^3 \otimes \mathbb{C}^2) \to \mathcal{L}(\mathbb{C}^2)$ we have

$$F((\Lambda_A \otimes \Lambda_B)(\rho_{XYAB}), \Phi_{A'B'}^{+}) \leqslant \tfrac{1}{2}.$$

Since the registers $X$ and $Y$ are classical, instead of optimizing over the most general channels from $\mathcal{L}(\mathbb{C}^3 \otimes \mathbb{C}^2)$ to $\mathcal{L}(\mathbb{C}^2)$ it suffices to consider channels which first read the classical register and then apply a suitable qubit $[\mathcal{L}(\mathbb{C}^2) \to \mathcal{L}(\mathbb{C}^2)]$ channel (see Lemma 2 for details). Let $\Lambda_A^x$ be the qubit channel of Alice corresponding to the value of the classical register $X$ being $x$ and similarly let $\Lambda_B^y$ be the qubit channel of Bob corresponding to $Y$ having value $y$. Since the target state is pure, the fidelity equals the inner product, which implies

$$\begin{aligned} &F((\Lambda_A \otimes \Lambda_B)(\rho_{XYAB}), \Phi_{A'B'}^{+}) \\ &\quad = \langle (\Lambda_A \otimes \Lambda_B)(\rho_{XYAB}), \Phi_{A'B'}^{+} \rangle \\ &\quad = \sum_{xy} p_{xy} \big\langle \big(\Lambda_A^x \otimes \Lambda_B^y\big)\big(\rho_{AB}^{xy}\big), \Phi_{A'B'}^{+} \big\rangle. \end{aligned} \tag{C4}$$

The intuition behind the proof goes as follows: There are no extraction channels which perform well both on the frame and in the center. We make this intuition rigorous in two steps. The following proposition shows that if Alice and Bob perform well on the frame, then the channels $\Lambda_A^1$ and $\Lambda_B^1$ must significantly contract the Bloch sphere. Note that in the argument below only six points of the frame are used (we leave the remaining two points undefined).

*Proposition 4.* Let

$$\rho_{AB}^{xy} = \begin{cases} |11\rangle\langle 11| & \text{if } (x, y) = (0, 0) \\ \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|) & \text{if } (x, y) = (0, 1), (0, 2), (1, 0), (2, 0) \\ \frac{1}{2}(|01\rangle\langle 01| + |10\rangle\langle 10|) & \text{if } (x, y) = (2, 2). \end{cases}$$

For these six points for fixed extraction channels $\Lambda_A^x$ and $\Lambda_B^y$ define

$$\varepsilon_{xy} := \tfrac{1}{2} - \big\langle \big(\Lambda_A^x \otimes \Lambda_B^y\big)\big(\rho_{AB}^{xy}\big), \Phi_{A'B'}^{+} \big\rangle. \tag{C5}$$

Note that $\varepsilon_{xy} \geqslant 0$, since the states $\rho_{AB}^{xy}$ are separable. If

$$\omega_A := \Lambda_A^1\left(\frac{\mathbb{1}_2}{2}\right), \quad \omega_B := \Lambda_B^1\left(\frac{\mathbb{1}_2}{2}\right),$$

then

$$\lambda_{\min}(\omega_A) \leqslant 4[8\varepsilon_{00} + 16(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22}) + 3\varepsilon_{10}],$$
$$\lambda_{\min}(\omega_B) \leqslant 4[8\varepsilon_{00} + 16(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22}) + 3\varepsilon_{01}].$$

In particular, we have $\lambda_{\min}(\omega_A), \lambda_{\min}(\omega_B) \leqslant 248\varepsilon_{\text{wav}}$ for

$$\varepsilon_{\text{wav}} := \tfrac{1}{62}[8\varepsilon_{00} + 16(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22}) + 3\varepsilon_{01} + 3\varepsilon_{10}].$$

The fact that the channels $\Lambda_A^1$ and $\Lambda_B^1$ map the center of the Bloch sphere to a point close to the boundary means that the input states are to a large extent erased. It is therefore not surprising that applying such channels to a maximally entangled state annihilates most of its entanglement.

*Proposition 5.* Let $\Lambda_A$ and $\Lambda_B$ be qubit channels such that the smaller eigenvalues of the normalized qubit density matrices

$$\Lambda_A\left(\frac{\mathbb{1}_2}{2}\right), \quad \Lambda_B\left(\frac{\mathbb{1}_2}{2}\right)$$

are at most $\lambda$. Then, for any pair of maximally entangled two-qubit states $\Psi_1$ and $\Psi_2$ we have

$$\langle (\Lambda_A \otimes \Lambda_B)(\Psi_1), \Psi_2 \rangle \leqslant \tfrac{1}{2} + 2\lambda.$$

These two propositions immediately imply the main result.

*Proposition 6.* Let

$$\rho_{XYAB} = \sum_{x,y=0}^{2} p_{xy} |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \rho_{AB}^{xy},$$

where the states $\rho_{AB}^{xy}$ corresponding to the frame are specified in Proposition 4, the state $\rho_{AB}^{11} = \Psi$ is some pure maximally entangled state, and the probability distribution is given by

$$p_{00} = \tfrac{4}{31}(1 - v), \quad p_{01} = p_{10} = \tfrac{3}{62}(1 - v),$$
$$p_{02} = p_{20} = p_{22} = \tfrac{8}{31}(1 - v), \quad p_{11} = v, \quad p_{12} = p_{21} = 0$$

for $v = \frac{1}{597}$. This state satisfies

$$\Xi(\rho_{XYAB} \to \Phi_{A'B'}^{+}) = \tfrac{1}{2}.$$

*Proof.* From Eq. (C4) we have

$$\begin{aligned} &F((\Lambda_A \otimes \Lambda_B)(\rho_{XYAB}), \Phi_{A'B'}^{+}) \\ &\quad = \sum_{xy} p_{xy} \big\langle \big(\Lambda_A^x \otimes \Lambda_B^y\big)\big(\rho_{AB}^{xy}\big), \Phi_{A'B'}^{+} \big\rangle \end{aligned}$$

$$= \sum_{(x,y)\neq(1,1)} p_{xy}\big\langle\big(\Lambda_A^x \otimes \Lambda_B^y\big)\big(\rho_{AB}^{xy}\big), \Phi^+_{A'B'}\big\rangle$$
$$+ p_{11}\big\langle\big(\Lambda_A^1 \otimes \Lambda_B^1\big)(\Psi), \Phi^+_{A'B'}\big\rangle.$$

The inner product in the first term can be written in terms of $\varepsilon_{xy}$ defined in Proposition 4. A direct calculation gives

$$\sum_{(x,y)\neq(1,1)} p_{xy}\big\langle\big(\Lambda_A^x \otimes \Lambda_B^y\big)\big(\rho_{AB}^{xy}\big), \Phi^+_{A'B'}\big\rangle$$
$$= \sum_{(x,y)\neq(1,1)} p_{xy}\left(\frac{1}{2} - \varepsilon_{xy}\right)$$
$$= \frac{1}{2}(1-v) - (1-v)\varepsilon_{\text{wav}} = (1-v)\left(\frac{1}{2} - \varepsilon_{\text{wav}}\right)$$

for $\varepsilon_{\text{wav}}$ defined in Proposition 4. Combining Propositions 4 and 5 leads to

$$\big\langle\big(\Lambda_A^1 \otimes \Lambda_B^1\big)(\Psi), \Phi^+_{A'B'}\big\rangle \leqslant \tfrac{1}{2} + 596\varepsilon_{\text{wav}}.$$

Adding the two immediately yields

$$F((\Lambda_A \otimes \Lambda_B)(\rho_{XYAB}), \Phi^+_{A'B'})$$
$$\leqslant (1-v)\big(\tfrac{1}{2} - \varepsilon_{\text{wav}}\big) + v\big(\tfrac{1}{2} + 596\varepsilon_{\text{wav}}\big)$$
$$= \tfrac{1}{2} + (597v - 1)\varepsilon_{\text{wav}} = \tfrac{1}{2}.$$

■

The value $p_{11} = \frac{1}{597}$ plugged into Eq. (C3) gives the CHSH violation of $\beta \approx 2.0014$.

### 2. Proof details

In this section we prove Propositions 4 and 5 used in the main argument. To do that we first need to prove three auxiliary lemmas.

The first lemma is a triangle-type inequality for the inner product of (finite-dimensional) density matrices.

*Lemma 1.* For finite-dimensional density matrices $\rho_0$, $\rho_1$, and $\sigma$ we always have

$$\langle\rho_0, \rho_1\rangle \geqslant 2(\langle\rho_0, \sigma\rangle + \langle\rho_1, \sigma\rangle) - 3.$$

In particular, if

$$\langle\rho_0, \sigma\rangle \geqslant 1 - \delta_0,$$
$$\langle\rho_1, \sigma\rangle \geqslant 1 - \delta_1,$$

then

$$\langle\rho_0, \rho_1\rangle \geqslant 1 - 2(\delta_0 + \delta_1).$$

*Proof.* The triangle inequality for the Schatten 2-norm (the Frobenius norm) implies that

$$\|\rho_0 - \rho_1\|_2 \leqslant \|\rho_0 - \sigma\|_2 + \|\sigma - \rho_1\|_2,$$

which can be written as

$$\sqrt{\langle\rho_0, \rho_0\rangle + \langle\rho_1, \rho_1\rangle - 2\langle\rho_0, \rho_1\rangle}$$
$$\leqslant \sqrt{\langle\rho_0, \rho_0\rangle + \langle\sigma, \sigma\rangle - 2\langle\rho_0, \sigma\rangle}$$
$$+ \sqrt{\langle\rho_1, \rho_1\rangle + \langle\sigma, \sigma\rangle - 2\langle\rho_1, \sigma\rangle}.$$

Since both sides are non-negative, we can square the inequality to obtain

$$-\langle\rho_0, \rho_1\rangle \leqslant \langle\sigma, \sigma\rangle - \langle\rho_0, \sigma\rangle - \langle\rho_1, \sigma\rangle + \sqrt{(\langle\rho_0, \rho_0\rangle + \langle\sigma, \sigma\rangle - 2\langle\rho_0, \sigma\rangle)(\langle\rho_1, \rho_1\rangle + \langle\sigma, \sigma\rangle - 2\langle\rho_1, \sigma\rangle)}.$$

The fact that for an arbitrary density matrix $\tau$ we have $\langle\tau, \tau\rangle \leqslant 1$ gives

$$-\langle\rho_0, \rho_1\rangle \leqslant 1 - \langle\rho_0, \sigma\rangle - \langle\rho_1, \sigma\rangle$$
$$+ 2\sqrt{(1 - \langle\rho_0, \sigma\rangle)(1 - \langle\rho_1, \sigma\rangle)}.$$

We bound the last term using the mean inequality $\sqrt{ab} \leqslant (a + b)/2$, which leads to

$$-\langle\rho_0, \rho_1\rangle \leqslant 3 - 2(\langle\rho_0, \sigma\rangle + \langle\rho_1, \sigma\rangle).$$

■

The second lemma formalizes the intuition that an arbitrary channel acting jointly on a classical and quantum register can be replaced by a channel that reads the classical register and acts on the quantum register accordingly.

*Lemma 2.* Let $\mathcal{H}_C$, $\mathcal{H}_Q$, and $\mathcal{H}_A$ be Hilbert spaces of dimensions $d_C$, $d_Q$, and $d_A$, respectively. Let $\{|e_j\rangle\}_{j=1}^{d_C}$ be an orthonormal basis of $\mathcal{H}_C$ and we say that $R_{CQ}$ is a classical-quantum operator acting on $\mathcal{H}_C \otimes \mathcal{H}_Q$ if it can be written as

$$R_{CQ} = \sum_j |e_j\rangle\langle e_j| \otimes S_j \tag{C6}$$

for some linear operators $S_j \in \mathcal{L}(\mathcal{H}_Q)$. Then, for an arbitrary channel $\Lambda : \mathcal{L}(\mathcal{H}_C \otimes \mathcal{H}_Q) \rightarrow \mathcal{L}(\mathcal{H}_A)$ there exists a collection of $d_C$ channels $\Lambda_j : \mathcal{L}(\mathcal{H}_Q) \rightarrow \mathcal{L}(\mathcal{H}_A)$ such that for all operators of the form (C6) we have

$$\Lambda(R_{CQ}) = \sum_j \Lambda_j(S_j). \tag{C7}$$

*Proof.* We define the channel $\Lambda_j$ through its action on an arbitrary operator $X \in \mathcal{L}(\mathcal{H}_Q)$. Let

$$\Lambda_j(X) := \Lambda(|e_j\rangle\langle e_j| \otimes X),$$

which ensures that $\Lambda_j$ is completely positive and trace preserving. The equality (C7) holds by construction. ■

The last lemma shows that if a channel maps the maximally mixed state to a state which is close to being pure, then this channel must contract all the Pauli observables.

*Lemma 3.* Let $\Lambda$ be a qubit quantum channel, let

$$\omega := \Lambda\left(\frac{\mathbb{1}_2}{2}\right),$$

and suppose that $\text{spec}(\omega) = \{\lambda, 1 - \lambda\}$ for $\lambda \in [0, 1/2]$. Let $\Gamma$ be a $2 \times 2$ Hermitian operator satisfying $\Gamma^2 = \mathbb{1}$ and $\text{tr}\,\Gamma = 0$. Then

$$-2\sqrt{\lambda}\mathbb{1}_2 \leqslant \Lambda(\Gamma) \leqslant 2\sqrt{\lambda}\mathbb{1}_2.$$

*Proof.* Since the quantum channel is a positive map, we have $\Lambda(\mathbb{1}_2 \pm \Gamma) \geqslant 0$ or, equivalently, $-2\omega \leqslant \Lambda(\Gamma) \leqslant 2\omega$. We start by writing both operators in the eigenbasis of $\omega$,

$$\omega = \begin{pmatrix} \lambda & \\ & 1-\lambda \end{pmatrix}, \quad \Lambda(\Gamma) = \begin{pmatrix} t & y \\ y^* & -t \end{pmatrix}$$

for some $t \in \mathbb{R}$ and $y \in \mathbb{C}$. Note that we have implicitly used the fact that $\Lambda(\Gamma)$ is Hermitian and traceless. The condition $\Lambda(\Gamma) \geqslant -2\omega$ reads

$$\begin{pmatrix} 2\lambda + t & y \\ y^* & 2 - 2\lambda - t \end{pmatrix} \geqslant 0$$

and implies that

$$(2\lambda + t)(2 - 2\lambda - t) - |y|^2 \geqslant 0.$$

Similarly, the condition $\Lambda(\Gamma) \leqslant 2\omega$ leads to

$$(2\lambda - t)(2 - 2\lambda + t) - |y|^2 \geqslant 0.$$

Adding these two conditions gives

$$t^2 + |y|^2 \leqslant 4\lambda(1-\lambda) \leqslant 4\lambda.$$

As the eigenvalues of $\Lambda(\Gamma)$ are easily seen to be $\pm\sqrt{t^2 + |y|^2}$, the claim follows directly from the last inequality. ■

Equipped with these three auxiliary lemmas, we are ready to tackle the two propositions used in the main argument.

*Proof of Proposition 4.* The proof consists of three steps. We first consider the four corner points, i.e., $(x, y) \in \{(0, 0), (0, 2), (2, 0), (2, 2)\}$, and show that the channels $\Lambda_A^0$ and $\Lambda_B^0$ map the entire Bloch sphere to a small region close to the boundary. In the second step we consider the points $(x, y) \in \{(0, 1), (1, 0)\}$ to show that the channels $\Lambda_A^1$ and $\Lambda_B^1$ have the same property. In the last step we compute an explicit bound on the eigenvalues of $\omega_A$ and $\omega_B$.

For $b \in \{0, 1\}$ and $x, y \in \{0, 1, 2\}$ define

$$\sigma_b^x := \Lambda_A^x(|b\rangle\langle b|),$$
$$\tau_b^y := \left[\Lambda_B^y(|b\rangle\langle b|)\right]^{\mathsf{T}},$$

which implies that

$$\begin{aligned}&\left\langle\left(\Lambda_A^x \otimes \Lambda_B^y\right)(|b\rangle\langle b| \otimes |b'\rangle\langle b'|), \Phi^+_{A'B'}\right\rangle\\ &= \left\langle \Lambda_A^x(|b\rangle\langle b|) \otimes \Lambda_B^y(|b'\rangle\langle b'|), \Phi^+_{A'B'}\right\rangle\\ &= \tfrac{1}{2}\left\langle \sigma_b^x, \tau_{b'}^y\right\rangle.\end{aligned}$$

Therefore, Eq. (C5) imposes constraints on the inner products between the operators $\sigma_b^x$ and $\tau_b^y$. Considering points $(x, y) = (0, 0), (0, 2), (2, 0), (2, 2)$ gives

$$\left\langle \sigma_1^0, \tau_1^0\right\rangle = 1 - 2\varepsilon_{00}, \tag{C8}$$

$$\left\langle \sigma_0^0, \tau_0^2\right\rangle + \left\langle \sigma_1^0, \tau_1^2\right\rangle = 2 - 4\varepsilon_{02}, \tag{C9}$$

$$\left\langle \sigma_0^2, \tau_0^0\right\rangle + \left\langle \sigma_1^2, \tau_1^0\right\rangle = 2 - 4\varepsilon_{20}, \tag{C10}$$

$$\left\langle \sigma_0^2, \tau_1^2\right\rangle + \left\langle \sigma_1^2, \tau_0^2\right\rangle = 2 - 4\varepsilon_{22}. \tag{C11}$$

Plugging the upper bound $\langle \sigma_b^x, \tau_{b'}^y\rangle \leqslant 1$ into Eq. (C9) immediately gives

$$\left\langle \sigma_1^0, \tau_1^2\right\rangle \geqslant 1 - 4\varepsilon_{02},$$

which combined with Eq. (C8) by Lemma 1 gives

$$\left\langle \tau_1^0, \tau_1^2\right\rangle \geqslant 1 - 4(\varepsilon_{00} + 2\varepsilon_{02}). \tag{C12}$$

Similarly, Eqs. (C10) and (C11) imply

$$\left\langle \sigma_0^2, \tau_0^0\right\rangle \geqslant 1 - 4\varepsilon_{20},$$
$$\left\langle \sigma_0^2, \tau_1^2\right\rangle \geqslant 1 - 4\varepsilon_{22},$$

which gives

$$\left\langle \tau_0^0, \tau_1^2\right\rangle \geqslant 1 - 8(\varepsilon_{20} + \varepsilon_{22}).$$

Combining this with Eq. (C12) gives

$$\left\langle \tau_0^0, \tau_1^0\right\rangle \geqslant 1 - 8[\varepsilon_{00} + 2(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22})], \tag{C13}$$

which concludes the first step of the proof. This lower bound implies that the states $\tau_0^0$ and $\tau_1^0$ are close to each other and moreover that they are close to being pure. Since these two states result from applying the channel $\Lambda_B^0$ to two pure orthogonal states, we conclude that the channel must shrink the entire Bloch sphere to a small region close to the boundary.

Considering the point $(x, y) = (1, 0)$ gives

$$\left\langle \sigma_0^1, \tau_0^0\right\rangle + \left\langle \sigma_1^1, \tau_1^0\right\rangle = 2 - 4\varepsilon_{10}.$$

Define $a, b \geqslant 0$ such that

$$\left\langle \sigma_0^1, \tau_0^0\right\rangle = 1 - a, \tag{C14}$$

$$\left\langle \sigma_1^1, \tau_1^0\right\rangle = 1 - b, \tag{C15}$$

which implies that $a + b = 4\varepsilon_{10}$. Applying the inner-product inequality proven in Lemma 1 to Eqs. (C13)–(C15) gives

$$\left\langle \sigma_0^1, \sigma_1^1\right\rangle \geqslant 1 - 32[\varepsilon_{00} + 2(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22})] - 4a - 2b$$

or

$$\left\langle \sigma_0^1, \sigma_1^1\right\rangle \geqslant 1 - 32[\varepsilon_{00} + 2(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22})] - 2a - 4b,$$

depending on the order. Averaging over these two bounds gives

$$\begin{aligned}\left\langle \sigma_0^1, \sigma_1^1\right\rangle &\geqslant 1 - 32[\varepsilon_{00} + 2(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22})] - 3a - 3b\\ &= 1 - 32[\varepsilon_{00} + 2(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22})] - 12\varepsilon_{10}\\ &= 1 - 4[8\varepsilon_{00} + 16(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22}) + 3\varepsilon_{10}],\end{aligned}$$

which concludes the second step of the proof.

The density matrix $\omega_A$ defined in the proposition is given by

$$\omega_A = \Lambda_A^1\left(\frac{\mathbb{1}_2}{2}\right) = \frac{1}{2}\left(\sigma_0^1 + \sigma_1^1\right).$$

Clearly, $\operatorname{tr}\omega_A = 1$ and

$$\begin{aligned}\operatorname{tr}\omega_A^2 &= \tfrac{1}{4}\left[\left\langle \sigma_0^1, \sigma_0^1\right\rangle + \left\langle \sigma_1^1, \sigma_1^1\right\rangle + 2\left\langle \sigma_0^1, \sigma_1^1\right\rangle\right]\\ &= \tfrac{1}{4}\left[\left\langle \sigma_0^1, \sigma_0^1\right\rangle + \left\langle \sigma_1^1, \sigma_1^1\right\rangle - 2\left\langle \sigma_0^1, \sigma_1^1\right\rangle\right] + \left\langle \sigma_0^1, \sigma_1^1\right\rangle\\ &= \tfrac{1}{4}\left[\left\langle \sigma_0^1 - \sigma_1^1, \sigma_0^1 - \sigma_1^1\right\rangle\right] + \left\langle \sigma_0^1, \sigma_1^1\right\rangle \geqslant \left\langle \sigma_0^1, \sigma_1^1\right\rangle.\end{aligned}$$

We take advantage of the fact that for $2 \times 2$ Hermitian matrices we have $[\operatorname{tr}(M)]^2 = \operatorname{tr}(M^2) + 2\det(M)$. If $\lambda$ is the smaller eigenvalue of $\omega_A$, then

$$\begin{aligned}\lambda &\leqslant 2\lambda(1-\lambda) = 2\det(\omega_A)\\ &= [\mathrm{tr}(\omega_A)]^2 - \mathrm{tr}\left(\omega_A^2\right) = 1 - \mathrm{tr}\left(\omega_A^2\right)\\ &\leqslant 1 - \left\langle\sigma_0^1, \sigma_1^1\right\rangle\\ &\leqslant 4[8\varepsilon_{00} + 16(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22}) + 3\varepsilon_{10}],\end{aligned}$$

which concludes the last step of the proof of the first statement. The proof of the second statement is essentially the same. From Eqs. (C8) and (C10) we obtain

$$\left\langle\sigma_1^0, \sigma_1^2\right\rangle \geqslant 1 - 4(\varepsilon_{00} + 2\varepsilon_{20}),$$

whereas Eqs. (C9) and (C11) imply

$$\left\langle\sigma_0^0, \sigma_1^2\right\rangle \geqslant 1 - 8(\varepsilon_{02} + \varepsilon_{22}).$$

Combining these yields

$$\left\langle\sigma_0^0, \sigma_1^0\right\rangle \geqslant 1 - 8[\varepsilon_{00} + 2(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22})]$$

and by adding the point $(x, y) = (0, 1)$ we arrive at

$$\left\langle\tau_0^1, \tau_1^1\right\rangle \geqslant 1 - 4[8\varepsilon_{00} + 16(\varepsilon_{02} + \varepsilon_{20} + \varepsilon_{22}) + 3\varepsilon_{01}].$$

Finally, we note that $\omega_B^{\mathrm{T}} = (\tau_0^1 + \tau_1^1)/2$, but since the transpose does not affect the spectrum, the final calculation is precisely the same. ■

*Proof of Proposition 5.* Since the statement is invariant under local unitaries, we can without loss of generality assume that $\Psi_1$ is the usual maximally entangled state, i.e.,

$$\Psi_1 = \tfrac{1}{4}(\mathbb{1}\otimes\mathbb{1} + \mathsf{X}\otimes\mathsf{X} - \mathsf{Y}\otimes\mathsf{Y} + \mathsf{Z}\otimes\mathsf{Z}).$$

Note that $\Psi_1$ can be written as

$$\Psi_1 = \tau + \tfrac{1}{4}(\mathsf{X}\otimes\mathsf{X} + \mathsf{Z}\otimes\mathsf{Z}),$$

where $\tau = (\mathbb{1}\otimes\mathbb{1} - \mathsf{Y}\otimes\mathsf{Y})/4$. Since $\tau$ is a separable state, we have

$$\langle(\Lambda_A \otimes \Lambda_B)(\tau), \Psi_2\rangle \leqslant \tfrac{1}{2}.$$

To bound the other two terms we use Lemma 3, which in particular implies that

$$\Lambda_A(\mathsf{X}) \otimes \Lambda_B(\mathsf{X}) \leqslant 4\lambda\, \mathbb{1}_4.$$

Therefore,

$$\langle\Lambda_A(\mathsf{X}) \otimes \Lambda_B(\mathsf{X}), \Psi_2\rangle \leqslant 4\lambda.$$

The same argument applied to $\Lambda_A(\mathsf{Z}) \otimes \Lambda_B(\mathsf{Z})$ concludes the proof. ■

[1] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
[2] J. S. Bell, Physics **1**, 195 (1964).
[3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).
[4] B. S. Tsirelson, J. Soviet Math. **36**, 557 (1987).
[5] B. S. Tsirelson, Hadronic J. Suppl. **8**, 329 (1993).
[6] S. J. Summers and R. F. Werner, Commun. Math. Phys. **110**, 247 (1987).
[7] S. Popescu and D. Rohrlich, Phys. Lett. A **169**, 411 (1992).
[8] D. Mayers and A. Yao, *Proceedings of the 39th Annual Symposium on Foundations of Computer Science, Palo Alto, 1998* (IEEE, Washington, DC, 1998).
[9] D. Mayers and A. Yao, Quantum Inf. Comput. **4**, 273 (2004).
[10] R. Colbeck, Quantum and relativistic protocols for secure multiparty computation, Ph.D. thesis, University of Cambridge, 2006, arXiv:0911.3814.
[11] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, Nature (London) **464**, 1021 (2010).
[12] R. Colbeck and A. Kent, J. Phys. A: Math. Theor. **44**, 095305 (2011).
[13] U. Vazirani and T. Vidick, *Proceedings of the 44th Annual ACM Symposium on Theory of Computing, New York, 2012* (ACM, New York, 2012), pp. 61–76.
[14] C. A. Miller and Y. Shi, J. ACM **63**, 33 (2016).
[15] J. Bouda, M. Pawłowski, M. Pivoluska, and M. Plesch, Phys. Rev. A **90**, 032313 (2014).
[16] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).
[17] A. Acín, N. Gisin, and L. Masanes, Phys. Rev. Lett. **97**, 120405 (2006).
[18] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).
[19] B. W. Reichardt, F. Unger, and U. Vazirani, Nature (London) **496**, 456 (2013).
[20] U. Vazirani and T. Vidick, Phys. Rev. Lett. **113**, 140501 (2014).
[21] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick, Nat. Commun. **9**, 459 (2018).
[22] J. Silman, A. Chailloux, N. Aharon, I. Kerenidis, S. Pironio, and S. Massar, Phys. Rev. Lett. **106**, 220501 (2011).
[23] J. Kaniewski and S. Wehner, New J. Phys. **18**, 055004 (2016).
[24] J. Ribeiro, L. P. Thinh, J. Kaniewski, J. Helsen, and S. Wehner, Phys. Rev. A **97**, 062307 (2018).
[25] J. Ribeiro, G. Murta, and S. Wehner, arXiv:1609.08487.
[26] J. Ribeiro, G. Murta, and S. Wehner, Phys. Rev. A **97**, 022307 (2018).
[27] A. Ekert and R. Renner, Nature (London) **507**, 443 (2014).
[28] A. Acín, D. Cavalcanti, E. Passaro, S. Pironio, and P. Skrzypczyk, Phys. Rev. A **93**, 012319 (2016).
[29] M. N. Bera, A. Acín, M. Kuś, M. W. Mitchell, and M. Lewenstein, Rep. Prog. Phys. **80**, 124001 (2017).
[30] P. Sekatski, J.-D. Bancal, S. Wagner, and N. Sangouard, Phys. Rev. Lett. **121**, 180505 (2018).
[31] J.-D. Bancal, N. Sangouard, and P. Sekatski, Phys. Rev. Lett. **121**, 250506 (2018).
[32] M. O. Renou, J. Kaniewski, and N. Brunner, Phys. Rev. Lett. **121**, 250507 (2018).
[33] S. Wagner, J.-D. Bancal, N. Sangouard, and P. Sekatski, arXiv:1812.02628.

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

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

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

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

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

[39] M. McKague, New J. Phys. **18**, 045013 (2016).

[40] Y. Wang, X. Wu, and V. Scarani, New J. Phys. **18**, 025021 (2016).

[41] I. Šupić, R. Augusiak, A. Salavrakos, and A. Acín, New J. Phys. **18**, 035013 (2016).

[42] M. McKague, Quantum **1**, 1 (2017).

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

[44] A. Kalev and C. A. Miller, Quantum Sci. Technol. **3**, 015002 (2017).

[45] O. Andersson, P. Badziąg, I. Bengtsson, I. Dumitru, and A. Cabello, Phys. Rev. A **96**, 032119 (2017).

[46] I. Šupić, A. Coladangelo, R. Augusiak, and A. Acín, New J. Phys. **20**, 083041 (2018).

[47] A. Coladangelo and J. Stark, arXiv:1709.09267.

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

[49] 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).

[50] W.-H. Zhang, G. Chen, P. Yin, X.-X. Peng, X.-M. Hu, Z.-B. Hou, Z.-Y. Zhou, S. Yu, X.-J. Ye, Z.-Q. Zhou, X.-Y. Xu, J.-S. Tang, J.-S. Xu, Y.-J. Han, B.-H. Liu, C.-F. Li, and G.-C. Guo, npj Quantum Inf. **5**, 4 (2019).

[51] W.-H. Zhang, G. Chen, X.-X. Peng, X.-J. Ye, P. Yin, Y. Xiao, Z.-B. Hou, Z.-D. Cheng, Y.-C. Wu, J.-S. Xu, C.-F. Li, and G.-C. Guo, Phys. Rev. Lett. **121**, 240402 (2018).

[52] S.-Y. Lee, J. Park, J. Kim, and C. Noh, Phys. Rev. A **95**, 012134 (2017).

[53] J.-D. Bancal, M. Navascués, V. Scarani, T. Vértesi, and T. H. Yang, Phys. Rev. A **91**, 022115 (2015).

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

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

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

[57] X. Wu, J.-D. Bancal, M. McKague, and V. Scarani, Phys. Rev. A **93**, 062121 (2016).

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

[59] N. D. Mermin, Phys. Rev. Lett. **65**, 1838 (1990).

[60] R. D. Gill, Stat. Sci. **29**, 512 (2014).

[61] D. Elkouss and S. Wehner, npj Quantum Inf. **2**, 16026 (2016).

[62] P.-S. Lin, D. Rosset, Y. Zhang, J.-D. Bancal, and Y.-C. Liang, Phys. Rev. A **97**, 032309 (2018).

[63] J. Sikora, A. Varvitsiotis, and Z. Wei, Phys. Rev. Lett. **117**, 060401 (2016).

[64] Y.-C. Liang and A. C. Doherty, Phys. Rev. A **73**, 052116 (2006).

[65] A. Acín, S. Massar, and S. Pironio, Phys. Rev. Lett. **108**, 100402 (2012).

[66] L. Masanes, Phys. Rev. Lett. **97**, 050503 (2006).

[67] A. Peres, Found. Phys. **29**, 589 (1999).

[68] T. Vértesi and N. Brunner, Nat. Commun. **5**, 5297 (2014).

[69] M. Ardehali, Phys. Rev. A **46**, 5375 (1992).

[70] A. V. Belinskii and D. N. Klyshko, Phys. Usp. **36**, 653 (1993).

[71] A. Gheorghiu, P. Wallden, and E. Kashefi, New J. Phys. **19**, 023043 (2017).

[72] I. Šupić and M. J. Hoban, New J. Phys. **18**, 075006 (2016).

[73] S. Goswami, B. Bhattacharya, D. Das, S. Sasmal, C. Jebaratnam, and A. S. Majumdar, Phys. Rev. A **98**, 022311 (2018).

[74] A. Tavakoli, J. Kaniewski, T. Vértesi, D. Rosset, and N. Brunner, Phys. Rev. A **98**, 062307 (2018).

[75] M. Farkas and J. Kaniewski, Phys. Rev. A **99**, 032316 (2019).

[76] K. T. Goh, J. Kaniewski, E. Wolfe, T. Vértesi, X. Wu, Y. Cai, Y.-C. Liang, and V. Scarani, Phys. Rev. A **97**, 022104 (2018).

[77] J.-D. Bancal, N. Gisin, Y.-C. Liang, and S. Pironio, Phys. Rev. Lett. **106**, 250404 (2011).

[78] K. F. Pál and T. Vértesi, Phys. Rev. A **83**, 062123 (2011).

[79] T. Moroder, J.-D. Bancal, Y.-C. Liang, M. Hofmann, and O. Gühne, Phys. Rev. Lett. **111**, 030501 (2013).

[80] R. Arnon-Friedman and H. Yuen, in *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*, edited by I. Chatzigiannakis, C. Kaklamanis, D. Marx, and D. Sannella, Leibniz International Proceedings in Informatics (LIPIcs) (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2018), Vol. 107, pp. 11:1–11:12.

[81] R. Arnon-Friedman and J.-D. Bancal, New J. Phys. **21**, 033010 (2019).

[82] We acknowledge J. Ribeiro for the proof of Proposition 3 for mixed input states (private communication), while a proof for pure input states was found before by A. Dahlberg (private communication).

[83] https://github.com/timcp/Self-Testing_Pure_TwoQubit_States (2018).