



UNIwersytet Warszawski
Wydział Fizyki

GABRIEL PEREIRA ALVES
nr albumu: 435726

BELL NON-LOCALITY AND CERTIFICATION IN QUANTUM DEVICES
Nielokalność Bella i certyfikacja urządzeń kwantowych

Praca doktorska wykonana
pod kierunkiem
Dr hab. Jędrzej Kaniewski

Warszawa, wrzesień 2024

Bell non-locality is a fundamental concept in quantum mechanics that reveals the non-classical correlations between entangled quantum systems, which cannot be explained by any local hidden variable theory. This phenomenon is not just of theoretical interest but also plays a critical role in the practical certification of quantum devices. Specifically, Bell non-locality provides a powerful tool for certifying the operation of quantum devices in a device-independent manner, meaning that the certification does not rely on trusting the internal workings of the devices but rather on the statistical properties of the measurement outcomes.

This thesis comprises a collection of three papers that explore Bell non-locality and its most direct application, the certification of quantum devices. The first part of the thesis offers a brief introduction to these topics, while the second part presents each of the three studies.

The first of these studies presents a standard approach to both topics by constructing a family of Bell inequalities and characterising their optimal realisation. This includes subjecting the measurements of one of the parties to a device-independent certification protocol and self-testing the optimal shared state.

In the second study, we introduce a variation of the random access code (RAC) protocol, termed biased RACs. We then develop a semi-device-independent certification protocol for the decoding measurements that covers all cases of 2-bit encoding and some cases of 3-bit encoding.

Finally, the last study explores the use of general Machine Learning and data science techniques in the simplest Bell scenario and its variant for the correlation space, framing the problem as a classification task to distinguish non-signalling behaviours between quantum and not quantum. While Machine Learning models do not outperform existing methods, our exploration of behaviour classification techniques has led to the development of new variants of the see-saw and NPA algorithms, allowing us to explore non-exposed points on the boundary of the CHSH quantum set.

Nielokalność Bella jest fundamentalną koncepcją w mechanice kwantowej, która ujawnia nieklasyczne korelacje między splątanymi układami kwantowymi, których nie można wyjaśnić żadną lokalną teorią zmiennych ukrytych. Zjawisko to nie jest tylko przedmiotem zainteresowania teoretycznego, ale odgrywa również kluczową rolę w praktycznej certyfikacji urządzeń kwantowych. W szczególności nielokalność Bella zapewnia potężne narzędzie do certyfikacji działania urządzeń kwantowych w sposób niezależny od urządzenia, co oznacza, że certyfikacja nie polega na zaufaniu do wewnętrznych działań urządzeń, ale raczej na statystycznych właściwościach wyników pomiaru.

Niniejsza rozprawa składa się z trzech artykułów, które badają nielokalność Bella i jej najbardziej bezpośrednio zastosowanie, certyfikację urządzeń kwantowych. Pierwsza część rozprawy zawiera krótkie wprowadzenie do tych tematów, podczas gdy druga część przedstawia każde z trzech badań.

Pierwsze z badań omawia standardowe podejście do obu tematów poprzez skonstruowanie rodziny nierówności Bella i scharakteryzowanie ich optymalnej realizacji. Obejmuje to poddanie pomiarów jednej ze stron niezależnemu od urządzenia protokołowi certyfikacji i samodzielne testowanie optymalnego współdzielonego stanu.

W drugim badaniu wprowadzamy odmianę protokołu kodu losowego dostępu (RAC), określaną jako stronnicze RAC. Następnie opracowujemy półniezależny od urządzenia protokół certyfikacji dla pomiarów dekodowania, który obejmuje wszystkie przypadki kodowania 2-bitowego i niektóre przypadki kodowania 3-bitowego.

Ostatnie badanie przedstawia zastosowanie ogólnych technik uczenia maszynowego i nauki o danych w najprostszym scenariuszu Bella i jego wariacie dla przestrzeni korelacji, ujmując problem jako zadanie klasyfikacyjne w celu odróżnienia zachowań niesygnalizacyjnych między kwantowymi i niekwantowymi. Podczas gdy modele uczenia maszynowego nie przewyższają istniejących metod, nasze badanie technik klasyfikacji zachowań doprowadziło do opracowania nowych wariantów algorytmów huśtawki i NPA, co pozwala nam badać nieeksponowane punkty na granicy zbioru kwantowego CHSH.

Abstract	2
Streszczenie	3
Preface	5
1 General introduction	6
1.1 Bell’s theorem	7
1.2 Device-independence and certification	11
1.3 Geometrical structures in Bell non-locality	12
1.3.1 CHSH correlation space	14
2 Specific concepts	15
2.1 Optimality of any pair of incompatible rank-one projective measurements for some non-trivial Bell inequality	15
2.1.1 Self-testing	17
2.2 Semidefinite programming techniques	18
2.2.1 NPA hierarchy	19
2.2.2 See-saw optimisation	21
2.3 Quantum random access codes	22
3 Set of papers	24
3.1 Comments on “ <i>Optimality of any pair of incompatible rank-one projective measurements for some nontrivial Bell inequality</i> ”	24
3.2 Comments on “ <i>Biased random access codes</i> ”	39
3.3 Comments on “ <i>Machine Learning meets the CHSH scenario</i> ”	62
Concluding remarks	78
References	79

The objective of this thesis is to introduce the concepts used in each of the works that make up this collection of papers, namely

1. *Optimality of any pair of incompatible rank-one projective measurements for some non-trivial Bell inequality*, published in Physical Review A in September 2022;
2. *Biased random access codes*, published in Physical Review A in October 2023; and
3. *Machine Learning meets the CHSH scenario*, to be published.

In summary, each of these works can be categorised either into the field of Bell non-locality (papers 1 and 3) or certification of quantum devices (1 and 2). To the end of logically organising the content, we have divided the introduction into two chapters. In the first chapter, we discuss the concepts that are common to all three works. We start by the historical introduction to the phenomenon of Bell non-locality, passing to the explanation of Bell's seminal theorem in Sec. 1.1, and applications in Sec. 1.2, culminating with a geometrical interpretation of Bell non-locality provided in Sec. 1.3.

In Chapter 2, we introduce the concepts specific to each study, beginning with the idea of self-testing, essential for the understanding of the first paper. We then proceed to explore the SDP techniques employed in both the second and third papers, specifically examining the NPA hierarchy and the see-saw optimisation. The chapter concludes with an introduction to the quantum random access code protocol, a quantum device distinct from standard Bell scenarios and the main focus of the third paper.

In Chapter 3, each work is presented, followed by a short note on the author's contributions. We end our considerations in the Concluding Remarks section.

To better understand the topic of Bell non-locality, it is worth initiating with a short historical introduction to the subject. Our starting point will be the thought experiment proposed in Ref. [1] which is a simplification of that presented in the seminal paper by Einstein, Podolsky and Rosen (EPR) in 1935 [2]. In both cases, the objective is to show that the result of the experiment is inconsistent with the concept of local-realism, which leads the authors to conclude that quantum theory was incomplete.

To put it into perspective, in addition to encompassing local causality, i.e., no effect can propagate faster than the speed of light, the concept of local-realism includes the idea that the properties of a given physical system are well-defined prior and independently of any measurement outcome. In other words, the argument presented by EPR seeks to address the problem in quantum mechanics that, although the wave function evolves deterministically according to the Schrödinger equation, the measurement process does not. It is in this context that the so-called EPR paradox arises.

In a simplified version of it, we consider an entangled state $|\psi^-\rangle$ composed of two parts, in which each part consists of a two-level system:

$$|\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle), \quad (1.1)$$

where $|0\rangle$ represents the ground state and $|1\rangle$ represents the excited state. When the parts are separated from each other, a measurement is carried out on one side, collapsing the joint state into either $|01\rangle$ or $|10\rangle$. As a result, one part effectively steers the marginal state of the opposite part, regardless of how far apart the system has been split.

Although this effect may seem obvious at first glance, after all the state was initially anticorrelated, the paradox only appears when the experiment is analysed from a non-realistic perspective. That is, assuming a classical system, the result would not be surprising at all, as the measurement effect can always be traced back to the initial preparation. On the other hand, in quantum mechanics, the indeterminacy of the state before measurement does not allow the same conclusion, leaving no room for any explanation that avoids the violation of local causality.

In Refs. [1, 2], the authors continue to argue that, in the absence of local-realism, quantum mechanics does not provide a complete description of physical reality. This limitation leads to the proposal of incorporating mechanisms beyond quantum theory, commonly referred to as hidden variables, which aim to recover realism and subsequently restore local-causality within the theory. In fact, a few propositions were made in this direction, being the most notable of them the pilot wave theory, introduced by de Broglie, in 1927 [3], and later continued by Bohm [4, 5]. In a nutshell, this theory presupposes the existence of an element denoted pilot wave that would serve as a kind of guide for particle motion. In this context, the hidden variables consist of the positions of such particles moving on the pilot guide and are not available for empirical verification due to experimental constraints. The pilot wave, in turn, assumes the role of the wave function, and is described by the Schrödinger equation, making the predictions of this theory consistent with the standard Copenhagen interpretation. However, despite recovering realism, this interpretation failed to produce a local theory, since its pilot wave is still instantly affected in its entirety by local measurements. That is, as soon as the shape of the pilot wave is disturbed locally, the trajectories of the particles carried therein are immediately modified, breaking local causality. In conclusion, the pilot wave showed that it was possible to complete quantum mechanics, but it did not end the debate as it featured non-local causality in the same way as criticised by EPR.

Nonetheless, in the following years, this theory would still play an important role in the development of quantum mechanics. During the 1930s, the debate concerning the completeness of quantum mechanics revolved around two main pillars, namely the aforementioned EPR argument and von Neumann's impossibility theorem [6]. To put it briefly, von Neumann argued that any theory incorporating hidden variables would inherently conflict with quantum mechanics, reinforcing the already predominant Copenhagen interpretation. Although the pilot wave theory was introduced before von Neumann's result, de Broglie, its original proposer, ended up abandoning it [7]. This effectively diminished the impact of the EPR argument at that time, reducing it to a mere criticism rather than a substantial counterpoint to the more widely accepted Copenhagen interpretation. However, in 1952, Bohm independently came up with the same theoretical framework proposed by de Broglie, which shed light on the idea that von Neumann's impossibility theorem did not completely forbid the use of hidden variables. Although this fact had already been observed in 1935 by Grete Hermann [8], it gained notoriety through Bell in the 1960s, who not only highlighted the inconsistencies in von Neumann's argument but also proposed his most famous theorem [9].

In the following, we present this seminal concept for the simplest Bell scenario, also known as CHSH scenario, as a reference to Clauser, Horne, Shimony and Holt [10]. For this, we begin by deriving the locality condition on the probabilities describing the scenario, concluding with the introduction of the CHSH inequality as demonstrated by Bell in Ref. [11].

1.1 Bell's theorem

Similar to von Neumann's result, Bell's theorem is a no-go theorem about the existence of hidden variables in quantum mechanics. It comes from a reinterpretation of the thought experiment proposed by EPR, deriving an inequality that, when violated, implies that the correlations observed between the parts cannot be locally explained. The most fundamental aspect of the

theorem is that the proposed inequalities depend on the probabilities of the experiment, thus allowing them to be verified experimentally.

The core of the argument in Bell's theorem revolves around the causal structure behind the scenario, and is best summarised by the so-called Reichenbach's common cause principle [12]. In short, this principle states that, given two statistically dependent events A and B , the dependence can either be attributed to a direct cause (A causes B or vice versa) or to a third event C , which acts as a cause common to A and B . In the latter case, when conditioned on the common cause, A and B are independent, that is,

$$p(A, B | C) = p(A | C) p(B | C). \quad (1.2)$$

Analogously, in Bell's theorem we start from a bipartite scenario in which the parts, Alice and Bob, share a bipartite state $|\psi^-\rangle$ on which they are allowed to perform one of two dichotomic measurements per round. The experimenters' choice of measurements are encoded by variables $x, y \in \{0, 1\}$, respectively, whereas their outcomes are designated by $a, b \in \{\pm 1\}$. After a sufficient number of rounds, if any causal dependence between Alice and Bob's measurements is observed, the following relation holds

$$p(a, b | x, y) \neq p(a | x) p(b | y). \quad (1.3)$$

Then, by employing Reichenbach's principle, we assume that in each run of the experiment the parts are moved far enough apart so that the measurement events are space-like separated, thus ruling out the possibility of a direct cause. In this way, any causal relation must be the result of a common cause, here encoded by a hidden variable λ :

$$p(a, b | x, y, \lambda) = p(a | x, \lambda) p(b | y, \lambda). \quad (1.4)$$

In terms of the probabilities accessible to the observers, we now use Bayes' rule to write

$$p(a, b | x, y) = \sum_{\lambda} p(a, b, \lambda | x, y) = \sum_{\lambda} p(a, b | x, y, \lambda) p(\lambda | x, y).$$

Next, assuming the experimenters' free choice, translated here as the statistical independence between x and y and the hidden variable, that is, $p(\lambda | x, y) = p(\lambda)$ ¹, we get,

$$p(a, b | x, y) = \sum_{\lambda} p(a | x, \lambda) p(b | y, \lambda) p(\lambda), \quad (1.5)$$

a relation known as locality condition. In practice, Eq. (1.5) tells us that when the observed correlation results from a common cause, the joint probabilities must admit a decomposition in terms of the probability distribution of λ . However, although this distribution is in principle inaccessible, this decomposition effectively poses a restriction to the observed probabilities.

To make this restriction evident, consider the observables A_x and B_y associated to the choices

¹In other words, one can assume that, since the λ lives in the common past of Alice and Bob, there is no statistical dependence between the choice of measurements and the hidden variable, as λ precedes x and y . The denial of this assumption leads to an interpretation of quantum mechanics known as superdeterminism. For details, check Ref. [13].

x and y of Alice and Bob. Then, define the expected value related to the joint distribution

$$\langle A_x B_y \rangle := \sum_{a,b=\pm 1} ab p(a, b | x, y), \quad (1.6)$$

and the expected values associated to marginal distributions conditioned on λ ,

$$\langle A_{x,\lambda} \rangle := \sum_{a=\pm 1} a p(a | x, \lambda) \quad \text{and} \quad \langle B_{y,\lambda} \rangle := \sum_{b=\pm 1} b p(b | y, \lambda). \quad (1.7)$$

In the latter case, as the distributions are bounded, one can write

$$|\langle A_{x,\lambda} \rangle| \leq 1 \quad \text{and} \quad |\langle B_{y,\lambda} \rangle| \leq 1. \quad (1.8)$$

Then, by combining Eqs. (1.6) and (1.7) with Eq. (1.5), we write

$$\begin{aligned} \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle &= \sum_{\lambda} [\langle A_{1,\lambda} \rangle \langle B_{0,\lambda} \rangle - \langle A_{1,\lambda} \rangle \langle B_{1,\lambda} \rangle] p(\lambda) \\ &= \sum_{\lambda} [\langle A_{1,\lambda} \rangle \langle B_{0,\lambda} \rangle (1 \pm \langle A_{0,\lambda} \rangle \langle B_{1,\lambda} \rangle)] p(\lambda) \\ &\quad - \sum_{\lambda} [\langle A_{1,\lambda} \rangle \langle B_{1,\lambda} \rangle (1 \pm \langle A_{0,\lambda} \rangle \langle B_{0,\lambda} \rangle)] p(\lambda). \end{aligned} \quad (1.9)$$

By using Eq. (1.8), we can bound it to

$$\begin{aligned} |\langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle| &\leq \sum_{\lambda} [(1 \pm \langle A_{0,\lambda} \rangle \langle B_{1,\lambda} \rangle) + (1 \pm \langle A_{0,\lambda} \rangle \langle B_{0,\lambda} \rangle)] p(\lambda) \\ &\leq 2 \pm [\langle A_0 B_1 \rangle + \langle A_0 B_0 \rangle], \end{aligned}$$

or, in its best known form,

$$|\langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle| \leq 2. \quad (1.10)$$

The pair of inequalities in Eq. (1.10) is better known by the acronym CHSH and was first introduced as the local constraint of the homonymous scenario [10]. These inequalities are not unique, but they are part of an orbit that totals eight CHSH inequalities. The remaining elements can be obtained by cyclically permuting the signs, which is a consequence of being able to relabel the measurement choices and exchange Alice for Bob, thus keeping the scenario unchanged.

Now, to conclude the theorem, it remains to show a quantum realisation that violates one of such inequalities. To do so, consider the singlet state in Eq. (1.1), and the observables

$$A_0 = \sigma_z, \quad A_1 = \sigma_x, \quad B_0 = -\frac{\sigma_x + \sigma_z}{\sqrt{2}} \quad \text{and} \quad B_1 = \frac{\sigma_x - \sigma_z}{\sqrt{2}}, \quad (1.11)$$

where σ_x and σ_z correspond to the x and z Pauli matrices, respectively. By evaluating the above realisation, the value of $2\sqrt{2}$ is obtained, exceeding the upper bound in Eq. (1.10) and demonstrating the incompatibility of local hidden variable models with quantum mechanics.

After the publication of Bell's original paper in 1964, the next natural step was to seek for experimental evidence to support the theorem. In this sense, the first tests which successfully

demonstrated violation of a Bell inequality appeared in the following years. In 1972, Freedman and Clauser were the first to violate a variant of the CHSH inequality [14]. However, some loopholes were pointed out in their experiment, most notably what would later become known as the locality and detection loopholes. Briefly, the locality loophole concerns the experimental requirement that the measurement process be subject only to causal influences local to each part. The detection loophole raises the possibility that, in a test using pairs of photons, for instance, given a low detection efficiency, the successful detections could result in a sample which is not representative of the realisation, producing a fake violation. To overcome these issues, the first test that managed to bypass the locality loophole was due to Alain Aspect in a series of experiments carried out in the early 1980s [15, 16, 17]. In 1998, the results of Aspect were improved by a team in Innsbruck led by Anton Zeilinger [18], in which they enforced an actual space-like separation between the parts. The detection loophole was addressed shortly after, in 2001 [19], and the first so-called loophole-free tests appeared in 2015, resulting from three independent experiments [20, 21, 22] that simultaneously closed the locality and detection loopholes. For the experimental efforts in the search of a Bell inequality violation, John Clauser, Alain Aspect and Anton Zeilinger were awarded the 2022 Physics Nobel Prize.

Apart from the experimental breakthroughs mentioned above, it is also worth commenting on some important theoretical advances that have not yet been considered before we conclude this section. The first of them concerns the value of $2\sqrt{2}$ obtained through the realisation in Eq. (1.11). This is known to be the maximum value by which a quantum realisation can violate the CHSH inequality, a result demonstrated by Tsirelson in 1980 [23]. However, although quantum realisations cannot be arbitrarily non-local, a more general and strict characterisation of the quantum set is lacking. For the specific case where the marginal distributions of the outcomes are uniform for both parts, an analytical description is known and was also obtained first by Tsirelson [24]. The same result was achieved later and independently by Landau [25], Uffink [26] and Masanes [27], a reason why this description is known as TLM condition. Unlike the CHSH inequality, the TLM condition is not linear and can be written as

$$|\arcsin \langle A_0 B_0 \rangle + \arcsin \langle A_0 B_1 \rangle + \arcsin \langle A_1 B_0 \rangle - \arcsin \langle A_1 B_1 \rangle| \leq \pi. \quad (1.12)$$

As before, they also consist of an orbit containing eight elements which can be obtained by the cyclic permutation of signs.

For more general cases than the CHSH scenario, besides the numerical approaches that will be discussed in the next chapter, there are two results that apply to the quantum realisations of any Bell scenario. First and foremost, entanglement is necessary to produce non-locality. The verification of this statement is straightforward, as if we consider a separable state $\rho_{AB} = \sum_i p_i \rho_A^i \otimes \rho_B^i$, the distributions arising from it can be factorised as in Eq. (1.5), i.e.,

$$\begin{aligned} p(a, b | x, y) &= \text{tr} (\rho_{AB} A_x \otimes B_y) \\ &= \text{tr} \left[\left(\sum_i p_i \rho_A^i \otimes \rho_B^i \right) A_x \otimes B_y \right] \\ &= \sum_i p_i \text{tr} (\rho_A^i A_x) \text{tr} (\rho_B^i B_y). \end{aligned}$$

Secondly, measurement incompatibility is also required to produce non-local behaviour. The

demonstration of this result is due to Arthur Fine [28], and shows that the statistics describing a Bell scenario are local if and only if there is a joint probability distribution for the results of all measurements carried out by one part. Then, if the observables describing these measurements are joint measurable, such a distribution must exist, and, as a consequence, the observed statistics decompose as in Eq. (1.5), being necessarily local.

Lastly, it is important to note that the results discussed so far represent the early development of Bell non-locality, during a time when its study was primarily driven by its fundamental significance to quantum mechanics. At the beginning of the 90s, this paradigm would be changed with more application-driven results, in particular with Artur Ekert's celebrated work [29], regarded as one of the founding articles of quantum cryptography. In essence, Ekert's idea is based on the observation that in a non-local model, the randomness in the measurement outcomes can be used to devise a cryptographic key distribution protocol, secured by the violation of the CHSH inequality. This idea is the basis of device-independent cryptography, in which the violation of a given Bell inequality ensures the production of secure cryptographic keys, without relying on physical assumptions of the devices. In what follows, we further discuss the device-independent paradigm that emerged from Ekert's pioneering work.

1.2 Device-independence and certification

Quantum cryptography originates from the idea of exploiting the properties of quantum mechanics to ensure secure communication against external actors. The BB84 protocol, introduced in 1984 by Bennet and Brassard [30], is considered, along with Ekert's work, as one of the foundational results in the field. However, while the security of BB84 depends on the ability of the parts to implement measurements on complementary bases, the protocol proposed by Ekert introduces the concept that would later become known as device-independence, which we explore below. In a nutshell, Ekert's protocol relies on the secrecy of a bit string key shared by the parts aiming to establish a secure communication channel. The encryption (decryption) algorithm can be thought of as a publicly known function which receives a pair of inputs – message and key – and returns an encrypted (decrypted) message. For this reason, the effort in creating this channel is resumed to the generation and distribution of the key, which is the final goal of the protocol.

That said, Ekert's key distribution scheme can be described as a bipartite Bell scenario in which the parts share a singlet state, as in Eq. (1.1). Subsequently, each part conducts one of three possible measurements, given by,

$$A_0 = \sigma_z, \quad A_1 = \frac{1}{\sqrt{2}}(\sigma_z + \sigma_x) \quad \text{and} \quad A_2 = \sigma_x, \quad (1.13)$$

for Alice, and

$$B_0 = \frac{1}{\sqrt{2}}(\sigma_z + \sigma_x), \quad B_1 = \sigma_x \quad \text{and} \quad B_2 = \frac{1}{\sqrt{2}}(\sigma_x - \sigma_z), \quad (1.14)$$

for Bob. Notice that if the observables corresponding to A_1 and B_1 are excluded, the remaining set of measurements matches exactly with the one in Eq. (1.11), with a sign flip at B_0 . Hence, after a sufficiently long series of executions, when Alice and Bob publicly disclose their

measurement choices, they can verify the value of

$$S_{\text{CHSH}} = \langle A_0 B_0 \rangle - \langle A_0 B_2 \rangle + \langle A_2 B_0 \rangle + \langle A_2 B_2 \rangle.$$

If S_{CHSH} is not violated, it indicates to both parts that their state is not as entangled as $|\psi^-\rangle$ or that their measurement devices do not match those in Eqs. (1.13) and (1.14). Consequently, they conclude that either the devices are not reliable or the protocol has been influenced by a third part. In contrast, when maximal violation occurs, the outcomes obtained when both parts chose the same measurement bases (A_1, B_0) and (A_2, B_1) are perfectly anticorrelated, and are then used to create the key. In the more realistic case where the violation is not maximal, an alleged eavesdropper could have access to a fraction of the key, and the parts would need to further resort to post-processing protocols such as information reconciliation and privacy amplification before completing the key distribution scheme. In other words, Ekert's approach only requires the devices to be pre-tested. Even if such devices are not ideally implemented, the CHSH violation ensures that an eavesdropper cannot gain complete access to the key.

Although the first work that can be said to be device-independent is attributed to Ekert, this aspect would only be fully recognised later. The first time this term was actually employed was in Ref. [31]², referring to a protocol that builds its trust based on the violation of a Bell inequality, thus dismissing any pre-defined physical assumptions on the devices (e.g. dimensionality). In other words, the idea of device-independence is closely related to that of certification, which in turn is often used as the act of validating a quantum process. It is important to note, however, that certification has a broader meaning and is used to designate validation protocols based on weaker assumptions than those typically used by a device-independent scheme.

In the next chapter, we provide a brief introduction to another device-independent protocol known as self-testing [33, 34]. Briefly, the idea of self-testing refers to the device-independent scheme in which the device's statistics allows the certification of its underlying state and measurements. This idea is one of the central points of the first paper referred to in this thesis, which mixes self-testing with other certification techniques to produce the main result. But before we go any further, we finish this chapter with a fundamental concept which is recurrently used in Bell non-locality, which concerns the interpretation of this phenomenon in the geometrical context of polytope theory.

1.3 Geometrical structures in Bell non-locality

Recent works on Bell non-locality frequently use a geometric interpretation, independently attributed to Froissart in 1981 [35], and subsequently to Garg and Mermin [36], and Pitowsky [37, 38, 39]. This section does not explore the details of these studies; instead, it offers a brief introduction to the geometric concepts that have emerged from them. To do so, we use the notation introduced by Tsirelson in Ref. [40] and resort again to the CHSH scenario as our starting point.

Defining a Bell scenario involves specifying the number of parts involved in the experiment, the number of measurements per part, and the outcomes for each measurement. Not all Bell scenarios are uniform; each part may have a different number of measurements, and each measurement may have a distinct set of outcomes. In this context, the CHSH scenario is the simplest,

²In Ref. [32], the author explains the historical appearance of the term in his App. B.

as it is composed by two parts, Alice and Bob, who each have access to two dichotomic measurements. Within this setup, a *behaviour*, alternatively denoted as *correlation*, is characterised by a tuple of 16 probabilities $p(a, b | x, y)$, with $a, b, x, y \in \{0, 1\}$, organised as,

$$\mathbf{p} := [p(00|00), \dots, p(11|00), \dots, \dots, p(00|11), \dots, p(11|11)]. \quad (1.15)$$

If written in this way, the behaviour of a given CHSH experiment can be thought of as a point belonging to a region of \mathbb{R}^{16} delimited by,

$$p(a, b | x, y) \geq 0, \quad \forall a, b, x, y, \quad (1.16a)$$

$$\text{and } \sum_{a,b} p(a, b | x, y) = 1, \quad \forall x, y, \quad (1.16b)$$

which is denoted as the CHSH probability space. Although the inputs of every such behaviours are guaranteed to be probability distributions, they do not necessarily correspond to an actual physical implementation of the scenario. For this, we resort to correlation sets that are more restrictive than the probability space.

The first of these sets is called non-signalling and is often abbreviated as \mathcal{NS} . It incorporates the behaviours in which the marginal distributions are locally constrained, i.e., $p(a | x, y) \equiv p(a | x)$ and $p(b | x, y) \equiv p(b | y)$, thus ensuring no direct causal dependence between parts. In terms of the joint probabilities $p(a, b | x, y)$, these constraints are written as

$$\begin{aligned} \sum_b p(a, b | x, y) &= \sum_b p(a, b | x, y'), \quad \forall a, x, y, y' \in \{0, 1\}, \text{ with } y \neq y', \\ \sum_a p(a, b | x, y) &= \sum_a p(a, b | x', y), \quad \forall b, x, x', y \in \{0, 1\}, \text{ with } x \neq x', \end{aligned} \quad (1.17)$$

and are named non-signalling conditions. Due to Eqs. (1.16b) and the additional non-signalling conditions, the non-signalling set effectively belongs to an 8-dimensional subspace of the original probability space. A few important properties of the non-signalling set are that it is convex, implying that the convex combination of two or more behaviours in \mathcal{NS} remains non-signalling, and it is bounded by the 16 inequalities shown in Eq. (1.16a). In addition, as the convex hull of these inequalities defines the boundaries of \mathcal{NS} , the non-signalling set forms a polytope, i.e., it can be generated by the intersection of a finite number of half-spaces. Equally important, is that non-signalling behaviours admit a parametrisation alternative to that in Eq. (1.15), written in terms of its marginals and correlators: $[\langle A_x \rangle, \langle B_y \rangle, \langle A_x B_y \rangle]_{x,y}$. These parameters are respectively defined as:

$$\langle A_x \rangle = \sum_a a p(a | x), \quad \langle B_y \rangle = \sum_b b p(b | y) \quad \text{and} \quad \langle A_x B_y \rangle = \sum_{a,b} ab p(a, b | x, y), \quad (1.18)$$

where outcomes a and b are labelled as ± 1 .

When considering behaviours that admit a local decomposition as in Eq. (1.5), we end up with a more restrictive set than \mathcal{NS} which is named local set and indicated as \mathcal{L} . Like the non-signalling set, the local set is convex and bounded, also forming a polytope. It is important to note that, in both cases, the inequalities in Eqs. (1.16a) and Eq. (1.10) serve as the facets of these polytopes. For the non-signalling set, these facets are represented by the 16 non-negativity

inequalities in Eq. (1.16a), whereas for the local set, they also include the eight CHSH inequalities described as in Eq. (1.10).

Another restriction to \mathcal{NS} is the so-called quantum correlations set, also indicated as \mathcal{Q} . It is composed of behaviours the entries of which admit the following realisation:

$$p(a, b | x, y) = \text{tr}(\rho_{AB} M_{a|x} \otimes N_{b|y}), \quad (1.19)$$

where ρ_{AB} is a state acting in the joint Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ of Alice and Bob, and $M_{a|x}$ and $N_{b|y}$ are operators acting on \mathcal{H}_A and \mathcal{H}_B , respectively. The sets $\{M_{a|x}\}_a$ and $\{N_{b|y}\}_b$ are also referred to as POVMs (positive operator valued measures), and each of its elements, denominated effects, are positive semi-definite operators summing to identity. Each effect $M_{a|x}$ is associated to the outcome a of the x -th measurement of Alice, as well as $N_{b|y}$ is associated to b -th outcome of Bob's measurement choice y . Together with \mathcal{NS} and \mathcal{L} , the quantum set obeys a strict inclusion relation in the CHSH scenario, given by $\mathcal{L} \subset \mathcal{Q} \subset \mathcal{NS}$. Nevertheless, unlike the first two sets, \mathcal{Q} does not form a polytope; it is also a bounded convex set, but it cannot be finitely generated.

It is important to note that the formulation of quantum correlations presented in Eq. (1.19) is not unique, and is commonly referred to as the tensor product paradigm. For finite-dimensional Hilbert spaces, it is equivalent to define quantum correlations as those generated by measurements acting on a single, larger Hilbert space, where $[M_{a|x}, N_{b|y}] = 0, \forall a, b, x, y$. This alternative definition is known in the literature as the commuting paradigm. For infinite-dimensional Hilbert spaces, it is known that the commuting paradigm is more comprehensive than the tensor product [40], and the question related to the equivalence of both definitions was conjectured in a problem known as Tsirelson's problem [41]. In 2020, the question was resolved as negative in Ref. [42]. Here, as we only use finite-dimensional Hilbert spaces, we stick to the tensor product formulation unless otherwise noticed.

1.3.1 CHSH correlation space

One last topic worth mentioning within the geometrical structures in non-locality is the region of the CHSH scenario obtained by considering non-signalling behaviours satisfying $\langle A_x \rangle = \langle B_y \rangle = 0, \forall x, y$. This region is known as correlation space, and it is reasonably better understood than the entire scenario³. All three sets \mathcal{L} , \mathcal{Q} and \mathcal{NS} have an analytical description of their boundaries in this region, and are limited by the 4-cube given by $|\langle A_x B_y \rangle| \leq 1, \forall x, y$. In addition, as in the entire CHSH scenario, the local set is further constrained by the eight CHSH inequalities, while \mathcal{Q} is delimited by the eight variants of the TLM conditions in Eq. (1.12). It is important to note that for the entire CHSH space, as well as for any other non-locality scenario, there is no known analytical description of the quantum boundary, which makes the TLM inequalities the only case in which the boundary of the quantum set is known analytically.

³See Ref. [43] for a comprehensive review about the correlation space.

This chapter is dedicated to presenting the specific concepts contained in the works that make up this thesis that have not yet been addressed in the text. The content is divided into three sections, each of which relates to one of the papers referred in the Preface.

In Sec. 2.1, we provide a brief summary of the results contained in the first paper, followed by the introduction of the concept of self-testing, necessary for deducing Theorem 2 of that work. Similarly, Sec. 2.2 covers the SDP techniques employed to generate the data used by the Machine Learning models in the article *Machine Learning Meets the CHSH Scenario*. We conclude this chapter in Sec. 2.3, where we discuss the quantum random access codes protocol and a novel variant, named after the work *Biased random access codes*.

It is important to note that the main goal of this chapter is not to provide a summary of each article, but rather to address the conceptual gaps they may leave. While certain concepts are thoroughly introduced in the texts, like the preliminary section on Machine Learning in the last paper, other concepts assume the reader's prior knowledge. This chapter is focused on covering those latter topics.

2.1 Optimality of any pair of incompatible rank-one projective measurements for some non-trivial Bell inequality

In the article sharing the same title as this section, here denoted as Ref. [44], we study the bipartite Bell scenario that was introduced previously, in Ref. [45], characterised by an integer $d \geq 2$. Here, we have two important distinctions when compared to our previous example, the CHSH scenario. Firstly, this scenario is not uniform (see Fig. 2.1a), meaning that the two parts, Alice and Bob, do not have access to the same number of measurements and outcomes. For a fixed d , Alice chooses among d^2 measurements, labelled by a two-character string $x := x_1x_2$, where $x_1, x_2 \in \{1, \dots, d\}$. In contrast, Bob has only two measurement choices, represented by $y \in \{1, 2\}$. Alice's measurements outcomes are then given by $a \in \{1, 2, \perp\}$, whereas Bob's outcomes are represented by $b \in \{1, \dots, d\}$.

Secondly, the inequality investigated in Ref. [44] is not a facet Bell inequality, but it is tailored

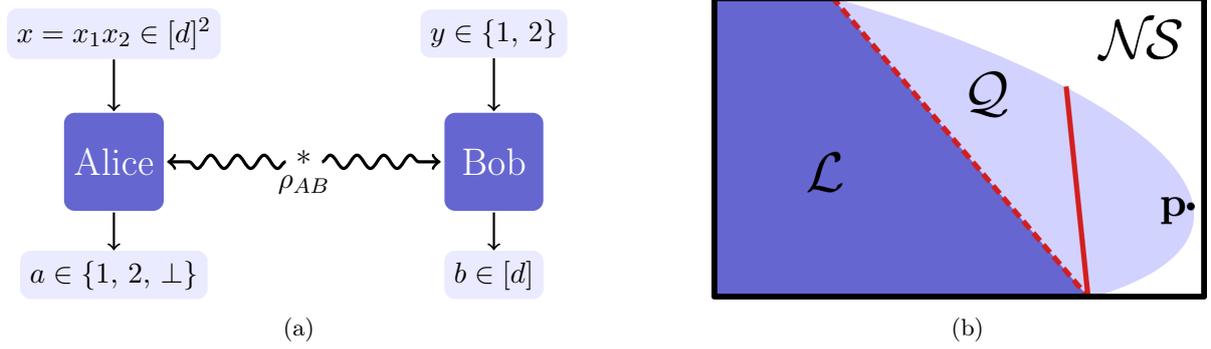


Fig. 2.1: (a) The Bell scenario investigated in Ref. [44], for a given $d \geq 2$. Each box represents one of the parts sharing a state ρ_{AB} , with the upper arrows designating the measurement choices and the lower arrows representing the alphabet of outcomes. The variables belong to the following sets: $b, x_1, x_2 \in \{1, \dots, d\}$, $y \in \{1, 2\}$ and $a \in \{1, 2, \perp\}$. (b) A 2D representation of a facet (dashed line) *vs.* non-facet (continuous line) Bell inequality. The non-facet inequality is tilted such that it is optimised by the correlation point \mathbf{p} .

to a specific pair of measurements. By definition, a facet, or tight Bell inequality, is a facet of the local polytope which is not shared by \mathcal{NS} , defining the boundaries of the local set that admit quantum violation. While non-facet inequalities do not provide necessary and sufficient criteria for non-locality, they are useful in certification tasks as they can be tailored to be optimised by a specific correlation point, as illustrated in Fig. 2.1b. In other words, the inequality in Ref. [44] is constructed in a way that it is maximally violated by a realisation that implements a given pair of measurements.

To write these measurements analytically, we start with the rank-one projective pair arising from the orthonormal bases $\{|e_j\rangle\}_{j=1}^D$ and $\{|f_k\rangle\}_{k=1}^D$ in \mathbb{C}^D , and denote $\mathcal{O}_{jk} := |\langle e_j | f_k \rangle|$ as the overlap between the elements of each basis. If the bases have any common vectors, the measurements $\{|e_j\rangle\langle e_j|\}_j$ and $\{|f_k\rangle\langle f_k|\}_k$ share at least a one-dimensional subspace in \mathbb{C}^D where they are compatible. In our construction, we truncate the original Hilbert space to \mathbb{C}^d , with $d < D$, ensuring that all bases elements are distinct, and the condition $\mathcal{O}_{jk} < 1$ is satisfied. Notice that, since we are in the rank-one projective paradigm, if the measurements are incompatible, we are guaranteed that $d \geq 2$. In this way, we assume that this procedure has already been implemented, and we are given a pair of orthonormal bases $\{|e_j\rangle\}_{j=1}^d$ and $\{|f_k\rangle\}_{k=1}^d$ with overlaps satisfying $\mathcal{O}_{jk} < 1$ and $d \geq 2$.

The inequality in the scenario of Fig. 2.1a is then split into two components and can be thought of as a game where Bob tries to guess Alice's character x_y depending on the value of his input y . We define the first of these components as the correlation score, given by

$$\mathcal{C}_d := \sum_{x_1, x_2} \sum_y \lambda_x [p(a = y, b = x_y | x, y) - p(a = \bar{y}, b = x_y | x, y)] \quad (2.1)$$

where $\lambda_x := \sqrt{1 - \mathcal{O}_{x_1 x_2}}$, and \bar{y} means that the value of y is flipped from 1 to 2 or vice versa. Note that the only rounds producing a contribution different of zero in \mathcal{C}_d correspond to the cases where $b = x_y$ and $a \neq \perp$. In those cases, \mathcal{C}_d increases when Alice outputs $a = y$ and decreases when $a = \bar{y}$. Conversely, when Alice produces $a = \perp$, the value of \mathcal{C}_d remains unchanged,

motivating the introduction of an additional term and giving rise to the final score:

$$\mathcal{F}_d := \mathcal{C}_d - \frac{1}{2} \sum_{x_1 x_2} \lambda_x^2 [p(a = 1 | x) + p(a = 2 | x)]. \quad (2.2)$$

This additional term can be thought of as a penalty, which balances the cases where $a \in \{1, 2\}$, when Alice plays the game, and $a = \perp$, when Alice does not play the game.

After defining the final score in Eq. (2.2), Ref. [44] proceeds to demonstrate some results related to its optimal realisability. The first of them corresponds to the bound

$$\mathcal{F}_d \leq d - 1, \quad (2.3)$$

which is obtained by using the Cauchy-Schwarz inequality. This result is followed by the introduction of an optimal realisation that saturates this bound, in which Bob's measurement operators are given by $P_{x_1} = |e_{x_1}\rangle\langle e_{x_1}|$, for $y = 1$, and $Q_{x_2} = |f_{x_2}\rangle\langle f_{x_2}|$, for $y = 2$ ¹. Then, two certification statements are derived. The first leverages the saturation condition of Cauchy-Schwarz to certify the measurements $\{P_{x_1}\}_{x_1}$ and $\{Q_{x_2}\}_{x_2}$, while the second consists of a self-testing statement for the optimal state, which is a maximally entangled pair in $\mathbb{C}^d \otimes \mathbb{C}^d$ given by:

$$|\Phi_d^+\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^d |i, i\rangle. \quad (2.4)$$

Lastly, the main result in Ref. [44] establishes that if $\mathcal{O}_{x_1 x_2} \leq 1$ and $d \geq 2$, the bound given in Eq. (2.3) cannot be locally achieved. This implies that, for all incompatible pairs $\{P_{x_1}\}_{x_1}$ and $\{Q_{x_2}\}_{x_2}$, there exists a functional \mathcal{F}_d maximised by these measurements that cannot be maximised by any local realisation.

While the methodology for most of these results is outlined in the content of Ref. [44], the approach to the self-testing statement is presented without any prior introduction. Therefore, in the next subsection, we provide a concise summary of the self-testing protocol applied for states.

2.1.1 Self-testing

As already mentioned before, self-testing refers to the device-independent scheme in which certain correlations in a Bell experiment determine its underlying state and measurements. The earliest results in this direction were obtained in Refs. [46, 47], stating that the CHSH maximal value of $2\sqrt{2}$ can only be obtained by a maximally entangled pair of two qubits. The concept was then formalised in 2004 by Mayers and Yao [34], and applies only to the extremal points of the quantum set. However, since achieving these points in real-world scenarios is challenging, a robust to noise version of this protocol was developed later in 2006 [48]. Here, we present the weak definition of the self-testing scheme for the case of bipartite pure states, as performed in Theorem 2.2 of Ref. [44]. A review with more comprehensive definitions can be found in Ref. [49].

We start by assuming that a certain correlation \mathbf{p} is observed, the entries of which are given by joint probabilities $p(a, b | x, y)$, for all a, b, x and y . From this point, a few standard assumptions

¹The subscripts x_1 and x_2 are used instead of b , since we only get a non-zero contribution to \mathcal{C}_d when $b = x_y$.

are made, specifically that there exists a realisation such that

$$p(a, b | x, y) = \text{tr}(\rho_{AB} M_{a|x} \otimes N_{b|y}), \quad (2.5)$$

where ρ_{AB} is the state acting on the joint Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, and the operators $M_{a|x}$ and $N_{b|y}$ are the POVM effects acting on the Hilbert spaces \mathcal{H}_A and \mathcal{H}_B of Alice and Bob, respectively. These operators are known as the physical realization of \mathbf{p} , where ρ_{AB} represents the physical state, and $\{M_{a|x}\}$ and $\{N_{b|y}\}$ denote the physical measurements. Furthermore, it is assumed that the runs of the experiment are independent and identically distributed, so the physical realisation is valid for every round and \mathbf{p} characterises the behaviour of the devices.

Now, suppose we are interested in self-testing a pure state $|\Phi\rangle_{AB}$, with its density matrix represented by $\Phi_{AB} := |\Phi\rangle\langle\Phi|_{AB}$. This state, denoted as reference, cannot be uniquely identified due to two properties: Firstly, any realisation in Eq. (2.5) is invariant under local unitaries. That is, given unitaries U and V acting on \mathcal{H}_A and \mathcal{H}_B , respectively, the realisation obtained by the state $U \otimes V \rho_{AB} U^\dagger \otimes V^\dagger$ and measurement operators $UM_{a|x}U^\dagger$ and $VN_{b|y}V^\dagger$ yield the same statistics as Eq. (2.5). Secondly, if an ancillary state $\sigma_{A'B'}$ is coupled to ρ_{AB} , the operators $M_{a|x} \otimes \mathbb{1}_{A'}$ and $N_{b|y} \otimes \mathbb{1}_{B'}$ are also capable to reproduce the probabilities in Eq. (2.5).

Then, to incorporate these degeneracies in the realisation of \mathbf{p} , we resort to the concept of local isometry. An isometry is a transformation between metric spaces that preserves the distance among its elements. Therefore, an isometry which acts locally in the Hilbert space of Alice must be a unitary operator, thus preserving the inner product, and, to account for additional degrees of freedom, it must increase the dimension of the input space. In other words, if V_A is a local isometry, then $V_A : \mathcal{H}_A \rightarrow \mathcal{H}_A \otimes \mathbb{C}^{d'}$, where d' is the ancillary dimension. Similarly, we define a local isometry on \mathcal{H}_B as the unitary $V_B : \mathcal{H}_B \rightarrow \mathcal{H}_B \otimes \mathbb{C}^{d'}$. In this way, we say that \mathbf{p} self-test the reference state $|\Phi\rangle_{AB}$, if, for any physical state ρ_{AB} compatible with \mathbf{p} ,

$$(V_A \otimes V_B) \rho_{AB} (V_A^\dagger \otimes V_B^\dagger) = \Phi_{AB} \otimes \rho', \quad (2.6)$$

where ρ' is the uncharacterised part of ρ_{AB} .

In the Theorem 2.2 of Ref. [44], we self-test the state $|\Phi_d^+\rangle$ in Eq. (2.4) by explicitly constructing isometries V_A and V_B . This is done for cases where the overlap matrix $\mathcal{O}_{x_1 x_2}$ has at least one row (or column) the elements of which are all strictly positive. Additionally, we provide an example of a matrix \mathcal{O} that fails to meet this condition. In this case, a state distinct from $|\Phi_d^+\rangle$ is found, which, along with the optimal measurements $\{P_{x_1}\}_{x_1}$ and $\{Q_{x_2}\}_{x_2}$, achieves the maximum violation of \mathcal{F}_d .

The following subsection presents a brief discussion of two numerical methods commonly used in conjunction to determine maximum values of a given Bell inequality. Both tools rely on a convex optimisation technique called semidefinite programming and were featured in the article *Machine Learning Meets the CHSH Scenario*, here referred to as Ref. [50].

2.2 Semidefinite programming techniques

In Ref. [50], we explore many distinct Machine Learning (ML) classification models to address the membership problem of non-signalling correlations in the quantum set. Specifically, we leverage ML techniques to decide whether a given non-signalling correlation can be classified as quantum

or not. The key idea of this work is to use the correlation space of the CHSH scenario as a reference for examining the complete scenario. As explained in Sec. 1.3.1, the correlation space in CHSH consists of the non-signalling region where $\langle A_x \rangle = \langle B_y \rangle = 0$, $\forall x, y$, and the boundary of the quantum set is analytically described by the TLM conditions. Therefore, the available analytical description offers a solid foundation for the models and enables us to estimate the limits of the ML techniques when extending our methods to the entire quantum set.

While much of the methodology in Ref. [50] is outlined in its *Preliminaries* and *Data generation* sections, the labelling of the correlation points is carried out by two complementary methods widely used in the study of non-locality, known as the NPA hierarchy and see-saw optimisation. Specifically, the NPA hierarchy is used to derive upper bounds for the violation of Bell inequalities, whereas the see-saw method is efficient in obtaining optimal quantum realisations, thereby establishing reliable lower bounds for the inequality values. In both instances, the methods are based on a class of convex optimisation problems referred to as semidefinite programming (SDP). Considering the importance of these two methods, we provide here a brief introduction to the concept of SDPs, followed by an explanation of the NPA hierarchy and see-saw techniques.

A semidefinite program is an optimisation problem in which the objective function is optimised within the intersection of the positive semidefinite cone of matrices in the problem space. Every SDP can be generically formulated as

$$\begin{aligned} \min_X \quad & \text{tr } AX \\ \text{s.t.} \quad & \mu_i(X) = B_i, \quad i = 1, \dots, n \\ & \nu_j(X) \preceq C_j, \quad j = 1, \dots, m \\ & X \succeq 0, \end{aligned} \tag{2.7}$$

where A , B_i and C_j are given square matrices, generally complex, and μ_i and ν_j are maps which are linear on X , the square matrix to be optimised. The symbols \succeq and \preceq denote positive and negative semidefinite, respectively. As in every convex optimisation problem, a dual problem can be formulated from the primal in Eq. (2.7). When the primal or dual problem are strictly feasible², they are said to satisfy a property known as strong duality, which guarantees that the optimal value of both primal and dual problems are necessarily equal.

There are several efficient algorithms available for solving SDPs, with the solver MOSEK [51] being particularly notable. Additionally, Python libraries such as CVXPY [52, 53] and PICOS [54] provide a collection of solvers and offer a simplified interface for problem formulation. Here, we do not explore any specific strategy for solving SDPs; instead, we move on to the description of the aforementioned methods based on SDPs, starting with the NPA hierarchy.

2.2.1 NPA hierarchy

Originally introduced in 2007 by Navascués, Pironio and Acín [55], the NPA optimisation is an SDP hierarchy based on a similar method developed previously by Lasserre, in 2001 [56]. Its central idea relies on the fact that there are necessary, though not sufficient, conditions that can be tested via an SDP to verify whether a given correlation belongs to the quantum set.

²For the primal problem in Eq. (2.7), we say that it is strictly feasible when there exists a matrix X' such that $X' \succ 0$ and $\nu_j(X') \prec C_j$, for all j . For the dual problem, the definition of strict feasibility is analogous for its dual variables.

The main advantage of this method is that, in addition to providing outer approximations to the quantum set, the SDP can be recast hierarchically to obey increasingly tighter conditions. In this way, each level n of the hierarchy defines a set of correlations, indicated as \mathcal{Q}_n , which satisfies $\mathcal{Q}_n \supseteq \mathcal{Q}_{n+1}$ and converges to the quantum set in the limit $n \rightarrow \infty$ [57, 58]. It is worth noting that, for the formulation of the hierarchy, it is considered the commuting paradigm, where the measurement operators belonging to different parts act in a single Hilbert space and obey commutation relations. In this way, exceptionally here, we adopt this paradigm in the description of the problem.

Suppose that a quantum correlation \mathbf{p} is obtained in a bipartite scenario by a realisation consisting of a pure state $|\psi\rangle$ and projective measurements $\{M_{a|x}\}_a$ and $\{N_{b|y}\}_b$ of Alice and Bob, respectively. As we do not wish to impose any constraints on the dimension of the Hilbert space, the state can be assumed to be pure and the measurements, projective, without any loss of generality. Therefore, if the projectors $M_{a|x}$ and $N_{b|y}$ correspond to the realisation \mathbf{p} , the following properties are necessarily satisfied:

$$\begin{aligned} M_{a|x}M_{a'|x} &= \delta_{a'a}M_{a|x}, \\ N_{b|y}N_{b'|y} &= \delta_{b'b}N_{b|y}, \\ \sum_a M_{a|x} &= \sum_b N_{b|y} = 1, \\ M_{a|x}^\dagger &= M_{a|x}, \quad N_{b|y}^\dagger = N_{b|y}, \\ \text{and } [M_{a|x}, N_{b|y}] &= 0. \end{aligned} \tag{2.8}$$

In parallel, by defining the set of vectors $\mathcal{V}_1 := \{M_{a|x}|\psi\rangle, N_{b|y}|\psi\rangle\}_{a,b,x,y}$, we write the Gram matrix Γ_1 associated to this set, which is positive semidefinite by construction. This matrix is referred to as the moment matrix, and it has the property that some of its entries correspond to the probabilities in the correlation \mathbf{p} , whereas some others have no physical meaning.

Now, notice that if \mathbf{p} is a quantum correlation, then Γ_1 necessarily exists. Moreover, given a behaviour $\mathbf{p} \in \mathcal{NS}$, the problem of verifying the existence of a moment matrix can be formulated as an SDP, since the non-physical entries of Γ_1 are restricted by linear relations which can be derived from Eqs. (2.8). Although the solution to this optimisation problem does not fully resolve the quantum set, it defines the set \mathcal{Q}_1 of which \mathcal{Q} is a subset. A hierarchical structure can then be derived by defining the set \mathcal{V}_2 formed by products of at most two operators $M_{a|x}$ and $N_{b|y}$ acting on $|\psi\rangle$:

$$\mathcal{V}_2 := \mathcal{V}_1 \cup \{M_{a|x}M_{a'|x'}|\psi\rangle, N_{b|y}N_{b'|y'}|\psi\rangle, M_{a|x}N_{b|y}|\psi\rangle\}_{a,a',b,b',x,x',y,y'}.$$

As before, the existence of a second-order moment matrix Γ_2 associated to \mathcal{V}_2 can be checked via an SDP for a given correlation. In the affirmative case, since Γ_1 is a block of Γ_2 , the existence of a first-order moment matrix is also confirmed. The generalisation to higher orders is then straightforward; if \mathcal{Q}_n defines the set of correlations for which a moment matrix Γ_n exists, then $\mathcal{Q}_n \supseteq \mathcal{Q}_{n+1}$.

In the literature, the NPA hierarchy is widely used in the study of non-locality in two primary ways. The first, discussed in the above paragraphs, involves checking whether a given non-signalling correlation belongs to one of the supersets \mathcal{Q}_n . In Ref. [50], this application is used to generate labels for correlations uniformly distributed in the non-signalling set of the CHSH

scenario. This approach is particularly effective in the CHSH correlation space, as the hierarchy converges to the quantum set already at level one. Another key application is determining upper bounds for Bell expressions, which can be framed as an NPA optimisation due to their linearity in the probabilities of the correlation vector. In Ref. [50], we propose a variation of this second method where the optimisation is conducted within a direction in the probability space rather than a Bell expression. The benefit of this approach is that, as the optimisation is constrained to a given direction, it enables the exploration of specific points on the boundary of the NPA supersets.

2.2.2 See-saw optimisation

The so-called see-saw optimisation refers to an iterative SDP technique used to obtain interior approximations for the optimal value of Bell expressions. It was introduced in 2007 by Liang and Doherty [59] and is based on a similar method proposed by Werner and Wolf in 2001 [60], in which the optimisation is performed explicitly. The main intuition behind the see-saw method is that although finding the optimal state and measurements for a given Bell expression is difficult, the problem becomes significantly easier when either the state or the measurements are fixed, allowing the problem to be tackled using SDPs.

The main difference from this technique to the aforementioned NPA hierarchy is that here the optimisation is constrained to a certain Hilbert space dimension. For instance, suppose that, in a given bipartite scenario, we want to maximise a Bell expression defined by the functional

$$\mathcal{G} = \sum_{a,b,x,y=0} \alpha_{ab|xy} M_{a|x} \otimes N_{b|y}, \quad (2.9)$$

where $\alpha_{ab|xy}$ are real coefficients, and $M_{a|x}$ and $N_{b|y}$ represent the measurement operators of Alice and Bob, respectively. That is, we are essentially interested to find the set of state and observables that maximise the value of $\text{tr}(\mathcal{G}\rho)$. However, as this objective function is not linear, we can fix the state and measurements alternately and divide the procedure into simpler optimisation steps.

In the initial step, all measurements are fixed to a randomly chosen set of operators $\{\tilde{M}_{a|x}\}_{a,x}$ and $\{\tilde{N}_{b|y}\}_{b,y}$, which act on Hilbert spaces of a specified dimension. Under these conditions, the state ρ^* that maximises $\text{tr}(\mathcal{G}\rho)$ can be explicitly determined, as it corresponds to the eigenvector associated with the largest eigenvalue of the operator \mathcal{G} with these fixed measurements. The optimisation then continues by fixing ρ^* and optimising each measurement in turn³. For Alice's measurement choice $x = 0$, this can be performed via an SDP that maximises the terms of \mathcal{G} involving the operators $M_{a|0}$, i.e.,

$$\begin{aligned} \max_{\{M_{a|0}\}} \quad & \sum_{a,b,y} \alpha_{ab|0y} \text{tr}(\rho^* M_{a|0} \otimes \tilde{N}_{b|y}) \\ \text{s.t.} \quad & \sum_a M_{a|0} = \mathbb{1} \quad \text{and} \quad M_{a|0} \succeq 0. \end{aligned} \quad (2.10)$$

Note that each of these steps incrementally increases the value of the objective function. Then, after the state and measurements have been optimised once, the procedure can be repeated at the discretion of a pre-established stopping condition, like minimal changes in the function value

³Alternatively, due to invariance under local unitary transformations, one might retain one of the randomly chosen initial measurements for each part, for instance $x, y = 0$, and optimise the remaining measurements only.

after a certain number of iterations. By doing so, the procedure converges to a local maximum, which can be further improved by resetting the initial random realisation multiple times or by increasing the dimension of the local Hilbert spaces.

In Ref. [50], we employed a novel approach similar to the one used for the NPA hierarchy, where see-saw optimisation is carried out towards a specific direction within the probability space. The advantage of this technique is that, in the CHSH scenario, it is known that local dimension 16 suffices to reproduce any quantum correlation [61], allowing the method to explore previously uncharted regions at the boundary of the CHSH quantum set. However, the downside is that this directional constraint makes the optimisation computationally intensive, thus reducing its practical usability. In the work explored in the next section, here referred to as Ref. [62], the see-saw technique is also used as the main optimisation tool, but it is tailored to a protocol distinct from non-locality scenarios known as random access code.

2.3 Quantum random access codes

While the earliest protocols for certification of quantum devices emerged in the context of Bell non-locality, their application was extended beyond Bell scenarios in the last years. Many additional schemes have been developed for different quantum devices, such as those involving EPR steering [63], contextuality [64], and prepare-and-measure scenarios [65], to name some recent examples. In Ref. [62], we present a few certification statements for a variation of the quantum random access code (QRAC) protocol, which we briefly introduce here. A QRAC is a communication task that allows a sender to encode a classical message consisting of multiple characters into a quantum state, such that a receiver can recover any of the characters with a probability of success higher than a classical code. Initially introduced by Stephen Wiesner as conjugate coding [66] in quantum communication complexity, QRACs were later rediscovered by Ambainis et al. [67, 68] in their investigation of quantum finite automata.

A typical QRAC can be understood as the quantum implementation of a classical protocol denoted simply as RAC, which involves the canonical parts Alice and Bob. In a RAC, Alice is given a n -character string $\mathbf{x} = x_0x_1 \dots x_{n-1}$, in which each character is chosen from an alphabet of cardinality d , and is asked to encode \mathbf{x} into a single character μ . The integers n , and d completely define the RAC scenario, which is denoted by $n^d \mapsto 1$. The message μ and an input y are then sent to Bob, who is asked to recover the character x_y of \mathbf{x} , as illustrated in Fig. 2.2. Hence, the encoding process can be interpreted as a map where Alice takes the string \mathbf{x} and outputs the message μ , while the decoding process involves n maps that, upon receiving μ , produce a character b . Whenever $b = x_y$ the protocol is considered successful. The figure of merit commonly used to evaluate the performance of the protocol is the *average success probability* \bar{P} , given by

$$\bar{P} = \frac{1}{nd^n} \sum_{\mathbf{x}, y} p(b = x_y | \mathbf{x}, y), \quad (2.11)$$

where $p(b = x_y | \mathbf{x}, y)$ denotes the probability of a successful decoding when the character x_y must be recovered from \mathbf{x} . The factor $\frac{1}{nd^n}$ reflects the assumption that the inputs \mathbf{x} and y are selected at random by both parts.

Thus, a QRAC consists of a quantum realisation of this protocol, which involves a set of preparations $\rho_{\mathbf{x}}$ and measurement operators $\{M_y^b\}_b$ such that $p(b = x_y | \mathbf{x}, y) = \text{tr}(\rho_{\mathbf{x}} M_y^{x_y})$, i.e.,

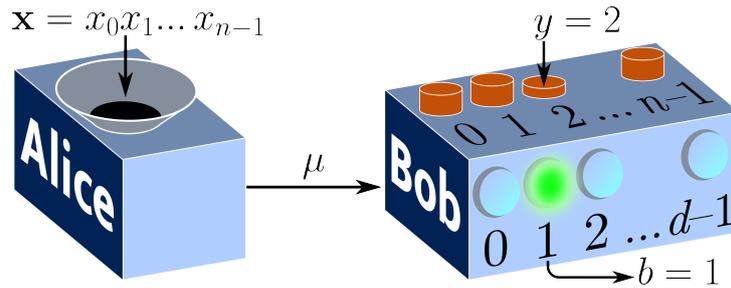


Fig. 2.2: Representation of a $n^d \rightarrow 1$ RAC scenario. The integers μ , b and x_i , where $i \in \{0, 1, \dots, n-1\}$ belong to the set $\{0, 1, \dots, d-1\}$. In the figure example, the protocol is successful if $x_2 = 1$.

Alice encodes \mathbf{x} in the state of a quantum system, and Bob decodes it via quantum measurements. In any QRAC scenario, the average success probability is always at least as high as that of its classical counterpart. Take the $2^2 \rightarrow 1$ scenario as an example: the optimal classical strategy for maximising \bar{P} requires Alice to always send one of her characters, enabling Bob to guess the other character when required, resulting in an average success of $\frac{3}{4}$. On the other hand, the QRAC implementation can exceed this by implementing a pair of decoding measurements constructed from mutually unbiased bases [69], which ensures an average success probability of $\frac{1}{2} + \frac{1}{2\sqrt{2}} \approx 0.854$.

In the literature, many variations of the RAC protocol have been developed, such as sharing a state between the parties to create entanglement-assisted RACs [70], increasing the size of the message μ [67, 68, 71], or recovering more than one character [72]. In Ref. [62], we propose to study yet another variation of the RAC protocol, which we name biased RACs, in which the distributions of \mathbf{x} and y are not necessarily uniform. In this case, figure of merit is given by

$$\mathcal{F} = \sum_{\mathbf{x}, y} \alpha_{\mathbf{x}y} p(b = x_y | \mathbf{x}, y), \quad (2.12)$$

where $\alpha_{\mathbf{x}y}$ denotes the components of an $n+1$ -order tensor, satisfying $\alpha_{\mathbf{x}y} > 0$ and $\sum_{\mathbf{x}, y} \alpha_{\mathbf{x}y} = 1$. The motivation for this approach comes from biased [73] and tilted [74] Bell inequalities, where modifying a given inequality affects its optimal value and changes the classical and quantum strategies that achieve it. The analogous idea, when applied to the RAC protocol, enables the certification of a more extensive class of measurements for certain scenarios.

The methodology adopted in Ref. [62] involved tackling the problem of optimising biased RACs through both numerical and analytical approaches. Analytically, our focus was on finding upper bounds for a few biased RAC scenarios, while numerically, we aimed to optimise the functional in Eq. (2.12) using a see-saw algorithm adapted to fit the RAC protocol. We show that attaining these upper bounds certifies a set of rank-one projective measurements for the biased scenario where $n = 2$ and 3 bits are encoded. This study also led to the development of a Python package [75] that allows users to easily set up a biased RAC scenario using its tensor $\alpha_{\mathbf{x}y}$. The package can then either find its optimal classical value by conducting an enhanced search over classical strategies, or obtain lower bounds for its optimal quantum value through see-saw optimisation.

3.1 Comments on “*Optimality of any pair of incompatible rank-one projective measurements for some nontrivial Bell inequality*”

As indicated by the title, this paper constructs a family of non-trivial Bell inequalities, meaning that its maximum value achieved by quantum theory is strictly higher than that obtained through local theories. In addition, these inequalities are optimised by a realisation that includes a pair of rank-one, projective measurements. The objective of this study is to explore how the incompatibility of these set of measurements relates to the non-locality they may generate regarding the proposed inequalities. The main result is then outlined in Corollary 2, where it is shown that each pair of such incompatible rank-one projective measurements is capable of maximising a non-trivial Bell inequality.

The student contributed to the development of the analytical results, including Theorem 1 and Lemma 1, and also proposed the main ideas behind Theorem 3 and Lemma 2. For these last two results, optimisation was required to determine the reported bounds, which were obtained from a code developed by him using the Python library *scipy.optimize*. In addition, the student was responsible for writing the appendices and the bulk of the main text.

Optimality of any pair of incompatible rank-one projective measurements for some nontrivial Bell inequality

Gabriel Pereira Alves^{*} and Jędrzej Kaniewski[†]

Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland

 (Received 17 February 2022; accepted 30 August 2022; published 23 September 2022)

Bell nonlocality represents one of the most striking departures of quantum mechanics from classical physics. It shows that correlations between spacelike separated systems allowed by quantum mechanics are stronger than those present in any classical theory. In a recent work [A. Tavakoli, M. Farkas, D. Rosset, J.-D. Bancal, and J. Kaniewski, *Sci. Adv.* **7**, eabc3847 (2021)], a family of Bell functionals tailored to mutually unbiased bases (MUBs) was proposed. For these functionals, the maximal quantum violation is achieved if the two measurements performed by one of the parties are constructed out of MUBs of a fixed dimension. Here, we generalize this construction to an arbitrary incompatible pair of rank-one projective measurements. By constructing a new family of Bell functionals, we show that for any such pair there exists a Bell inequality that is *maximally* violated by this pair. Moreover, when investigating the robustness of these violations to noise, we demonstrate that the realization which is most robust to noise is not generated by MUBs.

DOI: [10.1103/PhysRevA.106.032219](https://doi.org/10.1103/PhysRevA.106.032219)

I. INTRODUCTION

The discovery of Bell nonlocality lies among the most fundamental results of 20th-century physics. While the theoretical description was proposed by Bell in 1964 [1], the first experimental demonstration was performed by Freedman and Clauser [2], followed by the seminal work of Aspect *et al.* [3–5]. In 2015, three independent groups performed the Bell experiment in a loophole-free manner [6–8], meaning that they managed to eliminate several issues that could falsify the conclusions of the experiment.

In a nutshell, Bell nonlocality states that correlations between spatially separated parties allowed by quantum theory are stronger than those allowed by local-realistic theories [1,9]. It is easy to see that entanglement [10] and incompatibility of measurements [11] constitute two necessary resources to generate nonlocal correlations [12]. Conversely, conditions sufficient to generate nonlocality are known only for a restricted class of states or measurements. For a bipartite Bell scenario, all pure entangled states can generate nonlocality, a statement referred to as Gisin’s theorem [13]. A result of similar generality for incompatibility of measurements was obtained in Ref. [14], in which it is proved that every incompatible pair of projective measurements enables the violation of a Bell inequality. Here, we continue the study of how useful a pair of measurements is for the purpose of generating nonlocality. However, our focus is not merely on observing a Bell violation, but on producing the maximal violation allowed by quantum mechanics.

In this paper, we focus on pairs of rank-one projective measurements and generalize a framework originally presented in

Ref. [15] in which a family of bipartite Bell functionals is tailored to mutually unbiased bases (MUBs) [16,17]. In other words, in Ref. [15] the maximum value obtainable by any quantum realization of the functionals—here referred to as quantum value and denoted by β_Q —can be achieved if one of the parties implements a pair of rank-one projective measurements with uniform overlaps. Here, we study the same Bell scenario (shown in Fig. 1) as in Ref. [15], parametrized by an integer $d \geq 2$. However, our new functionals are tailored to a more general pair of rank-one projective measurements. More specifically, the measurement operators are constructed out of a pair of orthonormal bases $\{|e_j\rangle\}_{j=1}^d$ and $\{|f_k\rangle\}_{k=1}^d$ on \mathbb{C}^d and we denote the resulting overlap matrix by

$$O_{jk} := |\langle e_j | f_k \rangle|. \quad (1)$$

The only assumption we make is that these measurements are incompatible since this is a necessary condition for nonlocality.

The family of Bell functionals introduced in this paper is designed so that both the quantum value and a realization that achieves it can be written down explicitly. We also demonstrate that our functionals are nontrivial, meaning that the maximum value achievable by any classical realization—referred to as local value and denoted by β_L —is strictly smaller than the quantum value. This result is obtained by deriving a nontrivial certification statement for the measurements performed by one of the parties. Finally, for a wide class of functionals, we obtain a device-independent certification of the maximally entangled state of local dimension d .

To make our paper relevant for experiments, we investigate the robustness to noise of the proposed optimal realizations. In the standard model of noise, in which the state is replaced by an isotropic state, it can be shown that the bigger the gap between β_Q and β_L , the more robust to noise the optimal realization is. For even d , we found that the largest gap is

^{*}gpereira@fuw.edu.pl

[†]jkaniewski@fuw.edu.pl

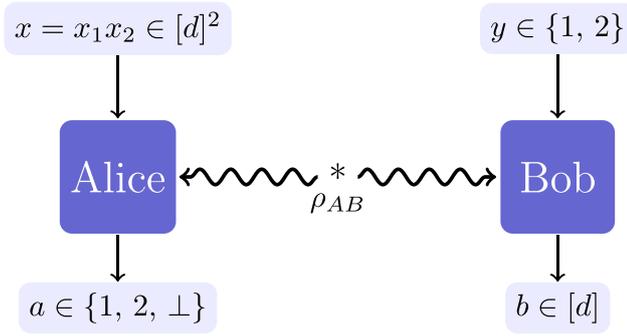


FIG. 1. A Bell scenario parametrized by $d \geq 2$. The inputs and outputs belong to the following sets: $b, x_1, x_2 \in \{1, \dots, d\}$, $y \in \{1, 2\}$, and $a \in \{1, 2, \perp\}$.

achieved if and only if the original measurements correspond to a direct sum of qubit MUBs, i.e., all the nonzero elements of the matrix O must be equal to $1/\sqrt{2}$. This contradicts our initial guess that the most noise-robust realization (in a fixed dimension) would correspond to the one generated by MUBs. For odd d , we could not determine the largest possible gap, but we have derived an upper bound on $\beta_Q - \beta_L$ which is valid for all $d \geq 3$. Furthermore, we show analytically that for every odd d there exist measurements that give rise to a realization that is more robust to noise than the realization obtained from MUBs.

II. A FAMILY OF BELL FUNCTIONALS

Our goal is to show that every pair of incompatible rank-one projective measurements is optimal for some Bell functional. To achieve this, it is convenient to first apply a simple preprocessing, which simply discards the subspace in which the measurements are compatible (which is not useful for the purpose of generating nonlocality). Then, we introduce a bipartite Bell scenario parametrized by an integer d and construct a family of functionals the quantum value of which can be computed analytically. Lastly, we present a realization that saturates the quantum value.

Consider a pair of orthonormal bases on $\mathbb{C}^{d'}$, which we denote by $\{|e_j\rangle\}_{j=1}^{d'}$ and $\{|f_k\rangle\}_{k=1}^{d'}$, and let $O'_{jk} := |\langle e_j|f_k\rangle|$ be the overlap matrix. If the two bases share a vector this results in a one-dimensional subspace of $\mathbb{C}^{d'}$ in which the measurements are compatible. We remove all such subspaces by truncating the original Hilbert space appropriately, which leads to a pair of rank-one projective measurements acting on \mathbb{C}^d for some $d \leq d'$ (since we assume that the original measurements are incompatible, we are guaranteed that $d \geq 2$). For these new measurements, the overlap matrix is guaranteed to satisfy $O_{jk} < 1$. Moreover, these new measurements can be implemented by performing the original measurements on a quantum state with appropriately chosen local support. From now, we will assume that this process has already been performed and we will restrict our attention to measurements acting on effective dimension d , and overlaps satisfying $O_{jk} < 1$.

Now, consider a bipartite Bell scenario characterized by an integer $d \geq 2$, the parties of which are named Alice and Bob.

For each d , Alice is given a two-character string denoted by $x := x_1x_2$, for $x_1, x_2 \in \{1, \dots, d\}$ and outputs $a \in \{1, 2, \perp\}$. Bob has two possible inputs labeled by $y \in \{1, 2\}$ and outputs $b \in \{1, \dots, d\}$ (see Fig. 1). We are given a pair of orthonormal bases $\{|e_j\rangle\}_{j=1}^d$ and $\{|f_k\rangle\}_{k=1}^d$ the overlap matrix of which is given by $O_{jk} := |\langle e_j|f_k\rangle|$ and satisfies $O_{jk} < 1$. We define the correlation score \mathcal{C}_d as

$$\mathcal{C}_d := \sum_{x_1, x_2=1}^d \sum_{y=1}^2 \lambda_x [p(a = y, b = x_y|x, y) - p(a = \bar{y}, b = x_y|x, y)], \quad (2)$$

where $\lambda_x := \sqrt{1 - O_{x_1x_2}^2}$ and the notation \bar{y} means that the value of y is flipped from 1 to 2 or vice versa. Note that the weights λ_x depend on the overlap matrix and, as we will later see, this ensures that the functional is tailored to the orthonormal bases $\{|e_j\rangle\}_{j=1}^d$ and $\{|f_k\rangle\}_{k=1}^d$. The only terms that produce a nonzero contribution to \mathcal{C}_d correspond to the cases when $b = x_y$ and $a \neq \perp$. In those cases, the score is increased (decreased) if Alice outputs $a = y$ ($a = \bar{y}$). In order to balance the cases where $a \in \{1, 2\}$ (Alice plays the game) and $a = \perp$ (Alice does not play the game), a penalty is introduced,¹ giving rise to our final score:

$$\mathcal{F}_d := \mathcal{C}_d - \frac{1}{2} \sum_{x_1, x_2=1}^d \lambda_x^2 [p(a = 1|x) + p(a = 2|x)]. \quad (3)$$

It turns out that the quantum value of \mathcal{F}_d does not depend on the overlap matrix O , as shown in the lemma below.

Lemma 1. For $d \geq 2$, the quantum value of \mathcal{F}_d is $d - 1$.

The proof of Lemma 1 consists of finding an upper bound on the value of \mathcal{F}_d and demonstrating that it can be saturated. Proving the upper bound relies on many elementary steps, which we present in Appendix A 1. For our purposes, it suffices to show how to construct a quantum realization that saturates the quantum value of \mathcal{F}_d .

We choose the measurements of Bob corresponding to inputs $y = 1$ and 2 to be $P_{x_1} = |e_{x_1}\rangle\langle e_{x_1}|$ and $Q_{x_2} = |f_{x_2}\rangle\langle f_{x_2}|$, respectively.² For the measurements of Alice, it is more convenient to work with Hermitian observables rather than measurement operators. For every input x of Alice, we express the three measurement operators $A_x^{(1)}$, $A_x^{(2)}$, and $A_x^{(\perp)}$ in a compact manner by defining the observable A_x as

$$A_x := (+1) \times A_x^{(1)} + (-1) \times A_x^{(2)} + 0 \times A_x^{(\perp)} \\ = A_x^{(1)} - A_x^{(2)}. \quad (4)$$

It is easy to see that if the measurement is projective, then the spectrum of A_x belongs to the set $\{\pm 1, 0\}$. In addition, any Hermitian operator the spectrum of which is contained in this set can be interpreted as an observable arising from a projective measurement, so the two representations are equivalent.

¹In the end of Appendix A 1 there is a clarification of why this penalty is necessary.

²We use x_1 (x_2) as the subscript for P (Q) because we only get a nonzero contribution to the functional when $b = x_y$.

Then, let us choose A_x as

$$A_x = \frac{1}{\lambda_x} (P_{x_1} - Q_{x_2})^\top, \quad (5)$$

where $(\cdot)^\top$ denotes the transposition in the computational basis. Because P_{x_1} and Q_{x_2} are rank one, the difference $P_{x_1} - Q_{x_2}$ is a rank-two operator. With a simple calculation, it is possible to check that the spectrum of $P_{x_1} - Q_{x_2}$ is contained in $\{\pm\lambda_x, 0\}$. Thus, since $O_{x_1x_2} < 1$ and $\lambda_x = \sqrt{1 - O_{x_1x_2}^2}$, we have $\lambda_x > 0$, and the normalizing factor ensures the spectrum of A_x is contained in $\{\pm 1, 0\}$.³

Finally, we pick the shared state as the maximally entangled state belonging to $\mathbb{C}^d \otimes \mathbb{C}^d$:

$$|\Phi_d^+\rangle = \frac{1}{\sqrt{d}} \sum_{k=1}^d |k, k\rangle. \quad (6)$$

Then, let us quickly verify that the resulting realization indeed achieves the quantum value of \mathcal{F}_d . Evaluating \mathcal{C}_d given in Eq. (2) leads to

$$\mathcal{C}_d = \sum_x \lambda_x \langle \Phi_d^+ | A_x \otimes (P_{x_1} - Q_{x_2}) | \Phi_d^+ \rangle, \quad (7)$$

where the summation over y has already been performed to give $A_x \otimes (P_{x_1} - Q_{x_2})$. Plugging in Eq. (5) and exploiting the fact that for any linear operator L we have $L \otimes \mathbb{1} | \Phi_d^+ \rangle = \mathbb{1} \otimes L^\top | \Phi_d^+ \rangle$ leads to

$$\mathcal{C}_d = \sum_x \langle \Phi_d^+ | \mathbb{1} \otimes (P_{x_1} - Q_{x_2})^2 | \Phi_d^+ \rangle. \quad (8)$$

Proceeding similarly for \mathcal{F}_d , we get

$$\mathcal{F}_d = \mathcal{C}_d - \frac{1}{2} \sum_x \lambda_x^2 \langle \Phi_d^+ | (A_x^{(1)} + A_x^{(2)}) \otimes \mathbb{1} | \Phi_d^+ \rangle. \quad (9)$$

Since the measurement operators of Alice are projective, the term in parentheses can be identified as $A_x^{(1)} + A_x^{(2)} = A_x^2$ and

$$\begin{aligned} \mathcal{F}_d &= \mathcal{C}_d - \frac{1}{2} \sum_x \lambda_x^2 \langle \Phi_d^+ | A_x^2 \otimes \mathbb{1} | \Phi_d^+ \rangle \\ &= \frac{1}{2} \sum_x \langle \Phi_d^+ | \mathbb{1} \otimes (P_{x_1} - Q_{x_2})^2 | \Phi_d^+ \rangle. \end{aligned} \quad (10)$$

Finally, since P_{x_1} and Q_{x_2} are projectors, we can evaluate the sum:

$$\sum_x (P_{x_1} - Q_{x_2})^2 = 2(d-1)\mathbb{1}, \quad (11)$$

which leads to $\mathcal{F}_d = d-1$ for the desired realization.

III. CHARACTERIZING THE OPTIMAL REALIZATION

Having computed the quantum value of \mathcal{F}_d , it would be natural to determine its local value and show that our functionals are indeed nontrivial. However, since β_L does not have a closed-form expression, let us postpone it to the next section. Fortunately, it turns out that proving $\beta_Q > \beta_L$ is possible by

characterizing the realizations that achieve β_Q . Therefore, to demonstrate that \mathcal{F}_d is nontrivial, we continue our analysis with the device-independent certification of the measurements and the state.

Theorem 1. For any $d \geq 2$ and overlap matrix satisfying $O_{x_1x_2} < 1$, under the assumption that Bob's marginal state is full rank, for any quantum realization which achieves the quantum value of \mathcal{F}_d , the following hold.

(1) Bob's measurement operators denoted by P_{x_1} and Q_{x_2} associated with inputs $y = 1$ and 2 , respectively, must satisfy

$$\begin{aligned} O_{x_1x_2}^2 P_{x_1} &= P_{x_1} Q_{x_2} P_{x_1} \quad \text{and} \\ O_{x_1x_2}^2 Q_{x_2} &= Q_{x_2} P_{x_1} Q_{x_2} \quad \forall x_1, x_2. \end{aligned} \quad (12)$$

(2) If O has a row (or column) the entries of which are all nonzeros, then there exist local isometries $V_A : \mathcal{H}_A \rightarrow \mathbb{C}^d \otimes \mathcal{H}_A$ and $V_B : \mathcal{H}_B \rightarrow \mathbb{C}^d \otimes \mathcal{H}_B$ that, when acting on the unknown state ρ_{AB} , yield

$$(V_A \otimes V_B) \rho_{AB} (V_A^\dagger \otimes V_B^\dagger) = \Phi_d^+ \otimes \rho_{\text{aux}}, \quad (13)$$

where Φ_d^+ is the density matrix of $|\Phi_d^+\rangle$ and ρ_{aux} corresponds to the uncharacterized part of ρ_{AB} .

Sketch of the proof. Here we present a sketch just for the first part (see Appendix B for a complete proof). To prove an upper bound on the quantum value of \mathcal{F}_d the Cauchy–Schwarz inequality is used several times. Saturating these inequalities allows us to deduce that for all $x = x_1x_2$ the action of Alice's and Bob's operators on ρ_{AB} satisfies the following relation:

$$\lambda_x A_x \otimes \mathbb{1} \rho_{AB} = \mathbb{1} \otimes (P_{x_1} - Q_{x_2}) \rho_{AB}. \quad (14)$$

Since our goal is to certify Bob's measurements, we can assume that Alice's measurements are projective, which implies that $A_x^3 = A_x$. Combining this identity with Eq. (14) yields a polynomial equation in terms of operators P_{x_1} and Q_{x_2} and the marginal state on Bob's side ρ_B . Assuming that ρ_B is full rank allows us to remove the state dependence. Finally, by examining Eq. (11), we conclude that saturating the quantum value on a state having full-rank ρ_B requires Bob's measurements to be projective. This leads to the desired Eqs. (12). ■

Equations (12) are derived purely from the fact that we observe the quantum value of \mathcal{F}_d . To prove that the functional \mathcal{F}_d is nontrivial, it suffices to argue that these relations cannot be satisfied by any deterministic strategy.

Theorem 2. For any $d \geq 2$ and overlap matrix satisfying $O_{x_1x_2} < 1$, the quantum value is strictly bigger than the local value.

Proof. A deterministic strategy can always be written as a quantum strategy where the local Hilbert spaces are one dimensional. If Bob outputs $b = u$ for $y = 1$ and $b = v$ for $y = 2$, the associated projectors correspond to $P_{x_1} = \delta_{x_1u}$ and $Q_{x_2} = \delta_{x_2v}$. If this strategy saturates the quantum value, Eqs. (12) would imply that $O_{x_1x_2}^2 \delta_{x_1u} = \delta_{x_1u} \delta_{x_2v} \delta_{x_1u}$, for all x_1, x_2 . However, choosing $x_1 = u$ and $x_2 = v$ leads to $O_{uv} = 1$, which is a contradiction. ■

At this point, we can see why the assumption $O_{x_1x_2} < 1$ was necessary. It is straightforward to check that \mathcal{F}_d can be defined for any overlap matrix, the quantum value is always equal to $d-1$, and the conclusion given in Eqs. (12) holds. However, without the assumption that $O_{x_1x_2} < 1$, the measurement certification statement is no longer sufficient

³Recall that taking the transposition does not affect the spectrum.

to deduce that the functional is nontrivial. In fact, in Appendix C2, we show that whenever $O_{x_1x_2} = 1$ for some x , then there exists a deterministic strategy achieving β_Q , i.e., $\beta_L = \beta_Q$.

As explained at the beginning, every pair of incompatible rank-one projective measurements can be preprocessed to give measurements the overlap matrix of which satisfies $O_{x_1x_2} < 1$ and, hence, gives rise to a nontrivial functional. Therefore, we obtain the following corollary.

Corollary 1. Every pair of incompatible rank-one projective measurements on a finite-dimensional Hilbert space is capable of producing nonlocal correlations and is optimal for some nontrivial Bell inequality.

IV. ROBUSTNESS AGAINST NOISE

To investigate how robust to noise the realizations presented above are, consider a specific noise model in which the measurements are kept unchanged while the state is replaced by the isotropic state

$$\rho_v = v\Phi_d^+ + (1-v)\frac{\mathbb{1} \otimes \mathbb{1}}{d^2}, \quad (15)$$

where $v \in [0, 1]$ is the visibility. A simple calculation shows that performing the optimal measurements on the isotropic state gives the value of $(d-1)(2v-1)$. We can use this result to compare the noise robustness of different functionals \mathcal{F}_d . More specifically, we are interested in discovering which functional enables a violation for the smallest v . To address this question, we first need to compute the local value.

For a fixed overlap matrix O , calculating the local value reduces to analyzing the deterministic strategies for \mathcal{F}_d . If we choose the strategy of Bob in which he outputs $b = u$ if $y = 1$ and $b = v$ if $y = 2$, the value becomes

$$\mathcal{F}_d = \sum_x \lambda_x (\delta_{x_1u} - \delta_{x_2v}) A_x - \frac{1}{2} \sum_x \lambda_x^2 A_x^2, \quad (16)$$

where $A_x \in \{\pm 1, 0\}$ describes the deterministic strategy of Alice. Let us now show that the optimal strategy of Alice can be explicitly determined. The sum can be split into three distinct cases: $R_\pm = \{x \in [d]^2 | (\delta_{x_1u} - \delta_{x_2v}) = \pm 1\}$ and $R_0 = [d]^2 \setminus (R_+ \cup R_-)$. For the R_0 terms, the optimal strategy of Alice is clearly not to play the game (this would lead to a negative contribution), so the optimal choice is $A_x = 0$.

For the R_\pm terms the optimal choice is $A_x = \delta_{x_1u} - \delta_{x_2v}$, which is better than $A_x = -(\delta_{x_1u} - \delta_{x_2v})$ (because $\lambda_x > 0$) and better than $A_x = 0$ (because $\lambda_x - \frac{1}{2}\lambda_x^2 > 0$). Plugging in the optimal strategy of Alice into Eq. (16) leads to the following auxiliary function:

$$s(u, v) := \sum_{x \in R_\pm} \left(\lambda_x - \frac{1}{2}\lambda_x^2 \right), \quad (17)$$

where the dependence on u and v is hidden inside the definitions of R_\pm . Then, β_L can be written as

$$\beta_L(O) = \max_{u, v} [s(u, v)]. \quad (18)$$

The choice of coefficients of our functionals, made at the beginning, ensures that the quantum value does not depend on the overlap matrix. It is a convenient choice because all the

dependence of O is contained in the local value. It is easy to see that in the noise model presented above, the largest robustness corresponds to the lowest local value. Searching for highly robust functionals leads to the following theorem.

Theorem 3. For $d \geq 2$ and any overlap matrix, $\beta_L(O) \geq d + \sqrt{2} - 5/2$.

Sketch of the proof (See Appendix C2 for details). Primarily, note that a lower bound for any strategy $s(u, v)$ is also a lower bound for $\beta_L(O)$. Because we can always relabel the measurements outputs, we assume, for simplicity, that O_{11} is the largest element of O , i.e., $O_{11} \geq O_{uv}$, $\forall u, v \in \{1, \dots, d\}$. Then, we lower bound $\beta_L(O)$ using the first strategy:

$$\beta_L(O) = \max_{u, v} [s(u, v)] \geq s(1, 1). \quad (19)$$

Next, as $s(1, 1)$ depends only on overlaps squared, we define a new variable $t_{x_1x_2} := O_{x_1x_2}^2$. This makes some terms in $s(1, 1)$ strictly concave in $t_{x_1x_2}$, which allows us to obtain a lower bound that depends only on t_{11} . Minimizing that function concludes the proof. ■

Although Theorem 3 provides a lower bound on $\beta_L(O)$, it does not say whether this bound is achievable. Fortunately, by demanding the saturation of the strictly concave terms of $s(1, 1)$, we extract this information.

Lemma 2. For even $d \geq 2$, the lower bound of Theorem 3 is achievable and every O that saturates it corresponds to a direct sum of qubit MUBs; i.e., up to permutations, it can be written as

$$O = \bigoplus_{i=1}^{d/2} \frac{1}{\sqrt{2}} J_2, \quad (20)$$

where J_2 is the 2×2 matrix of ones. For odd $d \geq 3$, the lower bound of Theorem 3 cannot be achieved by any overlap matrix.

The complete proof of Lemma 2 can be found in Appendix C3. Let us briefly comment on the difference between even and odd d . Our proof of Theorem 3 is based solely on the fact that the rows and columns of O are normalized. However, from Eq. (1), we know that O should be obtained by taking the entrywise absolute value of a unitary matrix, which, in general, is more restrictive than just imposing the normalization condition. For even d , there exists a valid overlap matrix which saturates the lower bound obtained in Theorem 3, but for odd d this is not the case. For a more detailed explanation, see Appendix C4.

V. DISCUSSION

In this paper, we have demonstrated how to tailor a Bell functional to a specific pair of incompatible rank-one projective measurements. More specifically, we have shown that for every such pair there exists a nontrivial Bell functional for which this pair is optimal. We have also proved that for these functionals a certain degree of certification is possible. If the quantum value of \mathcal{F}_d is achieved, the measurements of Bob must satisfy some simple polynomial relations and, under some additional mild conditions, a maximally entangled state can be extracted.

We have also investigated how robust to noise these functionals are, as quantified by the difference $\beta_Q - \beta_L$. For even d , the largest robustness to noise arises when the measurements are chosen as direct sums of qubit MUBs. For odd d , we have a conjecture for $d = 3$, but no analytical proof of optimality. It is worth pointing out, however, that for $d \geq 3$ the optimal noise robustness is not achieved by d -dimensional MUBs.

There are several open questions arising from this paper. An aspect that should be clarified is the precise conditions under which state certification is possible. We have shown that if the overlap matrix contains a non-null row or column, then a maximally entangled state can be extracted. On the other extreme, there are overlap matrices that decompose as a direct sum for which, as proved in Appendix B 3, state certification is not possible. Hence, the question about state certification is open for intermediate cases. Another question concerns the optimal noise robustness for odd dimensions. Our candidate for $d = 3$ exhibits a mathematically elegant structure and it would be interesting to see if the same happens for larger dimensions.

Lastly, it is known that all pure entangled states can generate nonlocality, and we have demonstrated that incompatible pairs of rank-one projective measurements also do so. By assembling these results, is it possible to achieve necessary and sufficient conditions for, at least, some limited class of states and measurements?

ACKNOWLEDGMENTS

We acknowledge fruitful discussions with Máté Farkas and Nicolas Gigena. The project ‘‘Robust certification of quantum devices’’ is carried out within the HOMING program of the Foundation for Polish Science cofinanced by the European Union under the European Regional Development Fund.

APPENDIX A: THE QUANTUM VALUE OF \mathcal{F}_d

In this Appendix, we provide a complete proof of Lemma 1. In the main text, we specified a quantum realization for the functional \mathcal{F}_d which achieves the value $d - 1$. Here, in the first part of this Appendix, we provide a matching upper bound, which shows that the quantum value of \mathcal{F}_d equals $d - 1$. In the second part, we present a natural extension of our functional to a set of $N > 2$ rank-one projective measurements.

1. Proof of Lemma 1

At this point, we are only interested in deriving an upper bound; hence, we could, without loss of generality, restrict ourselves to realizations where the state is pure and measurements are projective. However, we will later reuse this argument to derive some certification statements, so we do not want to introduce unnecessary assumptions. As we are not interested in characterizing the measurements of Alice, we will assume that they are projective, but we make no assumptions about the measurements of Bob and the shared state. Thus, let us represent the state by ρ_{AB} , an arbitrary density matrix, and reuse the symbols A_x , P_{x_1} , and Q_{x_2} to denote the observables of Alice and the two measurements of Bob, respectively.

We start again by using the Born rule to rewrite \mathcal{C}_d as

$$\mathcal{C}_d = \sum_x \lambda_x \operatorname{tr} \{ [A_x \otimes (P_{x_1} - Q_{x_2})] \rho_{AB} \}. \quad (\text{A1})$$

Taking the absolute value of every term in the summation leads to

$$\mathcal{C}_d \leq \sum_x |\lambda_x \operatorname{tr} \{ [A_x \otimes (P_{x_1} - Q_{x_2})] \rho_{AB} \}|. \quad (\text{A2})$$

In the next step, we apply the Cauchy-Schwarz inequality. For the Hilbert-Schmidt inner product, the Cauchy-Schwarz inequality reads $|\operatorname{tr}(X^\dagger Y)| \leq \sqrt{\operatorname{tr}(X^\dagger X)} \sqrt{\operatorname{tr}(Y^\dagger Y)}$, where X and Y are arbitrary operators acting in a given Hilbert space. Therefore, taking $X = (A_x \otimes \mathbb{1}) \rho_{AB}^{1/2}$ and $Y = [\mathbb{1} \otimes (P_{x_1} - Q_{x_2})] \rho_{AB}^{1/2}$, where $\rho_{AB}^{1/2}$ is the positive semidefinite square root of ρ_{AB} , we can apply the Cauchy-Schwarz inequality to all of the terms of \mathcal{C}_d , leading to

$$\mathcal{C}_d \leq \sum_x \lambda_x \sqrt{\operatorname{tr}[(A_x^2 \otimes \mathbb{1}) \rho_{AB}]} \sqrt{\operatorname{tr}[(\mathbb{1} \otimes (P_{x_1} - Q_{x_2})^2) \rho_{AB}]}. \quad (\text{A3})$$

The second application of Cauchy-Schwarz, now for the standard inner product of real vectors, leads to

$$\mathcal{C}_d \leq \sqrt{\sum_x \lambda_x^2 \operatorname{tr}[(A_x^2 \otimes \mathbb{1}) \rho_{AB}]} \sqrt{\sum_x \operatorname{tr}[(\mathbb{1} \otimes (P_{x_1} - Q_{x_2})^2) \rho_{AB}]}. \quad (\text{A4})$$

Let us now compute a universal upper bound on the second factor. For any measurement operators P_{x_1} and Q_{x_2} , it is true that $P_{x_1}^2 \leq P_{x_1}$ and $Q_{x_2}^2 \leq Q_{x_2}$ (recall that for Hermitian operators X and Y the inequality $X \geq Y$ is equivalent to $X - Y \geq 0$). Thus,

$$\begin{aligned} \sum_{x_1, x_2} (P_{x_1} - Q_{x_2})^2 &= \sum_{x_1, x_2} (P_{x_1}^2 + Q_{x_2}^2 - \{P_{x_1}, Q_{x_2}\}) \\ &\leq \sum_{x_1, x_2} (P_{x_1} + Q_{x_2} - \{P_{x_1}, Q_{x_2}\}) = 2(d-1)\mathbb{1}. \end{aligned} \quad (\text{A5})$$

Therefore, the second factor is upper bounded by $\sqrt{2(d-1)}$. Using the assumption that the measurement operators of Alice are projective, we can rewrite the observable as $A_x^2 = A_x^{(1)} + A_x^{(2)}$. Therefore,

$$\mathcal{C}_d \leq \sqrt{2(d-1)} \sum_x \lambda_x^2 \operatorname{tr} [((A_x^{(1)} + A_x^{(2)}) \otimes \mathbb{1}) \rho_{AB}]. \quad (\text{A6})$$

Finally, if we define

$$\gamma := \sum_x \lambda_x^2 \operatorname{tr} [((A_x^{(1)} + A_x^{(2)}) \otimes \mathbb{1}) \rho_{AB}], \quad (\text{A7})$$

it is easy to see that \mathcal{F}_d is upper bounded by

$$\mathcal{F}_d \leq \sqrt{2(d-1)} \gamma - \frac{1}{2} \gamma. \quad (\text{A8})$$

The last step is to maximize the right-hand side of Eq. (A8) over $\gamma = [0, d(d-1)]$ [it is easy to verify that the maximal value of γ equals $\sum_x \lambda_x^2 = d(d-1)$]. Using elementary calculus we conclude that the right-hand side is maximized only

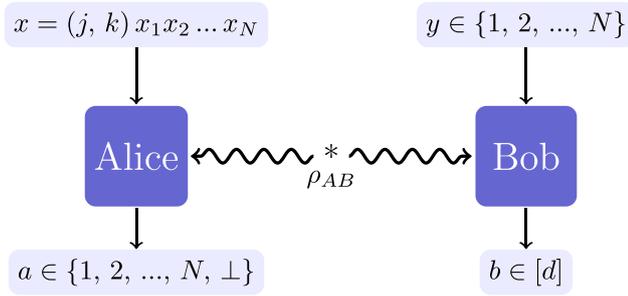


FIG. 2. A Bell scenario extended to a set of $N > 2$ projective measurements. Now, Alice is given a string with N digits x_1, x_2, \dots, x_N ranging from 1 to d and an ordered pair (j, k) , such that $j \in [N] - \{N\}$ and $k \in [N] - \{1\}$. Similarly, Alice's output now ranges from 1 to N and includes \perp . Bob takes an input $y \in [N]$ and produces output $b \in [d]$.

when $\gamma = 2(d - 1)$, which leads to the final bound:

$$\mathcal{F}_d \leq d - 1. \quad (\text{A9})$$

At this point, the reader can better appreciate why we needed to introduce a penalty in Eq. (3). In Eq. (A8), γ is maximized over the interval $[0, d(d - 1)]$. Then, the maximal γ can be found at $2(d - 1)$, giving us the quantum value above. Now, note that if our functional had no penalty, it would be written as $\mathcal{F}_d = \mathcal{C}_d$ and the maximization of \mathcal{C}_d in terms of γ would produce

$$\mathcal{C}_d \leq \sqrt{2(d - 1)\gamma}. \quad (\text{A10})$$

In this case, γ is maximal at $d(d - 1)$, in the edge of the interval, so \mathcal{C}_d is upper bounded by $\sqrt{2d}(d - 1)$. On the other hand, by comparing Eqs. (8) and (11), we obtain that

$$\mathcal{F}_d^{(j,k)} := \sum_{x_j, x_k=1}^d \lambda_x^{(j,k)} \left(\sum_{y \in \{j,k\}} [p(y, x_y|x, y) - p(\bar{y}, x_y|x, y)] - \frac{1}{2} \lambda_x^{(j,k)} [p(a = j|x) + p(a = k|x)] \right), \quad (\text{A12})$$

where \bar{y} flips the value of y from j to k and vice versa. Then, \mathcal{F}_d can be generalized by considering

$$\mathcal{F}_d^N = \sum_{j < k} \mathcal{F}_d^{(j,k)}. \quad (\text{A13})$$

If written so, Eq. (A13) is clearly upper bounded by

$$\mathcal{F}_d^N \leq \frac{1}{2} N(N - 1)(d - 1), \quad (\text{A14})$$

as we can simply use Eq. (A9) to sum the $\binom{N}{2}$ terms of Eq. (A13).

This upper bound can be saturated if the N orthonormal bases $\{|e_x^{(i)}\rangle\}_{x_i=1}^d$ in dimension d are such that $\lambda_x^{(j,k)} > 0$, for all j, k, x and $i = 1, 2, \dots, N$. In this case, we can mirror the optimal realization proposed in the main text by taking the same state as in Eq. (6), $P_{x_i} = |e_{x_i}^{(i)}\rangle\langle e_{x_i}^{(i)}|$ when Bob is given $y = i$ and

$$A_x^{(j,k)} = \frac{1}{\lambda_x^{(j,k)}} (P_{x_j} - Q_{x_k})^\top. \quad (\text{A15})$$

the realization produced by the maximally entangled state is simply $\mathcal{C}_d = 2(d - 1)$. In a nutshell, introducing the penalty over the outputs $a = 1$ and 2 makes the realization by the maximally entangled state the same as the upper bound for \mathcal{F}_d so that this state turns out to be optimal.

In addition, note that, without penalty, Alice never outputs $a = \perp$. Then, the inclusion of a third output is justified. It allows Alice to play the game and be penalized or not to play the game and keep the value of the realization unchanged. Adding a third output also guarantees that the spectrum of A_x is proportional to the spectrum of $P_{x_1} - Q_{x_2}$, since $\text{spec}(P_{x_1} - Q_{x_2}) \in \{\pm\lambda_x, 0\}$.

2. An extension to $N > 2$ rank-one projective measurements

The original functionals tailored to MUBs in Ref. [15] have recently been extended to $N > 2$ bases [18] and it turns out that an analogous extension can be constructed for non-MUB measurements. In this subsection, we present this generalization.

Suppose a bipartite Bell scenario (see Fig. 2) in which Alice is given a string $x = (j, k) x_1 x_2 \dots x_N$, in which $x_i \in [d]$, for $i = 1, \dots, N$, and (j, k) is an ordered pair such that $j \in [N] - \{N\}$, $k \in [N] - \{1\}$, and $j < k$. Similarly to the previous case, Alice now outputs $a \in \{1, 2, \dots, N, \perp\}$. On the other hand, the outputs of Bob are kept unchanged, while the possible inputs are $y \in [N]$.

Now, consider N d -dimensional orthonormal bases $\{|e_{x_1}^{(1)}\rangle\}_{x_1=1}^d, \{|e_{x_2}^{(2)}\rangle\}_{x_2=1}^d, \dots, \{|e_{x_N}^{(N)}\rangle\}_{x_N=1}^d$, where $d \geq 2$, and define

$$\lambda_x^{(j,k)} := \sqrt{1 - \left| \langle e_{x_j}^{(j)} | e_{x_k}^{(k)} \rangle \right|^2}. \quad (\text{A11})$$

For a fixed pair (j, k) , consider the following functional:

Moreover, according to Theorem 2 the functional in Eq. (A13) is nontrivial, if, for the N orthonormal bases, the condition $\lambda_x^{(j,k)} > 0, \forall j, k, x$ is satisfied.

Lastly, an interesting problem can be posed if, instead of providing N orthonormal bases, we specify $\lambda_x^{(j,k)}$ and try to recover the bases. In this case, it is not clear if this problem has a solution for any N or arrangements of $\lambda_x^{(j,k)}$. For instance, if we consider

$$\left| \langle e_{x_j}^{(j)} | e_{x_k}^{(k)} \rangle \right|^2 = \frac{1}{d} \quad \forall j, k \in [N] \text{ and } x_j, x_k \in [d], \quad (\text{A16})$$

we fix the values of $\lambda_x^{(j,k)}$ to be uniform. The resulting functionals are the same as the ones derived in Ref. [18] up to a factor of $\sqrt{1 - 1/d}$. The numerical results of Ref. [18] confirm what is already known in the literature: there is no more than $d + 1$ MUBs in dimension d and Zauner's conjecture [19] holds.

APPENDIX B: CERTIFICATION

In what follows, we present the proof of Theorem 1. First, we demonstrate the device-independent certification of the measurements of Bob, based solely on the condition that β_Q is saturated. Recall that, as in Appendix A 1, for the certification of the measurements of Bob, we can assume that the measurements of Alice are projective since we are not interested in their characterization. Next, we construct an isometry capable of extracting a maximally entangled state of dimension d . Lastly, we discuss some exceptional cases in which state certification cannot be achieved.

1. Bob's measurements (proof of the first part of Theorem 1)

If Eq. (A9) is saturated, all of the inequalities used in Appendix A 1 should be tight. This enables us to get some information about the state and measurements capable of achieving the quantum value.

The first step we must consider is going from Eq. (A1) to Eq. (A2), where the terms of \mathcal{C}_d are bounded by the absolute value:

$$\begin{aligned} \mathcal{C}_d &= \sum_x \lambda_x \text{tr}[(A_x \otimes (P_{x_1} - Q_{x_2}))\rho_{AB}] \\ &\leq \sum_x |\lambda_x \text{tr}[(A_x \otimes (P_{x_1} - Q_{x_2}))\rho_{AB}]|. \end{aligned} \quad (\text{B1})$$

If the above inequality is tight, the argument of the absolute value must be non-negative. Since, by construction, $\lambda_x > 0$, we conclude that $\text{tr}[A_x \otimes (P_{x_1} - Q_{x_2})\rho_{AB}] \geq 0$.

The saturation condition of the Cauchy-Schwarz inequality in Eq. (A3), in turn, implies that

$$\lambda_x A_x \otimes \mathbb{1} \rho_{AB}^{1/2} = \mu_x \mathbb{1} \otimes (P_{x_1} - Q_{x_2}) \rho_{AB}^{1/2}, \quad \forall x = x_1 x_2, \quad (\text{B2})$$

where $\mu_x \in \mathbb{C}$ is the proportionality constant. Right multiplying it by $\rho_{AB}^{1/2}$ and left multiplying by $\mathbb{1} \otimes (P_{x_1} - Q_{x_2})$ leads to

$$\lambda_x A_x \otimes (P_{x_1} - Q_{x_2}) \rho_{AB} = \mu_x \mathbb{1} \otimes (P_{x_1} - Q_{x_2})^2 \rho_{AB}. \quad (\text{B3})$$

If we trace both sides of this equation and use the fact that $\text{tr}[A_x \otimes (P_{x_1} - Q_{x_2})\rho_{AB}] \geq 0$ (deduced above), we conclude that μ_x is real and non-negative.

The saturation condition of Eq. (A4) implies that

$$\lambda_x^2 A_x^2 \otimes \mathbb{1} \rho_{AB} = \eta \mathbb{1} \otimes (P_{x_1} - Q_{x_2})^2 \rho_{AB} \quad \forall x = x_1 x_2, \quad (\text{B4})$$

where η is independent of x since the Cauchy-Schwarz in Eq. (A4) is applied to the entire summation. Now, let us multiply Eq. (B2) by its Hermitian conjugate and take the trace:

$$\lambda_x^2 \text{tr}[A_x^2 \otimes \mathbb{1} \rho_{AB}] = \mu_x^2 \text{tr}[\mathbb{1} \otimes (P_{x_1} - Q_{x_2})^2 \rho_{AB}]. \quad (\text{B5})$$

Comparing Eqs. (B5) and (B4) we conclude that $\mu_x = \sqrt{\eta}$, for all x . So, let us drop the index of μ_x and use just μ from now on.

The last but not less important saturation condition is implicit in Eq. (A5) where we upper bound the following

quantity:

$$\begin{aligned} &\sum_x \text{tr}[(\mathbb{1} \otimes (P_{x_1}^2 + Q_{x_2}^2 - \{P_{x_1}, Q_{x_2}\}))\rho_{AB}] \\ &\leq \sum_x \text{tr}[(\mathbb{1} \otimes (P_{x_1} + Q_{x_2} - \{P_{x_1}, Q_{x_2}\}))\rho_{AB}]. \end{aligned} \quad (\text{B6})$$

After a short algebraic manipulation, we get

$$\begin{aligned} &\text{tr}\left[\left(\mathbb{1} \otimes \sum_{x_1} (P_{x_1} - P_{x_1}^2)\right)\rho_{AB}\right] \\ &+ \text{tr}\left[\left(\mathbb{1} \otimes \sum_{x_2} (Q_{x_2} - Q_{x_2}^2)\right)\rho_{AB}\right] \geq 0. \end{aligned} \quad (\text{B7})$$

If we trace over Alice's subsystem, the term-by-term saturation of the above inequality implies

$$\begin{aligned} \text{tr}[(P_{x_1} - P_{x_1}^2)\rho_B] &= 0, \quad \forall x_1 \\ \text{tr}[(Q_{x_2} - Q_{x_2}^2)\rho_B] &= 0, \quad \forall x_2. \end{aligned} \quad (\text{B8})$$

Since both $P_{x_1} - P_{x_1}^2$ (or $Q_{x_2} - Q_{x_2}^2$) and ρ_B are positive semidefinite operators and their product has a null trace, they must be orthogonal:

$$\begin{aligned} (P_{x_1} - P_{x_1}^2)\rho_B &= 0, \\ (Q_{x_2} - Q_{x_2}^2)\rho_B &= 0. \end{aligned} \quad (\text{B9})$$

Assuming that Bob's marginal state is full rank, it is possible to eliminate ρ_B out of Eqs. (B9) by right multiplying it by ρ_B^{-1} . This way, we conclude that the saturation of Eq. (A5) implies that both measurements of Bob are projective. We can use this fact to sum Eq. (B5) over x to obtain

$$\sum_x \lambda_x^2 \text{tr}[A_x^2 \otimes \mathbb{1} \rho_{AB}] = \mu^2 \sum_x \text{tr}[\mathbb{1} \otimes (P_{x_1} - Q_{x_2})^2]. \quad (\text{B10})$$

Note that, if we assume the measurements of Alice to be projective, the right-hand side of the above equation can be identified as γ , in Eq. (A7). On the other hand, the left-hand side can be summed to $2\mu^2(d-1)$. The only value of γ which allows for the quantum value in Eq. (A8) is $2(d-1)$, so we conclude that $\mu = 1$.

Having obtained the value of the proportionality constants, let us right multiply Eq. (B2) by $\rho_{AB}^{1/2}$ and use $\mu = 1$ to rewrite the following relation:

$$\lambda_x A_x \otimes \mathbb{1} \rho_{AB} = \mathbb{1} \otimes (P_{x_1} - Q_{x_2}) \rho_{AB}. \quad (\text{B11})$$

Again, assuming that the measurements of Alice are projective, we can use Eq. (4) to write $A_x^3 = A_x$. Combining it with Eq. (B11), we get

$$\frac{1}{\lambda_x} \mathbb{1} \otimes (P_{x_1} - Q_{x_2}) \rho_{AB} = \frac{1}{\lambda_x^3} \mathbb{1} \otimes (P_{x_1} - Q_{x_2})^3 \rho_{AB}. \quad (\text{B12})$$

Since we have eliminated Alice's observable out of the equation, we can derive a relation involving only Bob's measurement operators by tracing out Alice's subsystem and right multiplying by the inverse of Bob's marginal state:

$$\lambda_x^2 (P_{x_1} - Q_{x_2}) = (P_{x_1} - Q_{x_2})^3. \quad (\text{B13})$$

After some algebraic manipulation,

$$O_{x_1 x_2}^2 (P_{x_1} - Q_{x_2}) = P_{x_1} Q_{x_2} P_{x_1} - Q_{x_2} P_{x_1} Q_{x_2}. \quad (\text{B14})$$

Now, if we sum over x_1 ,

$$\sum_{x_1} O_{x_1 x_2}^2 P_{x_1} - \sum_{x_1} O_{x_1 x_2}^2 Q_{x_2} = \sum_{x_1} (P_{x_1} Q_{x_2} P_{x_1} - Q_{x_2} P_{x_1} Q_{x_2}) \quad (\text{B15})$$

$$\sum_{x_1} O_{x_1 x_2}^2 P_{x_1} - Q_{x_2} = \sum_{x_1} (P_{x_1} Q_{x_2} P_{x_1}) - Q_{x_2}^2. \quad (\text{B16})$$

Because $Q_{x_2} = Q_{x_2}^2$,

$$\sum_{x_1} O_{x_1 x_2}^2 P_{x_1} = \sum_{x_1} (P_{x_1} Q_{x_2} P_{x_1}). \quad (\text{B17})$$

Finally, since the P_{x_1} projectors are orthogonal, this equality must hold term by term:

$$O_{x_1 x_2}^2 P_{x_1} = P_{x_1} Q_{x_2} P_{x_1}. \quad (\text{B18})$$

The summation over x_2 gives us the complementary equation:

$$O_{x_1 x_2}^2 Q_{x_2} = Q_{x_2} P_{x_1} Q_{x_2}. \quad (\text{B19})$$

Lastly, let us show that these two equations are complete, i.e., if for some finite-dimensional measurements Eqs. (B18) and (B19) hold, then it is possible to construct a realization that achieves the quantum value of \mathcal{F}_d . This means that the conditions derived are, at least for finite-dimensional measurements, necessary and sufficient, and no tighter characterization can be obtained.

Consider an incompatible pair of measurements $\{P_{x_1}\}_{x_1=1}^d$ and $\{Q_{x_2}\}_{x_2=1}^d$ acting on a D -dimensional Hilbert space, with $D < \infty$, satisfying Eqs. (B18) and (B19), for some set of overlaps $0 < O_{x_1 x_2} < 1$. Then, both P_{x_1} and Q_{x_2} are projectors and have equal traces, for all $x_1, x_2 \in \{1, \dots, d\}$. The projectivity can be demonstrated by simply summing Eqs. (B18) and (B19) over x_1 and x_2 to get $P_{x_1} = P_{x_1}^2$ and $Q_{x_2} = Q_{x_2}^2$, respectively. Showing that $\text{tr}(P_{x_1}) = \text{tr}(Q_{x_2})$ relies on demonstrating that the projectors are isomorphic. Defining $W_{x_1 x_2} = \frac{1}{O_{x_1 x_2}} P_{x_1} Q_{x_2}$ we can see that, for $O_{x_1 x_2} > 0$,

$$W_{x_1 x_2} W_{x_1 x_2}^\dagger = P_{x_1} \quad \text{and} \quad W_{x_1 x_2}^\dagger W_{x_1 x_2} = Q_{x_2}. \quad (\text{B20})$$

For a finite-dimensional Hilbert space this implies that $\text{tr}(P_{x_1}) = \text{tr}(Q_{x_2}) = n$, for some fixed $n \in \mathbb{N}$. Because the dimension of the Hilbert space is finite, we rewrite D as simply

$$D = \text{tr} \mathbb{1} = \sum_{x_1} \text{tr} P_{x_1} = dn. \quad (\text{B21})$$

Then, we construct a Hermitian operator $B_{x_1 x_2} = P_{x_1} - Q_{x_2}$. By computing an explicit expression for $B_{x_1 x_2}^3$, one can show that it satisfies a relation analogous to Eq. (B13):

$$B_{x_1 x_2}^3 = \lambda_x^2 B_{x_1 x_2} \quad \forall x_1, x_2 \in \{1, \dots, d\}. \quad (\text{B22})$$

Clearly, the spectrum of $B_{x_1 x_2}$ belongs to $\{0, \pm \lambda_x\}$. Because the trace of $B_{x_1 x_2}$ is null, the multiplicity of its non-null eigenvalues must be the same. So, let us calculate

$$\begin{aligned} \text{tr}(B_{x_1 x_2}^2) &= \text{tr}(P_{x_1} + Q_{x_2} - \{P_{x_1}, Q_{x_2}\}) \\ &= 2n - \text{tr}(\{P_{x_1}, Q_{x_2}\}) \\ &= 2n - 2\text{tr}(P_{x_1} Q_{x_2} P_{x_1}) = 2n\lambda_x^2, \end{aligned} \quad (\text{B23})$$

where we have used that $\text{tr}(P_{x_1} Q_{x_2}) = \text{tr}(P_{x_1} Q_{x_2} P_{x_1})$. Thus, both $+\lambda_x$ and $-\lambda_x$ must have multiplicity n .

Now, let us define a realization of \mathcal{F}_d in which the measurements of Bob are represented by P_{x_1} and Q_{x_2} , the observables of Alice are given by $A_x = 1/\lambda_x B_{x_1 x_2}^\dagger$, and the state is the D -dimensional maximally entangled state $|\Phi_D^+\rangle$. Like in Eq. (5), the spectrum of A_x also belongs to $\{0, \pm 1\}$. Then, evaluating the correlation score gives

$$\begin{aligned} C_d &= \sum_x \lambda_x \langle \Phi_D^+ | A_x \otimes B_{x_1 x_2} | \Phi_D^+ \rangle = \sum_x \langle \Phi_D^+ | \mathbb{1} \otimes B_{x_1 x_2}^2 | \Phi_D^+ \rangle \\ &= \frac{1}{D} \sum_x \text{tr}(B_{x_1 x_2}^2) = \frac{1}{dn} \sum_x 2n(1 - O_{x_1 x_2}^2) = 2(d-1), \end{aligned} \quad (\text{B24})$$

where we used the fact that the local state of Bob is $\mathbb{1}/D$. For the final score, we obtain

$$\begin{aligned} \mathcal{F}_d &= C_d - \frac{1}{2} \sum_x \lambda_x^2 \langle \Phi_D^+ | A_x^2 \otimes \mathbb{1} | \Phi_D^+ \rangle \\ &= C_d - \frac{1}{2} \sum_x \langle \Phi_D^+ | \mathbb{1} \otimes B_{x_1 x_2}^2 | \Phi_D^+ \rangle \\ &= C_d - \frac{1}{2D} \sum_x \text{tr}(B_{x_1 x_2}^2) = d-1, \end{aligned} \quad (\text{B25})$$

as desired.

2. The state (proof of the second part of Theorem 1)

Here, we present the argument for the certification of the state. We show that, if the overlap matrix O has at least one column or row the entries of which are nonzero, then for any realization that saturates the quantum value of \mathcal{F}_d we can construct local isometries $V_A : \mathcal{H}_A \rightarrow \mathbb{C}^d \otimes \mathcal{H}_A$ and $V_B : \mathcal{H}_B \rightarrow \mathbb{C}^d \otimes \mathcal{H}_B$ such that

$$(V_A \otimes V_B) \rho_{AB} (V_A^\dagger \otimes V_B^\dagger) = \Phi_d^+ \otimes \rho_{\text{aux}}, \quad (\text{B26})$$

where $\rho_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is the shared state, Φ_d^+ is the d -dimensional maximally entangled state, and ρ_{aux} corresponds to the uncharacterized part of ρ_{AB} . In particular, note that the isometries depend on the measurement operators of Bob, as explained below.

Let us start by defining the isometry on Bob's side. To do this, we need to assume that the j th column of O is nonzero, i.e., $O_{i,j} \neq 0$, for all i . Then, we introduce the isometry $V_B := SR$, where $R : \mathcal{H}_B \rightarrow \mathbb{C}^d \otimes \mathcal{H}_B$ and $S : \mathbb{C}^d \otimes \mathcal{H}_B \rightarrow \mathbb{C}^d \otimes \mathcal{H}_B$ are defined by

$$R := \sum_i |i\rangle \otimes P_i, \quad (\text{B27})$$

$$S := \sum_k |k\rangle \langle k| \otimes U_k. \quad (\text{B28})$$

U_k is a unitary operator defined as

$$U_k := \sum_i \frac{1}{O_{i,j} O_{i+k,j}} P_i Q_j P_{i+k}. \quad (\text{B29})$$

Note that the subscripts in both P_{i+k} and $O_{i+k,j}$ are taken as sum modulo d , where the result is in the set $\{1, \dots, d\}$. Then,

the action of U_k produces a cyclic shift on $\{P_i\}_{i=1}^d$:

$$U_k P_i U_k^\dagger = P_{i-k}. \quad (\text{B30})$$

Finally, V_B is rewritten as

$$V_B = \sum_i |i\rangle \otimes U_i P_i = \frac{1}{O_{d,j}} \sum_i \frac{1}{O_{i,j}} |i\rangle \otimes P_d Q_j P_i. \quad (\text{B31})$$

For Alice's side, let us first define a pair of operators acting in \mathcal{H}_A analogously to the action of P_{x_1} and Q_{x_2} on Bob's side:

$$\begin{aligned} \tilde{P}_{x_1} &:= \frac{1}{d} \left(\mathbb{1} + \sum_{x_2} \lambda_{x_2} A_{x_2} \right), \\ \tilde{Q}_{x_2} &:= \frac{1}{d} \left(\mathbb{1} - \sum_{x_1} \lambda_{x_1} A_{x_1} \right). \end{aligned} \quad \forall x = x_1, x_2 \quad (\text{B32})$$

If written as in the above equation, \tilde{P}_{x_1} and \tilde{Q}_{x_2} satisfy

$$\begin{aligned} (\tilde{P}_{x_1} \otimes \mathbb{1}) \rho_{AB} &= (\mathbb{1} \otimes P_{x_1}) \rho_{AB}, \\ (\tilde{Q}_{x_2} \otimes \mathbb{1}) \rho_{AB} &= (\mathbb{1} \otimes Q_{x_2}) \rho_{AB}, \end{aligned} \quad (\text{B33})$$

for all ρ_{AB} , $\{P_{x_1}\}_{x_1=1}^d$ and $\{Q_{x_2}\}_{x_2=1}^d$ that saturate the quantum value. The new operators \tilde{P}_{x_1} and \tilde{Q}_{x_2} satisfy similar algebraic relations as their counterparts on Bob's side. By taking the partial trace over Bob's marginal state, we can easily verify from Eqs. (B33) that

$$\begin{aligned} \tilde{P}_{x_1} \rho_A &= \tilde{P}_{x_1}^2 \rho_A, \\ \tilde{Q}_{x_2} \rho_A &= \tilde{Q}_{x_2}^2 \rho_A. \end{aligned} \quad (\text{B34})$$

Also, an analogous version of Eqs. (B18) and (B19) must be satisfied by \tilde{P}_{x_1} and \tilde{Q}_{x_2} :

$$\begin{aligned} (\tilde{P}_{x_1} \otimes \mathbb{1}) \rho_{AB} &= (\mathbb{1} \otimes P_{x_1}) \rho_{AB} = \frac{1}{O_{x_1, x_2}^2} (\mathbb{1} \otimes P_{x_1} Q_{x_2} P_{x_1}) \rho_{AB} \\ &= \frac{1}{O_{x_1, x_2}^2} (\tilde{P}_{x_1} \tilde{Q}_{x_2} \tilde{P}_{x_1} \otimes \mathbb{1}) \rho_{AB}, \end{aligned} \quad (\text{B35})$$

which leads to

$$\begin{aligned} O_{x_1, x_2}^2 \tilde{P}_{x_1} &= \tilde{P}_{x_1} \tilde{Q}_{x_2} \tilde{P}_{x_1}, \\ O_{x_1, x_2}^2 \tilde{Q}_{x_2} &= \tilde{Q}_{x_2} \tilde{P}_{x_1} \tilde{Q}_{x_2}. \end{aligned} \quad (\text{B36})$$

Lastly, defining \tilde{R} , \tilde{S} , and \tilde{U}_k analogously to Eqs. (B27), (B28), and (B29), respectively, by replacing $P_{x_1} \rightarrow \tilde{P}_{x_1}$ and $Q_{x_2} \rightarrow \tilde{Q}_{x_2}$, we get a similar expression for V_A :

$$V_A := \sum_i |i\rangle \otimes \tilde{U}_i \tilde{P}_i. \quad (\text{B37})$$

Now, let us evaluate the action of V on the state. Defining $\rho_{\text{out}} := (V_A \otimes V_B) \rho_{AB} (V_A^\dagger \otimes V_B^\dagger)$ gives

$$\rho_{\text{out}} = \sum_{i, i'} \sum_{j, j'} |i, i'\rangle \langle j, j'| \otimes (\tilde{U}_i \tilde{P}_i \otimes U_{i'} P_{i'}) \rho_{AB} (\tilde{P}_j \tilde{U}_j^\dagger \otimes P_j U_j^\dagger) \quad (\text{B38})$$

and because of Eqs. (B33)

$$\begin{aligned} &(\tilde{U}_i \tilde{P}_i \otimes U_{i'} P_{i'}) \rho_{AB} (\tilde{P}_j \tilde{U}_j^\dagger \otimes P_j U_j^\dagger) \\ &= (\tilde{U}_i \otimes U_{i'} P_{i'}) (\tilde{P}_i \otimes \mathbb{1}) \rho_{AB} (\tilde{P}_j \otimes \mathbb{1}) (\tilde{U}_j^\dagger \otimes P_j U_j^\dagger) \end{aligned}$$

$$\begin{aligned} &= (\tilde{U}_i \otimes U_{i'} P_{i'}) (\mathbb{1} \otimes P_i) \rho_{AB} (\mathbb{1} \otimes P_j) (\tilde{U}_j^\dagger \otimes P_j U_j^\dagger) \\ &= \delta_{i, i'} \delta_{j, j'} (\tilde{U}_i \otimes U_i P_i) \rho_{AB} (\tilde{U}_j^\dagger \otimes P_j U_j^\dagger). \end{aligned} \quad (\text{B39})$$

Thus, ρ_{out} turns out to be

$$\rho_{\text{out}} = \sum_{i, j} |i, i\rangle \langle j, j| \otimes (\tilde{U}_i \otimes U_i P_i) \rho_{AB} (\tilde{U}_j^\dagger \otimes P_j U_j^\dagger). \quad (\text{B40})$$

Using Eqs. (B33) it is possible to show that

$$(\tilde{U}_k \otimes \mathbb{1}) \rho_{AB} = (\mathbb{1} \otimes U_k^\dagger) \rho_{AB}, \quad (\text{B41})$$

and

$$\begin{aligned} \rho_{\text{out}} &= \sum_{i, j} |i, i\rangle \langle j, j| \otimes (\mathbb{1} \otimes U_i P_i U_i^\dagger) \rho_{AB} (\mathbb{1} \otimes U_j P_j U_j^\dagger) \\ &= \Phi_d^\dagger \otimes d (\mathbb{1} \otimes P_d) \rho_{AB} (\mathbb{1} \otimes P_d). \end{aligned} \quad (\text{B42})$$

Finally, as ρ_{out} must constitute a normalized state, we must have

$$\text{tr}[(\mathbb{1} \otimes P_d) \rho_{AB}] = \frac{1}{d}. \quad (\text{B43})$$

Equation (B42) leads us to two conclusions. First, it is possible to certify the shared state whenever O is a matrix possessing at least one column filled with nonzero elements, as in Eq. (B29). Second, as P_d is arbitrarily labeled, the maximal violation also implies that the marginal distribution of $\{P_{x_1}\}_{x_1=1}^d$ is uniform. In fact, replacing U_k in Eq. (B28) by U_{k+l} , where $l \in \{1, \dots, d\}$, leads to

$$\text{tr}[(\mathbb{1} \otimes P_{x_1}) \rho_{AB}] = \frac{1}{d} \quad \forall x_1 \in [1, \dots, d]. \quad (\text{B44})$$

In addition, note that P_i and Q_j in Eqs. (B27) and (B29) can be interchanged. If one does so, the state certification can be made whenever O has at least one nonzero row, and, in this case, the marginal distribution of $\{Q_{x_2}\}_{x_2=1}^d$ is guaranteed to be uniform:

$$\text{tr}[(\mathbb{1} \otimes Q_{x_2}) \rho_{AB}] = \frac{1}{d} \quad \forall x_2 \in [1, \dots, d]. \quad (\text{B45})$$

3. Exceptional cases for state certification

The presented method for certification works for a wide class of overlap matrices, but it still requires that O must have at least one column or row the entries of which are nonzero. This case does not include, for instance, overlap matrices as in Eq. (20). In this section, let us provide an example of a state different than $|\Phi_d^+\rangle$ that also achieves β_Q for such cases.

To introduce it, suppose a realization of \mathcal{F}_4 characterized by $P_{x_1} = |x_1\rangle \langle x_1|$ and $Q_{x_2} = |f_{x_2}\rangle \langle f_{x_2}|$, where $x_1, x_2 = 1, \dots, 4$, and

$$\begin{aligned} |f_1\rangle &= \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle), & |f_2\rangle &= \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle), \\ |f_3\rangle &= \frac{1}{\sqrt{2}}(|3\rangle + |4\rangle), & |f_4\rangle &= \frac{1}{\sqrt{2}}(|3\rangle - |4\rangle). \end{aligned} \quad (\text{B46})$$

If considered so, the overlap matrix is given by

$$O = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}, \quad (\text{B47})$$

that is, it has the form of Eq. (20), for $d = 4$. In the same way as before, the observables of Alice are defined as in Eq. (5). On the other hand, the state is given by

$$\rho_{AB} = \sum_{k=1}^2 (\mathbb{1} \otimes \Pi_k) \Phi_4^+ (\mathbb{1} \otimes \Pi_k) \quad (\text{B48})$$

where Φ_4^+ is the four-dimensional maximally entangled state and

$$\begin{aligned} \Pi_1 &= |1\rangle\langle 1| + |2\rangle\langle 2|, \\ \Pi_2 &= |3\rangle\langle 3| + |4\rangle\langle 4|. \end{aligned} \quad (\text{B49})$$

If written in this way, it is easy to see that

$$\begin{aligned} \Pi_1 P_{x_1} \Pi_1 &= \begin{cases} P_{x_1} & \text{if } x_1 = 1, 2, \\ 0 & \text{if } x_1 = 3, 4 \end{cases} \quad \text{and} \\ \Pi_2 P_{x_1} \Pi_2 &= \begin{cases} 0 & \text{if } x_1 = 1, 2, \\ P_{x_1} & \text{if } x_1 = 3, 4, \end{cases} \end{aligned} \quad (\text{B50})$$

which can be generalized to both measurements of Bob and expressed in short as

$$\sum_{k=1}^2 \Pi_k P_{x_1} \Pi_k = P_{x_1} \quad \text{and} \quad \sum_{k=1}^2 \Pi_k Q_{x_2} \Pi_k = Q_{x_2}. \quad (\text{B51})$$

Now, note that this realization not only saturates the quantum value of \mathcal{F}_4 , but preserves the same statistics that would be produced by substituting the four-dimensional maximally entangled state. To check it, suppose an arbitrary observable Γ of Alice, then

$$\begin{aligned} \text{tr}[(\Gamma \otimes P_{x_1}) \Phi_4^+] &= \sum_k \text{tr}[(\mathbb{1} \otimes \Pi_k) (\Gamma \otimes P_{x_1}) (\mathbb{1} \otimes \Pi_k) \Phi_4^+] \\ &= \text{tr} \left[(\Gamma \otimes P_{x_1}) \sum_k (\mathbb{1} \otimes \Pi_k) \Phi_4^+ (\mathbb{1} \otimes \Pi_k) \right] \\ &= \text{tr}[(\Gamma \otimes P_{x_1}) \rho_{AB}], \end{aligned} \quad (\text{B52})$$

where we just used Eq. (B51) for P_{x_1} and the cyclicity of the trace.

To verify that there is a clear difference between ρ_{AB} and Φ_4^+ , consider an isometry $V := V_A \otimes V_B$ such that the action of V_A is defined as

$$\begin{aligned} V_A |1\rangle_A &= |1\rangle_A |1\rangle_{A'}, \\ V_A |2\rangle_A &= |1\rangle_A |2\rangle_{A'}, \\ V_A |3\rangle_A &= |2\rangle_A |1\rangle_{A'}, \\ V_A |4\rangle_A &= |2\rangle_A |2\rangle_{A'}, \end{aligned} \quad (\text{B53})$$

where $|\cdot\rangle_A$ denotes the marginal state of Alice, and $|\cdot\rangle_{A'}$ denotes the marginal state of an ancillary subsystem of Alice. The action of V_B is defined in the same way for the marginal state of Bob and its ancillary subsystem. Then, by applying such an isometry to ρ_{AB} one gets

$$V \rho_{AB} V^\dagger = \frac{1}{2} (|1\rangle\langle 1|_A \otimes |1\rangle\langle 1|_B + |2\rangle\langle 2|_A \otimes |2\rangle\langle 2|_B) \otimes (\Phi_2^+)_{A'B'}, \quad (\text{B54})$$

where $(\Phi_2^+)_{A'B'}$ is the two-dimensional maximally entangled state between subsystems A' and B' . In other words, one can

show that, up to local isometries, ρ_{AB} is equivalent to a perfectly correlated classical random bit combined with a single copy of $(\Phi_2^+)_{A'B'}$.

One can extend this simple example to the case where the matrix of overlaps has a block-diagonal structure with K blocks of dimension d_k , where $k = 1, \dots, K$. In this case, consider the state

$$\rho_{AB} = \sum_{k=1}^K (\mathbb{1} \otimes \Pi_k) \Phi_d^+ (\mathbb{1} \otimes \Pi_k), \quad (\text{B55})$$

where Π_k projects into the d_k -dimensional subspace of the k th block of O . Using a proper definition of V , one can show that

$$V \rho_{AB} V^\dagger = \sum_{k=1}^K p_k |k\rangle\langle k|_A \otimes |k\rangle\langle k|_B \otimes (\Phi_{d_k}^+)_{A'B'}, \quad (\text{B56})$$

where $p_k = d_k/d$ and $(\Phi_{d_k}^+)_{A'B'}$ is the d_k -dimensional maximally entangled state between subsystems A' and B' . That is, it is possible to show that, up to local isometries, ρ_{AB} is equivalent to a convex combination of maximally entangled states of various dimensions, where the classical registers tell Alice and Bob which state they share.

APPENDIX C: THE LOCAL VALUE OF \mathcal{F}_d

In this Appendix, we proceed as follows: first, we obtain a universal lower bound for the local value. This constitutes a proof for Theorem 3. Next, we verify the existence of a d -dimensional overlap matrix that achieves this lower bound only for the even d cases. For odd d , we show that there is no such matrix.

1. Preliminaries for Theorem 3

Before starting with the demonstration of Theorem 3, some short auxiliary results are required. This section develops the solutions to two minimization problems arising while in the demonstration.

a. The constrained probability simplex

In the proof of Theorem 3, we are required to minimize a concave function over a polytope. In this subsection, let us characterize the extremal points of this polytope. We start by representing the convex set by variables $\{t_i\}_{i=1}^n$ such that $t_i \geq 0$, $\sum_i t_i = 1$ and $t_i \leq \tau$ for a fixed $\tau \in (0, 1)$. The first two constraints simply give us the probability simplex. The last one can be interpreted as hyperplanes that cut off the vertices of this simplex. Therefore, we refer to the resulting set as the constrained probability simplex.

Let the n -tuple $\mathbf{t} = (t_1, t_2, \dots, t_n)$ represent a point inside of the probability simplex. Its extremal points are given by the deterministic distributions, i.e., the permutations of

$$\mathbf{t} = (1, 0, 0, \dots, 0). \quad (\text{C1})$$

Naturally, if we require that $t_i \leq \tau$, these vertices no longer belong to the set. Then, let us show that the new extremal points of the constrained probability simplex are given by

permutations of

$$\mathbf{t} = \left(\underbrace{\tau, \tau, \dots, \tau}_{\times \lfloor \frac{1}{\tau} \rfloor}, 1 - \left\lfloor \frac{1}{\tau} \right\rfloor \times \tau, 0, 0, \dots, 0 \right). \quad (\text{C2})$$

To see that there are no other extremal points, one can wonder how many coordinates are admitted to take a value other than zero or τ . Let us suppose that t_j and t_{j+1} are two coordinates satisfying this condition, that is, $0 < t_j, t_{j+1} < \tau$. An arbitrary point \mathbf{t}_{arb} with components t_j and t_{j+1} must be, up to permutations,

$$\mathbf{t}_{\text{arb}} = (\tau, \tau, \dots, \tau, t_j, t_{j+1}, 0, 0, \dots, 0). \quad (\text{C3})$$

If \mathbf{t}_{arb} is not extremal, it is possible to decompose it into a convex sum. Changing this distribution for a quantity ϵ , we get

$$\mathbf{t}_+ = (\tau, \tau, \dots, \tau, t_j + \epsilon, t_{j+1} - \epsilon, 0, 0, \dots, 0) \quad (\text{C4})$$

or, in the inverse way,

$$\mathbf{t}_- = (\tau, \tau, \dots, \tau, t_j - \epsilon, t_{j+1} + \epsilon, 0, 0, \dots, 0). \quad (\text{C5})$$

Note that, for a small enough ϵ , \mathbf{t}_- and \mathbf{t}_+ are still valid points. However, $\mathbf{t}_{\text{arb}} = 1/2(\mathbf{t}_+ + \mathbf{t}_-)$, so \mathbf{t}_{arb} is not extremal. Thus, any extremal point must have zero or one coordinate in the open interval $0 < t_i < \tau$. The first case appears only when $1/\tau$ is an integer. Otherwise, we find ourselves in the second case. In either case, the extremal points must have the form of the vector \mathbf{t} given in Eq. (C2), up to permutations.

Having shown that all the extremal points have the form given above, let us argue that all those points are in fact extremal. To do so, it suffices to show that no single point from \mathbf{t} can be written as a convex combination of the remaining points. More specifically, suppose that \mathbf{t}_j , with $j = 1, \dots, J$, are points of the form of \mathbf{t} , in Eq. (C2). Then, let us show

that those points, in particular \mathbf{t}_1 , cannot be decomposed into a convex sum of \mathbf{t}_j , for $j \neq 1$. Without loss of generality, assume that, for some strictly positive weights w_j , where $\sum_j w_j = 1$, \mathbf{t}_1 can be written as a convex sum:

$$\mathbf{t}_1 = \sum_{j \in \mathcal{J}} w_j \mathbf{t}_j, \quad (\text{C6})$$

where \mathcal{J} is the set of indices j for which $w_j > 0$. Then, suppose that the first component of \mathbf{t}_1 is $[\mathbf{t}_1]_1 = \tau$, so

$$[\mathbf{t}_1]_1 = \sum_{j \in \mathcal{J}} w_j [\mathbf{t}_j]_1 = \tau. \quad (\text{C7})$$

By hypothesis, $[\mathbf{t}_j]_1 \leq \tau$, for all j , and

$$[\mathbf{t}_1]_1 = \sum_{j \in \mathcal{J}} w_j [\mathbf{t}_j]_1 \leq \tau \sum_{j \in \mathcal{J}} w_j = \tau. \quad (\text{C8})$$

In other words, Eq. (C8) is saturated, so $[\mathbf{t}_j]_1 = \tau$, for all $j \in \mathcal{J}$. We can repeat this same argument to all of the components of \mathbf{t}_1 equal to τ , leading to the same conclusion. In addition, using that the components of \mathbf{t}_j are non-negative, a similar argument can be used to show that if $[\mathbf{t}_1]_k = 0$, then $[\mathbf{t}_j]_k = 0$, for all $j \in \mathcal{J}$. The last component of \mathbf{t}_1 equals $1 - \lfloor \frac{1}{\tau} \rfloor \times \tau$, which is fixed by normalization, and so are those of \mathbf{t}_j , for all $j \in \mathcal{J}$. Therefore, \mathbf{t}_1 cannot be decomposed into a convex sum of \mathbf{t}_j , for $j \neq 1$. As \mathbf{t}_1 is an arbitrary point among all of the \mathbf{t}_j points, for all j , then all of the points of the form of \mathbf{t} , in Eq. (C2), are extremal.

b. Minimization of a specific function

In the proof of Theorem 3, we are required to find the minimum value of the following function:

$$s(\tau) := 2 \left\lfloor \frac{1-\tau}{\tau} \right\rfloor (\sqrt{1-\tau} - 1) + 2\sqrt{\tau \left(1 + \left\lfloor \frac{1-\tau}{\tau} \right\rfloor \right)} - \tau - 2 \quad (\text{C9})$$

over $\tau \in (0, 1]$. To do so, let us first show that $s(\tau)$ is continuous. By looking at the function in Eq. (C9), it can be noted that the possible discontinuous points are those at which $(1-\tau)/\tau$ is an integer. So, let us evaluate the sided limits of $s(\tau)$ at points of the form of $\tau = 1/n$, where n is a positive integer:

$$\lim_{\epsilon \rightarrow 0^+} s(1/n - \epsilon) = 2(n-1)(\sqrt{1-1/n} - 1) - 1/n \quad (\text{C10})$$

and

$$\begin{aligned} \lim_{\epsilon \rightarrow 0^+} s(1/n + \epsilon) &= 2(n-2)(\sqrt{1-1/n} - 1) + 2(\sqrt{1-1/n} - 1) - 1/n \\ &= 2(n-1)(\sqrt{1-1/n} - 1) - 1/n. \end{aligned} \quad (\text{C11})$$

Since the sided limits coincide, $s(\tau)$ is continuous with respect to τ .

Now, note that, inside of the interval $\tau \in [\frac{1}{n+1}, \frac{1}{n}]$, $s(\tau)$ can be written as

$$s(\tau) = 2(n-1)(\sqrt{1-\tau} - 1) + 2\sqrt{n\tau} - \tau - 2. \quad (\text{C12})$$

If we treat n as a fixed parameter, it is easy to see that $s(\tau)$ is a concave function of τ (a linear combination of concave

terms with non-negative coefficients is concave). Moreover, for a single-variable concave function defined over a closed and bounded interval, the points that minimize this function are at the edges of the interval. In this case, we are looking for points of the form $\tau = 1/n$. Therefore, to minimize $s(\tau)$, we can discard the points in the interior and focus only at the edges.

Then, let us define a function $g(n)$ the domain of which is the set of positive integers:

$$g(n) := s\left(\tau = \frac{1}{n}\right) = 2(n-1)\left(\sqrt{1 - \frac{1}{n}} - 1\right) - \frac{1}{n}. \quad (\text{C13})$$

For a moment, assume that n is a continuous variable. This way, $g(n)$ can be minimized just by evaluating its derivative:

$$\frac{dg}{dn} = 2\left(\sqrt{1 - \frac{1}{n}} - 1\right) + \frac{n-1}{n^2\sqrt{1 - 1/n}} + \frac{1}{n^2}. \quad (\text{C14})$$

Setting $\frac{dg}{dn} = 0$ after some algebraic manipulation leads to

$$n^{*2} - n^* - 1 = 0 \Rightarrow n^* = \frac{1 + \sqrt{5}}{2} \approx 1.62, \quad (\text{C15})$$

which provides a single positive root for n^* . Moreover, the second derivative, $\frac{d^2g}{dn^2}$, is positive at $n = n^*$ and so n^* is indeed a minimum of $g(n)$. However, $g(n)$ is only defined for positive integers, and we have to restrict the analysis to this set. Since n^* is the only critical point of $g(n)$ and $\frac{dg}{dn}$ is positive for $n > n^*$ and negative for $n < n^*$, it suffices to check the closest integers of n^* , i.e., $n = 1$ and 2 .

Then, returning to $s(\tau)$, as $s(\tau = 1) = -1$ and $s(\tau = 1/2) = \sqrt{2} - 5/2$, $s(\tau)$ is minimized at $\tau = 1/2$.

2. The analysis of the local value in Eq. (18)

In this section, let us show some properties that can be extracted from the expression of the local value in Eq. (18). First, suppose that $O_{ij} = 1$, for some $i, j \in \{1, \dots, d\}$. Note that this assumption leads to $O_{i,x_2} = \delta_{x_2,j}$ and $O_{x_1,j} = \delta_{x_1,i}$, as the rows and columns of O are normalized. Next, Eq. (17) evaluated for the strategy $s(i, j)$ of Bob leads to

$$s(i, j) = \sum_{x_1 \neq i}^d \left[\sqrt{1 - O_{x_1,j}^2} - \frac{1}{2}(1 - O_{x_1,j}^2) \right] + \sum_{x_2 \neq j}^d \left[\sqrt{1 - O_{i,x_2}^2} - \frac{1}{2}(1 - O_{i,x_2}^2) \right] = d - 1. \quad (\text{C16})$$

Note that this is the largest value that can be achieved by any strategy, as the quantum value is also $d - 1$. Therefore, in this case, this must be the optimal strategy of Bob, and $\beta_L(O) = d - 1$. This is the reason why the assumption $O_{x_1,x_2} < 1$ is required for most of our analysis. If one desires to trivialize the Bell inequalities presented in this paper, it suffices to construct \mathcal{F}_d out of a matrix with a single overlap equal to 1.

Now, let us show that $\beta_L(O)$ can be lower bounded by a clever choice of the strategies of Bob. This constitutes a proof for Theorem 3. Without loss of generality, let us start by identifying the largest element of O as O_{11} . We can always do so, as it is always possible to relabel the outputs. Then, let us lower bound β_L by restricting the set of strategies from $u, v \in \{1, \dots, d\}$ to $u, v \in \{1\}$:

$$\beta_L(O) = \max_{u,v} [s(u, v)] \geq \max_{u,v \in \{1\}} [s(u, v)] = s(1, 1). \quad (\text{C17})$$

In other words, we choose to bound β_L by the strategy related to the largest element of O , which we defined to be O_{11} . This

particular choice circumvents the maximization and reduces the problem of lower bounding β_L to the calculation of a single strategy. Recall that, from Eq. (17), we are summing over the interval $R_{\pm} = \{x \in [d]^2 | (\delta_{x_{1u}} - \delta_{x_{2v}}) = \pm 1\}$, which means that the sum is performed over row u and column v , but it excludes the term (u, v) , as $(\delta_{uu} - \delta_{vv})$ is, obviously, zero. Then, by making this choice, we are lower bounding β_L by the strategy that excludes the largest term of O in the sum.

The strategy $s(1, 1)$, in turn, is given by

$$s(1, 1) = \sum_{x_1 \neq 1}^d \left[\sqrt{1 - O_{x_1,1}^2} - \frac{1}{2}(1 - O_{x_1,1}^2) \right] + \sum_{x_2 \neq 1}^d \left[\sqrt{1 - O_{1,x_2}^2} - \frac{1}{2}(1 - O_{1,x_2}^2) \right]. \quad (\text{C18})$$

Defining $t_{x_1,x_2} := O_{x_1,x_2}^2$ and using the normalization of rows and columns of O , we get

$$s(1, 1) = \sum_{x_1 \neq 1}^d \sqrt{1 - t_{x_1,1}} + \sum_{x_2 \neq 1}^d \sqrt{1 - t_{1,x_2}} - d - t_{11} + 2. \quad (\text{C19})$$

To continue, let us define an auxiliary function, h , such that

$$h(t_{21}, t_{31}, \dots, t_{d,1}) := \sum_{x_1=2}^d \sqrt{1 - t_{x_1,1}}, \quad (\text{C20})$$

i.e., it corresponds to the first term of Eq. (C19). Note that h is a Schur-concave function with respect to the $d - 1$ variables $(t_{21}, t_{31}, \dots, t_{d,1})$. This is of particular importance because any Schur-concave function, when minimized over a compact set, achieves its minimum on some extremal point of this set. Thus, we can use this fact to minimize both functions.

Then, let us express the $d - 1$ variables of h as a vector, $\mathbf{t} := (t_{21}, t_{31}, \dots, t_{d,1})$. For the elements of \mathbf{t} , we have

$$t_{x_1,1} \geq 0, \quad \sum_{x_1 \neq 1} t_{x_1,1} = 1 - t_{11} \quad \text{and} \quad t_{x_1,1} \leq t_{11}. \quad (\text{C21})$$

The last condition comes from the fact that O_{11} (and, by consequence, t_{11}) is the largest element. The solution to the minimization of $h(\mathbf{t})$ is presented in Sec. C where the extremal points \mathbf{t}_{ext} of the set in Eqs. (C21) are given by permutations of

$$\mathbf{t}_{\text{ext}} = \left(\underbrace{t_{11}, \dots, t_{11}}_{\times \lfloor \frac{1-t_{11}}{t_{11}} \rfloor}, 1 - t_{11} - \left\lfloor \frac{1-t_{11}}{t_{11}} \right\rfloor \times t_{11}, 0, \dots, 0 \right). \quad (\text{C22})$$

An identical solution is obtained for the minimization of the second term of Eq. (C19), so $s(1, 1)$ can be lower bounded by

$$s(1, 1) \geq 2 \left[\frac{1-t_{11}}{t_{11}} \right] (\sqrt{1-t_{11}} - 1) + 2\sqrt{t_{11}} \left(1 + \left\lfloor \frac{1-t_{11}}{t_{11}} \right\rfloor \right) - t_{11} + d - 2. \quad (\text{C23})$$

Now, we have a single-variable function, which can be easily minimized. This is precisely the function analyzed in Sec. V

and the minimum is proven to occur at $t_{11} = 1/2$. Thus,

$$s(1, 1) \geq d + \sqrt{2} - \frac{5}{2}, \quad (\text{C24})$$

which implies that $\beta_L(O) \geq d + \sqrt{2} - 5/2$, for any d -dimensional O , with $d \geq 2$.

3. Proof of Lemma 2

In this section, we investigate the overlap matrices that saturate the lower bound on β_L . Let us start with the following lemma.

Lemma 3. If a $d \times d$ matrix O^* is such that $\beta_L(O^*) = d + \sqrt{2} - 5/2$, then a 2×2 MUB block can be separated from the remainder.

Proof. Assume that O^* is a $d \times d$ matrix such that $\beta_L(O^*) = d + \sqrt{2} - 5/2$. From the argument developed in the last section, we start by assigning $1/2$ to t_{11} , or, in this case, assuming that $O_{11}^* = 1/\sqrt{2}$. Because of the strict concavity of $h(\mathbf{t})$ in Eq. (C20), Eq. (C23) is tight if and only if the vector that minimizes $h(\mathbf{t})$ is given by permutations of

$$\mathbf{t}_{\text{ext}} = \left(\frac{1}{2}, 0, \dots, 0 \right); \quad (\text{C25})$$

that is, to saturate the lower bound, we take $O_{12}^* = O_{21}^* = 1/\sqrt{2}$. Besides that, we must also guarantee that the overlap matrix corresponds to a valid unitary. Recall that, from Eq. (1), the overlaps are taken as the absolute value of the inner product between bases $\{e_{x_1}\}_{x_1=1}^d$ and $\{f_{x_2}\}_{x_2=1}^d$. In other words, the elements of O are the absolute value of the elements of some unitary. It is easy to see that this implies that $O_{22}^* = 1/\sqrt{2}$, so that the partial form of O^* is given by

$$O^* = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & \dots & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 & \dots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & & \\ 0 & 0 & & & O_{\text{rest}} \end{bmatrix}, \quad (\text{C26})$$

where O_{rest} represents the uncharacterized elements of O^* . ■

Now, to continue with the proof of Lemma 2, we have to obtain some characterization of $\beta_L(O_{\text{rest}})$. In fact, we will show that

$$\beta_L(O_{\text{rest}}) = (d - 2) + \sqrt{2} + \frac{5}{2}, \quad (\text{C27})$$

which coincides with the lower bound for dimension $d - 2$.

Because O_{rest} is a square matrix of dimension $d - 2$, Theorem 3 already implies that $\beta_L(O_{\text{rest}})$ is lower bounded by $(d - 2) + \sqrt{2} + 5/2$. On the other hand, the upper bound is a consequence of Lemma 3. Because $\beta_L(O^*) = d + \sqrt{2} - 5/2$, all strategies of O^* must be upper bounded by this same amount. In particular, note that for each strategy labeled by outputs inside of the O_{rest} block, in Eq. (C26), there are four zero terms, contributing in total with $+2$. Then, when considering only O_{rest} , we have $\beta_L(O_{\text{rest}}) \leq (d - 2) + \sqrt{2} + 5/2$.

Thus, for dimension $d - 2$, the block O_{rest} also fulfils the assumptions of Lemma 3, so we can also extract a 2×2 MUB block from it. Naturally, the indefinite iteration of the above

argument will lead us to two cases: either d is even and

$$O^* = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & \dots & 0 & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & \dots & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \quad (\text{C28})$$

or d is odd and

$$O^* = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & \dots & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 & \dots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & & & 1 \end{bmatrix}, \quad (\text{C29})$$

where the last block corresponds to the element $O_{d,d}^* = 1$. However, while Eq. (C28) is a correct form of O^* in the even case, for odd d , this is not true. First, it turns out that the matrix in Eq. (C29) has one overlap equal to 1, which we discarded from our analysis. Second, we know, from Appendix C 2, that this leads to $\beta_L(O^*) = d - 1$, which contradicts the initial statement of Lemma 3, that $\beta_L(O^*) = d + \sqrt{2} - 5/2$. Thus, for the odd d case, there is no O^* that saturates the lower bound in Theorem 3.

4. Some considerations for the odd d case

Finally, let us quickly explain why the lower bound derived in Theorem 3 is not tight for odd d . Consider the entrywise squared version of O , which we refer to as T . If written so, T assumes the form of a *unistochastic* matrix; i.e., if $U_{x_1 x_2}$ are the elements of a unitary U , for $x_1, x_2 \in \{1, \dots, d\}$, then $T_{x_1 x_2} = |U_{x_1 x_2}|^2$. A unistochastic matrix is also a particular case of a *bistochastic* matrix—a non-negative matrix the columns and rows of which add up to one. For 2×2 arrays, it happens that the bistochastic and the unistochastic sets of matrices coincide, but for larger dimensions the unistochastic set is a proper subset of the bistochastic set.

It is clear that our original intention was to minimize $\beta_L(O)$ over the set of overlap matrices. However, the proof of Theorem 3 relies only on the fact that T is bistochastic and, in fact, the derived lower bound corresponds exactly to the lowest value achievable by a bistochastic matrix. For even d , among the optimal bistochastic matrices, there exist some which are also unistochastic and, hence, the resulting bound is tight.

For odd d , however, if we try to simultaneously saturate the bound and enforce unistochasticity, we reach a contradiction, as shown in the previous section. On the other hand, the lower bound can be saturated by a bistochastic matrix, e.g.,

$$T^* = \begin{bmatrix} 1/2 & 0 & 0 & \dots & 0 & 1/2 \\ 1/2 & 1/2 & 0 & \dots & 0 & 0 \\ 0 & 1/2 & 1/2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1/2 & 1/2 \end{bmatrix}, \quad (\text{C30})$$

which is valid for all $d \geq 2$. However, none of the optimal bistochastic matrices for odd d happen to be unistochastic.

Now, let us focus on the case $d = 3$. An aspect that makes the optimization over the set of unistochastic matrices for

$d = 3$ difficult is the fact that this set is not convex, but only star convex. Also, the unistochastic and bistochastic sets are both centered at $\frac{1}{3}J_3$, where J_3 is the 3×3 matrix of ones [20]. A proper notion of center can be acquired if one considers a uniformly weighted convex combination of the extremal points of the permutation matrices. The permutation matrices are the only extremal points of the bistochastic set, and they are also extremal for the unistochastic set.

We have numerically implemented a function that calculates the local value for a given overlap matrix. By performing a local search over a large number of random starting points we have reached the conjecture that the smallest value of β_L for $d = 3$ is achieved for

$$O_{\text{conj}} = \begin{bmatrix} 1/3 & 2/3 & 2/3 \\ 2/3 & 1/3 & 2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}. \quad (\text{C31})$$

To see why this conjecture is reasonable, consider the entrywise squared version of O_{conj} , which we refer to as T_{conj} . An analytic condition derived in Ref. [21] allows us to check that T_{conj} is not only unistochastic, but also lies at the boundary of the unistochastic set. Furthermore, consider a permutation of the three-dimensional matrix in Eq. (C30):

$$T_3^* = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}. \quad (\text{C32})$$

Then,

$$T_{\text{conj}} = \frac{1}{3}(\frac{1}{3}J_3) + \frac{2}{3}T_3^*. \quad (\text{C33})$$

That is, if one considers the line segment connecting the center of the unistochastic set and the optimal bistochastic matrix, T_{conj} can be found at the intersection of this segment with the boundary of the unistochastic set.

Lastly, let us show that O_{conj} provides a smaller local value than MUBs. By calculating $\beta_L(O_{\text{conj}})$, we get

$$\beta_L(O_{\text{conj}}) = \frac{1}{9}[6(\sqrt{8} + \sqrt{5}) - 13] \approx 1.9319. \quad (\text{C34})$$

For d -dimensional MUBs, we obtain

$$\beta_L\left(\frac{1}{\sqrt{d}}J_d\right) = 2(d-1)\sqrt{\frac{d-1}{d}} - \frac{(d-1)^2}{d}. \quad (\text{C35})$$

If evaluated for $d = 3$, then $\beta_L(J_3/\sqrt{3}) = 1.9327$, which is slightly bigger than $\beta_L(O_{\text{conj}})$. In fact, O_{conj} allows us to construct counterexamples for all odd dimensions. For odd $d \geq 3$, consider the following matrix:

$$O_d^\oplus = \left(\bigoplus_{i=1}^{\lfloor d/2 \rfloor - 1} \frac{1}{\sqrt{2}}J_2 \right) \oplus O_{\text{conj}}. \quad (\text{C36})$$

The local value obtained for O_d^\oplus is given by

$$\beta_L(O_d^\oplus) = d - 3 + \frac{1}{9}[6(\sqrt{8} + \sqrt{5}) - 13] \approx d - 1.0681. \quad (\text{C37})$$

If one takes the derivative of $\beta_L(J_d/\sqrt{d}) - \beta_L(O_d^\oplus)$, by a simple analytic argument it is possible to conclude that this derivative is positive for all odd $d \geq 3$. Because $\beta_L(J_d/\sqrt{d}) - \beta_L(O_d^\oplus)$ is positive for $d = 3$, then it must also be positive for all odd $d \geq 3$. Therefore, this shows that for all odd $d \geq 3$, the realization of \mathcal{F}_d that is most robust to noise does not correspond to MUBs in dimension d .

-
- [1] J. S. Bell, *Phys. Phys. Fiz.* **1**, 195 (1964).
[2] S. J. Freedman and J. F. Clauser, *Phys. Rev. Lett.* **28**, 938 (1972).
[3] A. Aspect, P. Grangier, and G. Roger, *Phys. Rev. Lett.* **47**, 460 (1981).
[4] A. Aspect, P. Grangier, and G. Roger, *Phys. Rev. Lett.* **49**, 91 (1982).
[5] A. Aspect, J. Dalibard, and G. Roger, *Phys. Rev. Lett.* **49**, 1804 (1982).
[6] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, *Nature (London)* **526**, 682 (2015).
[7] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin *et al.*, *Phys. Rev. Lett.* **115**, 250401 (2015).
[8] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor *et al.*, *Phys. Rev. Lett.* **115**, 250402 (2015).
[9] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Rev. Mod. Phys.* **86**, 419 (2014).
[10] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
[11] T. Heinosaari, T. Miyadera, and M. Ziman, *J. Phys. A: Math. Theor.* **49**, 123001 (2016).
[12] A. Fine, *Phys. Rev. Lett.* **48**, 291 (1982).
[13] N. Gisin, *Phys. Lett. A* **154**, 201 (1991).
[14] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, *Phys. Rev. Lett.* **103**, 230402 (2009).
[15] A. Tavakoli, M. Farkas, D. Rosset, J.-D. Bancal, and J. Kaniewski, *Sci. Adv.* **7**, eabc3847 (2021).
[16] J. Schwinger, *Proc. Natl. Acad. Sci. USA* **46**, 570 (1960).
[17] T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski, *Int. J. Quantum Inf.* **08**, 535 (2010).
[18] M. P. Colomer, L. Mortimer, I. Frérot, M. Farkas, and A. Acín, *Quantum* **6**, 778 (2022).
[19] G. Zauner, Quantendesigns, grundlege einer nichtkommutativen designtheorie, Ph.D. thesis, University of Vienna, 1999.
[20] I. Bengtsson, A. Ericsson, M. Kuś, and K. Życzkowski, *Commun. Math. Phys.* **259**, 307 (2005).
[21] A. Fedullo, *Nuovo Cimento B* **107**, 1413 (1992).

3.2 Comments on “*Biased random access codes*”

In *Biased random access codes*, we describe a variation of the RAC protocol, named accordingly, in which both the string to be encoded and the character to be recovered are not uniformly distributed. This bias in the input distribution significantly affects the encoding and decoding strategies that optimise the protocol’s performance in both quantum and classical settings. We then tackle the problem of optimising these biased RACs using both numerical and analytical methods. For the numerical approach, we present the RAC-tools Python package, which implements algorithms that compute the exact classical value and lower bounds for the quantum value of any biased RAC. Analytically, we derive upper bounds for the optimal performance based on projective measurements, specifically for RACs where the encoded strings consist of either two characters from a d -length alphabet or n bits.

The student’s key contribution to the project was developing the numerical analysis for the scenarios studied, including the creation and documentation of the Python package RAC-tools. Additionally, the student contributed to the analytical portion of the research, with a particular focus on the proposition of Lemma 1. In terms of manuscript preparation, the main contributions involved writing Secs. II and III of the main text, developing Appendices A, B, and D, creating all figures and plots, and providing input on the remaining parts of the text.

Biased random access codesGabriel Pereira Alves^{1,*}, Nicolas Gigena^{1,†}, and Jędrzej Kaniewski^{1,‡}*Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland*

(Received 28 March 2023; accepted 25 September 2023; published 13 October 2023)

A random access code (RAC) is a communication task in which the sender encodes a random message into a shorter one to be decoded by the receiver so that a randomly chosen character of the original message is recovered with some probability. Both the message and the character to be recovered are assumed to be uniformly distributed. In this paper, we extend this protocol by allowing more general distributions of these inputs, which alters the encoding and decoding strategies optimizing the protocol performance, with either classical or quantum resources. We approach the problem of optimizing the performance of these biased RACs with both numerical and analytical tools. On the numerical front, we present algorithms that allow a numerical evaluation of the optimal performance over both classical and quantum strategies and provide a Python package designed to implement them, called RAC-tools. We then use this numerical tool to investigate single-parameter families of biased RACs in the $n^2 \mapsto 1$ and $2^d \mapsto 1$ scenarios. For RACs in the $n^2 \mapsto 1$ scenario, we derive a general upper bound for the cases in which the inputs are not correlated, which coincides with the quantum value for $n = 2$ and in some cases for $n = 3$. Moreover, it is shown that attaining this upper bound self-tests pairs or triples of rank-1 projective measurements, respectively. An analogous upper bound is derived for the value of RACs in the $2^d \mapsto 1$ scenario, which is shown to be always attainable using mutually unbiased measurements if the distribution of input strings is unbiased.

DOI: [10.1103/PhysRevA.108.042608](https://doi.org/10.1103/PhysRevA.108.042608)**I. INTRODUCTION**

In the past decades several instances have been found in which quantum resources provide an advantage in the performance of a given task. Quantum computing algorithms [1–5], such as Shor’s factorization algorithm [6], are just one example of the power of quantum resources: Spatially separated parties can use entanglement [7] in a shared quantum state, for instance, to improve their performance in a nonlocal game [8–10], to quantum teleport [11–14] the state of a third system held by one of them, or to densely encode classical information to be sent via a quantum channel [15–17]. Quantum devices have also shown to be powerful resources for certain communication tasks in which a quantum state is prepared by one party and sent to another one, who performs a measurement to extract information. Such tasks are known as prepare-and-measure experiments, and they find application in quantum information processing protocols like quantum key distribution (QKD) [18–22], randomness certification [23–26], and quantum random access codes [27–30]. A random access code (RAC) is a communication task in which a string of characters, chosen at random from a given alphabet, is encoded into a shorter string in such a way that any of the characters in the original string can be recovered, with some probability, by means of a decoding strategy. Both the string to be encoded and the character to be recovered are uniformly distributed, with the encoding party not knowing in

advance which character should be retrieved by the decoding procedure. In that sense, the RAC can be understood as a form of nondeterministic data compression.

The implementation of the RAC protocol and its variations have been the subject of intense research, finding applications in cryptography [31,32], self-testing of measurements [33,34], foundational aspects of no-signaling correlations [35], and quantum communication complexity [36,37]. In this work, we introduce a generalization of the RAC protocol in which neither the string to be encoded nor the character to be recovered are uniformly distributed. Biasing the distribution of inputs has a nontrivial effect on the encoding and decoding strategies optimizing the performance of the protocol in both quantum and classical realizations. We approach here the problem of finding the optimal performance of such biased RACS, and the strategies attaining it, with both numerical and analytical techniques. In the numerical front we present the RAC-tools Python package, built to implement algorithms providing the exact classical value and lower bounds to the quantum value of an arbitrary biased RAC. On the analytical side, we derive upper bounds for the optimal performance over projective measurements of RACs in which the character strings to be encoded consist either of two characters to be chosen from length d alphabet or of n characters to be chosen from a length 2 alphabet. In the cases in which these upper bounds are attainable, we study the optimal quantum strategies achieving the optimal performance, paying special attention to the dependence of the optimal measurements on the biasing parameters and the regions in parameter space in which quantum strategies provide an advantage over classical ones. These analytical results are then compared with those produced by the numerical package.

*gpereira@fuw.edu.pl

†nicolas.gigena@fuw.edu.pl

‡jkaniewski@fuw.edu.pl

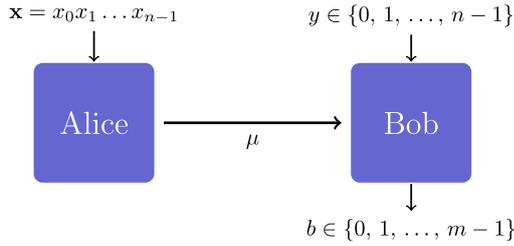


FIG. 1. The $n^m \mapsto 1$ RAC. Alice encodes her input $\mathbf{x} = x_0 x_1 \dots x_{n-1}$ into a message μ which is sent to Bob. Based on the message μ and his input y , Bob tries to guess the y th character of Alice. Each character x_i of \mathbf{x} ranges from 0 to $m - 1$, for $i = 0, 1, \dots, n - 1$.

II. BIASED RACS

A RAC scenario, as depicted in Fig. 1, involves two parties, Alice and Bob. A RAC scenario is parameterized by integers n, m, d , which we assume to be equal to or larger than 2. Alice is given an n -character string $\mathbf{x} \in S = \{0, \dots, m - 1\}^{\times n}$ and asked to encode it into a single character $\mu \in \{0, \dots, d - 1\}$, which she will later send to Bob. Bob, on the other hand, is asked to decode Alice’s message in order to retrieve the value of the y th character x_y in the original string. In order to do so Bob evaluates the image $b \in \{0, \dots, m - 1\}$ of μ under a previously chosen decoding function, represented in the figure by Bob’s box. We consider the task successful when Bob correctly guesses x_y from his decoding function, i.e., when $b = x_y$. We denote a scenario like the one described above by $n^m \mapsto 1$ RAC, where we allow the message and the characters in the string to belong to alphabets with different cardinalities, $m \neq d$. Whenever $m = d$, however, we will use the notation $n^d \mapsto 1$, since it is the usual notation in the literature.

The figure of merit that is commonly used to study a RAC is the *average success probability* \bar{P} , which is simply an average of the winning probabilities over all combinations of \mathbf{x} and y :

$$\bar{P} = \frac{1}{nm^n} \sum_{\mathbf{x}, y} p(b = x_y | \mathbf{x}, y), \tag{1}$$

where $p(b = x_y | \mathbf{x}, y)$ denotes the probability of a successful decoding when string \mathbf{x} is encoded and character x_y is to be recovered, and the factor $\frac{1}{nm^n}$ reflects the assumption that both \mathbf{x} and y are uniformly distributed. The optimal pairs of encoding-decoding strategies implemented by Alice and Bob will therefore be those maximizing \bar{P} . Note that the trivial strategy of outputting a fixed (or random) value of b achieves average success probability of $\frac{1}{m}$, so we will be interested only in strategies that outperform this value.

In this work we study a more general class of RACs in which the distribution of Alice’s and Bob’s inputs, \mathbf{x} and y , respectively, is not necessarily uniform. We refer to these as *biased RACs*, or *b-RACs*. In fact, let us start with the most general real linear functional, i.e., a tensor α_{xyb} of order $(n + 2)$, which attributes a specific weight to each combination of inputs \mathbf{x} and y and output b . The value of this functional on

a probability distribution $p(b | \mathbf{x}, y)$ equals

$$\mathcal{F} = \sum_{\mathbf{x}, y, b} \alpha_{xyb} p(b | \mathbf{x}, y). \tag{2}$$

Since the probability distribution is normalized an additive shift in the coefficients of the tensor results in an additive shift in the value. Thus, we can focus on tensors which are non-negative. Similarly, by normalizing the coefficients we can, without loss of generality, consider tensors satisfying $\sum_{\mathbf{x}, y, b} \alpha_{xyb} = 1$. Now, we would like to focus our attention on functionals that to some extent resemble the standard random access code, where the winning condition reads $b = x_y$. In this spirit, we will consider only tensors whose coefficients vanish whenever $b \neq x_y$. Due to non-negativity and normalization such linear functionals can be interpreted as probability distributions over \mathbf{x} and y , and let us denote them by α_{xy} . Then the value of the functional equals

$$\mathcal{F} = \sum_{\mathbf{x}, y} \alpha_{xy} p(b = x_y | \mathbf{x}, y), \tag{3}$$

and it should be clear that this is precisely the same as the original RAC except that the distribution of inputs might be nonuniform (in the RAC we have $\alpha_{xy} = \frac{1}{nm^n}$). Hence, biasing the input distribution of a RAC is analogous to biasing or tilting functionals in a Bell nonlocality scenario [38–40], as the consequence of adopting bias is to modify the functional in Eq. (1) and, in turn, the optimal realization. An interesting aspect of this generalization is that the inputs of Alice and Bob are not necessarily independent. To the best of our knowledge such scenarios were first analyzed in Ref. [41]. An alternative direction would be to change the goal of the decoding function from recovering a given character to a more general function of the input string, a generalization that has recently been explored in Ref. [42], but we do not consider such scenarios in this work.

As is the case with Bell scenarios, *b-RACs* can be fundamentally interpreted as an experiment in which a particular behavior of the involved devices, specified by the conditional probabilities $\{p(b = x_y | \mathbf{x}, y)\}$, determines a certain value for the previously specified figure of merit \mathcal{F} in Eq. (3). In the case of *b-RACs* the behavior is determined by the encoding and decoding strategies implemented by Alice and Bob. So far we have focused on classical strategies in which, upon receiving her input \mathbf{x} , Alice computes its image μ under an encoding function $E : \{0, \dots, m - 1\}^{\times n} \mapsto \{0, \dots, d - 1\}$ and sends it to Bob. Bob takes μ as the argument of his y th decoding function $D_y : \{0, \dots, d - 1\} \mapsto \{0, \dots, m - 1\}$ producing $b = D_y(\mu)$. The strategies we just described are deterministic, i.e., the probability of Bob’s output being b is given by

$$p(b | \mathbf{x}, y) = \begin{cases} 1 & \text{if } b = D_y(E(\mathbf{x})) \\ 0 & \text{otherwise} \end{cases}, \quad \forall \mathbf{x}, y, b. \tag{4}$$

Clearly, if Alice and Bob decide to employ some nondeterministic strategy (even if we allow them to share classical randomness), the corresponding behavior $\{p(b = x_y | \mathbf{x}, y)\}$ will belong to the convex hull of classical deterministic behaviors. It follows then that the *b-RAC* functional in Eq. (3) attains its maximum \mathcal{F}_C for one of these deterministic

strategies, which can be found through an exhaustive search, as is the case for the local value of Bell functionals. It should nonetheless be noted that in a bipartite scenario with n measurement settings per party and m measurements per setting the number of deterministic strategies scales as m^{2n} , whereas in the $n^m \mapsto 1$ b -RAC scenario this number grows double exponentially, according to $d^{m^n} \times m^{dn}$. Therefore, computing the classical value through exhaustive search becomes infeasible even for small values of n , d , and m . The following lemma provides a simpler and more efficient approach to the computation of the classical value.

Lemma 1. (a) For a $n^m \mapsto 1$ b -RAC and a fixed encoding function $E(\mathbf{x}) = \mu$, the optimal decoding functions are $D_y^*(\mu) = b$, where b is the character that maximizes the sum

$$\sum_{\mathbf{x} \in S} \alpha_{\mathbf{x}, y, b} \delta_{\mu, E(\mathbf{x})}. \quad (5)$$

(b) For a $n^m \mapsto 1$ b -RAC and fixed decoding functions $\{D_y\}_y$, the optimal encoding function is given by $E^*(\mathbf{x}) = \mu$, where μ is the character that maximizes the sum

$$\sum_{y=0}^{n-1} \alpha_{\mathbf{x}, y, D_y(\mu)}. \quad (6)$$

It follows from Lemma 1, which is proved in Appendix A, that we can reduce the exhaustive search to a search either over encoding functions only, statement (a), or over decoding functions only, statement (b). In the first case, the complexity of the problem reduces from $d^{m^n} \times m^{dn}$ to $d^{m^n} \times nmd$, whereas in the second it reduces to $d \times m^{n(d+1)}$. A similar treatment is usually implemented in Bell scenarios, where optimal measurement (response function in the classical case) can be computed if the remaining components are fixed.

A. Quantum value of a biased RAC

As an information processing task, the RAC can be generalized to quantum strategies. A quantum strategy involves quantum devices, which transforms the b -RAC scenario into a particular case of a prepare-and-measure experiment [43]. Such scenarios are often encountered in many new quantum technologies such as quantum communication [44] and quantum cryptography [45]. In a quantum strategy Alice, upon receiving a string \mathbf{x} , encodes it no longer in a classical character μ , but in a qudit density operator $\rho_{\mathbf{x}}$ over a d -dimensional Hilbert space \mathcal{H} that she then sends to Bob. Once Bob receives Alice's preparation and his input y , he performs a decoding measurement described by operators $\{M_y^b\}_{b=1}^m$, producing output b . From Born's rule it follows then that the probability of a successful decoding is $p(b = x_y | \mathbf{x}, y) = \text{tr}(\rho_{\mathbf{x}} M_y^{x_y})$, and the ensuing value of the figure of merit in Eq. (3) reads

$$\mathcal{F} = \sum_{\mathbf{x}, y} \alpha_{\mathbf{x}, y} \text{tr}(\rho_{\mathbf{x}} M_y^{x_y}). \quad (7)$$

The optimal quantum encoding-decoding strategies will be those for which the functional in Eq. (7) attains its maximum value, which we will denote by \mathcal{F}_Q to distinguish it from the optimal value over the set of classical behaviors, which we will denote from now on by \mathcal{F} . The optimization problem

over both preparations $\rho_{\mathbf{x}}$ and measurements $\{M_y^{x_y}\}$ involved in the determination of \mathcal{F}_Q is in general hard, but it can be approached numerically by means of a *see-saw* algorithm, as we will describe below. This method, which relies on the fact that for fixed preparations optimal measurements can be found efficiently and vice versa, is a known numerical technique for obtaining lower bounds for the quantum value of Bell inequalities [46,47].

In a nutshell, the see-saw algorithm consists in the repeated implementation of a two-step optimization procedure, since we need to optimize the b -RAC functional over preparations and over measurements. We start with a set of randomly chosen measurements, for which we can find the optimal preparations by noting that the functional \mathcal{F} can be written as¹

$$\mathcal{F} = \sum_{\mathbf{x}, y, b} \alpha_{\mathbf{x}, y, b} \text{tr}(\rho_{\mathbf{x}} M_y^b) = \sum_{\mathbf{x}} \text{tr} \left(\rho_{\mathbf{x}} \sum_{y, b} \alpha_{\mathbf{x}, y, b} M_y^b \right), \quad (8)$$

becoming thus apparent that the density matrix $\rho_{\mathbf{x}}$ maximizing the trace is given by a state associated with the largest eigenvalue of the positive semidefinite operator $\sum_{y, b} \alpha_{\mathbf{x}, y, b} M_y^b$. This means that if we are interested in computing the quantum value we can without loss of generality assume $\rho_{\mathbf{x}}$ to be a pure state. Once the preparations are determined the algorithm proceeds to the second step, which is finding the ensuing optimal measurements for these preparations. In order to do so it is convenient to again rewrite the functional \mathcal{F} as

$$\mathcal{F} = \text{tr} \left(\sum_{\mathbf{x}, y, b} M_y^b \alpha_{\mathbf{x}, y, b} \rho_{\mathbf{x}} \right) = \text{tr} \left(\sum_{y, b} M_y^b \varrho_{y, b} \right), \quad (9)$$

where

$$\varrho_{y, b} := \sum_{\mathbf{x}} \alpha_{\mathbf{x}, y, b} \rho_{\mathbf{x}} \quad (10)$$

is a subnormalized density matrix. It is then easily seen that the optimization over measurements takes the form of a semidefinite program (SDP)

$$\begin{aligned} \max_{\{M_y^b\}} \quad & \text{tr} \left(\sum_b M_y^b \varrho_{y, b} \right) \\ \text{s.t.} \quad & M_y^b \geq 0, \quad \forall b \\ \text{and} \quad & \sum_b M_y^b = \mathbb{1}, \end{aligned} \quad (11)$$

the solution of which, for all n measurements, completes the second step in the optimization procedure and one iteration in the see-saw algorithm, with the newly found measurements becoming the starting point of the next iteration. Note that the problem in Eq. (11) is one of minimum-error discrimination [48] of the states $\varrho_{y, b}$, i.e., the optimal measurements $\{M_y^b\}_{b=1}^m$ are those minimizing the error in the discrimination of $\varrho_{y, b}$, $\forall b$.

Before closing this section some remarks are in order. First, it should be noted that while the iterative procedure described

¹We relax here the RAC condition $b = x_y$ to highlight the fact that the argument holds as it is for more general functionals.

above will always converge in value to some maximum of \mathcal{F} , there is no guarantee this is a global maximum. This will highly depend on the starting point, which is chosen at random. Second, it is worth noting that the see-saw procedure described above can be also implemented to find the optimal performance over classical strategies: Since advantage of quantum strategies is rooted in the possibility of performing incompatible measurements, restricting all measurement operators to be diagonal in the computational basis will reduce the optimization procedure to a maximization over classical strategies. This restriction is easily imposed by just initializing the algorithm with random diagonal matrices as seeds for measurements, which ensures that all the states and measurements arising during the see-saw procedure will be diagonal in the same basis. Note, however, that, unlike the exhaustive search previously discussed, this is a completely heuristic method.

As a final observation, note that in the first step of the see-saw procedure, when the decoding strategy is fixed, the optimal value is given by

$$\max_{\{\rho_x\}} \mathcal{F} = \sum_{\mathbf{x}} \lambda_{\max} \left(\sum_{y,b} \alpha_{xyb} M_y^b \right), \quad (12)$$

where $\lambda_{\max}(\mathcal{O})$ denotes the largest eigenvalue of operator \mathcal{O} . That is, finding the optimal preparations $\{\rho_x\}$ for a fixed set of measurements is an eigenvalue problem, which can be solved analytically, and therefore we are left to look only for the optimal measurements. In spite of this simplification, finding the quantum value remains a hard problem in general. Nonetheless, as we will see later, there are cases in which it can be approached analytically.

III. THE RAC-TOOLS PYTHON PACKAGE

In this section we introduce a Python package [49] that implements the numerical methods described in the previous section. Our goal was to construct a tool that allows the user to easily determine basic properties of a b -RAC, such as its classical and quantum value. This tool requires some standard Python packages like *numpy* and *scipy*, as well as the *cvxpy* package to solve SDPs. Among the different solvers that can be used with *cvxpy*, we have found *MOSEK* [50,51] to be the most reliable. It is also available at no cost for academic use. Throughout this section, we provide a succinct description of how the package works, which is further elaborated in Appendix B.

The RAC-tools package is written to implement both the exhaustive search and see-saw algorithms, which were discussed in Sec. II. For the exhaustive search method, users can invoke the `perform_search` function, while the `perform_seesaw` function is employed for the see-saw optimization. Since our main focus is on biased RACs, an essential part of the package deals with the specification of the biasing tensor α_{xyb} , as defined by the functional in Eq. (2). Note that using this definition instead of the one in (3) allows the user to define functionals in a class larger than that of RAC functionals, which corresponds to the condition $b = x_y$. The desired biasing tensor can be written explicitly and passed to both `perform_search` and `perform_seesaw` in the form of

a Python dictionary. Alternatively, the user can opt for any of the built-in biasing tensors provided by the package via the `generate_bias` function and input only a reduced set of parameters.

Before entering into the more technical details of `perform_search` and `perform_seesaw`, let us provide an example to better illustrate how `generate_bias` works. Consider a $2^2 \mapsto 1$ RAC in which the Alice's input \mathbf{x} is uniformly distributed, but Bob is asked to retrieve the first character of \mathbf{x} with probability $w \in [0, 1]$. The bias tensor defining this b -RAC is given by

$$\alpha_{xy} = \begin{cases} \frac{1}{4}w & \text{if } y = 0, \\ \frac{1}{4}(1-w) & \text{otherwise,} \end{cases} \quad (13)$$

where the factor $\frac{1}{4}$ results from \mathbf{x} being uniformly distributed. We can compute its classical and quantum value by passing to `perform_search` and `perform_seesaw`, respectively, the string `bias='Y_ONE'` and the float `weight=w`. While the variable `weight` encodes the amount of bias desired, the variable `bias` encodes the type of bias that the user wants to compute. For example, if the user passes `weight=0.75` as an argument, `generate_bias` builds a biasing tensor with components

$$\alpha_{xy} = \begin{cases} \frac{3}{16} & \text{if } y = 0, \\ \frac{1}{16} & \text{otherwise.} \end{cases} \quad (14)$$

In a similar manner to that demonstrated in the above example, we can also consider biases that exclusively affect Alice's input \mathbf{x} , or alternatively affect both \mathbf{x} and \mathbf{y} simultaneously. RAC-tools includes several built-in bias families, including the one shown above, which we describe in more detail in Appendix B 1. Moving forward, we proceed now to the description of the operational details of the functions `perform_search` and `perform_seesaw`.

A. The `perform_search` function

The main use of the function `perform_search` is to compute the optimal classical performance of an $n^m \mapsto 1$ b -RAC. This function takes as argument the integers n , d , and m , encoded by the analogous Python parameters `n`, `d`, and `m`, as well as the bias tensor α_{xyb} . The latter can be entered via the Python dictionary `bias_tensor` or as the aforementioned reduced set of parameters `bias` and `weight`, for one of the built-in bias families. While these parameters fix a particular b -RAC scenario, the variable `method` defines the searching approach that should be employed by `perform_search`. When setting `method=1` and `method=2`, the corresponding implementations correspond to the approaches described in statements (a) and (b) of Lemma 1, respectively. On the other hand, `method=0` implements a purely exhaustive search, where neither the encoding nor decoding functions are fixed. It is worth noting that the value of m is set by default to be the same as that of d , and therefore there is no need to declare it when studying $n^d \mapsto 1$ b -RACs.

When `perform_search` finishes the execution, it generates a report like the one in Fig. 2, for the case of the $2^2 \mapsto 1$ RAC. As can be seen in the figure, the report provides not only the optimal value of the functional but also an encoding-

```

> perform_search(n=2, d=2, method=0)

=====
                        RAC-tools v1.0
=====

----- Summary of computation -----

Total time of computation: 0.001859 s
Total number of encoding/decoding functions: 256
Average time per function: 7e-06 s

----- Analysis of the optimal realization -----

Computation of the classical value for the  $2^2 \mapsto 1$  RAC:
0.75
Number of functions achieving the computed value: 24

First functions found achieving the computed value
Encoding:
E: [0, 0, 0, 1]
Decoding:
D0: [0, 1]
D1: [0, 1]

----- End of computation -----

```

FIG. 2. Report produced by the function `perform_search` for the unbiased $2^2 \mapsto 1$ RAC. In addition to the summary of the computation, the user is provided with the first strategy found attaining the optimal value and the number of such equivalent strategies. For the encoding function $E(\mathbf{x})$, the result is shown in a tuple organized in ascending order of \mathbf{x} , $[E(00\dots 0), E(00\dots 1), \dots, E((m-1)\dots(m-1))]$. The decoding functions $D_y(\mu)$ follow a similar pattern: each row corresponds to a distinct input y , and the result is organized in ascending order of μ .

decoding pair attaining this value, along with information about computing time and total number of encoding-decoding functions. This report can be disabled by setting the variable `verbose=False`, in which case `perform_search` still returns the information displayed in the report, but in the form of a dictionary, allowing the user to manipulate this information. A more detailed description of this function is given in Appendix B 2.

B. The `perform_seesaw` function

This function implements the see-saw algorithm as described in Sec. II. As mentioned before, the algorithm always converges to a maximum of the functional described in Eq. (7). However, it is not guaranteed that this maximum represents the global maximum due to the random initialization of the algorithm. Hence, in addition to providing the integers n , d , and m , along with the tensor α_{xyb} that specifies the scenario, the user is required to enter the number of random initializations through the parameter `seeds` when invoking the function. It should be noted that since increasing the number of initializations increases the computation time, deciding which is the best value for `seeds` is a problem on its own; In Appendix B 3 we provide, as a guide, Table II, which contains the number of seeds used for generating the numerical results presented in this work.

In addition, as noted in the description of the see-saw algorithm, it can also be used to compute the classical value

of a b -RAC by restricting the measurements and preparations to be diagonal in the computational basis. This condition can be passed to the `perform_seesaw` function via the extra variable `diagonal`. If `diagonal=True`, the function initializes the see-saw algorithm with random diagonal measurements and the retrieved value corresponds to an estimation of the classical value. The default value of this variable is `False`.

After finishing the computation, `perform_seesaw` prints a report including a short analysis of the measurement operators attaining the optimal value found, such as whether the operators are projective or if they are mutually unbiased. Moreover, the report informs also about the computation time and the number of random starting points used by the code. A detailed description about this data can be found in Appendix B 3. Figure 3 shows an example of the report printed by the function in the case of the $2^2 \mapsto 1$ RAC. As before, this report can be disabled by setting the variable `verbose=False`, in which case the information displayed in it is returned in the form of a Python dictionary.

IV. ANALYTICAL RESULTS FOR THE $n^2 \mapsto 1$ RAC

As discussed in Sec. II, whether classical or quantum the strategies maximizing the b -RAC functional are in general hard to find analytically. An exception is provided by some RACs whose output is a single bit since, as we will see below, this greatly simplifies the solution of the corresponding optimization problems. Since we are interested in the advantage provided by quantum strategies, let us start by discussing the optimal classical performance.

A. Classical

We are now interested in finding the optimal encoding-decoding strategies for general biased $n^2 \mapsto 1$ RAC. The case of unbiased RACs has been already studied in Ref. [30], and as we show below, the authors' analysis can be extended to the general case with slight modifications.

A salient feature of the optimal classical encoding-decoding strategies for b -RACs is that in some cases they ignore part of the input. Excluding part of the input in the search for the optimal encoding and decoding strategies reduces the complexity of the problem, since it involves fewer bits; thus the optimal value is easier to compute if the subset of bits to be ignored is known beforehand. Unfortunately this knowledge does not seem to be available in advance, and in order to find the set of bits ignored by the best strategy we need to compare the values for all possible options, which makes the evaluation computationally hard. We will return to this issue when discussing optimal quantum strategies.

The following lemma shows how to find the optimal classical strategies under the assumption that no bit is ignored. If the actual optimal strategy for a given b -RAC does ignore part of the input string \mathbf{x} , the result applies to the set of bits taken into account by the strategy.

Lemma 2. The optimal non-bit-ignoring strategies for the $n^2 \mapsto 1$ b -RAC comprise a weighted majority encoding function and identity map for decoding.

Proof. We already know from Lemma 1 how to find the optimal decoding function for a given fixed encoding, and

```

> perform_seesaw(n=2, d=2, seeds=5)

=====
                        RAC-tools v1.0
=====

----- Summary of computation -----

Number of random seeds: 5
Average time for each seed: 0.14852 s
Average number of iterations: 3
Seeds 1e-13 close to the best value: 5

----- Analysis of the optimal realization for seed #1 -----

Estimation of the quantum value for the 22→1 RAC:
0.853553390593

Measurement operator ranks
M[0] ranks: 1 1
M[1] ranks: 1 1

Measurement operator projectiveness
M[0, 0]: Projective          6.44e-15
M[0, 1]: Projective          6.44e-15
M[1, 0]: Projective          6.78e-15
M[1, 1]: Projective          6.78e-15

Mutual unbiasedness of measurements
M[0] and M[1]: MUB          5.91e-14

----- End of computation -----

```

FIG. 3. Report produced by the function `perform_seesaw` for the unbiased $2^2 \mapsto 1$ RAC. In the first part of the report, the function produces a small summary of the computation, displaying information such as the number of random starting points, average time per starting point, etc. The subsequent part presents the optimal value found along with a short analysis of the measurements attaining this value. The notation $M[y]$ refers to the y th measurement, while $M[y, b]$ refers to the operator yielding output b . In this way the item *Measurement operator ranks* presents the computed ranks for the operators of the y th measurement, arranged in ascending order of b . Similarly, the item *Measurement operator projectiveness* indicates whether the identified operators can be considered projective or not. The numerical value presented in the second column serves as a measure for projectiveness, with a value approaching zero indicating that this operator is close of being projective. Last, the item *Mutual unbiasedness of measurements* analyzes the possibility of constructing each pair of measurements out of mutually unbiased bases. Analogous to the previous item, the second column provides a measure of proximity for a given pair. Further technical details regarding the `perform_seesaw` function can be found in Appendix B.

the optimal encoding function for a fixed decoding. In the particular case of the $n^2 \mapsto 1$ b -RAC, given a fixed encoding $E : \{0, 1\}^{\times n} \mapsto 0, 1$ mapping input \mathbf{x} into a bit μ , i.e., $E(\mathbf{x}) = \mu$, it implies that the optimal decoding function for the y th must satisfy

$$D_y(\mu) = \begin{cases} 0 & \text{if } \sum_{\mathbf{x}} \alpha_{\mathbf{x}|y} \delta_{x_y, 0} \geq \sum_{\mathbf{x}} \alpha_{\mathbf{x}|y} \delta_{x_y, 1}, \\ 1 & \text{otherwise,} \end{cases} \quad (15)$$

where $\alpha_{\mathbf{x}|y} = \alpha_{\mathbf{x}y}/r_y$, with $r_y = \sum_{\mathbf{x}} \alpha_{\mathbf{x}y}$, can be interpreted as the probability of string \mathbf{x} being Alice's input given that the y th bit is to be recovered by Bob.

Because both $D_y(\mu)$ and its argument μ are bits there are only four possible decoding functions: two constant maps $D_y(\mu) = 0, 1$, the identity map $D_y(\mu) = \mu$, and a flip of the input $D_y(\mu) = 1 - \mu$. Now suppose the optimal decoding in Eq. (15) corresponds to one of the constant functions, e.g., $D_y(\mu) \equiv 0$. This implies $\sum_{\mathbf{x}} \alpha_{\mathbf{x}|y} \delta_{x_y, 0} \geq \sum_{\mathbf{x}} \alpha_{\mathbf{x}|y} \delta_{x_y, 1}$ for both $\mu = 0, 1$. Because $\sum_{\mathbf{x}} \alpha_{\mathbf{x}|y} \delta_{x_y, 0} + \sum_{\mathbf{x}} \alpha_{\mathbf{x}|y} \delta_{x_y, 1} = 1$, it follows from the previous relation that $\sum_{\mathbf{x}} \alpha_{\mathbf{x}|y} \delta_{x_y, 1} \leq \frac{1}{2}$, meaning that $x_y = 0$ is a more probable event in the inputs than $x_y = 1$. We can interpret this result as “ignoring the encoding” being the best decoding strategy Bob can implement for that particular bit, in which case it makes no sense for Alice to consider it in the encoding to begin. On the other hand, it is not hard to see that if an encoding strategy ignores the y th bit, the ensuing optimal decoding is a constant function mapping the input to a constant value, which is the most frequent for x_y . By identifying constant decoding functions with parts of the input \mathbf{x} that are ignored by the optimal strategy, we are left with only two possible decoding maps for those bits that are taken into account. This two maps are actually equivalent since it is easy to check that if $D_y(\mu) = 1 - \mu$ is optimal for the encoding E , then $D_y(\mu) = \mu$ is optimal for the encoding $E' = E \circ \neg_y$, where $\neg_y : S \mapsto S$ is the function flipping the y th bit of a given string in S . It follows then that the optimal decoding strategy can always be chosen to be the identity map.

Now we can move on to discuss the optimal encoding function. We have seen that whenever D_y is a constant function the optimal encoding ignores the y th bit, so we can focus here on the case where the identity map is the optimal decoding function. In that case, it follows from the second statement in Lemma 1 that the optimal encoding strategy should satisfy

$$\mathcal{F} = \sum_{\mu, y, \mathbf{x} \in S_\mu} \alpha_{\mathbf{x}y} \delta_{x_y, \mu} = \sum_{\mu, \mathbf{x} \in S_\mu} \alpha_{\mathbf{x}} \sum_y r_{y|\mathbf{x}} \delta_{x_y, \mu}, \quad (16)$$

where $r_{y|\mathbf{x}} := \alpha_{\mathbf{x}y}/\alpha_{\mathbf{x}}$, with $\alpha_{\mathbf{x}} := \sum_y \alpha_{\mathbf{x}y}$. We can think of this condition as determining which value (0 or 1) has the greater weight in the string \mathbf{x} , that is, the optimal strategy corresponds to a *weighted majority* encoding. ■

Note that both Eqs. (15) and (16) are easily obtained by application of statements (a) and (b) of Lemma 1, which may raise the question of why this solution does not extend straightforwardly to the general case, where Bob can output more than two possible values. The key feature of the argument above is, as we have shown, that the reduced number of possible decoding functions allows us to state that for every bit there are only two options: either the bit is ignored by the strategy, or it is decoded using the identity map. Whenever $d > 2$ or $m > 2$ this is no longer true, as there exist decoding functions which are neither constant nor permutations.

B. Quantum

Having found a procedure to determine the optimal classical strategies, we can now move on to explore the optimal quantum strategies and the cases in which these can provide an advantage. As explained at the end of Sec. II A, for a fixed decoding strategy $\{M_y^b\}$ the optimal encoding of input \mathbf{x} is determined by the largest eigenvalue of the operators $\sum_{\mathbf{x}y} \alpha_{\mathbf{x}y} M_y^{x_y}$. Although simplified, this problem is still hard,

since search of the optimal measurements $\{M_y^b\}$ is to be carried over the set of all possible measurement operators. The following lemma, which applies to any RAC with two outcomes, greatly simplifies this search, by restricting it to the subclass of projective measurements.

Lemma 3. The value \mathcal{F}_Q of the $n^2 \mapsto 1$ b -RAC can always be reached by a decoding strategy consisting only of projective measurements.

Proof. Let $\{M_k^0, M_k^1\}$ be the measurement operators over the Hilbert space \mathcal{H} describing Bob's decoding map for the k th bit. It follows from the completeness relation $M_k^0 + M_k^1 = \mathbb{1}$ that the objective function in SDP (11) can be written as

$$\begin{aligned} \mathcal{F}_k &= \text{tr}[\varrho_{k,1} + (\varrho_{k,0} - \varrho_{k,1})M_k^0] \\ &= \text{tr}[\varrho_{k,0} - (\varrho_{k,0} - \varrho_{k,1})M_k^1] \\ &= \frac{1}{2} + \frac{1}{2}\text{tr}(\varrho_{k,0} - \varrho_{k,1})M_k, \end{aligned} \quad (17)$$

where in the last line we have taken the average of the expressions in the first two and have written $M_k = M_k^0 - M_k^1$. Hermiticity of $(\varrho_{k,0} - \varrho_{k,1})$ implies that we can find orthogonal subspaces \mathcal{H}_+ and \mathcal{H}_- , such that $\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-$, spanned by its eigenvectors associated with non-negative and negative eigenvalues, respectively. If X_{\pm} denotes the projectors onto these subspaces, it is then apparent that the optimal value for \mathcal{F}_k is attained for $M_k = X_+ - X_-$, from which it follows that M_k^0 and M_k^1 can be chosen to be projectors. ■

It is worth noting that either M_k^0 or M_k^1 in the proof above could equal the identity operator. In such a case one of the measurement operators would be a projector over the entire Hilbert space \mathcal{H} , which corresponds to a decoding strategy in which Bob always guesses 0 (or 1) for the k th bit regardless of Alice's encoding, i.e., a constant decoding map. As already seen in the discussion of optimal classical strategies, if a constant decoding function is optimal for a given bit, we can find an optimal encoding that ignores that bit. Indeed, assume without loss of generality that the optimal k th decoding strategy requires $M_k^0 = \mathbb{1}$. Because the optimal preparations $\rho_{\mathbf{x}}$ are eigenstates associated with the largest eigenvalue of the operators $\sum_y \alpha_{xy} M_y^{x_y}$, which we now can write as $(1 - x_k)\alpha_{xk} + \lambda_{\max}(\sum_{y \neq k} \alpha_{xy} M_y^{x_y})$, the quantum value of the b -RAC value can be expressed as

$$\begin{aligned} \mathcal{F}_Q &= f_k + \sum_{\mathbf{x}} \lambda_{\max} \left(\sum_{y \neq k} \alpha_{xy} M_y^{x_y} \right), \\ f_k &= \sum_{\mathbf{x}} (1 - x_k) \alpha_{xk}, \end{aligned} \quad (18)$$

where f_k is the contribution to the value of the trivial decoding of the k th bit, and the second term in \mathcal{F}_Q is the quantum value of b -RAC with input strings of $n - 1$ bits. It follows then that the optimal preparations are eigenstates of the operators $\sum_{y \neq k} \alpha_{xy} M_y^{x_y}$, meaning that the encoding strategy ignores the k th bit as stated above.

The argument above is easily generalized to the case in which the optimal strategy ignores any number of bits. It should be noted here, as we did in discussing classical strategies in Lemma 2, that the knowledge of the bits that should be ignored by the optimal strategy makes easier the evaluation of

\mathcal{F}_Q . However, such information does not seem to be available in advance and can be obtained only by comparing the values of all possible bit-ignoring strategies, making the evaluation of \mathcal{F}_Q computationally hard. If we denote by s a given subset of $I_n = \{0, 1, \dots, n - 1\}$, we can write the value associated with an encoding-decoding strategy ignoring the bits in s as

$$\mathcal{F}_Q^s = \sum_{k \in s} f_k + F_Q^s, \quad (19)$$

where f_k is again given by Eq. (18) and

$$F_Q^s = \max_{\{M_y^{x_y}\}} \sum_{\mathbf{x}} \lambda_{\max} \left(\sum_{y \notin s} \alpha_{xy} M_y^{x_y} \right) \quad (20)$$

is the contribution from the bits that are not ignored by the encoding-decoding strategy. With this notation we can formally write the quantum value as

$$\mathcal{F}_Q = \max_s \mathcal{F}_Q^s. \quad (21)$$

It follows from this discussion that finding the b -RAC optimal value over quantum strategies reduces to solving, for all $s \subset I_n$, the optimization problem in the second term in Eq. (19), i.e., finding the optimal strategies involving all bits of the input strings. In what follows, we explore the cases where an analytical solution to this problem is available.

1. Qubit strategies

Let us now consider $d = m = 2$ and focus on optimal strategies involving all the bits in the input string \mathbf{x} . The decoding strategies in this scenario involve two-outcome measurements on qubits, which we can always write as convex combinations of rank-1 projective and trivial measurements, i.e., $M_y^{x_y} \in \{0, \mathbb{1}\}$. Since trivial measurements are associated with bit-dropping strategies, which are assumed here to be suboptimal, it follows that the quantum value can be attained only with rank-1 projective measurements. By restricting the decoding strategies to rank-1 projective measurements we can express the b -RAC value as

$$\begin{aligned} F &= \sum_{\mathbf{x}} \text{tr} \rho_{\mathbf{x}} \left(\sum_y \alpha_{xy} M_y^{x_y} \right) \\ &= \frac{1}{2} + \sum_{\mathbf{x}} \text{tr} \rho_{\mathbf{x}} \left(\sum_y \alpha_{xy} (-1)^{x_y} \mathbf{m}_y \cdot \boldsymbol{\sigma} \right) \\ &\leq \frac{1}{2} + \sum_{\mathbf{x}} \left| \sum_y \alpha_{xy} (-1)^{x_y} \mathbf{m}_y \right|, \end{aligned} \quad (22)$$

where in the second line we have expanded the projectors $M_y^{x_y}$ in the Pauli basis, $M_y^{x_y} = \frac{1}{2} \mathbb{1} + (-1)^{x_y} \mathbf{m}_y \cdot \boldsymbol{\sigma}$, $\mathbf{m}_y \in \mathbb{R}^3$, $|\mathbf{m}_y| = \frac{1}{2}$, and in the third line we used that the value of the trace is upper bounded by the largest eigenvalue of the traceless operator in the argument, an upper bound that is attained when $\rho_{\mathbf{x}}$ is an eigenstate associated with this eigenvalue. It follows from this result that the quantum value of the

$n^2 \mapsto 1$ b -RAC is given by

$$F_Q = \frac{1}{2} + \max_{\{\mathbf{m}_y\}} \sum_{\mathbf{x}} \alpha_{\mathbf{x}} \left| \sum_y r_{y|\mathbf{x}} (-1)^{x_y} \mathbf{m}_y \right|, \quad (23)$$

with $\alpha_{\mathbf{x}} = \sum_y \alpha_{\mathbf{x}y}$ and $r_{y|\mathbf{x}} = \alpha_{\mathbf{x}y}/\alpha_{\mathbf{x}}$. Note that, as suggested in Ref. [30] for the case of unbiased RACs, the value F_Q can be thought of as the (weighted) average distance traveled by a random walker in \mathbb{R}^3 (up to some scaling and shift). Moreover, it can be checked by direct calculation that if the vectors \mathbf{m}_y are constrained to be parallel, F_Q reduces to the optimal RAC value over classical strategies.

While the optimization problem in Eq. (22) is in general hard to solve for arbitrary bias tensors, it greatly simplifies if we restrict ourselves to the subclass of *factorizable* biases. Consider a bias such that $\alpha_{\mathbf{x}y} = \alpha_{\mathbf{x}} r_y$, where $\sum_y r_y = \sum_{\mathbf{x}} \alpha_{\mathbf{x}} = 1$ because of normalization. This means that the inputs \mathbf{x} and y are independent random variables. It is easy to see in this case that, in the sum over input strings in Eq. (23), the term associated with the string \mathbf{x} has the same value as that associated with string $\bar{\mathbf{x}}$ if the latter can be obtained from \mathbf{x} by flipping all of its bits. Taking this into account we can rewrite F_Q as the sum over only half of the input strings

$$\begin{aligned} F_Q &= \frac{1}{2} + \max_{\{\mathbf{m}_y\}} \sum_{\mathbf{x}} p_{\mathbf{x}} \left| \sum_y r_y (-1)^{x_y} \mathbf{m}_y \right| \\ &= \frac{1}{2} + \max_{\{G\}} \sum_{\mathbf{x}} p_{\mathbf{x}} \sqrt{\text{tr} G \mathbf{v}_{\mathbf{x}} \mathbf{v}_{\mathbf{x}}^T}, \end{aligned} \quad (24)$$

with $p_{\mathbf{x}} = \alpha_{\mathbf{x}} + \alpha_{\bar{\mathbf{x}}}$, $\mathbf{v}_{\mathbf{x}}$ an n -dimensional tuple with components $(\mathbf{v}_{\mathbf{x}})_y = (-1)^{x_y} r_y$ and G the Gram matrix of the measurement vectors \mathbf{m}_y , i.e., $G_{ij} = \langle \mathbf{m}_i, \mathbf{m}_j \rangle$. We can now think of the sum over \mathbf{x} in Eq. (24) as a scalar product between two 2^{n-1} -dimensional tuples, one with components $p_{\mathbf{x}}$ and the other with components $\sqrt{\text{tr} G \mathbf{v}_{\mathbf{x}} \mathbf{v}_{\mathbf{x}}^T}$. Using the Cauchy-Schwarz inequality we can upper bound F_Q by

$$\begin{aligned} F_Q &\leq \frac{1}{2} + \sqrt{\sum_{\mathbf{x}} p_{\mathbf{x}}^2} \max_G \sqrt{\text{tr} G \sum_{\mathbf{x}} \mathbf{v}_{\mathbf{x}} \mathbf{v}_{\mathbf{x}}^T} \\ &= \frac{1}{2} + \sqrt{2^{n-3}} \sqrt{\sum_{\mathbf{x}} p_{\mathbf{x}}^2} \sqrt{\sum_y r_y^2}, \end{aligned} \quad (25)$$

where in the last line we used that $(\sum_{\mathbf{x}} \mathbf{v}_{\mathbf{x}} \mathbf{v}_{\mathbf{x}}^T)_{ij} = 2^{n-1} r_i^2 \delta_{ij}$, as can be checked by direct calculation. It is easy to check that this expression reduces, for $\alpha_{\mathbf{x}y} = \frac{1}{n2^n}$, to the upper bound $F_Q \leq \frac{1}{2} + (2\sqrt{n})^{-1}$ previously derived in Ref. [30]. It is also worth remarking that the upper bound we just derived depends on $p_{\mathbf{x}}$ rather than directly on $\alpha_{\mathbf{x}}$, and this feature holds for the value associated with any quantum or classical strategy, since it is a consequence of the independence of inputs \mathbf{x} and y .

The bound in Eq. (25) was obtained via the Cauchy-Schwarz inequality between two tuples with components $\alpha_{\mathbf{x}}$ and $\sqrt{\text{tr} G \mathbf{v}_{\mathbf{x}} \mathbf{v}_{\mathbf{x}}^T}$, respectively, which are all real and non-negative. Thus, in order to saturate the bound following 2^{n-1} conditions (one per input string) must be satisfied

$$\frac{p_{\mathbf{x}}}{\sqrt{\text{tr} G \mathbf{v}_{\mathbf{x}} \mathbf{v}_{\mathbf{x}}^T}} = \frac{1}{\sqrt{2^{n-3}}} \sqrt{\frac{\sum_{\mathbf{x}'} p_{\mathbf{x}'}^2}{\sum_y r_y^2}}, \quad (26)$$

where as before we have written $p_{\mathbf{x}} = \alpha_{\mathbf{x}} + \alpha_{\bar{\mathbf{x}}}$. Note that the quantity on the right-hand side is a constant characterizing the particular b -RAC under study. Since projectivity of measurements implies the norms of the vectors \mathbf{m}_y is maximal, $|\mathbf{m}_y| = \frac{1}{2}$, we can use that $\text{tr} G \mathbf{v}_{\mathbf{x}} \mathbf{v}_{\mathbf{x}}^T = \frac{1}{4} \sum_{i,j} (-1)^{x_i+x_j} r_i r_j \cos(\theta_{ij})$ to rewrite Eq. (26) as a condition to be satisfied by the angles θ_{ij} between vectors,

$$\sum_{i < j} (-1)^{x_i+x_j} r_i r_j \cos(\theta_{ij}) = \frac{1}{2} \left(\sum_y r_y^2 \right) \left(\frac{2^{n-1} p_{\mathbf{x}}^2}{\sum_{\mathbf{x}'} p_{\mathbf{x}'}^2} - 1 \right). \quad (27)$$

Now fix a pair of indices (i, j) , with $j \neq i$, and define S_{ij} as the subset of input strings satisfying $x_i = x_j$. Then summing over the strings in S_{ij} in Eq. (27) we arrive at

$$\cos(\theta_{ij}) = \frac{1}{2r_i r_j} \left(\frac{\sum_y r_y^2}{\sum_{\mathbf{x}'} p_{\mathbf{x}'}^2} \right) \left(\sum_{\mathbf{x} \in S_{ij}} p_{\mathbf{x}}^2 - \sum_{\mathbf{x} \notin S_{ij}} p_{\mathbf{x}}^2 \right). \quad (28)$$

If instead of the vectors $\{\mathbf{m}_y\}$ we parametrize the set of optimal measurements by the cosine of the angles between them, then Eq. (28) shows how to construct the optimal decoding strategy from the biasing parameters when the upper bound (25) is attained. On the other hand, we can use the relation to quickly discard the possibility of the bound being attained: If for a given b -RAC the cosines generated by Eq. (28) are inconsistent, then the upper bound cannot be attained. Consistency here turns out to be equivalent to (1) $\cos(\theta_{ij}) \in [-1, 1]$ and (2) the matrix with elements

$$\tilde{G}_{ij} = \begin{cases} \frac{1}{4} & \text{if } i = j \\ \frac{1}{4} \cos(\theta_{ij}) & \text{if } i \neq j \end{cases} \quad (29)$$

being positive semidefinite.

It turns out that in the $2^2 \mapsto 1$ scenario whenever quantum strategies can provide some advantage over their classical counterpart, this upper bound is actually attainable, as shown in the following lemma.

Lemma 4. The quantum value of a $2^2 \mapsto 1$ b -RAC with biasing strategy $\alpha_{\mathbf{x}y} = \alpha_{\mathbf{x}} r_y$ is

$$F_Q = \max \left\{ F_C, \frac{1}{2} + \frac{1}{\sqrt{2}} \sqrt{\sum_{\mathbf{x}} p_{\mathbf{x}}^2} \sqrt{\sum_y r_y^2} \right\}, \quad (30)$$

where $F_C = \frac{1}{2} + \max\{\frac{1}{2}p_{00} + \frac{1}{2}p_{01}(r_0 - r_1), \frac{1}{2}p_{00}(r_0 - r_1) + \frac{1}{2}p_{01}\}$, with $p_{\mathbf{x}} = \alpha_{\mathbf{x}} + \alpha_{\bar{\mathbf{x}}}$, is the optimal performance over classical strategies.

Proof. Since $n = 2$ the sum over y in Eq. (24) has only two terms. This means, since vectors \mathbf{m}_y have maximal norm, that there is only one parameter to optimize over, which is the angle θ between \mathbf{m}_0 and \mathbf{m}_1 . Looking for critical values, we find that the b -RAC functional can attain a maximal value only for θ satisfying $\sin(\theta) = 0$ or

$$\cos(\theta) = \left(\frac{r_0^2 + r_1^2}{2r_0 r_1} \right) \left(\frac{p_{00}^2 - p_{01}^2}{p_{00}^2 + p_{01}^2} \right), \quad (31)$$

which is just the condition in Eq. (28), implying that F_Q is then given by Eq. (25). On the other hand, if θ equals 0 or π , then F_Q reads $\frac{1}{2} + \frac{1}{2}p_{00} + \frac{1}{2}p_{01}(r_0 - r_1)$ and $\frac{1}{2} + \frac{1}{2}p_{00}(r_0 - r_1)$,

respectively. These two quantities can be checked to correspond to the two options present in the expression for the classical value. ■

Lemma 4 shows not only that the upper bound given in Eq. (25) can be attained, but that it will be attained with any bias for which quantum strategies provide any advantage over classical strategies, thus providing a complete solution to the $2^2 \mapsto 1$ *b*-RAC value problem. Moreover, it follows from the proof that attaining the upper bound in Eq. (25) self-tests the angle θ between the Bloch vectors defining the optimal measurements. Unfortunately such a simple result does not hold for a larger number of bits, with solutions becoming more complex already in the case of 3-bit input strings. For this particular case, nonetheless, the upper bound in Eq. (25) still can be attained for many biasing tensors, as we show in the following lemma, which extends the proof of Lemma 4.

Lemma 5. The quantum value F_Q of a $3^2 \mapsto 1$ *b*-RAC is given by

$$F_Q = \frac{1}{2} + \sqrt{\sum_{\mathbf{x}} p_{\mathbf{x}}^2} \sqrt{\sum_y r_y^2}, \quad (32)$$

whenever there exists an optimal decoding strategy for which the Bloch vectors $\{\mathbf{m}_y\}$ are linearly independent.

Proof. If the vectors are linearly independent, then the ensuing Gram matrix G is full rank. As we did in Eq. (24), we can write the optimal performance in terms of the elements of $G_{i,j} = \langle \mathbf{m}_i, \mathbf{m}_j \rangle = \frac{1}{4} \cos(\theta_{ij})$ and in doing so parametrize F_Q by the angles θ_{ij} , which are all independent. As a consequence, a critical point will satisfy $\partial_{\theta_{ij}} F_Q = \partial_{G_{ij}} F_Q \sin \theta_{ij} = 0$, which implies $\partial_{G_{ij}} F_Q = 0$ because of the independence of the measurement operators. A direct calculation shows then that a critical point $\{\theta_{ij}\}$ should be such that the following relation is satisfied $\forall i, j$:

$$\sum_{\mathbf{x}} \frac{p_{\mathbf{x}}}{\sqrt{\text{tr} G \mathbf{v}_{\mathbf{x}} \mathbf{v}_{\mathbf{x}}^T}} (-1)^{x_i+x_j} = 0. \quad (33)$$

This equation can be rewritten as a matrix equation $A \cdot b = 0$ with

$$b_{\mathbf{x}} = \frac{p_{\mathbf{x}}}{\sqrt{\text{tr} G \mathbf{v}_{\mathbf{x}} \mathbf{v}_{\mathbf{x}}^T}}, \quad (34)$$

and $A \in \mathbb{R}^{3 \times 4}$ a matrix with entries $A_{(i,j),\mathbf{x}} = (-1)^{x_i+x_j}$. It is straightforward to check that the null space of A is $\text{Null}(A) = \text{span}\{(1, 1, 1, 1)\}$, implying that in a critical point the quotient $\frac{p_{\mathbf{x}}}{\sqrt{\text{tr} G \mathbf{v}_{\mathbf{x}} \mathbf{v}_{\mathbf{x}}^T}}$ takes the same value for all strings \mathbf{x} , and that this value is $\sqrt{\frac{\sum_{\mathbf{x}} p_{\mathbf{x}}^2}{2^{n-3} \sum_y r_y^2}}$, thus proving that the only critical point in F_Q satisfies condition (26) and F_Q is given by the upper bound in Eq. (25). ■

The above lemma shows that the cosines built up from the biasing parameters via the condition in Eq. (28) will indeed provide the optimal measurements if, in addition to being consistent, the matrix \tilde{G} of Eq. (29) is full rank. Moreover, it is clear that attaining the upper bound in Eq. (25) self-tests the angles satisfying Eq. (28) since the optimal performance is achieved only by satisfying these relations. We can give a geometric interpretation to these conditions by noting that positive semidefiniteness of \tilde{G} is ensured if

its determinant is non-negative. We can write this last condition as $1 - \sum_{i<j} \cos^2(\theta_{ij}) + 2 \prod_{i<j} \cos(\theta_{ij}) \geq 0$, which is the equation of an “inflated tetrahedron” centered at the origin. Note that this origin corresponds to $\cos(\theta_{ij}) = 0 \forall i \neq j$, i.e., the angles associated with mutually unbiased measurements, which are in turn the optimal decoding strategy for the case of unbiased input strings \mathbf{x} , as can be easily checked in Eq. (28). Therefore, we have that the upper bound in Eq. (25) becomes the quantum value whenever the optimal measurements, as described by the three cosines $\{\cos(\theta_{ij}) = 4 \langle \mathbf{m}_i, \mathbf{m}_j \rangle\}$, live inside the inflated tetrahedron.

The geometrical picture introduced above turns out to be very helpful in understanding how different measurements become optimal as the bias tensor α_{xy} departs from $\alpha_{xy} = \frac{1}{32^3}$, which is the unbiased case. In what follows we study the solution when the bias tensor corresponds to one of the built-in biasing functions described in Sec. III. A similar analysis for other of these built-in biases can be found in Appendix C.

2. The X_ONE and Y_ALL bias family

We can have a better understanding about how different decoding strategies become optimal for different biases by analyzing a few examples. Consider the case of an input string bias of the form

$$\alpha_{\mathbf{x}} = \begin{cases} w & \text{if } \mathbf{x} = 000 \\ \frac{1-w}{7} & \text{otherwise} \end{cases}, \quad 0 \leq w \leq 1 \quad (35)$$

in combination with an arbitrary distribution $\{r_y\}$ for the requested bit. A numerical analysis suggests that for all members of the family the optimal quantum strategy involves all three bits in the input string, in which case the optimal performance is given by Eq. (23). In what follows then we will restrict our attention to those strategies.

The optimal decoding strategy for w in the vicinity of $\frac{1}{8}$ is expected to be determined by Eq. (28). We find, for a bias of this form, that the upper bound in Eq. (25) is attainable if the biasing coefficients are such that the three cosines

$$\cos[\theta_{ij}(w)] = \frac{\sum_y r_y^2}{2r_i r_j} h(w), \quad (36)$$

where

$$h(w) = \frac{32w^2 + 20w - 3}{48w^2 - 12w + 13} \quad (37)$$

is an increasing function for $w \in [0, 1]$, and $h(\frac{1}{8}) = 0$, are found to be consistent. Note that from Eq. (36) it follows that the optimal strategy $(\cos[\theta_{01}(w)], \cos[\theta_{02}(w)], \cos[\theta_{12}(w)]) \in \mathbb{R}^3$ continuously departs from the center of the tetrahedron as w increases, in a direction that is specified by both the sign of $h(w)$ and the fixed but otherwise arbitrary choice of the $\{r_y\}$ bias. In what follows we will focus on the case of $h(w) \geq 0$, i.e., $w \in [1/8, 1]$, since it is enough to understand the behavior of the solutions in the entire interval.

For uniform $\{r_y\}$ we see that the cosines in Eq. (36) are the same $\forall i, j$ and consistent for $\frac{1}{8} \leq w \leq 5/12$, increasing from 0 to 1 as $w \in [1/8, 1]$ increases. Within this region, quantum strategies have an advantage over the classical ones except for $w = 5/12$, in which case the two values coincide, $F_Q = F_C$. The optimal decoding strategy moves from the center of the

inflated tetrahedron described above to its $(1, 1, 1)$ vertex as w increases within $[\frac{1}{8}, \frac{5}{12}]$. For biasing parameter $w > \frac{5}{12}$ the cosines determined by Eq. (36) stop being consistent and the optimal strategy is expected to remain in the vertex of the tetrahedron, as follows from Lemma 5 and the symmetry of the solution. The optimal performance resulting from this analysis is shown in Fig. 4, in comparison with the numerical value obtained with the see-saw procedure described in Sec. II A.

Now let $\{r_y\}$ be biased such that $r_0 > r_1 = r_2$. In this situation the condition in Eq. (36) implies $\cos[\theta_{01}(w)] = \cos[\theta_{02}(w)] < \cos[\theta_{12}(w)]$. As in the previous case the point

$$\begin{aligned} F_Q(w > w_c) &= \frac{1}{2} + \frac{r_0}{2}(p_{010} + p_{001}) + \max_{\{\mathbf{m}_0, \mathbf{m}_1\}} \{p_{000}|r_0\mathbf{m}_0 + 2r_1\mathbf{m}_1| + p_{100}|r_0\mathbf{m}_0 - 2r_1\mathbf{m}_1|\} \\ &\leq \frac{1}{2} + \frac{r_0}{2}(p_{010} + p_{001}) + \frac{1}{\sqrt{2}}\sqrt{r_0^2 + 4r_1^2}\sqrt{p_{000}^2 + p_{100}^2} \end{aligned} \quad (38)$$

with the upper bound in the second line being attained for $\cos[\theta_{01}] = \cos[\theta_{02}] = c_{\text{opt}}$, where

$$c_{\text{opt}} = \left(\frac{r_0^2 + 4r_1^2}{4r_0r_1} \right) \left(\frac{p_{000}^2 - p_{100}^2}{p_{000}^2 + p_{100}^2} \right) \quad (39)$$

follows from Eq. (31). As already discussed, F_Q coincides with this upper bound when the cosines produced by Eq. (39) as long as $-1 < c_{\text{opt}} < 1$, becoming equal to the optimal classical value otherwise. The piecewise-defined optimal performance,

$$F_Q = \begin{cases} \frac{1}{2} + \sqrt{\sum_x p_x^2} \sqrt{\sum_y r_y^2} & \text{if } w \leq w_c \\ \frac{1}{2} + \frac{r_0}{2}(p_{010} + p_{001}) + \frac{1}{\sqrt{2}}\sqrt{r_0^2 + 4r_1^2}\sqrt{p_{000}^2 + p_{100}^2} & \text{if } w \geq w_c, \end{cases} \quad (40)$$

results from this analysis and is depicted in Fig. 5, together with the results obtained from the numerical package and the upper bound value in Eq. (25), for the case $r_0 = \frac{1}{2}$ and $r_1 = r_2 = \frac{1}{4}$.

For any other bias in $\{r_y\}$, the solution will have a similar behavior as a function of w , departing from the origin as the parameter increases but reaching the boundary of the inflated tetrahedron somewhere over one of the curved faces at a given critical value w_c of the biasing parameter. For $w > w_c$ the optimal measurements are no longer independent and the solution remains on the boundary, moving towards the vertex as w approaches 1 (in the limit $w \mapsto 1$ only the contribution from $\mathbf{x} = 000$ string is relevant, and therefore all the angles between measurements tend to zero). Finally, note that if we choose an analogous bias in which the weight w in Eq. (35) is assigned to a string different from 000, we will reach a similar conclusion except that some of the cosines in Eq. (36) might become negative, and for $w \mapsto 1$ the optimal solution might converge to a different vertex of the tetrahedron.

V. ANALYTICAL RESULTS FOR THE $2^d \mapsto 1$ RAC

After analyzing the quantum value of different b -RACs in the qubit setting, here we explore the strategies attaining the quantum value of the $2^d \mapsto 1$ b -RAC. In this scenario, if we consider a factorizable bias tensor $\alpha_{xy} = \alpha_x r_y$, and a quantum realization $\{\rho_{x_0x_1}, M_0^{x_0}, M_1^{x_1}\}$, the b -RAC value is given by

$$\mathcal{F} = \sum_{x_0x_1} \alpha_{x_0x_1} \text{tr}[\rho_{x_0x_1}(r_0M_0^{x_0} + r_1M_1^{x_1})], \quad (41)$$

in \mathbb{R}^3 representing the optimal strategy moves away from the coordinate origin as w increases, but now towards a point located on the edge of the tetrahedron connecting the $(1, 1, 1)$ and $(-1, -1, 1)$ vertices, and reaching it for a given value w_c of the biasing parameter satisfying $h(w_c) = \frac{2r_1^2}{r_0^2 + 2r_1^2}$. For larger values of w , F_Q is no longer given by Eq. (25), but numerics suggest that the optimal solution remains on the edge of the tetrahedron, i.e., $\cos[\theta_{12}(w)] = 1$ for $w > w_c$, moving towards the $(1, 1, 1)$ vertex as w increases. We can find the ensuing value F_Q by imposing $\mathbf{m}_1 = \mathbf{m}_2$ in the first line of Eq. (24), which leads to

with $x_i \in \{0, \dots, d-1\}$, and where $\rho_{x_0x_1}$ and $M_y^{x_i}$ are operators over a d -dimensional Hilbert space.

For the particular case of $\alpha_{x_0x_1} = \frac{1}{d^2}$ and $r_y = \frac{1}{2}$ it is known that the quantum value can be attained only with rank-1 projective measurement operators [34]. On the other hand, the results produced by the RAC-tools package for $2^d \mapsto 1$ b -RACs with $d \leq 6$ (see Appendix D) suggest that optimizing over projective measurements might already be enough to find the quantum value. In the following, we will then restrict ourselves to finding the optimal quantum strategies using projective measurements. As we will see in the following lemma, it is possible in this scenario to derive an upper bound analogous to Eq. (25) found for the $n^2 \mapsto 1$ b -RAC.

Lemma 6. The optimal value over projective measurements, \mathcal{F}_P , of the $2^d \mapsto 1$ b -RAC defined by the bias tensor $\alpha_{xy} = \alpha_x r_y$ satisfies

$$\mathcal{F}_P \leq \frac{1}{2} + \frac{1}{2}\sqrt{d^2 - 4d(d-1)r_0r_1} \sqrt{\sum_{x_0x_1} \alpha_{x_0x_1}^2}. \quad (42)$$

Proof. We begin by noting that $r_0M_0^{x_0} + r_1M_1^{x_1}$ is positive semidefinite, and as a consequence the value of the functional in Eq. (41) is upper bounded by

$$\mathcal{F}_P \leq \sum_{x_0x_1} \alpha_{x_0x_1} \lambda_{x_0x_1}, \quad (43)$$

where $\lambda_{x_0x_1}$ denotes the largest eigenvalue of $r_0M_0^{x_0} + r_1M_1^{x_1}$. Because the measurements are assumed to be projective, by Jordan's Lemma there is a basis in which the operators $M_0^{x_0}$

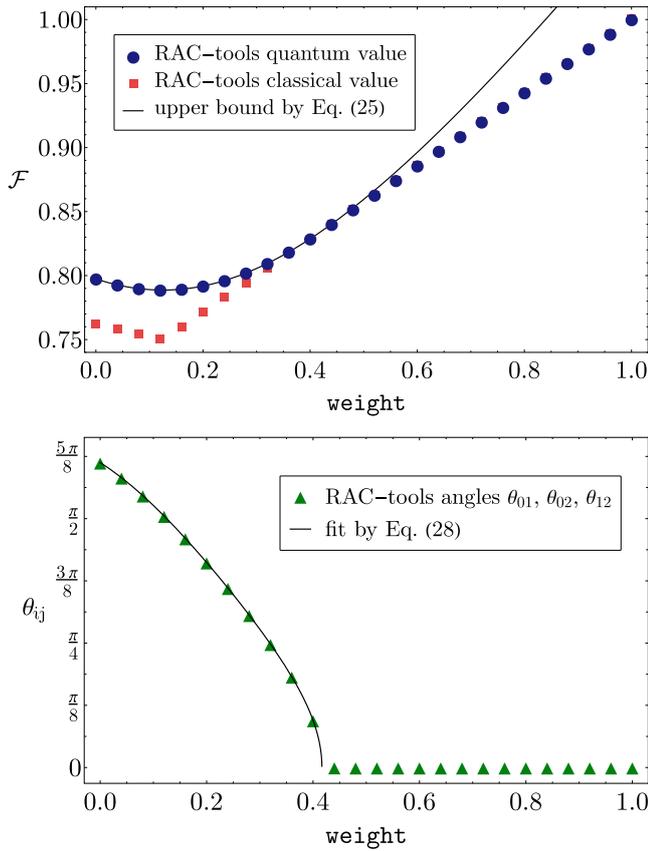


FIG. 4. Top: Optimal performance of the $3^2 \mapsto 1$ RAC, with X_{ONE} bias, over classical (red squares) and quantum (blue dots) encoding-decoding strategies. For $w \in [1/8, 5/12]$ the quantum value is given by the upper bound in Eq. (25) (solid line) and is strictly larger than its classical counterpart. For larger values of the biasing parameter the two values coincide. Bottom: Angles θ_{ij} between the Bloch vectors defining the measurement operators in the optimal decoding strategy. As expected from the symmetry of the bias tensor, the three angles coincide and have positive values in the quantum advantage region $w \in [1/8, 5/12]$, vanishing for $w \geq 5/12$.

and $M_1^{x_1}$ are jointly block-diagonal, with blocks of dimension 1 or 2. The restriction of these projectors to the k th Jordan block, P_k and Q_k , respectively, are rank-1 projectors regardless of the block dimension. The angle between the pure states on which they project, defined by $\cos^2(\theta_k) = \text{tr} P_k Q_k$, is one of the *principal angles* between $M_0^{x_0}$ and $M_1^{x_1}$. The principal angle defines the coefficients of P_k and Q_k when the block is two-dimensional, which are given by

$$P_k = \frac{1}{2} \begin{bmatrix} 1 + \cos(\theta_k) & \sin(\theta_k) \\ \sin(\theta_k) & 1 - \cos(\theta_k) \end{bmatrix}, \quad (44a)$$

$$Q_k = \frac{1}{2} \begin{bmatrix} 1 + \cos(\theta_k) & -\sin(\theta_k) \\ -\sin(\theta_k) & 1 - \cos(\theta_k) \end{bmatrix}. \quad (44b)$$

It follows then that we can write

$$r_0 M_0^{x_0} + r_1 M_1^{x_1} = \sum_k r_0 P_k + r_1 Q_k, \quad (45)$$

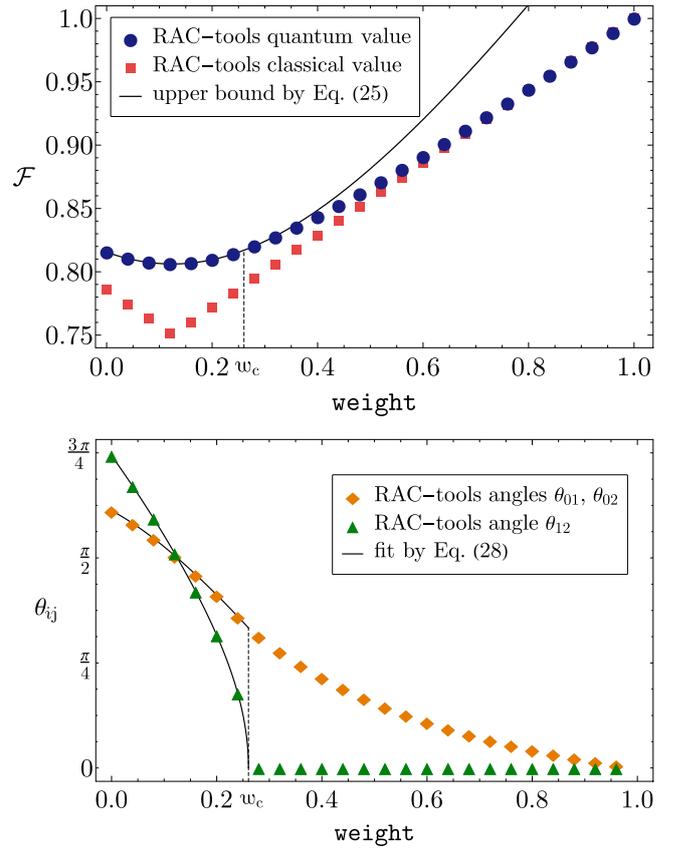


FIG. 5. Top: Optimal performance of the $3^2 \mapsto 1$ RAC with X_{ONE} bias and $r_0 = 0.5$, $r_1 = r_2 = 0.25$. As a result of the bias in the requested bit, one of the conditions in Eq. (36) is saturated before the other two, giving rise to a region in which the optimal quantum performance (blue dots) is strictly larger than the optimal classical value (red squares) but nonetheless strictly smaller than the upper bound in Eq. (25) (solid line). This region is found to be $w > \frac{1}{12}(7\sqrt{3} - 9)$. Bottom: Angles θ_{ij} parametrizing the optimal decoding strategy. Because of the asymmetry in the bias, one of the angles, θ_{12} (orange diamonds), is different from the other two (green triangles) and decreases faster as a function of the biasing parameter w , vanishing exactly at $w = \frac{1}{12}(7\sqrt{3} - 9)$.

implying that $\lambda_{x_0 x_1}$ is the largest eigenvalues of one of the blocks $r_0 P_k + r_1 Q_k$. If the corresponding block is two-dimensional, a direct calculation shows it is given by

$$\begin{aligned} \lambda_{x_0 x_1} &= \frac{1}{2} \left[1 + \max_k \sqrt{1 - 4r_0 r_1 \sin^2(\theta_k)} \right] \\ &\leq \frac{1}{2} \left[1 + \sqrt{1 + 4r_0 r_1 (\text{tr} M_0^{x_0} M_1^{x_1} - 1)} \right] \end{aligned} \quad (46)$$

where, in the last line, we used that $\cos^2(\theta_k) = \text{tr} P_k Q_k$ and that $\text{tr} P_k Q_k \leq \text{tr} M_0^{x_0} M_1^{x_1} \forall k$. If the block is one-dimensional, then it is easy to see that $\lambda_{x_0 x_1} \in \{0, r_0, r_1, 1\}$, in which case the upper bound in Eq. (46) also holds. Combining this upper bound with the Cauchy-Schwarz inequality we arrive at

$$\mathcal{F}_P \leq \frac{1}{2} + \frac{1}{2} \sqrt{d^2 - 4d(d-1)r_0 r_1} \sqrt{\sum_{x_0, x_1} \alpha_{x_0 x_1}^2}, \quad (47)$$

where we have used the completeness relation satisfied by the measurement operators to write $\sum_{x_0 x_1} \text{tr} M_0^{x_0} M_1^{x_1} = d$. ■

As follows from the proof given above, attaining the upper bound of Lemma 6 is possible if there exist projectors $M_y^{x_y}$ such that the upper bound in Eq. (46) is saturated and condition

$$\sqrt{1 + 4r_0 r_1 (\text{tr} M_0^{x_0} M_1^{x_1} - 1)} \propto \alpha_{x_0 x_1} \quad (48)$$

is satisfied, where the proportionality constant is easily found to be

$$C = \sqrt{\frac{d^2 - 4r_0 r_1 d(d-1)}{\sum_{x_0 x_1} \alpha_{x_0 x_1}^2}}. \quad (49)$$

Now note that the measurement operators satisfy $\text{tr} M_y^{x_y} = \text{rank}(M_y^{x_y})$, since we are considering projective measurements. Then, squaring equation Eq. (48) and summing over either x_0 or x_1 we arrive at

$$\text{rank}(M_y^{x_y}) = \frac{1}{4r_0 r_1} \left[d(4r_0 r_1 - 1) + C^2 \sum_{x_{1-y}} \alpha_{x_0 x_1}^2 \right]. \quad (50)$$

It is then easy to check that for $\alpha_{x_0 x_1} = \frac{1}{d^2}$ Eq. (50) implies that the upper bound in Eq. (42) can be attained only with rank-1 measurement operators. In fact, by virtue of Eq. (48) we have that the optimal measurements are mutually unbiased, thus recovering the solution reported in [34] for the case $r_0 = r_1 = \frac{1}{2}$, and extending it to arbitrary biases on Bob's input. Figure 6 shows the agreement of the optimal value provided by the numerical package and the upper bound in Lemma 6 for the cases $d = 3$, $d = 4$, and $d = 5$.

The upper bound in Eq. (42) can also be attained for more general b -RACs. Indeed, since for any pair of rank-1 projective measurements it holds that $\text{tr} M_0^{x_0} M_1^{x_1} = |U_{x_0 x_1}|^2$, with U a $d \times d$ unitary matrix, it follows from the condition in Eq. (48) that the upper bound in Eq. (42) will be attainable with rank-1 projectors if there exists a unistochastic matrix B satisfying

$$B_{x_0 x_1} = 1 + \frac{1}{4r_0 r_1} (C^2 \alpha_{x_0 x_1}^2 - 1). \quad (51)$$

Last, it is worth remarking that for some particular biases the optimal measurement operators may satisfy $\text{rank}(M_y^{x_y}) \neq 1$, as suggested by Eq. (50). Actually, as follows from the discussion above, for any pair of projective measurements saturating inequality Eq. (46) we can find b -RAC such that its optimal value is attained by these measurements. By summing over x_0 and x_1 in Eq. (48) we find that the entries of the bias tensor specifying this b -RAC are given by

$$\alpha_{x_0 x_1} = \frac{\sqrt{1 + 4r_0 r_1 (\text{tr} M_0^{x_0} M_1^{x_1} - 1)}}{\sum_{x_0 x_1} \sqrt{1 + 4r_0 r_1 (\text{tr} M_0^{x_0} M_1^{x_1} - 1)}}. \quad (52)$$

It should be noted that the b -RACs defined in this way are not necessarily interesting from the perspective of studying the advantages of quantum resources, since there is no guarantee regarding the distance of the upper bound to the classical value; e.g., if the measurements we have chosen commute, then the upper bound will coincide with the classical value.

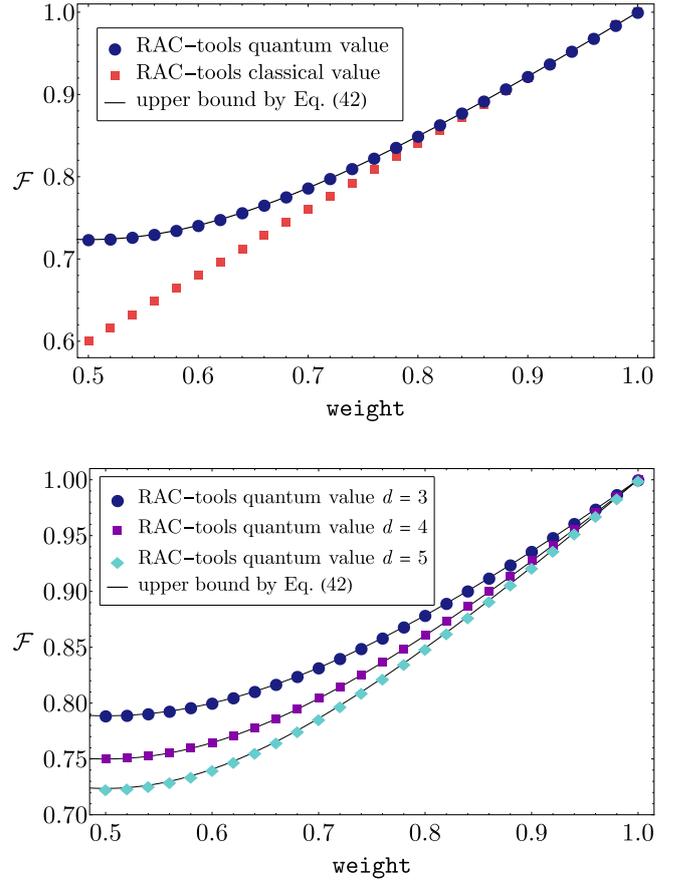


FIG. 6. Top: Optimal performance over quantum (blue dots) and classical strategies (red squares) of the $2^5 \mapsto 1$ b -RAC defined by the bias tensor $\alpha_{xy} = \frac{1}{d^2} r_y$ corresponding to the Y_{ONE} family, as computed by the RAC-tools package. The numerical results for the quantum value are compared with the upper bound in Eq. (42) (solid line). Bottom: Numerical results for the quantum value of the $2^d \mapsto 1$ b -RAC for $d = 3$ (blue dots), $d = 4$ (purple squares), and $d = 5$ (cyan diamonds).

This procedure can be used to build b -RACs tailored to specific pairs of projective measurements, in which the operators' rank is not restricted to 1.

VI. CONCLUSIONS

In this work, we have presented b -RACS as a generalization of the RAC protocol in which the distribution of inputs to the parties is not necessarily uniform. Introducing a bias on these distributions has a profound impact on both the optimal value of the RAC functional and encoding-decoding strategies achieving it, and also on the capacity of quantum devices to provide an advantage in the protocol performance. Understanding how to optimize the performance of a given biased RAC is therefore a step in improving our understanding of the advantages of quantum resources.

The problem of optimizing the performance of an arbitrary b -RAC can be approached numerically with the aid of the algorithms we have presented here, which can be implemented by means of the RAC-tools Python package we produced for

that purpose. The package allows the user to define arbitrary biases in the input distribution, and compute the classical and quantum value of the ensuing RAC functional, along with the encoding and decoding strategies attaining these values. We have used the package to study the b -RAC performance for different biases in the $n^2 \mapsto 1$ and $2^d \mapsto 1$ scenarios, focused in the case of uncorrelated inputs. For these examples we also provide analytical results for the quantum value and the measurements attaining it, showing how these are determined by the chosen input bias.

In the $n^2 \mapsto 1$ scenario, we have found that both classical and quantum optimal strategies may actually ignore part of the input strings. For quantum strategies it is first observed that optimal decoding can always be done with projective measurements. This allows the derivation of a simple upper bound which coincides with the quantum value for $2^2 \mapsto 1$ b -RACs and, in some cases, for $3^2 \mapsto 1$ b -RACs. Moreover, it is shown that attaining this upper bound self-tests the angles between the optimal measurement operators and, in particular, for the case of uniformly distributed input strings the optimal b -RAC performance certifies that the measurements correspond to MUBs. The argument in the derivation of this upper bound can be extended to the $2^d \mapsto 1$ scenario, providing thus an upper bound to the optimal performance achievable with projective measurements. This bound is shown to be always attainable using mutually unbiased measurements if the distribution of input strings is unbiased, regardless of the bias on the distribution of requested characters. For more general biases the upper bound will in general not be attainable, but we have shown that there are several instances in which this value is achievable. It is still not clear at the time of writing if, as suggested by our numerical results, the optimal b -RAC performance in this scenario is attainable only with projective measurements. In that case the upper bound we derived would coincide with the quantum value, and it would be worth to investigate the possibility of extending the self-testing results previously derived for the unbiased RAC in this scenario.

We have focused the discussion of analytical results in this work, almost completely, on the case of biased RACs in which the inputs of both parties are uncorrelated, since introducing correlations between them departs from the original spirit of the RAC protocol. Nevertheless, investigating how correlations in the inputs affect the performance of the protocol is an interesting next step for which the numerical tools we have developed are applicable.

ACKNOWLEDGMENTS

We acknowledge fruitful discussions with Máté Farkas. The project ‘‘Fundamental aspects of the quantum set of correlations’’ (Grant No. 2019/35/D/ST2/02014) is carried out within the SONATA project of the National Science Centre, Poland.

APPENDIX A: PROOF OF LEMMA 1

Proof. For statement (a), assume that a particular encoding function $E(\mathbf{x}) = \mu$ is fixed. If so, the performance of \mathcal{F} in Eq. (2) for an arbitrary set of decoding functions $\{D_y\}_{y=0}^{n-1}$ is

given by

$$\mathcal{F} = \sum_{\mathbf{x}, y} \alpha_{\mathbf{x}, y, D_y(E(\mathbf{x}))} = \sum_{\mathbf{x}, y, \mu} \alpha_{\mathbf{x}, y, D_y(\mu)} \delta_{\mu, E(\mathbf{x})}, \quad (\text{A1})$$

where we used Eq. (4) to compute the statistics. Then the maximization of \mathcal{F} over the set of decoding functions is equivalent to the maximization of the image of $D_y(\mu)$:

$$\max_{\{D_y\}_y} \mathcal{F} = \sum_{\mu, y} \max_{D_y(\mu)} \left\{ \sum_{\mathbf{x}} \alpha_{\mathbf{x}, y, D_y(\mu)} \delta_{\mu, E(\mathbf{x})} \right\}, \quad (\text{A2})$$

which yields, for some y and μ , $D_y^*(\mu) = b$, where b is the optimal image of $D_y(\mu)$.

Now, for statement (b), we proceed similarly by assuming that the decoding functions $D_y(\mu)$ are fixed for all y . Then the value assumed by \mathcal{F} for an arbitrary encoding function $E(\mathbf{x})$ is written as

$$\mathcal{F} = \sum_{\mathbf{x}, y} \alpha_{\mathbf{x}, y, D_y(E(\mathbf{x}))}. \quad (\text{A3})$$

Analogous to statement (a), the maximization of \mathcal{F} over the encoding function is equivalent to maximize the image of $E(\mathbf{x})$:

$$\max_E \mathcal{F} = \sum_{\mathbf{x}} \max_{E(\mathbf{x})} \left\{ \sum_y \alpha_{\mathbf{x}, y, D_y(E(\mathbf{x}))} \right\}, \quad (\text{A4})$$

which produces $E^*(\mathbf{x}) = \mu$, where μ is the optimal image for $E(\mathbf{x})$. ■

As explained in the main text, this lemma is useful in reducing the inherent complexity associated with the exhaustive search algorithm. To provide further clarity, we can maximize Eqs. (A2) and (A4) with respect to the encoding and decoding functions, respectively. This additional maximization yields the classical value of \mathcal{F} for both case, as follows:

$$\mathcal{F}_C = \max_E \left\{ \sum_{\mathbf{x}, y} \alpha_{\mathbf{x}, y, D_y^*(E(\mathbf{x}))} \right\}, \quad (\text{A5a})$$

$$\mathcal{F}_C = \max_{\{D_y\}_{y=0}^{n-1}} \left\{ \sum_{\mathbf{x}, y} \alpha_{\mathbf{x}, y, D_y(E^*(\mathbf{x}))} \right\}. \quad (\text{A5b})$$

That is, Eqs. (A5a) and (A5b) introduce a two-step maximization method that yields the precise classical value. For Eq. (A5a), we first optimize over the decoding functions and then over the encoding functions. Conversely, for Eq. (A5b), we follow the reverse order. This simple modification avoids the maximization over all combinations of E and $\{D_y\}_{y=0}^{n-1}$. Furthermore, since the RAC protocol is asymmetrical with respect to Alice and Bob, the difference between Eqs. (A5a) and (A5b) relies only on the computational complexity for each case.

APPENDIX B: RAC-TOOLS USER GUIDE

In the main text, we introduced the functions that make up the RAC-tools package. In this Appendix, we provide a more detailed description of these functions and their features.

1. The `generate_bias` function

Since our interest in this work is to study the quantum and classical value of biased RACs, the main feature of the RAC-tools package is that it allows the user to introduce bias in the RAC functional, which will be optimized by either the `perform_search` or `perform_seesaw` functions. One way of doing this is by building a custom bias tensor and passing it as an argument to either of these functions as a Python dictionary. However, as constructing a bias tensor requires some effort, we provide a functionality that allows the user to choose from several simple and natural families of bias tensors. Each of these families takes one or more parameters, and it is particularly interesting to study the behavior of RACs as we vary the parameter. An example of such a construction was given at the beginning of Sec. III. In what follows we describe in detail the built-in bias options that the user can access via the `generate_bias` function.

The goal of `generate_bias`, in short, is to construct a properly normalized bias tensor using only a few previously specified parameters. This function is not intended to be called by the user, who should in turn specify the parameters defining the desired bias tensor as arguments of the optimization functions. In order to do so, the value of two variables, `bias` and `weight`, must be specified. The variable `bias` is a string determining the structure of the bias to be generated, whereas the variable `weight` is a real-valued parameter (or a vector of parameters) that determines the actual weights given to different terms in the objective function.

As an example, we can consider a general version of the `Y_ONE` bias family already introduced in the main text. This is a family of bias tensors in which the input strings \mathbf{x} are distributed uniformly, but there is bias in Bob's input, as one of the characters of \mathbf{x} , e.g., x_k , is requested more (or less) frequently than the others. If we call w the parameter defining how often Bob is asked to recover x_k , then the bias tensor takes the form

$$\alpha_{xy} = \begin{cases} \frac{1}{m^n} w & \text{if } y = 0, \\ \frac{1}{m^n} \frac{(1-w)}{n-1} & \text{otherwise.} \end{cases} \quad (\text{B1})$$

In order to build this bias tensor via the `generate_bias` function, we need to pass as arguments of either `perform_search` or `perform_seesaw` the following string and float: `bias='Y_ONE'` and `weight=w`. By symmetry, the `Y_ONE` family considers only biasing the first character against the rest, as biasing other values of y produces analogous results. It is possible, nevertheless, to introduce a bias on the frequency with which any of the characters x_y is requested from Bob. This can be done by setting `bias='Y_ALL'` and `weight=List`, where `List` is a list (or a tuple) of floats of length n adding up to one. In this case, the bias tensor obtained from `generate_bias` takes the form

$$\alpha_{xy} = \frac{1}{m^n} w_y, \quad (\text{B2})$$

where w_y is the weight corresponding to the y th character in the input string \mathbf{x} and the factor $\frac{1}{m^n}$ results from the input strings \mathbf{x} being uniformly distributed.

For introducing biases in the distribution of input strings, the package offers several one-parameter families, which we enumerate below:

(1) `X_ONE`. Analogous to the `Y_ONE` family, it biases the input $\mathbf{x} = 0^{\times n}$ against the $m^n - 1$ remaining strings. The user is allowed to define the weight w that will be given to this first input, which will be used to generate a bias tensor of the form $\alpha_{xy} = \alpha_x \frac{1}{n}$, where

$$\alpha_{\mathbf{x}} = \begin{cases} w & \text{if } \mathbf{x} = 0^{\times n}, \\ \frac{1-w}{m^n-1} & \text{otherwise.} \end{cases} \quad (\text{B3})$$

(2) `X_DIAG`. This family of biases gives a special weight to input strings of the form $\mathbf{x} = i^{\times n}$, where $i = 0, \dots, m-1$. Since there are m of these strings, in terms of the parameter w controlled by the user, the distribution of input strings takes the form

$$\alpha_{\mathbf{x}} = \begin{cases} \frac{w}{m} & \text{if } \mathbf{x} = i^{\times n}, \\ \frac{1-w}{m^n-m} & \text{otherwise.} \end{cases} \quad (\text{B4})$$

(3) `X_CHESS`. In this case, the input strings are split into two classes depending on whether $\sum_j x_j$ is odd or even. Since the parity of the total number of strings is the same as that of m , when the latter is even, half of the strings go into each of the classes defined above. In that case, in terms of the weight w chosen by the user, the ensuing distribution of input strings is given by

$$\alpha_{\mathbf{x}} = \begin{cases} \frac{2w}{m^n} & \text{if } \sum_j x_j \text{ is even,} \\ \frac{2(1-w)}{m^n} & \text{otherwise.} \end{cases} \quad (\text{B5})$$

On the other hand, if m is odd, the number of strings satisfying $\sum_j x_j$ odd is $\frac{m^n-1}{2}$. In that case, the distribution of input strings reads

$$\alpha_{\mathbf{x}} = \begin{cases} \frac{2w}{m^n-1} & \text{if } \sum_j x_j \text{ is odd,} \\ \frac{2(1-w)}{m^n+1} & \text{otherwise.} \end{cases} \quad (\text{B6})$$

For $n = 2$, we can think of the elements of $\alpha_{\mathbf{x}}$ as the entries of a matrix, in which case the biased elements are arranged in a pattern that resembles a chess board.

(4) `X_PLANE`. As before, the idea of this type of bias is to split the set of strings into two classes defined by the condition $x_0 = 0$. This corresponds to biasing just the first bit of the input string. Since there are m^{n-1} strings satisfying $x_0 = 0$, in terms of the parameter w , the ensuing distribution of input strings reads

$$\alpha_{\mathbf{x}} = \begin{cases} \frac{w}{m^{n-1}} & \text{if } x_0 = 0, \\ \frac{1-w}{m^n-m^{n-1}} & \text{otherwise.} \end{cases} \quad (\text{B7})$$

All the biases introduced so far depend only on \mathbf{x} or only on y . In the next step, we could take one bias of each kind and combine them, which would lead to a product distribution over \mathbf{x} and y . However, as mentioned in the introduction, there is no reason why we should restrict ourselves to product distributions. If we go back to the linear functional given in Eq. (2), it is natural to consider the case where the coefficients of the functional depend only on b , i.e., the answer that Bob is expected to give. In Appendix C we discuss such scenarios and refer to them as `B_ONE` and `B_ALL` biases. In our usual language such cases correspond to nonfactorizable

```

> perform_search(n=2, d=2, method=0)

=====
                        RAC-tools v1.0
=====

----- Summary of computation -----

Total time of computation: 0.001859 s
Total number of encoding/decoding functions: 256
Average time per function: 7e-06 s

----- Analysis of the optimal realization -----

Computation of the classical value for the  $2^2 \mapsto 1$  RAC:
0.75
Number of functions achieving the computed value: 24

First functions found achieving the computed value
Encoding:
E: [0, 0, 0, 1]
Decoding:
D0: [0, 1]
D1: [0, 1]

----- End of computation -----

```

FIG. 7. Report produced by the `perform_search` function for the unbiased $2^2 \mapsto 1$ RAC. The first part of the report provides the user with information about the computation, whereas the second part provides the user with the RAC classical value and a realization attaining it. In this example, the value is found through an exhaustive search of the best value over both encoding and decoding maps.

distributions of inputs \mathbf{x} and y . The first bias of this kind, `bias='B_ONE'`, corresponds to biasing the first outcome of Bob, $b = 0$, against the remaining $d - 1$ outputs

$$\alpha_{xy} = \begin{cases} \frac{1}{n} \frac{1}{m^{d-1}} w & \text{if } x_y = 0, \\ \frac{1}{n} \frac{1}{m^{d-1}} \frac{(1-w)}{m-1} & \text{otherwise,} \end{cases} \quad (\text{B8})$$

where $\frac{1}{n} \frac{1}{m^{d-1}}$ is the normalization factor. If a more general bias of the outputs is required, then the user can enter `bias='B_ALL'` as an option, in which case they should input as weight a Python list (or tuple). The `generate_bias` function will then output a bias tensor of the form

$$\alpha_{xy} = \frac{1}{n} \frac{1}{m^{d-1}} w_{x_y}, \quad (\text{B9})$$

where w_{x_y} is the weight on the character x_y – and consequently on the b th output of Bob since $b = x_y$.

2. The `perform_search` function

The goal of `perform_search` is to exactly compute the best classical performance of a given $n^m \mapsto 1$ RAC. The function can perform this computation either via a complete exhaustive search or by means of the less expensive approach that follows from Eqs. (A5a) and (A5b). To operate `perform_search` it is enough to specify in its argument the three integers defining the scenario, n , d , and m , and the search method. The latter can be introduced by declaring either `method=0` for a pure exhaustive search, `method=1` for a search over encoding maps as in Eq. (A5a), or `method=2` for a search over decoding maps as in Eq. (A5b), which is the

default method. Furthermore, the value of m is set by default to coincide with that of d , so that users are not expected to declare it unless they require these numbers to be different.

An example of how this function operates can be seen in Fig. 7, in which the user desires to estimate the classical value of the $2^2 \mapsto 1$ unbiased RAC. The function is called passing as arguments `n=2`, `d=2`, and `method=0`, and once the procedure is finished the report in Fig. 7 is printed. The *Summary of computation* section of the report informs the user the total time of computation as well as the total number of encoding and decoding functions analyzed for the chosen search method. For the case of `method=0`, this latter information corresponds to the total number of combinations of encoding and decoding functions, i.e., $d^{m^n} \times m^{dn}$. In addition, the average time taken to iterate over each function (or combination of encoding and decoding functions, if `method=0`) is displayed at *Average time per function*.

In the second part of the report, the user can see the computed classical value and the number of functions that achieve this value. Also, the report provides the user with a particular pair of encoding and decoding strategies attaining the optimal value. For the encoding function $E(\mathbf{x})$, the result is displayed in a tuple that is organized in ascending order of \mathbf{x} , i.e., $[E(00 \dots 0), E(00 \dots 1), \dots, E((m-1) \dots (m-1))]$. For the decoding functions $D_y(\mu)$, each row corresponds to a distinct input y and it is organized in ascending order of μ . As it is expected from Lemma 2, the optimal strategy reported in Fig. 7 consists of a majority encoding function and identity map for decoding.

Before closing, we would like to recall that the exhaustive search method will necessarily require more computation time than that required by either of the two approaches following from Lemma 1, since in the first case the search is performed over both encoding and decoding maps. Table I presents a comparison in terms of computation time for all of the three methods. The table shows that the method in Eq. (A5b) is the best in terms of computation time for most part of the cases, except for a few cases in which d is the largest integer and $n = 2$. In those cases, the method in Eq. (A5a) is equivalent or better.

3. The `perform_seesaw` function

This function implements the see-saw algorithm described in Sec. II A, and its goal is to provide lower bounds to the quantum value of a given $n^m \mapsto 1$ b -RAC. As is the case with `perform_search`, the `perform_seesaw` function takes as argument the integers defining the scenario, n , d , and m and the bias tensor, either as a dictionary or via one of the aforementioned built-in options. The user is also asked to pass as an argument the number of starting points for the algorithm by means of the variable `seeds`. Moreover, it is possible to use this function to compute a lower bound to the classical value, by means of the variable `diagonal`. If `diagonal=True`, the function initializes the see-saw algorithm with random diagonal measurements, and the optimization is then restricted to operators which are diagonal in the computational basis. By default, `diagonal=False`, and the algorithm optimizes the functional value over POVM measurements.

TABLE I. Cases which can be executed in less than one hour for `method=0`. Computation time comparison with `method=1` and `method=2`. For the construction of this table, an octa-core CPU was used (4×3.2 GHz and 4×2.064 GHz). The first column contains the executable cases for a given $n^m \mapsto 1$ RAC. The second column contains the classical value computed for each unbiased case. The other columns contain the time taken to execute each procedure. Cases are sorted in ascending order by computation time for `method=0`.

$n^m \mapsto 1$	Classical value	Time (method=0)	Time (method=1)	Time (method=2)
$2^2 \mapsto 1$	3/4	1.8 ms	0.32 ms	0.28 ms
$2^2 \mapsto 1$	7/8	26 ms	2 ms	1.2 ms
$3^2 \mapsto 1$	3/4	0.14 s	11 ms	2.4 ms
$2^2 \mapsto 1$	1	0.18 s	7.8 ms	6.5 ms
$2^3 \mapsto 1$	5/9	0.22 s	21 ms	2.6 ms
$2^2 \mapsto 1$	1	1.4 s	19 ms	28 ms
$2^2 \mapsto 1$	1	12 s	40 ms	82 ms
$3^2 \mapsto 1$	19/24	21 s	0.16 s	23 ms
$2^3 \mapsto 1$	2/3	1 min 4 s	0.42 s	27 ms
$2^2 \mapsto 1$	1	1 min 29 s	61 ms	0.25 s
$2^4 \mapsto 1$	7/16	2 min 20 s	2.2 s	12 ms
$4^2 \mapsto 1$	11/16	5 min 18 s	2.4 s	21 ms
$2^2 \mapsto 1$	1	11 min 6 s	0.1 s	1 s
$3^2 \mapsto 1$	5/6	33 min 16 s	1.5 s	0.12 s

When called, `perform_seesaw` runs the see-saw algorithm as many times as the number of seeds specified by the user, generating a lower bound to the quantum value per starting point. The best value is therefore the largest among all these lower bounds, implying that the chances of the function providing the actual quantum value of the b -RAC increase with the number of seeds, as well as the computation time. In Table II, we provide the number of seeds used to generate the numerical results presented in the main text and later in Appendix C.

TABLE II. Number of seeds used for the numerical results presented in the figures. The first column contains the executable cases for a given $n^d \mapsto 1$ RAC followed by the bias family in the second column. The third column contains the number of seeds used followed by a link to the respective figure. For the cases where the realization varies smoothly according to the `weight`, only a few seeds are needed. This is the case of all bias families explored in this work expect for `X_PLANE`, in which there is a critical point for `weight` in which the realization starts to ignore a bit.

$n^d \mapsto 1$	Bias	Seeds	Figure
$3^2 \mapsto 1$	X_ONE	3	Fig. 4
$3^2 \mapsto 1$	X_ONE with $r_0 = 0.5$ and $r_1 = r_2 = 0.25$	3	Fig. 5
$2^3 \mapsto 1$	Y_ONE	3	Fig. 6
$2^4 \mapsto 1$	Y_ONE	3	Fig. 6
$2^5 \mapsto 1$	Y_ONE	3	Fig. 6
$4^2 \mapsto 1$	Y_ONE	3	Fig. 9
$2^2 \mapsto 1$	X_CHESS	3	Fig. 10
$2^2 \mapsto 1$	X_PLANE	10	Fig. 11
$3^2 \mapsto 1$	X_PLANE	10	Fig. 12

Because the see-saw algorithm is iterative, convergence criteria must be adopted to decide whether the optimal value for a given seed has been attained after a particular number of steps. In the `perform_seesaw` implementation of this algorithm, we impose two convergence criteria, and the procedure is finished whenever the two are satisfied. The first criterion is related to the convergence of the \mathcal{F} value. It is satisfied whenever the difference between two consecutive evaluations of \mathcal{F} is smaller than a value that can be set by the user via the variable `prob_bound`. The default value of this variable is set to 10^{-9} . The second stopping criterion considers the convergence of the measurements, and it focuses on the distance between the optimal measurement operators in two consecutive iterations of the algorithm. More precisely, we will say that the measurements converged if the condition

$$\max_{y,b} \|M_y^b - N_y^b\| < t \quad (\text{B10})$$

is satisfied, where $\|\cdot\|$ denotes the Frobenius norm, N_y^b and M_y^b denote two consecutive measurement operators associated with the same value of the y th character of the input \mathbf{x} , and t is a threshold that can be defined by the user via the variable `meas_bound`, which as a default takes the value 10^{-7} . For the evaluation of the condition in Eq. (B10), we use the function `norm`, from `numpy.linalg`, to implement the Frobenius norm.

The value of both `prob_bound` and `meas_bound` can be passed as an argument to `perform_seesaw`. In addition to the convergence criteria, we have imposed a limit to the number of iterations to be executed by the algorithm, so that if after 200 iterations either the value or the measurements fail to converge, the calculation stops. In this case, the message *maximum number of iterations reached* is displayed as a warning.

```

> perform_seesaw(n=2, d=2, seeds=5)

=====
                        RAC-tools v1.0
=====

----- Summary of computation -----

Number of random seeds: 5
Average time for each seed: 0.14852 s
Average number of iterations: 3
Seeds 1e-13 close to the best value: 5

---- Analysis of the optimal realization for seed #1 ----

Estimation of the quantum value for the 22->1 QRAC:
0.853553390593

Measurement operator ranks
M[0] ranks: 1 1
M[1] ranks: 1 1

Measurement operator projectiveness
M[0, 0]: Projective      6.44e-15
M[0, 1]: Projective      6.44e-15
M[1, 0]: Projective      6.78e-15
M[1, 1]: Projective      6.78e-15

Mutual unbiasedness of measurements
M[0] and M[1]: MUB      5.91e-14

----- End of computation -----

```

FIG. 8. Report produced by the function `perform_seesaw`. In the first part of the report, the function produces a summary of the computation, displaying information such as the number of random starting points, etc. In the second part, it displays the largest optimal value found among all seeds and an analysis of the optimal measurements obtained by this seed.

This limit can be modified by entering a different value to the variable `max_iterations` in the argument of the function.

An example of the operation of `perform_seesaw` can be seen in Fig. 8, in which the user wants to estimate the quantum value of the $2^2 \mapsto 1$ unbiased RAC. As in the case of `perform_search`, the user passes as arguments $n = 2$ and $d = 2$ to define the scenario, but now instead of choosing a search method the user introduces the number of starting points to be used by passing `seeds=5`. After finishing the procedure the function prints a report divided into two parts. In the *Summary of the computation*, it presents the number of random starting points, the average processing time, and the average number of iterations among all starting points. In addition, it shows how many starting points produced an optimal value that is close to the largest value obtained. The interval to consider two values produced by different starting points as close is the accuracy of the solver MOSEK, which is set to 10^{-13} . This informs the user how frequent it is to obtain such an estimation; if this number is much smaller than `seeds`, this indicates that the user should increase the number of starting points in case of a new execution.

In the second part, the estimation of the optimal value is reported, followed by information about the set of measurements attaining such value. Note that the reported value in Fig. 8 matches the one found by Ref. [30]. Next, the

report displays the rank of the optimal measurement operators, which is computed using the function `matrix_rank` of `numpy.linalg`. In addition, the user can check whether the measurement operators are projective. The number shown in the second column of *Measurement operator projectiveness* corresponds to the quantity

$$\| |(M_y^b)^2 - M_y^b | \| . \quad (\text{B11})$$

For both of these checks, rank and projectiveness, we preset a tolerance of 10^{-7} .

Last, in the case where at least two measurements are rank-one and projective, the function also computes whether each pair of measurements can be constructed out of mutually unbiased bases (MUBs). For a pair of rank-one projective measurements, let us say $\{P^a\}_{a=0}^{m-1}$ and $\{Q^b\}_{b=0}^{m-1}$, where a and b denote the a th and the b th outcome, it is enough [52, Appendix B] to check if

$$P^a = m P^a Q^b P^a \quad \text{and} \\ Q^b = m Q^b P^a Q^b \quad \forall a, b \in \{0, 1, \dots, m-1\}. \quad (\text{B12})$$

In this case, the number displayed in the second column of *Mutual unbiasedness of measurements* represents the quantity

$$\max_{a,b} \{ \| |m P^a Q^b P^a - P^a| \|, \| |m Q^b P^a Q^b - Q^b| \| \}. \quad (\text{B13})$$

For the cases in which the amount in Eq. (B13) is lower than `MUB_BOUND=5e-6`, the function prints MUB. Otherwise, it simply displays Not MUB.

APPENDIX C: ANALYSIS FOR OTHER BUILT-IN FAMILIES OF BIAS

In the main text, we have used the analytical results derived for the $2^n \mapsto 1$ scenario to study the quantum value of the b -RACs determined by the `X_ONE` bias family introduced above. Here, we offer a similar analysis for the b -RAC families determined by others of these built-in biases in the $2^n \mapsto 1$ scenario.

1. The `Y_ONE` bias family

We start by looking at the case where the bias is only on the requested bit y , i.e., $\alpha_x = \frac{1}{2^n}$, which leads to $p_x = \frac{1}{2^{n-1}} \forall x$. From p_x being constant follows, for $n = 2$ and $n = 3$, that $\cos(\theta_{ij}) = 0 \forall i \neq j$ in Eq. (28), i.e., the optimal measurements are mutually unbiased. The quantum value is therefore given by the upper bound in Eq. (25),

$$F_Q = \frac{1}{2} + \frac{1}{2} \sqrt{\sum_y r_y^2}. \quad (\text{C1})$$

For $n = 4$, the upper bound is not attainable when $r_y = \frac{1}{4}$, as it would require the four vectors $\{\mathbf{m}_y\}$ to be mutually orthogonal. For weak biases ($r_y \approx \frac{1}{4}$) satisfying the conditions in Eq. (28), it would still require these vectors to be linearly independent, and therefore the upper bound is still not attainable. However, if we consider a stronger bias such that the weight on one of the bits becomes negligible, we would expect the bound in Eq. (25) to be attainable again. We can realize

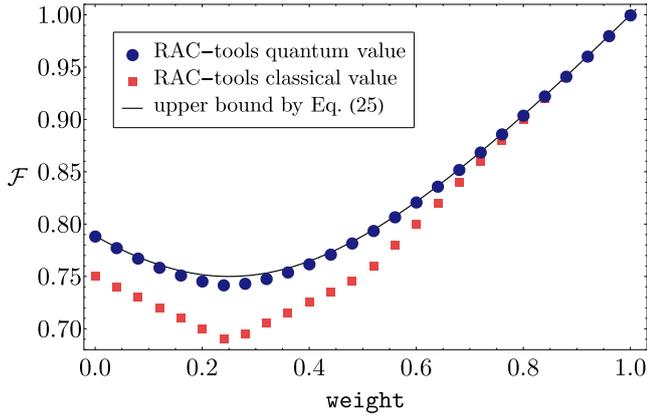


FIG. 9. Optimal performance of the $4^2 \mapsto 1$ b -RAC, with Y_{ONE} bias, over classical (red squares) and quantum (blue dots) encoding-decoding strategies.

such situation by defining

$$r_y = \begin{cases} w & \text{if } y = 0 \\ \frac{1-w}{3} & \text{otherwise} \end{cases}, \quad 0 \leq w \leq 1. \quad (\text{C2})$$

Clearly in this case $w = 0 \Rightarrow r_0 = 0$, and the value of the first bit is never requested from Bob to be decoded. Thus, there are only three Bloch vectors representing measurements that can be chosen to be orthogonal to each other, so that the bound in Eq. (25) is attained.

This is indeed the case, as is shown in Fig. 9, which depicts the results provided by the numerical package for the optimal performance of this $4^2 \mapsto 1$ b -RAC over both, quantum and classical encoding-decoding strategies. As can be seen in this plot, the numerical quantum values (blue dots) lie very close to upper bound (solid line), coinciding with it only at the extremal values ($w = 0, 1$) and exhibiting the largest difference at $w = \frac{1}{4}$. An inspection of the optimal measurements extracted from the numerical solutions shows that the angles θ_{ij} parametrizing the measurements naturally divide into two sets, $\Theta_0 = \{\theta_{0i}\}$ and $\Theta_{\bar{0}} = \{\theta_{i\bar{j}}\}$, with $i \neq j = 1, 2, 3$, and the angles in one of them being all $\frac{\pi}{2}$. As explained above, for $r_0 = 0$, the optimal solution involves only the vectors \mathbf{m}_i , $i = 1, 2, 3$, which can be chosen to be mutually orthogonal. For small values of r_0 , we would then expect these three vectors to remain orthogonal (or close to orthogonal), since they contribute the most to the functional value, and \mathbf{m}_0 to

be some linear combination of them. The solution provided by the numerical search shows that this intuition is correct since for $w \in [0, w_0]$, with $w_0 \approx 0.27415$, we find $\{\mathbf{m}_i\}$ to be an orthogonal set, with \mathbf{m}_0 being aligned with any of them, meaning that the angles in $\Theta_{\bar{0}}$ are all $\frac{\pi}{2}$. This solution has also been found by the authors of Ref. [30] in a numerical search for the case of $w = \frac{1}{4}$.

On the other hand, if $w = r_0 = 1$, only the first bit is to be retrieved by Bob, which can be done with probability 1 with a classical strategy. For $w \approx 1$ therefore we would expect \mathbf{m}_0 to lie orthogonal to the subspace spanned by $\{\mathbf{m}_i\}$, which is therefore bound to have dimension 2. This is indeed the case, as shown by our numerical results: For $w \in [w_0, 1]$ the optimal value is numerically attained with a decoding strategy in which \mathbf{m}_0 is orthogonal to all the \mathbf{m}_i 's, which therefore span a plane in \mathbb{R}^3 , implying that the angles in Θ_0 are all $\frac{\pi}{2}$, while the angles in $\Theta_{\bar{0}}$ can be chosen to be $\Theta_{\bar{0}} = \{\frac{\pi}{3}\}$. It follows then that the \mathbf{m}_i are uniformly distributed in the plane orthogonal to \mathbf{m}_0 .

2. The X_{CHESS} bias family

Let us now consider a different distribution α_x for the input strings, given by

$$\alpha_x = \begin{cases} \frac{w}{2^n} & \text{if } \sum_i x_i \text{ odd} \\ \frac{1-w}{2^n} & \text{otherwise} \end{cases}, \quad 0 \leq w \leq 1 \quad (\text{C3})$$

and an arbitrary distribution $\{r_y\}$ for the requested bit. As we will show now, this bias has no net effect on the b -RAC value when the number of bits is odd. Indeed, note first that there are 2^n input bit strings, half of which are such that $\sum_i x_i$ is even. Now if \bar{x} is the string obtained from x by flipping all of its bits, then the sum of bits has the same parity in both strings if n is even, whereas if n is odd this parity is different. As a result, it follows from Eq. (C3) that for odd n we have $p_x = \alpha_x + \alpha_{\bar{x}} = \frac{1}{2^{n-1}}$ and the functional value becomes

$$F_Q = \frac{1}{2} + \frac{1}{2^{n-1}} \max_{\{\mathbf{m}_y\}} \sum_x \left| \sum_y r_y (-1)^{x_y} \mathbf{m}_y \right|, \quad (\text{C4})$$

which is the same as that of the unbiased case. We can illustrate this feature by analyzing the $n = 2$ and $n = 3$ cases. A direct calculation shows that for $n = 2$ the value is given by

$$\begin{aligned} F_Q &= \frac{1}{2} + \max_{\{\mathbf{m}_0, \mathbf{m}_1\}} \frac{1-w}{2} |r_0 \mathbf{m}_0 + r_1 \mathbf{m}_1| + \frac{w}{2} |r_0 \mathbf{m}_0 - r_1 \mathbf{m}_1| \\ &\leq \frac{1}{2} + \max \left\{ \frac{1-w}{4} + \frac{w}{4} (r_0 - r_1), \frac{w}{4} + \frac{1-w}{4} (r_0 - r_1), \frac{1}{2\sqrt{2}} \sqrt{w^2 + (1-w^2)} \sqrt{r_0^2 + r_1^2} \right\}, \end{aligned} \quad (\text{C5})$$

where, in the second line, we used Lemma 4. In Fig. 10, we show this value as a function of the biasing parameter w , compared with the numerical results for the optimal value over both quantum and classical strategies for the case of $r_0 = r_1 = \frac{1}{2}$. As is easy to check from Eq. (C5), the optimal performance for quantum strategies is better than that over

the classical ones for $w \in (0, 1)$, becoming equal only for $w = 0$ (1) in which case only the strings 00 and 11 (01 and 10) are given to Alice for encoding.

In the case $n = 3$, as explained above, the value of both classical and quantum strategies becomes insensitive to variations of the biasing parameter w , coinciding with that of the

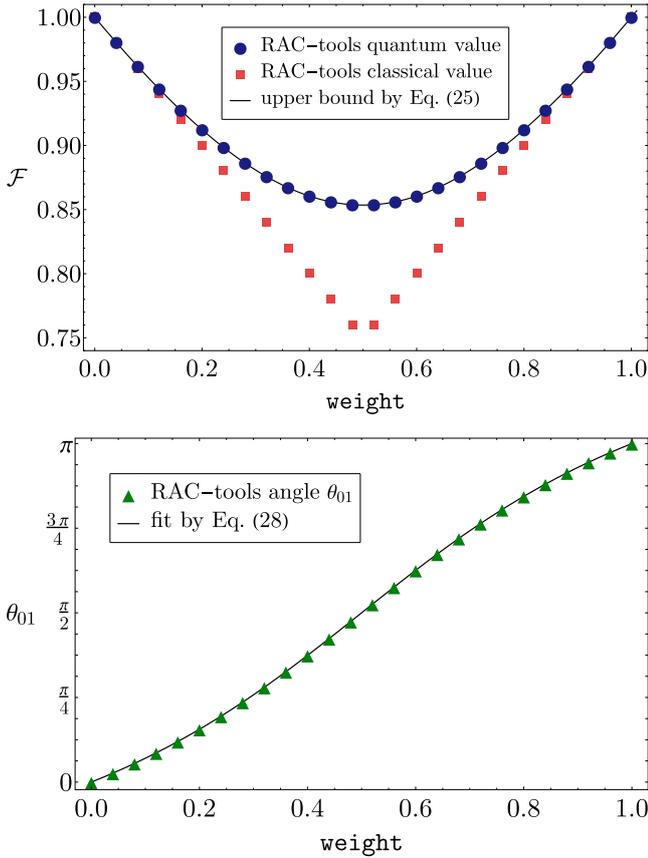


FIG. 10. Top: Optimal performance of the $2^2 \mapsto 1$ b -RAC with X_{CHESS} bias, in combination with $r_0 = r_1 = \frac{1}{2}$, over quantum strategies (blue dots) and their classical counterpart (red squares). The numerical results for the quantum value are seen to agree with the theoretical prediction (solid line) extracted from Lemma 4. Bottom: Angle between the Bloch vectors defining the optimal decoding strategy, as a function of the biasing parameter w . It is seen that these vectors are aligned only for the extremal values of w , meaning that the quantum value is strictly greater than the classical one for $w \in (0, 1)$.

unbiased RAC, which is given by

$$F_Q = \frac{1}{2} + \frac{1}{2} \sqrt{\sum_y r_y^2}, \quad (\text{C6})$$

as follows from the previous discussion. It is perhaps surprising that even though in the extreme cases of $w \in \{0, 1\}$ we are left with only four out of the original eight strings, the RAC task does not get any easier. This could serve as an indication that some subsets of strings are as difficult to compress as the set of all strings (regardless of whether the compression is classical or quantum).

3. The X_{PLANE} bias family

So far, we have been introducing biases in the distribution of input strings by setting $\alpha_{xy} = \alpha_x r_y$, where α_x is interpreted as the probability for input \mathbf{x} to be encoded. As a result, when considering the individual random variables corresponding to the characters in the input strings, they will, in general,

exhibit correlations. We can consider the case in which the input string characters are independently biased by defining $\alpha_{\mathbf{x}} = \prod_i \alpha_{x_i}$, where α_{x_i} denotes the probability of the i th character in the string being x_i . In particular, for the case of two-bit input strings the bias tensor is given by $\alpha_{x_0 x_1 y} = \frac{1}{2} \alpha_{x_0} \alpha_{x_1}$.

Now let $\alpha_{x_1=0} = \alpha_{x_1=1} = 1/2$, and keep α_{x_0} arbitrary. It then follows that $p_{00} = \alpha_{00} + \alpha_{11} = \frac{1}{2}$, and $p_{01} = \alpha_{01} + \alpha_{10} = \frac{1}{2}$, implying that the bias has no net effect on the value of strategies that do not drop bits, as can be checked directly from Eq. (24). Consequently, the optimal value of this $2^2 \mapsto 1$ b -RAC among quantum, non-bit-dropping strategies, is $F_Q = \frac{1}{2}(1 + \frac{1}{\sqrt{2}})$. We should note, however, that for extreme biases, the first character of the input string is either always 0 or always 1. In a situation as such there is no reason to include the first bit in the strategy, since the best performance can be obtained with a constant decoding function. Because this bit dropping could become optimal as a strategy for biases below the extremal value, in order to compute the quantum value we should compare F_Q with the best value attained by a bit-dropping strategy, which is easily found to be

$$\mathcal{F}_Q^1 = \begin{cases} \frac{1}{2}(1 + \alpha_{x_0=0}) & \text{if } \alpha_{x_0=0} > \alpha_{x_0=1}, \\ \frac{1}{2}(1 + \alpha_{x_0=1}) & \text{otherwise,} \end{cases} \quad (\text{C7})$$

where the symbol \mathcal{F}_Q^1 , as defined in Eq. (19), denotes that the strategy attaining this value does not encode one of the bits of the input string. Note that \mathcal{F}_Q^1 coincides with the classical value for this RAC, since by ignoring a bit we are left with only one to consider in the encoding strategy. It follows from Eq. (C7) that dropping the first bit becomes optimal whenever $\mathcal{F}_Q^1 > \frac{1}{2}(1 + \frac{1}{\sqrt{2}})$, which occurs for $\alpha_{x_0} > \frac{1}{\sqrt{2}}$ if $\alpha_{x_0} > \alpha_{x_1}$. In that case, the quantum value is therefore given by

$$\mathcal{F}_Q = \begin{cases} \frac{1}{2}(1 + \frac{1}{\sqrt{2}}) & \text{if } \alpha_{x_0} \leq \frac{1}{\sqrt{2}} \alpha, \\ \frac{1}{2}(1 + \alpha_{x_0}) & \text{otherwise.} \end{cases} \quad (\text{C8})$$

Figure 11 depicts the numerical results provided by the RAC-tools package, which agree with the analytical value provided above.

A completely analogous analysis can be carried for the $n = 3$ case for $\alpha_{x_0 x_1 x_2} = \frac{1}{4} \alpha_{x_0}$. As was the case for $n = 2$, in the vicinity of the uniform distribution we expect the optimal strategy to include all three bits in the input string, in which case a direct calculation shows that the bias has no effect on the quantum value, which is given by $F_Q = \frac{1}{2}(1 + \frac{1}{\sqrt{3}})$. As before we expect the optimal strategy to ignore the first bit when its value is strongly biased towards 0 or 1. If we take $\alpha_{x_0=0} \geq \alpha_{x_0=1}$, a direct calculation shows that the maximum value attained by a strategy ignoring the first bit is given by

$$\mathcal{F}_Q^1 = \frac{1}{3} \left[\alpha_{x_0=0} + 1 + \frac{1}{\sqrt{2}} \right], \quad (\text{C9})$$

which becomes larger than F_Q for $\alpha_{x_0} > \frac{1}{2}(1 + \sqrt{3} - \sqrt{2})$. The quantum value of this b -RAC is therefore given by the following piecewise function

$$\mathcal{F}_Q = \begin{cases} \frac{1}{2}(1 + \frac{1}{\sqrt{3}}) & \text{if } \alpha_{x_0} \leq \frac{1}{2}(1 + \sqrt{3} - \sqrt{2}), \\ \frac{1}{3} \left[\alpha_{x_0=0} + 1 + \frac{1}{\sqrt{2}} \right] & \text{otherwise.} \end{cases} \quad (\text{C10})$$

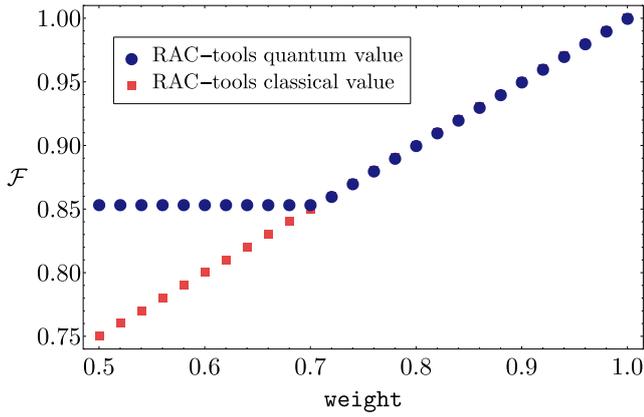


FIG. 11. Optimal performance of the $2^2 \mapsto 1$ b -RAC defined by the bias tensor $\alpha_{xy} = \frac{1}{4}\alpha_{x_0}$. For $\alpha_{x_0} \leq \frac{1}{\sqrt{2}}$, the optimal strategy encodes both bits in the input string, and the quantum value (blue dots) is $\mathcal{F}_Q = F_Q = \frac{1}{2} + (2\sqrt{2})^{-1}$ since the bias has no effect on the functional value. For $\alpha_{x_0} > \frac{1}{\sqrt{2}}$, the best strategy does not encode the first bit, which reduces the value of the functional to the maximum attainable with classical strategies (red squares). For the region in which both bits are encoded, the angles representing the optimal measurements, as obtained by the RAC-tools package, are all $\frac{\pi}{2}$.

Figure 12 depicts the results on the quantum and classical value of this b -RAC produced by the RAC-tools package, in full agreement with the analytical results provided above.

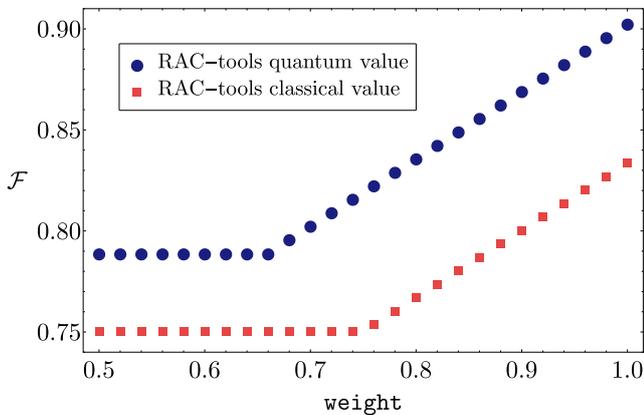


FIG. 12. Optimal performance of the $3^2 \mapsto 1$ b -RAC defined by the bias tensor $\alpha_{x_0x_1x_2} = \frac{1}{4}\alpha_{x_0}$. As observed in the $n = 2$ case, there is a threshold value, $\alpha_{x_0} = \frac{1}{2}(1 + \sqrt{3} - \sqrt{2})$, above which the optimal quantum strategy ignores the first bit. Below the threshold the quantum value (blue dots) is $\mathcal{F}_Q = F_Q = \frac{1}{2}(1 + \frac{1}{\sqrt{3}})$, which coincides with the value of the unbiased RAC. For stronger biases, it becomes convenient to ignore the first bit, in which case the quantum value can be written as a shifted and rescaled version of the unbiased $2^2 \mapsto 1$ RAC. As a result, unlike in the $n = 2$ case, the optimal value above the threshold does not coincide with the classical value (red squares). Similarly to the scenario illustrated in Fig. 11, in the region where there are no ignored bits, the obtained optimal measurements are mutually orthogonal. When x_0 is ignored, measurement 0 becomes proportional to identity, while measurements 1 and 2 remain orthogonal.

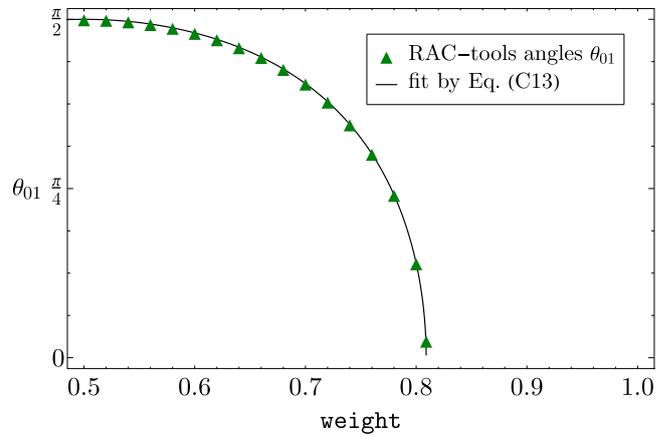
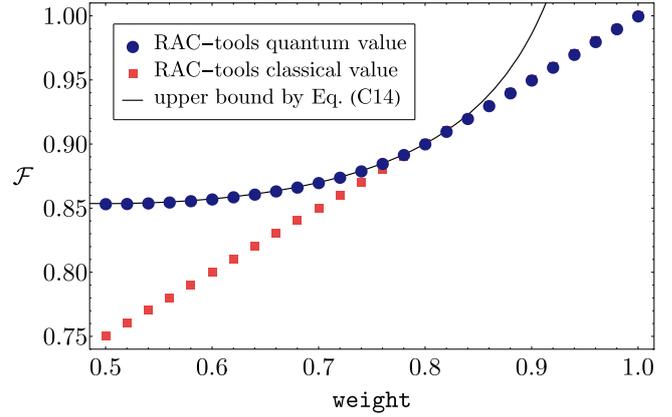


FIG. 13. Top: Numerical quantum value (blue dots) of the biased RAC defined by $\alpha_{xy} = \frac{w_{xy}}{4}$, as a function of parameter w_0 , compared to the corresponding classical value (red squares) and the value in Eq. (C14) (solid line). Quantum strategies provide an advantage for $w_0 \leq \frac{1+\sqrt{5}}{4}$. Bottom: Angle θ between the Bloch vectors parametrizing the optimal measurements. For w_0 within the region of quantum advantage, θ is given by Eq. (C13), becoming 0 outside of it, where $\mathcal{F}_Q = \mathcal{F}_C$.

4. The B_ONE bias family

As described above, the elements of a bias tensor in this family take the form $\alpha_{xy} = \frac{1}{n} \frac{1}{m^{n-1}} w_{xy}$. In the particular case of $n = 2$ and $d = m = 2$, which is addressed in Ref. [41], we have

$$\alpha_{xy} = \frac{w_{xy}}{4}, \quad (\text{C11})$$

and a direct calculation using Eq. (23) shows that, for this bias tensor, the b -RAC functional reads

$$\mathcal{F} \leq \frac{1}{2} + \frac{1}{4} \left[\left| \cos\left(\frac{\theta}{2}\right) \right| + \sqrt{1 - 4\mu \cos^2\left(\frac{\theta}{2}\right)} \right] \quad (\text{C12})$$

with $\mu = w_0 w_1$ and θ the angle between the Bloch vectors \mathbf{m}_0 and \mathbf{m}_1 characterizing the measurement operators. A search for critical points in Eq. (C12) shows that there are only two,

satisfying either $\sin(\theta) = 0$ or

$$\cos\left(\frac{\theta}{2}\right) = \frac{1}{\sqrt{4\mu + 16\mu^2}}. \quad (\text{C13})$$

Whenever θ satisfies Eq. (C13) the ensuing value of the functional reads

$$\mathcal{F}_Q = \frac{1}{2} + \frac{1}{8\sqrt{1+4\mu^2}} \left(\frac{1}{\sqrt{2}} + 4\sqrt{\mu} \right), \quad (\text{C14})$$

as previously reported in [41]. As observed in the previous cases the condition $\sin(\theta) = 0$ corresponds to commuting measurement operators, implying that in this case the value coincides with the classical value. The region in which this is the case is easily found to be $w_0 \in [0, \frac{3-\sqrt{5}}{4}] \cup [\frac{1+\sqrt{5}}{4}, 1]$. In Fig. 13, both the quantum and classical values found by the RAC-tools package are depicted and compared with \mathcal{F}_Q in Eq. (C14).

APPENDIX D: RANDOM BIASES FOR THE $2^d \mapsto 1$ RAC

In this Appendix, we provide numerical evidence to support the claim that optimizing over projective measurements should be enough to find the quantum value of $2^d \mapsto 1$ b -RACs. Briefly, we tried to find counter-examples of such b -RACs in which the optimal realization is achieved by non-projective measurements. In order to do that, we exhaustively evaluated the `perform_search` function for $2^d \mapsto 1$ b -RACs with $d \leq 6$. We considered two kinds of biases: fully random biases and random factorizable biases. For the first case, we simply made up a bias tensor α_{xy} the entries of which are uniformly distributed within the region $\alpha_{xy} \geq 0$, for all \mathbf{x}, y , and $\sum_{\mathbf{x}y} \alpha_{xy} = 1$. For the second case, we considered bias tensors such that $\alpha_{xy} = \alpha_x r_y$, where both α_x and r_y are uniformly distributed over the inputs \mathbf{x} and y , respectively.

TABLE III. Samples with random biases for $2^d \mapsto 1$ RACs. This table consists of a compilation of numerical results produced by the RAC-tools package. The first column specifies the integer d used for a given $2^d \mapsto 1$ RAC. The second column specifies one of the two kinds of biases used to generate this data set. The third column specifies how many of such samples we considered for each case. In the two last columns, we show how often we retrieve a realization the measurements of which are all projective (fourth column) or not all projective (fifth column). The number of seeds used for each sample was 3.

d	Bias	No. samples	No. P	No. NP
2	Full random	10 000	10 000	0
2	Factorizable random	10 000	10 000	0
3	Full random	10 000	10 000	0
3	Factorizable random	10 000	9999	1
4	Full random	5000	5000	0
4	Factorizable random	5000	4996	4
5	Full random	2500	2500	0
5	Factorizable random	2500	2500	0
6	Full random	1000	1000	0
6	Factorizable random	1000	1000	0

The numerical results of this computation can be found in Table III. Apart from some pathological examples, all of the obtained realizations make use of projective measurements. This was the case for the vast majority of the computed samples except for five cases of random factorizable biases (four cases for $d = 4$ and one case for $d = 3$) in which the random draw of r_y was almost deterministic. For these cases, since the weight in one of the measurements is almost zero, the classical and quantum values are numerically very close and due to insufficient numerical precision the optimization terminates with a nonprojective quantum strategy.

- [1] D. Deutsch and R. Jozsa, *Proc. R. Soc. London A* **439**, 553 (1992).
- [2] L. K. Grover, in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96 (ACM Press, New York, 1996), pp. 212–219.
- [3] E. Bernstein and U. Vazirani, *SIAM J. Comput.* **26**, 1411 (1997).
- [4] D. R. Simon, *SIAM J. Comput.* **26**, 1474 (1997).
- [5] A. Montanaro, *npj Quantum Inf.* **2**, 15023 (2016).
- [6] P. W. Shor, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society, Washington, 1994), pp. 124–134.
- [7] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
- [8] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Rev. Mod. Phys.* **86**, 419 (2014).
- [9] R. Uola, A. C. S. Costa, H. C. Nguyen, and O. Gühne, *Rev. Mod. Phys.* **92**, 015001 (2020).
- [10] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-A. Larsson, *Rev. Mod. Phys.* **94**, 045007 (2022).
- [11] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Phys. Rev. Lett.* **70**, 1895 (1993).
- [12] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, *Nature (London)* **390**, 575 (1997).
- [13] D. Boschi, S. Branca, F. De Martini, L. Hardy, and S. Popescu, *Phys. Rev. Lett.* **80**, 1121 (1998).
- [14] S. Pirandola, J. Eisert, C. Weedbrook, A. Furusawa, and S. L. Braunstein, *Nat. Photon.* **9**, 641 (2015).
- [15] C. H. Bennett and S. J. Wiesner, *Phys. Rev. Lett.* **69**, 2881 (1992).
- [16] K. Mattle, H. Weinfurter, P. G. Kwiat, and A. Zeilinger, *Phys. Rev. Lett.* **76**, 4656 (1996).
- [17] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, 2011).
- [18] C. H. Bennett and G. Brassard, in *Proceedings of the International Conference on Computers, Systems and Signal Processing* (Bangalore, 1984), pp. 175–179.
- [19] K. Inoue, E. Waks, and Y. Yamamoto, *Phys. Rev. Lett.* **89**, 037902 (2002).
- [20] F. Grosshans, G. Van Assche, J. Wenger, R. Brouri, N. J. Cerf, and P. Grangier, *Nature (London)* **421**, 238 (2003).
- [21] D. Stucki, N. Brunner, N. Gisin, V. Scarani, and H. Zbinden, *Appl. Phys. Lett.* **87**, 194108 (2005).

- [22] F. Xu, X. Ma, Q. Zhang, H.-K. Lo, and J.-W. Pan, *Rev. Mod. Phys.* **92**, 025002 (2020).
- [23] H.-W. Li, M. Pawłowski, Z.-Q. Yin, G.-C. Guo, and Z.-F. Han, *Phys. Rev. A* **85**, 052308 (2012).
- [24] J. Bowles, M. T. Quintino, and N. Brunner, *Phys. Rev. Lett.* **112**, 140407 (2014).
- [25] E. Passaro, D. Cavalcanti, P. Skrzypczyk, and A. Acín, *New J. Phys.* **17**, 113010 (2015).
- [26] A. Acín and L. Masanes, *Nature (London)* **540**, 213 (2016).
- [27] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, in *Proceedings of the Thirty-First Annual ACM Symp. on Theory of Computing* (ACM Press, New York, 1999), pp. 376–383.
- [28] A. Nayak, in *Proceedings of the 40th Annual Symp. on Foundations of Computer Science* (IEEE Computer Society, Washington, 1999), pp. 369–376.
- [29] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, *J. ACM* **49**, 496 (2002).
- [30] A. Ambainis, D. Leung, L. Mancinska, and M. Ozols, *arXiv:0810.2937* v3.
- [31] M. Pawłowski and N. Brunner, *Phys. Rev. A* **84**, 010302(R) (2011).
- [32] A. Chaturvedi, M. Ray, R. Veynar, and M. Pawłowski, *Quantum Inf. Proc.* **17**, 131 (2018).
- [33] A. Tavakoli, J. Kaniewski, T. Vértesi, D. Rosset, and N. Brunner, *Phys. Rev. A* **98**, 062307 (2018).
- [34] M. Farkas and J. Kaniewski, *Phys. Rev. A* **99**, 032316 (2019).
- [35] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, *Nature (London)* **461**, 1101 (2009).
- [36] H. Buhrman and R. de Wolf, in *Proceedings of the 16th Annual IEEE Conf. on Computational Complexity* (IEEE, New York, 2001), pp. 120–130.
- [37] D. Gavinsky, J. Kempe, O. Regev, and R. de Wolf, *SIAM J. Comput.* **39**, 1 (2009).
- [38] A. Acín, S. Massar, and S. Pironio, *Phys. Rev. Lett.* **108**, 100402 (2012).
- [39] T. H. Yang and M. Navascués, *Phys. Rev. A* **87**, 050102(R) (2013).
- [40] C. Bamps and S. Pironio, *Phys. Rev. A* **91**, 052111 (2015).
- [41] P. Kaczyńska, Bachelor’s thesis, Faculty of Physics, University of Warsaw (2021).
- [42] J. F. Doriguello and A. Montanaro, *Quantum* **5**, 402 (2021).
- [43] R. Gallego, N. Brunner, C. Hadley, and A. Acín, *Phys. Rev. Lett.* **105**, 230501 (2010).
- [44] N. Gisin and R. Thew, *Nat. Photon.* **1**, 165 (2007).
- [45] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, *Rev. Mod. Phys.* **74**, 145 (2002).
- [46] R. F. Werner and M. M. Wolf, *Quantum Inf. Comput.* **1**, 1 (2001).
- [47] T. Ito, H. Imai, and D. Avis, *Phys. Rev. A* **73**, 042109 (2006).
- [48] J. Bae and L.-C. Kwek, *J. Phys. A: Math. Theor.* **48**, 083001 (2015).
- [49] G. Pereira Alves, J. Kaniewski, and N. Gigena, <https://github.com/jedreky/QRAC-tools> (2022).
- [50] E. D. Andersen and K. D. Andersen, in *High Performance Optimization*, edited by H. Frenk, K. Roos, T. Terlaky, and S. Zhang (Springer US, Boston, 2000), pp. 197–232.
- [51] MOSEK-ApS, <https://www.mosek.com/> (2022).
- [52] A. Tavakoli, M. Farkas, D. Rosset, J.-D. Bancal, and J. Kaniewski, *Sci. Adv.* **7**, eabc3847 (2021).

3.3 Comments on “*Machine Learning meets the CHSH scenario*”

In this work, we investigate a variety of Machine Learning (ML) models to address the membership problem of quantum correlations within the CHSH non-signalling set. We approached this problem by framing it as a classification task and initially concentrated on the CHSH correlation space, where we created a toy model for our study. Given that the boundaries of the quantum set in this region are analytically known, the toy model served as an effective benchmark for assessing the accuracy of different ML models. We experimented with both simple models, like Naive Bayes and Decision Trees, as well as more sophisticated ones, such as Gaussian Processes and Dense Neural Networks. Among the models tested, Support Vector Machines and Neural Networks were the most reliable, showing good efficiency and accuracy. We then expanded these successful models to encompass the entire CHSH scenario, where a comprehensive analytical description of the quantum set boundaries is not available. The results indicated that, while achieving good average performance is relatively easy, training a model to handle more complex cases, such as correctly classifying points near the boundary between sets, is quite challenging.

The student’s contributions to this project include the development of data generation and classification techniques, with the steered see-saw algorithm being particularly noteworthy, as well as the creation of Support Vector Machine models for the 4D and 8D variants of the CHSH scenario. The student also significantly contributed to the writing of the text, which includes Secs. I, II, and III, parts of Sec. IV, and Appendix A, in addition to creating figures and plots.

Machine learning meets the CHSH scenario

Gabriel Pereira Alves^{1,*} Nicolas Gigena^{2,†} and Jędrzej Kaniewski^{1,‡}

¹*Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland and*

²*IFLP/CONICET and Departamento de Física, Universidad Nacional de La Plata, C.C. 67, La Plata (1900), Argentina.*

(Dated: July 22, 2024)

In this work, we perform a comprehensive study of the machine learning (ML) methods for the purpose of characterising the quantum set of correlations. As our main focus is on assessing the usefulness and effectiveness of the ML approach, we focus exclusively on the CHSH scenario, both the 4-dimensional variant, for which an analytical solution is known, and the 8-dimensional variant, for which no analytical solution is known, but numerical approaches are relatively well understood. We consider a wide selection of approaches, ranging from simple data science models to dense neural networks. The two classes of models that perform well are support vector machines and dense neural networks, and they are the main focus of this work. We conclude that while it is relatively easy to achieve good performance on average, it is hard to train a model that performs well on the "hard" cases, i.e., points in the vicinity of the boundary of the quantum set. Sadly, these are precisely the cases which are interesting from the academic point of view. In order to improve performance on hard cases one must, especially for the 8-dimensional problem, resort to a tailored choice of training data, which means that we are implicitly feeding our intuition and biases into the model. We feel that this is an important and often overlooked aspect of applying ML models to academic problems, where data generation or data selection is performed according to some implicit subjective criteria. In this way, it is possible to unconsciously steer our model, so that it exhibits features that we are interested in seeing. Hence, special care must be taken while determining whether ML methods can be considered objective and unbiased in the context of academic problems.

I. INTRODUCTION

Proposed in 1964 [1], Bell non-locality [2] establishes that the predictions of quantum mechanics are inconsistent with local hidden-variables (LHV) models, constituting a fundamental aspect of quantum correlations in spatially separated systems. The so-called Clauser-Horne-Shimony-Holt (CHSH) scenario [3] is the simplest scenario in which the correlations observed by two parties can be verified as not admitting a local-realistic explanation, assuming that the shared quantum state is entangled and that the pair of dichotomic measurements implemented by each party is incompatible [4]. Such correlations are identifiable through the violation of Bell inequalities, and assessing the extent of these violations has always been a question of broad interest. Tsirelson was one of the first to look into the topic, deriving, in 1980 [5], an upper bound for the violation of the CHSH inequality, a limit that would later come to be known as Tsirelson's bound. In the same direction, Tsirelson himself, in 1987 [6], followed by others [7–9], obtained a tighter characterisation in the form of a nonlinear inequality restricting the set of quantum correlations, currently known as the TLM inequality. The latter, even if satisfied by any quantum correlation, is sufficient only when the local outcomes of Alice and Bob have uniform distributions.

Exploiting the fact that the set of quantum correlations is convex, two complementary heuristic approaches

are often used to find maximum values of a given Bell inequality. On the one hand, outer approximations to the quantum set can be obtained through the NPA hierarchy [10] of semi-definite programs (SDPs), which is guaranteed to converge to the quantum set. In the CHSH scenario with uniformly distributed outcomes, the TLM inequalities can be recovered at the first level of this hierarchy, but it is currently unknown if it converges to the quantum set at a finite level in any other scenario. On the other hand, the see-saw optimisation [11] provides an inner approximation that can also be cast as an SDP [12]. In this latter procedure, from a random initialisation, state and measurements are optimised separately and iteratively, so that after a certain number of iterations the objective function converges to a local maximum. Unlike the NPA approach, the see-saw optimisation depends on the quantum state dimension, and due to its non-deterministic initialisation, the yielded output is also non-deterministic. Both techniques, NPA and see-saw, when applied to a given Bell functional, often converge to the same value [2], making this combination a valuable tool for exploring the boundary of the set of quantum correlations.

While the approaches mentioned above were specifically designed for the problem of quantum correlations, machine learning (ML) methods are based on the idea that by providing a sufficient amount of data, it is possible to approach a given problem without relying on any prior intuition. In fact, with the emergence of such tools, many areas of physics [13] and, more specifically, quantum information [14, 15], have benefited from their application in addressing problems without analytical solutions or even those the computation of which is costly. With Bell non-locality, the use of ML was no different,

* gpereira@fuw.edu.pl

† nicolas.gigena@fisica.unlp.edu.ar

‡ jkaniewski@fuw.edu.pl

and advances in quantifying non-locality in Bell scenarios [16], networks [17], or even to assist the solution of feasibility SDPs [18] were made, to mention a few.

Building upon the foundational work in non-locality and the advances in ML, this work aims to use ML techniques, specifically support vector machines (SVMs) and neural networks (NNs), to delve deeper into the exploration of the quantum set. Differently from previous methodologies that applied varied methods to specific problems, here we concentrate on the simplest Bell scenario, assessing distinct models to identify the most effective approaches. Our methodology involves using the subspace of the CHSH scenario where the TLM inequalities hold as a benchmark for exploring the entire quantum set, framing the task as a classification problem. As previously pointed out, this region of the quantum set has an analytical description, allowing us to develop a model to be compared with the results already known. By doing so, we aim to achieve an understanding of the models' efficacy in this confined scenario, subsequently extrapolating the trained models to the entire quantum set. This strategy does not only ensure a solid foundation for the models, thanks to the available analytical description, but also allows us to estimate the limits of our techniques in unexplored regions of the quantum set. In essence, we aim to find out which ML methods work best for quantum correlations, what kind of accuracy can be expected, and how expensive they are compared to the currently used methods.

This work is organized as follows. In Sec. II, we provide a preliminary introduction to the main concepts and notations regarding Bell non-locality, as well as present the basic notions of machine learning. Sec. III discusses the data generation methods for correlations only and also for the entire scenario. We conclude this section by commenting on a novel technique to close the gap between NPA and see-saw for non-exposed points in the boundary of the quantum set. Sec. IV discusses the ML models utilized, and the results obtained for both SVMs and NNs. We close our considerations in Sec. V.

II. PRELIMINARIES

Contrary to what is known for the set of non-signalling (\mathcal{NS}) and local (\mathcal{L}) correlations, which in every single-sourced Bell scenario are polytopes, the set of quantum correlations (\mathcal{Q}) has a more complex shape. Characterising its boundary is difficult, and a partial solution is known only in the CHSH scenario. In this section, we introduce the established concepts and notations regarding the CHSH scenario, followed by basics of ML.

A. Correlation sets in the CHSH scenario

The CHSH scenario consists of a bipartite setup where the two parties, Alice and Bob, share a state in which

they are allowed to perform two dichotomic measurements. We label Alice's and Bob's measurements by $x, y \in \{0, 1\}$, whereas their respective outcomes by $a, b \in \{\pm 1\}$. Given the measurements x and y , the probability of obtaining outcomes a and b , respectively, is written as $p(a, b | x, y)$. The 16-tuple of all joint probabilities is named *correlation* or *behaviour*, which we can think of as a point in \mathbb{R}^{16} . Because of normalisation and non-signalling constraints, the non-signalling polytope lives in an 8-dimensional subspace of \mathbb{R}^{16} , thus making it possible to represent it with a lower dimensional parametrisation. One of the most widely used is that in terms of *marginals* and *correlators* $\{\langle A_x \rangle, \langle B_y \rangle, \langle A_x B_y \rangle\}_{x,y}$, related to the probabilities via the following expressions:

$$\begin{aligned} \langle A_x B_y \rangle &:= \sum_{a,b \in \{\pm 1\}} ab p(a, b | x, y), \\ \langle A_x \rangle &:= \sum_{a \in \{\pm 1\}} a p_A(a | x) \quad \text{and} \\ \langle B_y \rangle &:= \sum_{b \in \{\pm 1\}} b p_B(b | y), \end{aligned} \quad (1)$$

where $p_A(\cdot)$ and $p_B(\cdot)$ denote the marginal distributions of $p(a, b | x, y)$ over a and b , respectively. In this case, the set of non-signalling behaviours is represented by a subset of the hypercube in \mathbb{R}^8 defined by the coordinates $\langle A_x \rangle, \langle B_y \rangle, \langle A_x B_y \rangle \in [-1, 1]$. Of particular interest for us is the subspace satisfying $\langle A_x \rangle = \langle B_y \rangle = 0$, i.e., the restriction of the set of behaviours to the *correlation space*, since for this subset the TLM inequalities and the positivity facets fully describe the boundary of the quantum set.

Quantum behaviours are those for which a bipartite state ρ_{AB} and local measurement operators $\{M_{a|x}, N_{b|y}\}$ exist such that

$$p(a, b | x, y) = \text{tr}(\rho_{AB} M_{a|x} \otimes N_{b|y}), \quad \forall a, b, x, y. \quad (2)$$

We can directly compute the correlators as the expected values of the observables, defined by $A_x = M_{1|x} - M_{-1|x}$ and $B_y = N_{1|y} - N_{-1|y}$, that is,

$$\langle A_x B_y \rangle = \text{tr}(\rho_{AB} A_x \otimes B_y). \quad (3)$$

We refer to the triple $\{A_x, B_y, \rho_{AB}\}$ as a *realisation* and \mathcal{Q} is defined as the set of behaviours for which such a realisation exists.

As already mentioned, the boundary of \mathcal{Q} restricted to the correlation space admits an analytical description in terms of the TLM inequalities, the elements of which can be obtained by cyclically permuting the signs of

$$\begin{aligned} &|\arcsin \langle A_0 B_0 \rangle + \arcsin \langle A_0 B_1 \rangle \\ &+ \arcsin \langle A_1 B_0 \rangle - \arcsin \langle A_1 B_1 \rangle| \leq \pi. \end{aligned} \quad (4)$$

This characterisation allows to randomly generate a set of correctly labelled behaviours that can be used to train

an ML model. Unfortunately, a similar description of the boundary for the entire quantum set is lacking, which prevents a straightforward classification of behaviours as quantum or not quantum. Nonetheless, before addressing the topic of data classification, which is discussed in detail in Sec. III, let us briefly introduce the fundamental concepts and notation of the field of ML.

B. Machine learning in a nutshell

Machine Learning is an interdisciplinary field at the intersection of computer science and statistics that is dedicated to developing methods that allow computers to learn and make decisions based on data [19]. A fundamental aspect of ML is its ability to build predictive models that can analyse data and identify patterns or trends. Such ability is advantageous in applications where traditional analysis methods fail to provide reliable predictions.

Conventionally, ML models are split into three categories, namely supervised learning, unsupervised learning, and reinforcement learning. For the purposes of this work, we will concentrate on supervised learning, i.e., a technique which is characterised by the use of labelled datasets to train the models [20]. These datasets consist of input-output pairs, where the model is provided with input data, often referred to as *features*, along with the correct *labels*, allowing it to learn how to predict the output from the input.

There are two main types of tasks in supervised learning, which basically differ in the nature of the output: regression tasks deal with continuous output, e.g., predicting house prices, stock market trends, or temperature forecasts, while classification tasks deal with predicting discrete outputs, typically categories or classes. Our work presents the quantum correlation problem as a classification task in which the features are given by the CHSH behaviour, as in Eq. (1), and the labels consist of a binary class dividing each feature as being quantum or not quantum.

From an operational point of view, constructing an ML model typically involves a few elementary steps, namely data preprocessing, model selection, training and testing. In our particular context, this paradigm shifts slightly as we do not use pre-existing data but rather generate it ourselves, a process detailed in the subsequent section. Following data preprocessing, further steps involve selecting and training a given ML model. The choice of model depends on the nature of the task (e.g., regression, classification), the characteristics of the data and the expected difficulty of the task. Common models in supervised learning include linear regression, logistic regression, support vector machines, and decision trees, among others. More complex problems may require advanced techniques such as neural networks.

Once the data and model are fixed, the subsequent step is to divide the dataset into training and testing

sets. This division is critical for evaluating the model’s performance on unseen data, which is a proxy for how it will perform in real-world scenarios. A common practice is to allocate a larger portion of the data for training (e.g., 70-80%) and a smaller portion for testing (e.g., 20-30%). Here, due to the nature of the chosen models, we also include a validation set, splitting the generated data into validation, training and testing sets, following a 15-70-15 proportion, respectively. After the data is split, the training process involves using the validation set to tune the hyperparameters of the model. This process is iteratively performed to find the optimal parameter values that minimise a predefined loss function, which quantifies the difference between the predicted and actual values.¹ The split between training and validation is crucial as it ensures that the evaluation obtained through the test set consists of a realistic estimate of the model’s performance, which brings us to the last step of model development.

The assessment of the model involves using evaluation metrics applied to the test set, which may vary based on the type of task. For instance, accuracy, precision, recall, and the F1 score are commonly used for classification tasks, while mean squared error or mean absolute error are used for regression tasks. The evaluation phase not only provides insights into the model’s performance but also helps in diagnosing issues such as overfitting, where the model performs well on the training data but poorly on previously unseen examples.

Although the description of the model development process may appear linear based on the sequence of outlined steps, in reality it rarely follows this path exactly. This divergence is evident in our approach, where some steps are skipped or modified. For this reason, we emphasise that the last paragraphs are not intended to describe our model construction specifically, but to provide a basic understanding of ML to a non-specialised audience. In the following, we introduce the two classes of models employed in this work, namely support vector machines and dense neural networks, as well as discuss why we consider this choice suitable for the problem of quantum correlations.

1. Support vector machines

Support vector machines are a class of supervised learning models used primarily for classification, but also for regression tasks [21]. Initially, they were designed to solve linearly separable problems [22], where the data can be split by a flat boundary, that is, a straight line, a plane

¹ It should be noted that the term “training” referred here is what in statistics is referred to as “fitting”, i.e., the optimisation of parameters in statistical models (like linear regression). ML’s “training” not only mirrors this aspect but also includes the feature of boosting model performance through data exposure.

or a hyperplane, depending on the dimension of the features. In this sense, the goal of the SVM algorithm is to find the optimal hyperplane separating classes, which is defined as the one maximally distant from data points of either class. It follows then that the model’s performance strictly depends on the data points closest to this optimal hyperplane, denoted as *support vectors*, which preserves the method’s efficiency as it scales to higher dimensions.

The real breakthrough in the utility of SVMs came with the realisation that they could also be adapted to solve non-linear problems through the kernel trick [23, 24]. This approach involves applying a kernel function to map the original non-linearly separable data into a higher-dimensional space, where it becomes linearly separable. Common kernel functions include the polynomial, radial basis function (RBF), and sigmoid kernels. By using this trick, SVMs can efficiently classify data that is non-linearly separable in its original feature space, which makes it a natural candidate for the problem tackled in this work. In other words, SVMs not only exhibit good scalability, which is desirable for Bell scenarios where the dimensionality of correlations is typically high, but they also handle the non-linear aspects of the boundary of the quantum set. This contrasts with linear ML models such as Bayesian or logistic regression, which generally fail in such contexts.

2. Neural networks

Inspired by the neural architecture of the human brain, neural networks [25] are a sophisticated class of ML models consisting of interconnected nodes, also known as *neurons*, which collectively process and interpret data inputs. The advantage of such models lies in their versatility to learn and adapt to a wide variety of tasks, ranging from image recognition to natural language processing.

In this work, we focus on one of the simplest types of neural networks denominated as multi-layer perceptrons (MLPs) [26], a type of feedforward neural network. The “feedforward” aspect refers to the unidirectional flow of data, i.e., the data supplied to a given neuron does not return to previous neurons, which guarantees a non-cyclical flow of information with no loops or backward connections. This architecture is particularly beneficial for tasks where the current output depends solely on the current input, with no dependence on the context provided by previous inputs.

In this sense, MLPs exemplify this feedforward structure, featuring layers of interconnected neurons divided into an input layer, one or more hidden layers and an output layer, as illustrated in FIG. 1. Each layer in an MLP is structured in a way which is commonly referred to as *dense*, where every neuron in a given layer is connected to all neurons in the next layer through a set of adjustable weights. During its operation, each neuron first computes a weighted sum of its inputs and then applies an activation function. The activation function is

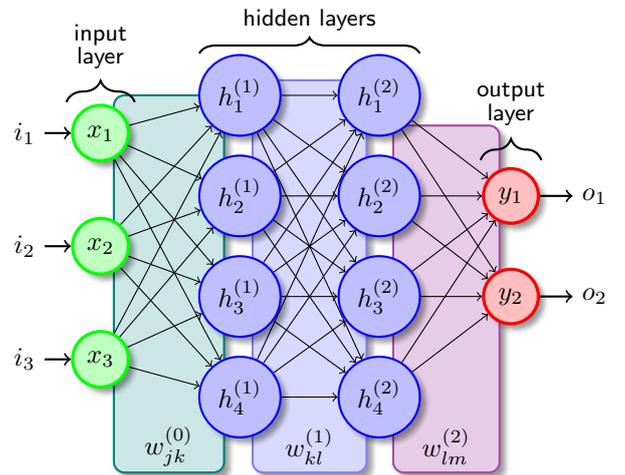


FIG. 1. A typical MLP diagram showing two hidden layers. The initial layer consists of three neurons x_j in which the inputs i_j are passed, for $j = 1, 2, 3$. Each input neuron is connected to the k -th neuron $h_k^{(1)}$ in the first hidden layer through the set of weights $w_{jk}^{(0)}$, for $k = 1, 2, 3, 4$. Similarly, the neurons in the first hidden layer are connected to $h_l^{(2)}$ through the weights $w_{kl}^{(1)}$, for $l = 1, 2, 3, 4$. The second hidden layer ends up in the output layer via $w_{lm}^{(2)}$, where $m = 1, 2$. At the end of the process, y_1 and y_2 yield output o_1 and o_2 , respectively.

crucial as it introduces non-linearity in the model, which allows the network to capture complex patterns from the data. Common activation functions include the sigmoid, hyperbolic tangent, and ReLU (Rectified Linear Unit) [27, 28].

The learning ability of MLPs relies on their method of iteratively refining weights during training. At this stage, the network undergoes a process denoted as forward pass, where each training example is fed through the network from the input to the output layer, generating a prediction. Then, the loss calculation phase assesses the accuracy of these predictions comparing them to the true values with a loss function, which quantifies the error in the previous step. The network then computes the gradient of the loss function with respect to each weight in a process known as backward pass or backpropagation. In the last step, an optimisation algorithm, often gradient descent, adjusts the weights to minimise the loss function. Completing all of these steps for all training examples constitutes one epoch, marking a full cycle through the training data. After multiple epochs, the resulting model should be able to yield refined predictions through the output layer, where neurons are configured to match the specific outputs needed. For classification tasks, the number of neurons in this layer typically corresponds to the number of classes the model should predict, whereas, for regression, a single neuron produces a continuous value. In other words, the goal of the output layer is to transform the outputs of the last hidden layer into a format suitable for the problem.

In summary, MLPs are versatile tools, being well-suited for a wide range of problems, including the one addressed in this work. Additionally, Ref. [29] points out that if all activation functions are convex and non-decreasing, and all weights are non-negative, the output is guaranteed to be a convex function of the inputs. Such an observation can be applied to MLPs by simply constraining the weights to be non-negative, since most activation functions are already convex and non-decreasing. Here, we use this idea to exploit the convexity of our problem and ensure that the MLP model produces a convex output. Nonetheless, before we enter into the specifics of fitting the model, as discussed in Sec. IV, let us shift our focus to the process of data generation.

III. DATA GENERATION

Model and data are the two main components of a typical ML problem. Distinct models have different computational capabilities and requirements, but most are generally adaptable for various purposes. Naturally, some models will perform better than others, in terms of accuracy and efficiency, given a particular dataset. For this reason, part of the challenge in the search for an ML solution consists of deciding which tool is better suited for the data at hand, which is usually a constraint determined by the problem. However, in the scenario considered here, there is no data available in advance, leaving the only alternative to generate it by ourselves. As this is not a trivial task, especially in distinguishing behaviours between quantum and not quantum, the following subsections are devoted to this discussion, presenting a few different ways to sample and classify non-signalling behaviours.

But before we go any further, let us point out that producing training data may seem counterintuitive here, as we need to classify behaviours first and this is also the task we want our model to perform. Nonetheless, it is still worth to look for an ML model for two reasons. Firstly, an ML model would effectively combine the strengths of the methods we use to generate data, such as see-saw and NPA, consolidating them into a single, unified approach. Secondly, once trained, the resulting model has the potential to be more efficient in terms of computation time and memory than the original methods. This makes exploring the development of a well-trained ML model from a relatively small set of classified examples highly worthwhile.

A. Uniform sampling

In our first approach to generate data, we create a uniform sample within the \mathcal{NS} polytope. This is achieved using an algorithm called *hit-and-run*, which employs a random walk starting from an interior point, as implemented in Ref. [30]. The hit-and-run method enables

the generation of uniformly distributed points across any convex bounded shape where the boundaries are known analytically. The sample is produced by inputting the hyperplane representation of \mathcal{NS} and a starting point, which we chose to be its geometric centre, corresponding to the origin of the coordinate system in both 4 and 8-dimensional variants of the CHSH scenario. For the correlation space, the hyperplane representation of \mathcal{NS} is given by the facets of a cube in dimension 4, whereas the same representation for the entire scenario is expressed by the non-signalling facets, i.e.,

$$(-1)^{a+b+1} \langle A_x B_y \rangle + (-1)^a \langle A_x \rangle + (-1)^b \langle B_y \rangle \leq 1, \\ \forall a, b, x, y, \quad (5)$$

where, here, the outcomes are labelled as $a, b \in \{0, 1\}$.

Once the sample is generated, the classification of points as quantum or not quantum is straightforward for correlations only, as the first level of NPA is enough to solve the membership problem. Nonetheless, for the entire CHSH scenario there is no analogous method allowing to easily classify correlations points, and the best alternative is to approximate \mathcal{Q} with the superset \mathcal{Q}_n defined by the n -th level of NPA. The drawback is that the SDP associated with the membership of a given level becomes costly as n increases, creating a compromise between the quantity of resources to be used and the quality of the approach. In FIG. 2, this question is addressed with the comparison between the relative volumes of the first NPA supersets, showing an advantage in terms of computation time to level $1 + AB$. Although NPA converges when $n \rightarrow \infty$, the quality of the approximation for large n does not improve drastically when compared to previous levels. Furthermore, the execution time increases in a way that makes the computation infeasible for datasets larger than 10^5 points, even for level $n = 5$. In this way, we aim to create datasets for training ML models to decide whether a given behaviour belongs to \mathcal{Q}_{1+AB} or not, which, in good approximation, would tell us if it belongs to \mathcal{Q} .

Also, it is worth noting that the volumes spanned by \mathcal{Q} and \mathcal{NS} in the CHSH scenario are almost the same, making the volume ratio close to one². A consequence of this feature is that a model that trivially predicts every point to be local can be considered good if accuracy with respect to a uniformly distributed sample is taken as a figure of merit to evaluate performance. Moreover, it is likely that an unbiased loss function would prioritise the larger class in the problem. Here we consider three ways in which this data imbalance can be handled. First, we use balanced loss functions, like the balanced binary cross entropy, for binary classification, and balanced figures of merit for further evaluation of the models. Secondly, we

² For the correlation space, this number is known, and it is approximately 0.925 [31, 32].

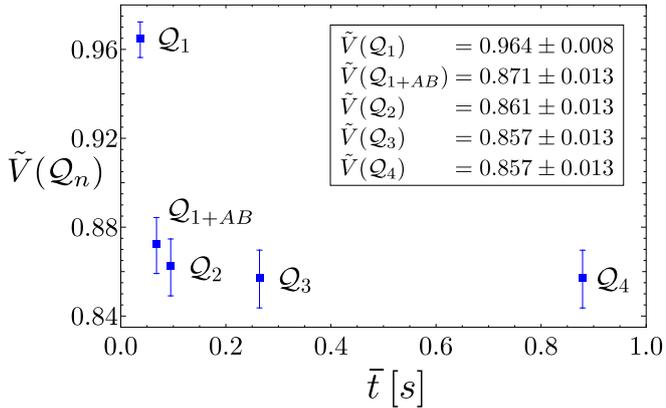


FIG. 2. Relative volumes of Q_n vs. average computation time \bar{t} . This plot was produced by computing the volume ratios of Q_n with respect to \mathcal{NS} using a simple Monte Carlo method. Briefly, a sample of points uniformly distributed in \mathcal{NS} is generated, and the membership is solved for each point. Note that here, the volumes of Q_n are considered regarding the 8-simplex violating one of the CHSH facets, i.e., the sample produced is composed only of non-local points. For each NPA level, an initial sample of 10 000 points is created and classified, a procedure repeated 50 times. The average CPU time \bar{t} per point is shown in the x -axis. It is possible to observe that levels higher than $1 + AB$ provide a small improvement towards the actual volume of Q at the cost of a significant increase in CPU time. The CPU clock speed used is 2.9 GHz (base) and 3.9 GHz (boost).

generate a balanced sample for the two classes, which would make a model that trivially predicts all points as belonging to one class perform badly in terms of accuracy. Lastly, we introduce a technique that samples the training data only near the boundary between classes. Intuitively, this seems a smart way to proceed, since focusing the training data near the boundary naturally leads to a more difficult problem, as points far away from the boundary are easier to predict. Consequently, achieving greater accuracy in such a model would imply a decision boundary that resembles the true boundary more accurately. To describe this idea in more detail, we move on to the next subsection.

B. Sampling near the quantum boundary

Briefly speaking, the idea behind this method is to find a correlation point on the boundary of Q , and, from this point, generate a pair of training points belonging to each of the classes. For the correlation space, since $Q_1 = Q$, the procedure is straightforward and is implemented with the first level of NPA. To do so, we begin with the origin in \mathbb{R}^4 as the starting point, and draw a random unit direction \mathbf{u} in the 4-sphere. Then, by solving the SDP,

$$\begin{aligned} \max \quad & \lambda \\ \text{s.t.} \quad & \lambda \mathbf{u} \in Q_1, \end{aligned} \quad (6)$$

we can find a solution λ^* which produces a point $\mathbf{p}_b = \lambda^* \mathbf{u}$ in the boundary of Q_1 . The training points are then created from \mathbf{p}_b by considering the pair

$$\mathbf{p}_{\pm} = (1 \pm \epsilon) \mathbf{p}_b, \quad (7)$$

where ϵ is a fixed offset. This ensures that the training set is balanced by construction, since the classification is done according to the offset sign, i.e., \mathbf{p}_- is always quantum, while \mathbf{p}_+ is not. Also, through direct numerical verification, it is possible to check that the classification in Eq. (7) compared to the analytical TLM conditions results in correctly classified samples for offsets greater than 10^{-10} , which is sufficient for our numerical purposes.

As for the entire quantum set, the procedure is analogous, beginning by choosing a direction in \mathbb{R}^8 . Since there is no known finite level of NPA that corresponds to the actual quantum set, Q_{1+AB} replaces Q_1 in Eq. (6). With the method detailed in the following subsection, we show in Appendix A that this approximation generates training points that are mostly correctly classified when $\epsilon \geq 10^{-3}$.

It is also worth mentioning that, while proximity of the training data to the NPA superset is ensured for small ϵ in Eq. (7), the actual distance of the training points may vary depending on \mathbf{u} . Additionally, since \mathbf{u} will be generally tilted with respect to the direction normal to the boundary in the vicinity of \mathbf{p}_b , the density of training points will vary from one region to another. In an attempt to mitigate these issues, we initially devised two variations to the method described in Eqs. (6) and (7) for the correlation space models. The first of them involved performing a second SDP to compute the distance between \mathbf{p}_{\pm} and Q_1 , thereby ensuring a uniform distance for all training points to the set. The second variation involved pre-selecting the directions \mathbf{u} to guarantee a uniform distribution of points \mathbf{p}_b on the hyper-surface of Q_1 . Nonetheless, although the training data produced by these variations differed slightly, their effects on the final ML models were negligible and boiled down to fluctuations in the values of the figures of merit. As a result, we decided to retain only the simplest model, which is outlined in Eqs. (6) and (7).

C. Steered see-saw

Similarly to the see-saw optimisation, the following method also consists of an iterative SDP technique. But here, instead of maximising a given linear functional, we compute the distance of a point \mathbf{p}_b , on the boundary of one of the NPA supersets, to the set of points admitting a quantum realisation of local dimension d , denoted as $Q_{d \times d}$. That is, the central idea essentially mirrors the task of computing the distance between a point and a convex set. However, while $Q_{d \times d}$ is not always convex [33], the procedure is divided so that each step is cast as

a convex problem.

In the first of these steps, we randomly select observables \hat{A}_x and \hat{B}_y , for $x, y \in \{0, 1\}$, acting in a Hilbert space of local dimension d , to write

$$\begin{aligned} \min_{\rho} \quad & \|\mathbf{q} - \mathbf{p}_b\| \\ \text{s.t.} \quad & \mathbf{q} \in \mathcal{Q}_{d \times d}. \end{aligned} \quad (8)$$

Here, the symbol $\|\cdot\|$ denotes the 2-norm, and \mathbf{q} represents a correlation computed out of the averages over \hat{A}_x , \hat{B}_y and the variable state ρ . By using the solution $\tilde{\rho}$ obtained from Eq. (8), we then perform an analogous optimisation for the observables of each part alternately. After completing all steps, the resulting distance is compared to a previously defined bound, set at 10^{-7} . If the distance is smaller than this threshold, the process concludes; otherwise, it iterates for a given number of seeds. That is, this procedure essentially acts as a classification test: \mathbf{p}_b belongs to \mathcal{Q} when the distance falls below the threshold, but no conclusion is drawn if it does not.

When compared to the standard see-saw, the approach outlined here has a couple of advantages. The first is that the optimisation is not carried out in relation to a particular Bell functional, but to a direction in the coordinate system, allowing to explore non-exposed points on the boundary of \mathcal{Q} . Second, even though the membership test may be inconclusive, for a sufficiently large dimension and number of seeds, it can be interpreted as negative, since $\mathcal{Q}_{d' \times d'} = \mathcal{Q}$, for $d' \geq 16$ [34]. On the other hand, the feasibility of this procedure for applications involving a sufficiently large dataset is far from practical. TABLE I presents a CPU time comparison for the methods discussed in this section to produce a sample of the same size. Clearly, the steered see-saw has a significantly

TABLE I. CPU time for each sampling method in the 8-dimensional CHSH space. The first column specifies the method utilised, while the second column shows the total CPU time to classify an initial sample of 10^4 points. The first and second rows show the case of a uniformly distributed sample when the classes are unbalanced and balanced, respectively. The sampling strategy is similar for both methods, with the difference that the balanced sample accumulates non-local points until the size of each class reaches half of the desired sample size. The two last rows show the time taken for the methods discussed in Secs. III B and C, respectively. It is worth noting that the steered see-saw consists of a classification test in \mathcal{Q} , while the other methods use \mathcal{Q}_{1+AB} as reference. For more details regarding the classification with steered see-saw, see Appendix A.

method	total CPU time
uniform sampling (unbalanced)	6 min 52 s
uniform sampling (balanced)	43 min 27 s
sampling near the boundary	3 min 23 s
steered see-saw	\gtrsim 158 days

larger cost than the previous methods. While the other procedures may typically require a couple of SDPs to solve the membership problem, the latter method subjects a single point to a series of SDPs, in addition to the need to iterate over d . The result is that, though we can use the steered see-saw in the small correlation space, producing a training set with this method for the entire 8-dimensional space of the CHSH scenario is computationally challenging.

Nonetheless, despite the fact that we cannot directly use this approach for our ML task, it can still be applied for some other questions. In Appendix A, we assess the quality of the approximation made in Sec. III B when using \mathcal{Q}_{1+AB} as the reference set for classification. In short, we draw a sample of 10^4 unit directions in \mathbb{R}^8 and use Eq. (6) to determine an optimal λ for \mathcal{Q}_{1+AB} . The results are then compared with the steered see-saw using $d = 6$ and 50 seeds. The outcome is that for around 99.46% of the directions starting from the origin in \mathbb{R}^8 , the points produced via Eq. (7) are correctly classified when $\epsilon \geq 10^{-3}$, showing that resolving \mathcal{Q} from \mathcal{Q}_{1+AB} numerically is difficult. Given that ML models are intrinsically probabilistic, this implies that even if a training sample using \mathcal{Q} as a reference is produced, the results obtained with such a dataset would be no different from those produced by using \mathcal{Q}_{1+AB} as a reference.

IV. FITTING THE MODEL

In this section, we discuss the training details of the chosen ML models, as well as present a performance analysis based on the evaluation metrics available for each. In the first part, we show the results of the fitting in the correlation space, subsequently moving to the entire scenario in the second part. Each subsection is divided according to one of the two models considered, namely SVMs and NNs.

It is worth to note that aside from the aforementioned models, we have also experimented with other types of models which can be found in Python's Scikit-learn library [35], such as Gaussian Processes, Naive Bayes, and Decision Trees. However, we have decided to discontinue some of the original approaches due to distinct reasons. On the one hand, Decision Trees were found to lack stability, with slight variations in training data leading to drastic model changes, making it difficult to be reproduced. Moreover, they were sensitive to the choice of coordinate system, i.e., given that the correlations allow for arbitrary choices of coordinates, we cannot predict in advance which coordinate system will lead to the most effective tree structure. In contrast, Naive Bayes and Gaussian Processes demonstrated to be more reliable but were discarded due to Naive Bayes' inferior accuracy and the extensive training time required by Gaussian Processes. Unlike SVMs, which are powered by the C++ library LIBSVM [36] in Scikit-learn, Gaussian Processes are fully implemented in Python. While this implemen-

tation offers great flexibility and ease of use, it may not achieve the same performance as systems optimised in lower-level languages, especially for large-scale problems. Therefore, it clearly became apparent that SVMs and NNs are the most promising directions, which we cover in the following.

A. Benchmark in the correlation space

Our initial interest in this particular subset comes from the fact that here the boundary of \mathcal{Q} is known to be given by the TLM inequalities, also coinciding with the boundary of the first NPA superset \mathcal{Q}_1 . In what follows, we explore the performance of SVMs in producing a decision boundary that approximates that of \mathcal{Q} , i.e., in solving the problem of classifying non-signalling points as quantum vs. not quantum.

1. SVMs on correlations only

As mentioned above, in the development of the models discussed here, the Scikit-learn library was used as the main modelling tool, employing the RBF kernel among the available kernel options. Based on the datasets appearing in the three first rows of TABLE I, we developed three distinct SVM models, which we refer to as unbalanced, balanced, and offset models, respectively. In each case, a sample of 10^4 points was created following the standard split of 70% for the training set and 15% for the test and validation sets each. In all three models, the adjustment of the SVM hyperparameters was performed through a grid search in the hyperparameter space, selecting those values that maximise accuracy in each validation set. For the offset model, in addition to the hyperparameter search, it was also necessary to define the offset parameter ϵ used to produce the sample. In this case, we started tentatively with $\epsilon = 10^{-2}$, obtaining the accuracies shown in TABLE II.

Initially, we considered the accuracy measures relative to three reference sets: the training and testing sets of each model, to assess generalisation within the same data

TABLE II. Accuracies for the correlation SVM models. The second and third columns contain the accuracies computed using the testing and training sets as a reference, respectively. The last column shows the accuracies when the models are tested using the unbalanced test set, i.e., when the test set is taken as a sample of points uniformly distributed within the correlation space of \mathcal{NS} .

model	test acc.	train acc.	unbalanced acc.
unbalanced	0.9926	0.9934	—
balanced	0.9840	0.9901	0.9796
offset ($\epsilon = 10^{-2}$)	0.9307	0.9961	0.9800

type, and the unbalanced test set, to determine generalisation across the entire \mathcal{NS} polytope. Despite the good performance of the three models, the small variations observed in TABLE II make it difficult to evaluate the influence of each dataset and, specifically, the effect of ϵ on the quality of the predictions nearby the boundary. To address this issue, we included an additional accuracy measure, denoted here as *spread test*, which also involves sampling the reference set near the boundary of \mathcal{Q} .

To generate this reference set, we adopted a different sampling method from that discussed in Sec. III B. To put it briefly, the distinction lies in how the initial boundary points are collected; rather than being sampled in a random unit direction, they are uniformly distributed along the hyper-surface defined by the boundary of \mathcal{Q}_1 . This is achieved by creating a uniformly distributed sample within a shell of very small thickness that encloses \mathcal{Q}_1 . The directions of each point in this sample are then used by the method in Sec. III B to find approximations on the boundary of \mathcal{Q}_1 . Lastly, each boundary point is shifted as $[1 + \mathcal{N}(0, \sigma)] \mathbf{p}_b$, where $\mathcal{N}(0, \sigma)$ is a normal distribution centred at zero and with standard deviation σ . Labels are assigned by directly checking the sign of $\mathcal{N}(0, \sigma)$.

The spread test results are detailed in TABLE III. The offset model exhibits good performance at $\sigma = 10^{-2}$, decreasing at $\sigma = 10^{-3}$, whereas the balanced and unbalanced models consistently perform poorly, given that 0.5 can be achieved with a model that guesses the labels randomly. Additionally, this test serves to highlight how the ϵ value influences prediction accuracy near the boundary. More specifically, with an offset at $\epsilon = 10^{-3}$, the accuracies listed in TABLE IV were observed. Although this smaller ϵ leads to a mild overfitting, as shown by the gap between training and testing accuracies, it improves the resolution at the boundary without disturbing the model’s ability to predict other regions. FIG. 3 illustrates the agreement across the unbalanced, balanced, and offset (with $\epsilon = 10^{-3}$) models in two correlation slices, where the decision boundary aligns closely with the analytical boundary shown by the dashed line.

To conclude, it remains to be mentioned that although the smaller offset leads to a better resolution at the boundary, it cannot be further improved. That is, in our experiments, as ϵ diminishes further, finding a good fitting becomes increasingly hard. The gap between train-

TABLE III. Accuracies produced for the spread test. Columns two and three present the values obtained when the test corresponds to boundary points shifted by a normal distribution $\mathcal{N}(0, \sigma)$ with $\sigma = 10^{-2}$ and 10^{-3} , respectively. The smaller the spread of the distribution, the closer the test points are to the boundary of the set.

model	spread acc. ($\sigma = 10^{-2}$)	($\sigma = 10^{-3}$)
unbalanced	0.5430	0.5030
balanced	0.5426	0.5027
offset ($\epsilon = 10^{-2}$)	0.8318	0.5920

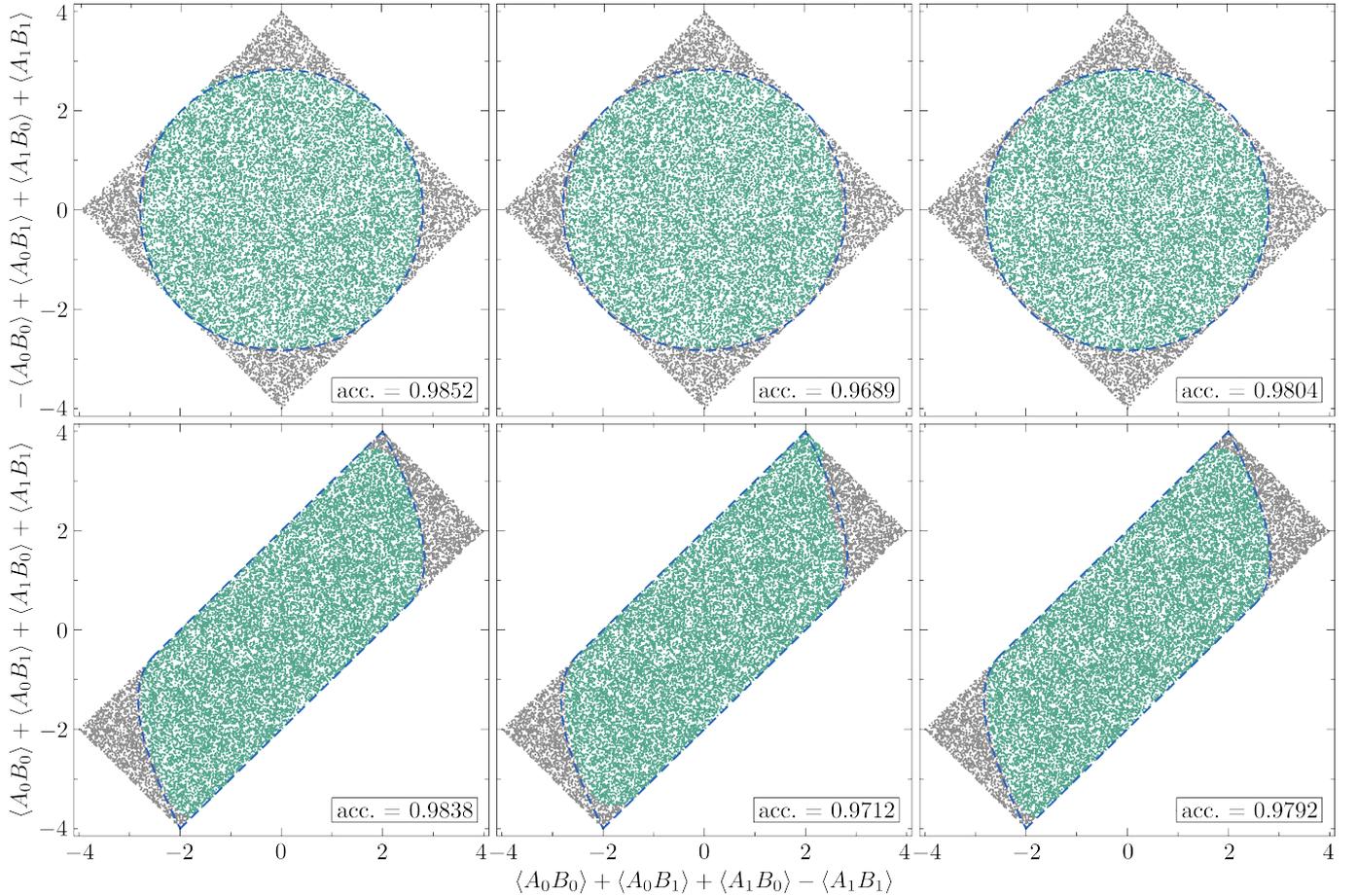


FIG. 3. Correlation slices for the unbalanced (right), balanced (centre) and offset (left) SVM models with $\epsilon = 10^{-3}$. The upper and lower slices were introduced in Refs. [37] and [38], respectively. The green dots depict the quantum class, while gray represent the points classified as outside \mathcal{Q} . The dashed line depicts the true boundary given by the TLM condition. In each plot, 2×10^4 points were generated and classified; the accuracies are indicated in the lower corner for each slice.

TABLE IV. Accuracies for the offset SVM model with $\epsilon = 10^{-3}$. Column one contains a summary of all accuracies appearing in TABLES II and III, and column two displays their values. Compared to $\epsilon = 10^{-2}$, although performance was reduced for the training and test sets, the values for the unbalanced sample improved, as well as for the spread test.

acc. measure	value
test acc.	0.7353
train acc.	0.8128
unbalanced acc.	0.9912
spread acc. ($\sigma = 10^{-2}$)	0.8923
spread acc. ($\sigma = 10^{-3}$)	0.6586

ing and testing accuracies increases while the performance on uniformly distributed data declines, suggesting that there is a limit to how much the offset value can be reduced while maintaining the ability to accurately classify correlation points.

2. NNs on correlations only

Apart from the SVM models mentioned above, we also approached the problem with the implementation of a feedforward neural network. This model was also executed in Python, using TensorFlow [39] and Keras [40]. We considered here the same datasets as in the previous discussion for SVMs. For each of them we configured a network with 64 neurons in the input layer, 16, and 4, respectively, for the first and second hidden layers, and two neurons in the output layer, corresponding to each class in our problem. The chosen activation function is ReLu for all neurons except for the last one, for which we chose Softmax. Note that, since we are dealing with two classes, we could instead use a single neuron in the last layer activated with a Sigmoid function, which is mathematically equivalent in terms of activation. We observed, however, that the first option results in more accurate models. We complemented this basic architecture with an extra feature: Since the boundary between classes is known to be convex, we enforced this constraint on the

decision boundary by forcing the neuron weights to be non-negative, as suggested in Ref. [29]. Nonetheless, we observed the impact of this constraint to be negligible in terms of model performance or training time. The architecture we chose for our network is the result of a trial and error approach, in which we looked for the simplest network that can provide a good model, as measured by our figures of merit.

Regarding the training of the network, a common choice of loss function for classification problems is the Binary Cross Entropy (BCE) function. We have found, however, that a variation of it known as Focal Loss results in models with a better performance. The Focal Loss function is defined in [41] as $FL(p) = -\alpha(1-p)^\gamma \log(p)$, where parameter α is introduced to compensate class imbalances, and the *focusing* parameter γ diminishes the weight of easy-to-classify behaviours on the loss function, so that the training is carried mostly on hard examples, which lie near the boundary between classes. We observe that a suitable choice of the value of these hyperparameters ($\alpha = 10^{-2}, \gamma = 2$) results in a noticeable improvement in terms of accuracy for the models trained with data uniformly sampled inside \mathcal{NS} .

In terms of evaluation, on the other hand, we decided to introduce a custom metric which is a balanced version of the Binary Accuracy metric, as a figure of merit to assess the performance of the trained model. The reason for the introduction of this custom metric is that it is expected to be better suited for assessing model accuracy in the presence of unbalanced datasets. We observed indeed that this metric can resolve better the differences between trained models, despite their relatively small magnitude ($\approx 1 - 10\%$). We trained the neural network using the same training sets that gave rise to the unbalanced, balanced and offset (with $\epsilon = 10^{-3}$) models described previously, with each set split into training, validation and test. In order to maximise the performance of the final model, we trained up to 10 models with different random initial values for the neuron weights, out of which we kept the model with higher accuracy. The results of the neural network training are shown in TABLE V, in which, as before, we show the accuracy values obtained for the test, train and unbalanced test sets. To further illustrate the model performance, we have considered a pair of slices of the CHSH correlation spaces and computed the model accuracy within these slices, which are shown in FIG. 4. To avoid repeating similar plots, in FIG. 4 we only present the results for unbalanced model; for the remaining models we simply show the plot accuracy in TABLE VI obtained for both slices.

We can see that the models trained with unbalanced and balanced data attain good scores for the training, validation and test sets, with a small difference in favour of the unbalanced data set. Good scores are also observed when assessing the model in the slices we have chosen. These results, however, are not observed for the offset model, for which the score is barely above one half. Interestingly, as shown in TABLE V, the performance of these

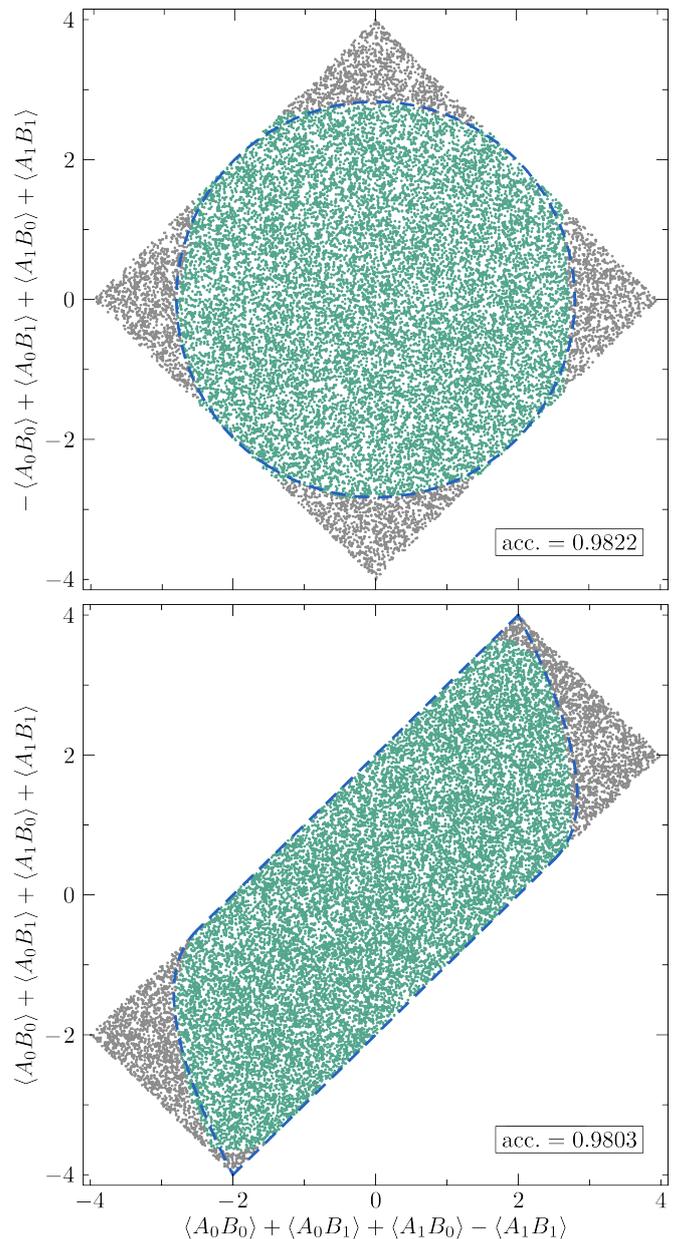


FIG. 4. Performance of the neural network model trained with the unbalanced dataset on two slices of the correlation space. The slice on top is defined by two orthogonal PR boxes, whereas the one on the bottom is defined by the relation $\langle A_0 B_0 \rangle = \langle A_0 B_1 \rangle = \langle A_1 B_0 \rangle$. The panels depict a set of 2×10^4 points on each slice, classified by the NN.

models on the slices is above 0.8 whereas when evaluated on the unbalanced set we find 0.80535, although it should be noted that the latter accuracies can significantly vary depending on the seed values of the weights. We interpret these differences in the results of model evaluation as a reflection of the difference in the amount of easy to classify points across training sets. Nonetheless, we still observe, contrary to our prior intuition, that the training sets consisting of examples near the boundary result

TABLE V. Accuracies obtained for the unbalanced, balanced and offset NN models. For all datasets, we trained the neural network using as function loss the Binary Cross Entropy (BCE) and a balanced version of it.

model	test acc.	train acc.	unbalanced acc.
unbalanced	0.99836	0.99581	0.99836
balanced	0.98501	0.99505	0.98306
offset ($\epsilon = 10^{-3}$)	0.50252	0.50328	0.80535

TABLE VI. Slice accuracies for all presented SVM and NN models in the correlation space. Slice 1 refers to the accuracy obtained in the upper slice (as in FIGs. 3 and 4) defined by the two orthogonal PR boxes. Slice 2 refers to the lower slice as presented for the first time in Ref. [38]. For each slice, the same sample of size 2×10^4 was generated and classified by the respective model, as specified in the first columns.

	model	slice 1	slice 2
SVM	unbalanced	0.9852	0.9838
	balanced	0.9689	0.9712
	offset ($\epsilon = 10^{-2}$)	0.9256	0.9212
	offset ($\epsilon = 10^{-3}$)	0.9804	0.9792
NN	unbalanced	0.9822	0.9803
	balanced	0.9972	0.9929
	offset ($\epsilon = 10^{-3}$)	0.8846	0.8477

in models that are significantly outperformed by those trained with uniformly distributed examples. In order to explore these differences in more depth, we implemented variations to the sampling method described in Sec. III B. These variations consisted of randomly distributing the examples created near the boundary, following either a uniform or a normal distribution. No appreciable difference in the obtained results was observed, which suggests that the sampling method is suboptimal for training of neural network models.

B. The entire CHSH space

Here, unlike the previous subsection, the models did not follow a similar thread of development that could be universally applied. Initially, our approach involved expanding the feature dimension across the complete CHSH space, transitioning from the classification previously based on level 1 of the NPA hierarchy to level $1 + AB$. Similar to the correlation models, a dataset comprising 10^4 points was generated and classified. This dataset was then used to feed the models, which performed poorly both in terms of accuracy metrics and slices. To overcome these issues, different strategies were implemented for each case, starting with the SVM models described below.

1. SVMs on the entire CHSH space

Our approach here leverages the local constraints within the CHSH scenario to tailor our datasets to the non-local region of the \mathcal{NS} polytope. This is done firstly by increasing the feature dimension in our earlier data generation methods to cover the whole CHSH space, and then excluding local points. For balanced and unbalanced models, this adaptation is straightforward, as we simply change the polytope from which the points are drawn using the hit-and-run technique. For this purpose, we employ the 8-simplex defined by the eight local vertices which saturate the inequality

$$\langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2, \quad (9)$$

and the box $\mathbf{p}_{\text{PR}} = [0, 0, 0, 0, 1, 1, 1, -1]$ which attains the geometrical bound in Eq. (9). Given the one-to-one correspondence between PR boxes and local facets in the CHSH scenario, the \mathcal{NS} polytope can be partitioned between the local set and eight disjoint and identical simplexes, which can be mapped to each other by relabelling settings, outcomes, and exchanging parts. Hence, for any non-signalling correlation, we can classify it as quantum if it meets any local constraint, and otherwise, the model trained on the simplex is applied. In addition, since the local polytope occupies about 94.12% of the total \mathcal{NS} volume [32], each simplex constitutes approximately 0.74% of the non-signalling volume, thus decreasing the redundancy in the model trained within this region. As for the offset model, such adaptation is made by filtering a set of random unit directions in \mathbb{R}^8 , leaving only those intersecting with the facet in Eq. (9). As before, this leads to a reduction in the model's training region, since in terms of the total solid angle comprised by the 8-sphere, the percentage of selected directions is approximately 0.21%³.

We then proceed with model development by generating samples containing 4×10^4 points based on the previously used data types: unbalanced, balanced, and offset, with ϵ values of 10^{-2} and 10^{-3} for the offset data. Similar to the correlation space, here we also employed the RBF kernel of Scikit-learn and maintained a 70-15-15 ratio for training, testing, and validation sets, respectively. The resulting models can be divided into two cases. For the offset data, developing a model that consistently performed well across all accuracy measures proved challenging, as shown in TABLE VII. Unlike its correlation space analogues, the reduction of the offset value here did not enhance performance measures; instead, it caused more significant overfitting in the model with $\epsilon = 10^{-3}$ compared to the one with $\epsilon = 10^{-2}$. In contrast, increasing

³ To the best of our knowledge, this value has not been previously reported in the literature. An estimate can easily be obtained by following the procedure outlined in the sentence preceding this note.

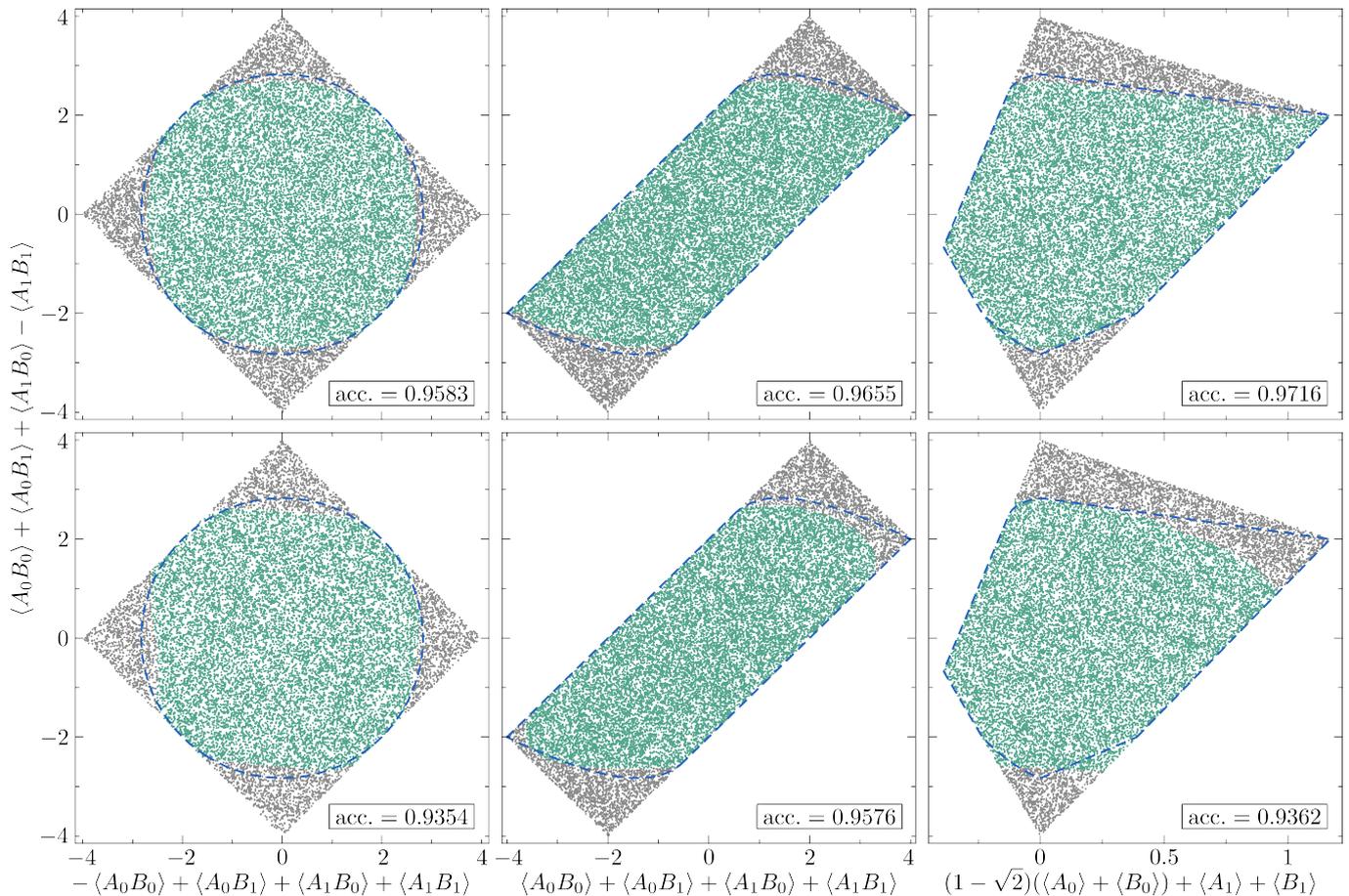


FIG. 5. Two-dimensional slices of the entire CHSH scenario. The plots on the upper row correspond to the results for the SVM models trained with balanced data. The lower row depicts the results obtained for NNs using unbalanced data. The plots on the left and in the centre of the image correspond to the same correlation slices shown in FIGs. 3 and 4. The right panels depict the slice introduced in FIG. 3 of Ref. [38], where we draw the boundary of the quantum set using level 3 of NPA.

TABLE VII. Accuracy measures for the SVM models in the entire CHSH space. The first column depicts the type of data used to train a given model, while the remaining columns present the values of the same accuracy measures used in the correlation models.

model	test acc.	train acc.	unbalanced acc.	spread acc. ($\sigma = 10^{-2}$)	spread acc. ($\sigma = 10^{-3}$)
unbalanced	0.9527	0.9527	—	0.5182	0.5013
balanced	0.9610	0.9637	0.9153	0.5547	0.5009
offset ($\epsilon = 10^{-2}$)	0.6417	0.7617	0.8266	0.6733	0.5157
offset ($\epsilon = 10^{-3}$)	0.2893	0.6462	0.7959	0.7600	0.5335

the feature dimension for the balanced and unbalanced models did not have a significant effect on the final model quality. Although their performance is still poor at points near the set boundary, as shown by the spread accuracy in TABLE VII, they show reasonable predictive capacity in other regions, as illustrated by the slices in the upper row of FIG. 5. The two first slices replicate the performance of the correlation models, which is extended to other regions of the representation space, as depicted in the rightmost plot.

2. NNs on the entire CHSH space

In the previous section, we have found that the training data sampled near the boundary is not well suited for neural network models. Consequently, for the 8-dimensional CHSH set, we consider only models trained with unbalanced data. We started by retraining the network we built for the benchmark on training sets of different sizes, starting with 10^4 points. We first observe that the performance of the training model improves with the size of the dataset, up to a size of $\sim 5 \times 10^4$

points. While in terms of global accuracy the results of the training seem as good as those obtained in the 4-dimensional correlation space, with differences within $\sim 10^{-2}$, we find that the differences in accuracy on the slices studied above is about 5 times larger. Interestingly, changes in the architecture of the network which increase the amount of resources like, for instance, doubling the amount of neurons in the input layer and/or adding extra hidden layers, do not seem to appreciably improve the performance of the trained models. It is not clear to us at the moment whether there is a combination of network architecture and hyperparameter values that may result in a better performing model, or if further improvement requires significantly larger training sets, which would represent an important practical limitation for these neural network models.

The global accuracy of the best model we could train in the conditions described above is 0.9967, and in the lower row of figure of FIG. 5 we show the performance of this model on the slices of the set previously introduced.

V. DISCUSSION

Using machine learning methods to solve problems in quantum physics is a relatively new phenomenon. Our goal was to conduct a comprehensive study of how these methods perform on a relatively simple and well-understood example. More specifically, we have chosen the problem of characterising the quantum set in the CHSH scenario, and we have explored various data science and machine learning models. It quickly became apparent that support vector machines and neural networks are the most promising candidates.

A key difference between the way ML is typically used and the way it is used in physics research is the source of data. In a research setting we are responsible for generating the data ourselves and one must be careful about how this is done exactly. In our case, uniform sampling leads to a highly imbalanced dataset. To fix this problem, we can either employ rejection sampling (which might be quite inefficient) or try to find a “smart” way of performing non-uniform sampling. However, one must be careful, because the latter might implicitly introduce some bias in the data resulting from our subjective opinion on what it means to be “smart”.⁴

Another observation that we made is that standard ML models are designed to perform well “on average”. However, in physics research the usual focus is on analysing some specific phenomenon, e.g. determining the quantum value of a Bell functional or analysing a specific

boundary region. Indeed, in research one almost exclusively cares about the “hard” cases, so average performance is not that relevant. For such tasks, the ML models do not seem to be a great fit.

Overall, both the SVMs and NNs exhibit reasonable performance, however, at some point adding more resources (e.g. adding more data or increasing model complexity) leads to diminishing returns. The slices we plotted for the 4-dimensional set look quite good, but one can see some imperfections. For the slices of the 8-dimensional set, the imperfections are clearly visible. Note that the input dimension of 4 or 8 is very low for ML methods. Nevertheless, generating a data sample that is representative for the 8-dimensional correlation space is highly non-trivial.

Finally, we would like to make a comment on the approach of forcing the neural networks to be convex by restricting the coefficients to be negative. Convexity is not a concept that often appears in real-world data.⁵ Hence, we were excited to be able to apply this trick to data where convexity is indeed relevant. However, this did not lead to a visible improvement, which might suggest that combining ML approaches with concepts from abstract mathematics is not always fruitful.

Overall, we have found that constructing an ML model that performs reasonably well on a relatively simple problem is easy. However, boosting its performance further is not that straightforward and at some point one has to start being smart about how the data is generated and how the model is chosen. At this point, however, there is a risk of injecting our own biases into the data without even realising it. Constructing models for larger problems is also challenging due to the sheer amount of data required to represent a high-dimensional space.

We believe that our findings will contribute to a better understanding of the types of problems where ML can provide an advantage over currently existing methods.

ACKNOWLEDGEMENTS

We acknowledge support from the National Science Centre (NCN), Poland under the SONATA project “Fundamental aspects of the quantum set of correlations” (grant No. 2019/35/D/ST2/02014). G.P.A. gratefully acknowledges Poland’s high-performance Infrastructure PLGrid (HPC Centre: ACK Cyfronet AGH) for providing computer facilities and support (computational grant No. PLG/2023/016878). N.G. acknowledges support from CONICET and CIC (R.R.) of Argentina, CONICET PIP Grant No. 11220200101877CO

⁴ Note that the same problem arises in the conventional setting where ML models are trained on pre-existing data, where the choice of how to clean and filter the data is also somewhat sub-

jective. However, in our case, we are dealing with a particularly severe variant of this problem.

⁵ In fact, we are not aware of a single real-world dataset in which convexity plays a role.

- [1] J. S. Bell, *Physics Physique Fizika* **1**, 195 (1964).
- [2] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Rev. Mod. Phys.* **86**, 419 (2014).
- [3] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969).
- [4] A. Fine, *Phys. Rev. Lett.* **48**, 291 (1982).
- [5] B. S. Tsirelson, *Lett. Math. Phys.* **4**, 93 (1980).
- [6] B. S. Tsirelson, *J. Math. Sci.* **36**, 557 (1987).
- [7] L. J. Landau, *Found. Phys.* **18**, 449 (1988).
- [8] J. Uffink, *Phys. Rev. Lett.* **88**, 230406 (2002).
- [9] L. Masanes, “Necessary and sufficient condition for quantum-generated correlations,” arXiv (2003).
- [10] M. Navascués, S. Pironio, and A. Acín, *Phys. Rev. Lett.* **98**, 010401 (2007).
- [11] R. F. Werner and M. M. Wolf, *Quantum Information & Computation* **1**, 1 (2001).
- [12] Y.-C. Liang and A. C. Doherty, *Phys. Rev. A* **75**, 042103 (2007).
- [13] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [14] V. Dunjko and H. J. Briegel, *Rep. Prog. Phys.* **81**, 074001 (2018).
- [15] K. Bharti, T. Haug, V. Vedral, and L.-C. Kwek, *AVS Quantum Science* **2**, 034101 (2020).
- [16] A. Canabarro, S. Brito, and R. Chaves, *Phys. Rev. Lett.* **122**, 200401 (2019).
- [17] T. Kriváchy, Y. Cai, D. Cavalcanti, A. Tavakoli, N. Gisin, and N. Brunner, *npj Quantum Information* **6**, 70 (2020).
- [18] T. Kriváchy, Y. Cai, J. Bowles, D. Cavalcanti, and N. Brunner, *New J. Phys.* **23**, 103034 (2021).
- [19] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, NY, 2006).
- [20] T. Hastie, R. Tibshirani, and J. Friedman, “Overview of supervised learning,” in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, NY, 2009) pp. 9–41.
- [21] W. S. Noble, *Nature Biotechnology* **24**, 1565 (2006).
- [22] V. Vapnik, *Estimation of Dependences Based on Empirical Data* (Springer, New York, NY, 2006) this book contains a translation of the 1982 Russian edition and the 2006 afterword “Empirical Inference Science”.
- [23] B. E. Boser, I. M. Guyon, and V. N. Vapnik, in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT ’92 (Association for Computing Machinery, New York, NY, USA, 1992) p. 144–152.
- [24] C. Cortes and V. Vapnik, *Machine Learning* **20**, 273 (1995).
- [25] S. Haykin, *Neural Networks: A Comprehensive Foundation*, International edition (Prentice Hall, 1999).
- [26] I. Goodfellow, Y. Bengio, and A. Courville, “Deep feed-forward networks,” in *Deep learning* (MIT press, 2016) Chap. 6.
- [27] X. Glorot, A. Bordes, and Y. Bengio, in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 15, edited by G. Gordon, D. Dunson, and M. Dudík (PMLR, Fort Lauderdale, FL, USA, 2011) pp. 315–323.
- [28] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for activation functions,” (2017), arXiv:1710.05941.
- [29] B. Amos, L. Xu, and J. Z. Kolter, in *Proceedings of the 34th International Conference on Machine Learning*, Vol. 70, edited by D. Precup and Y. W. Teh (PMLR, 2017) pp. 146–155.
- [30] F. Font-Clos, “Hit and run,” <https://github.com/fontclos/hitandrun> (2018), online; accessed July-2023.
- [31] A. Cabello, *Phys. Rev. A* **72**, 012113 (2005).
- [32] E. Wolfe and S. F. Yelin, *Phys. Rev. A* **86**, 012123 (2012).
- [33] K. F. Pál and T. Vértesi, *Phys. Rev. A* **80**, 042114 (2009).
- [34] J. M. Donohue and E. Wolfe, *Phys. Rev. A* **92**, 062120 (2015).
- [35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay, *J. of Machine Learning Research* **12**, 2825 (2011).
- [36] C.-C. Chang and C.-J. Lin, *ACM Trans. Intell. Syst. Technol.* **2** (2011), 10.1145/1961189.1961199.
- [37] C. Branciard, *Phys. Rev. A* **83**, 032123 (2011).
- [38] 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).
- [39] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (2016) pp. 265–283.
- [40] F. Chollet *et al.*, “Keras,” <https://keras.io> (2015).
- [41] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” (2018), arXiv:1708.02002 [cs.CV].

Appendix A: \mathcal{Q}_{1+AB} vs. $\mathcal{Q}_{6\times 6}$

In this appendix, we leverage the approaches described in Secs. III B and C to numerically compare the level $1+AB$ of NPA with the set of quantum realisations with local dimension 6, referred to as \mathcal{Q}_{1+AB} and $\mathcal{Q}_{6\times 6}$, respectively. The choice of a local dimension of 6 is based on its computational feasibility within the steered see-saw algorithm and the fact that, for $d \geq 4$, $\mathcal{L} \in \mathcal{Q}_{d\times d}$ [34]. This allows us to limit our focus to directions pointing towards one of the CHSH facets, since the boundaries of the two analysed sets coincide in other directions.

We then begin with a sample of 10^4 directions obtained by intersecting random directions with the local facet defined in Eq. (9). For each of the selected directions, we compute a pair of points: one on the boundary of \mathcal{Q}_{1+AB} , using the approach outlined in Eq. (6), and an interior estimate within $\mathcal{Q}_{6\times 6}$, using the steered see-saw with 50

random initializations or until the problem value is lower than 10^{-7} . The distances between each corresponding pair are then calculated and analysed. Among the total directions considered, 9220 showed a distance of less than 10^{-2} , with 6793 of these falling below 10^{-3} . For 5903 directions, the gap was smaller than the initially defined threshold of 10^{-7} . The plot in FIG. 6 illustrates the distribution of distances for the pairs that are at least 10^{-3} apart. As a result, considering that the simplex comprised by the facet in Eq. (9) covers only 0.21% of the total solid angle measured at \mathbb{R}^8 from its origin, we observe that the method described in Sec. III B correctly classifies points for about 99.87% of the total directions when $\epsilon = 10^{-2}$ and 99.46% when $\epsilon = 10^{-3}$.

It is worth noting that, since steered see-saw does not guarantee obtaining a point on the boundary of $\mathcal{Q}_{6 \times 6}$, the quality of the estimate depends on the number of random initialisations used. For this reason, in an attempt to obtain a more realistic number for the quantity of seeds required, we carried out the test that can be summarised by the data in TABLE VIII. In summary, we produce a sample of again 10^4 points uniformly distributed in the simplex defined by the local facet of Eq. (9) and the PR box corresponding to this facet. Then, we eliminate points that did not belong to the superset \mathcal{Q}_4 of NPA and classified the remaining points with steered see-saw using initially 5 seeds and a given local dimension. As before, if the problem value obtained is smaller than the threshold of 10^{-7} for a given dimension, the evaluated point is considered as belonging to $\mathcal{Q}_{d \times d}$. Initially, our idea was to compute the volume resulting from the intersection of the simplex with the sets $\mathcal{Q}_{d \times d}$, for $d \leq 16$.

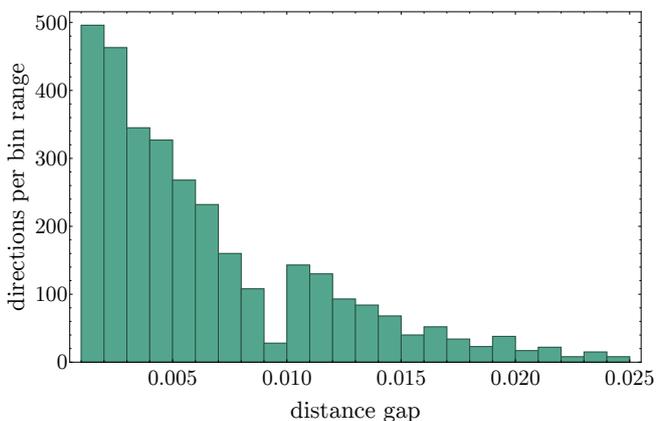


FIG. 6. The distribution of distances for the directions where the result was greater than 10^{-3} . Each bin has a length of also 10^{-3} . In the optimisation presented in Eq. (8), the minimised quantity is the distance between the variable encoded in the correlation \mathbf{q} and the target point $\mathbf{p}_b = \lambda^* \mathbf{u}$, with the direction \mathbf{u} not included in the problem constraints. Consequently, the interior estimate of \mathbf{p}_b does not necessarily align with \mathbf{u} , though the angle between the two points relative to the polytope's centre decreases proportionally with the gap between them.

However, this quickly proved to be more challenging than anticipated. As the dimension increased throughout the computation rounds, the process became unfeasible for $d = 10$. Nonetheless, during the following rounds, by decreasing the local dimension and increasing the number of seeds, we were able to classify most of the remaining points.

The data presented allows for two general conclusions. Firstly, the number of seeds used in the computation described at the beginning of this appendix is suboptimal, indicating that the superset \mathcal{Q}_{1+AB} is an even better approximation of \mathcal{Q} for developing an ML classification model. Secondly, numerically distinguishing between sets of quantum realisations with fixed local dimension for $d \geq 10$ is difficult. This is because it is not only computationally expensive but also due to the small volume difference between these sets in higher dimensions.

TABLE VIII. Classification of a sample of 10^4 points uniformly distributed in the simplex comprised over the local facet defined in Eq. (9) and the box $\mathbf{p}_{PR} = [0, 0, 0, 0, 1, 1, 1, -1]$. After removing points not belonging to the level 4 of NPA, the remaining 8563 points were classified in 17 rounds, varying the local dimension and the number of seeds. The second column shows the local dimension used, the third column shows the number of random initialisations, and the fourth column shows the number of points for which the quantum realisation found had a distance less than 10^{-7} from the target point. The last column indicates the total computation time per round, formatted as days:hours:mins. Note that in the 8th round, with $d = 9$, the computation time increased such that further increases in dimension became prohibitive. In subsequent rounds, the local dimension was reduced to $d = 6$ and 8, and the number of seeds was increased. The total CPU time was 158 days, 1 hour, and 26 minutes, with only 54 points remaining unclassified.

round #	local d	No. seeds	No. $\in \mathcal{Q}_{d \times d}$	CPU time
1	2	5	2271	01:07:30
2	3	5	2597	02:22:38
3	4	5	2097	01:07:45
4	5	5	496	01:12:32
5	6	5	390	03:10:23
6	7	5	171	04:12:04
7	8	5	94	06:07:27
8	9	5	56	17:11:05
9	6	50	158	11:03:25
10	8	50	37	16:20:39
11	6	100	77	12:11:57
12	6	150	24	11:04:53
13	6	200	14	11:05:17
14	6	250	9	12:06:02
15	6	300	9	14:20:05
16	6	350	7	14:04:17
17	6	400	2	15:01:20

Bell non-locality and certification of quantum devices are two interconnected fields that gained significant attention, particularly after the early 90s, with the initial development of quantum cryptography protocols. In this thesis, we aim to contribute to both areas, either through the development of certification protocols, as presented in Refs. [44] and [62], or by exploring Bell scenarios, as investigated in Refs. [44] and [50].

In Ref. [44], we present two general certification statements, with the first addressing a pair of rank-one projective measurements with d outcomes, and the other involving a maximally entangled state acting in a Hilbert space with local dimension d . While similar findings have been reported in the literature¹, the notable result of this work is the proof that, for rank-one projective measurements, incompatibility results in the violation of a Bell inequality. Although a similarly general result for projective measurements was obtained in Ref. [79], our study extends this by showing that when the measurements are rank-one, we can construct an inequality that is optimised by those measurements.

As for Ref. [62], we introduce a new variation of the random access code protocol, which we refer to as biased RACs. This approach has advantages from a certification perspective, as it does not require space-like separation for experimental implementation and enables the characterisation of a broad class of measurements. Nevertheless, such certification protocols must consider stronger assumptions, such as ensuring that one of the parts is trustworthy and that the system size is constrained. For this biased protocol, we successfully certified the measurements in all cases where a 2-bit string is encoded, as well as in some instances involving a 3-bit string.

Lastly, in Ref. [50], we perform an in-depth study of machine learning methods in the simplest non-locality scenario. We specifically chose to investigate the problem of characterising the CHSH quantum set, applying a range of data science and machine learning methods. Our study concludes that, although achieving good performance on average is fairly straightforward, training a model to accurately resolve the boundary of the quantum set is challenging. Conversely, we successfully created new methods for classifying quantum behaviours, particularly the steered see-saw and NPA techniques, which enable the exploration of unexposed points on the boundary of the quantum set.

¹See Refs. [76, 77] for the certification of a wide range of d -outcome measurements, and Ref. [78] for the certification of a multipartite state in any local dimension.

References

- [1] D. Bohm and Y. Aharonov. “Discussion of Experimental Proof for the Paradox of Einstein, Rosen, and Podolsky”. In: *Phys. Rev.* 108 (4 Nov. 1957), pp. 1070–1076. DOI: [10.1103/PhysRev.108.1070](https://doi.org/10.1103/PhysRev.108.1070).
- [2] A. Einstein, B. Podolsky, and N. Rosen. “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” In: *Phys. Rev.* 47 (10 May 1935), pp. 777–780. DOI: [10.1103/PhysRev.47.777](https://doi.org/10.1103/PhysRev.47.777).
- [3] L. de Broglie. *An introduction to the study of wave mechanics*. See chapters 6, 9 and 10. New York: E. P. Dutton and Company, Inc., 1930.
- [4] D. Bohm. “A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. I”. In: *Phys. Rev.* 85 (2 Jan. 1952), pp. 166–179. DOI: [10.1103/PhysRev.85.166](https://doi.org/10.1103/PhysRev.85.166).
- [5] D. Bohm. “A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. II”. In: *Phys. Rev.* 85 (2 Jan. 1952), pp. 180–193. DOI: [10.1103/PhysRev.85.180](https://doi.org/10.1103/PhysRev.85.180).
- [6] J. von Neumann. *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer-Verlag, 1932.
- [7] L. de Broglie. *Non-linear Wave Mechanics: A Causal Interpretation*. By Louis de Broglie. Translated [from the French] by Arthur J. Knodel and Jack C. Miller. Elsevier Publishing Company, 1960. URL: <https://books.google.pl/books?id=Gi7cwQEACAAJ>.
- [8] G. Hermann. “Die naturphilosophischen Grundlagen der Quantenmechanik”. In: *Grete Hermann — Between Physics and Philosophy*. Ed. by E. Crull and G Bacciagaluppi. Vol. 42. Studies in History and Philosophy of Science. Original work published in 1935. Springer, 2016. Chap. 15, pp. 239–278.
- [9] J. S. Bell. “On the Einstein Podolsky Rosen paradox”. In: *Physics Physique Fizika* 1 (3 Nov. 1964), pp. 195–200. DOI: [10.1103/PhysicsPhysiqueFizika.1.195](https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195).
- [10] J. F. Clauser et al. “Proposed Experiment to Test Local Hidden-Variable Theories”. In: *Phys. Rev. Lett.* 23 (Oct. 1969), pp. 880–884. DOI: [10.1103/PhysRevLett.23.880](https://doi.org/10.1103/PhysRevLett.23.880).
- [11] J. S. Bell. “Introduction to the hidden-variable question”. In: *49th International School of Physics “Enrico Fermi”: Foundations of Quantum Mechanics*. Varenna, Italy, June 1970, pp. 171–181. URL: <https://cds.cern.ch/record/400330>.

- [12] H. Reichenbach. *The Direction of Time*. Ed. by M. Reichenbach. Mineola, N.Y.: Dover Publications, 1956.
- [13] M. J. W. Hall. “The Significance of Measurement Independence for Bell Inequalities and Locality”. In: *At the Frontier of Spacetime*. Ed. by T. Asselmeyer-Maluga. Switzerland: Springer, 2016. Chap. 11, pp. 189–204. URL: <https://arxiv.org/abs/1511.00729>.
- [14] S. J. Freedman and J. F. Clauser. “Experimental Test of Local Hidden-Variable Theories”. In: *Phys. Rev. Lett.* 28 (14 Apr. 1972), pp. 938–941. DOI: [10.1103/PhysRevLett.28.938](https://doi.org/10.1103/PhysRevLett.28.938).
- [15] A. Aspect, P. Grangier, and G. Roger. “Experimental Tests of Realistic Local Theories via Bell’s Theorem”. In: *Phys. Rev. Lett.* 47 (7 Aug. 1981), pp. 460–463. DOI: [10.1103/PhysRevLett.47.460](https://doi.org/10.1103/PhysRevLett.47.460).
- [16] A. Aspect, P. Grangier, and G. Roger. “Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell’s Inequalities”. In: *Phys. Rev. Lett.* 49 (2 July 1982), pp. 91–94. DOI: [10.1103/PhysRevLett.49.91](https://doi.org/10.1103/PhysRevLett.49.91).
- [17] A. Aspect, J. Dalibard, and G. Roger. “Experimental Test of Bell’s Inequalities Using Time-Varying Analyzers”. In: *Phys. Rev. Lett.* 49 (25 Dec. 1982), pp. 1804–1807. DOI: [10.1103/PhysRevLett.49.1804](https://doi.org/10.1103/PhysRevLett.49.1804).
- [18] G. Weihs et al. “Violation of Bell’s Inequality under Strict Einstein Locality Conditions”. In: *Phys. Rev. Lett.* 81 (23 Dec. 1998), pp. 5039–5043. DOI: [10.1103/PhysRevLett.81.5039](https://doi.org/10.1103/PhysRevLett.81.5039).
- [19] M. A. Rowe et al. “Experimental violation of a Bell’s inequality with efficient detection”. In: *Nature* 409 (6822 Feb. 2001), pp. 791–794. DOI: [10.1038/35057215](https://doi.org/10.1038/35057215).
- [20] B. Hensen et al. “Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres”. In: *Nature* 526 (7575 Oct. 2015), pp. 682–686. DOI: [10.1038/nature15759](https://doi.org/10.1038/nature15759).
- [21] L. K. Shalm et al. “Strong Loophole-Free Test of Local Realism”. In: *Phys. Rev. Lett.* 115 (25 Dec. 2015), p. 250402. DOI: [10.1103/PhysRevLett.115.250402](https://doi.org/10.1103/PhysRevLett.115.250402).
- [22] M. Giustina et al. “Significant-Loophole-Free Test of Bell’s Theorem with Entangled Photons”. In: *Phys. Rev. Lett.* 115 (25 Dec. 2015), p. 250401. DOI: [10.1103/PhysRevLett.115.250401](https://doi.org/10.1103/PhysRevLett.115.250401).
- [23] B. S. Tsirelson. “Quantum generalizations of Bell’s inequality”. In: *Letters in Mathematical Physics* 4.2 (Mar. 1980), pp. 93–100. ISSN: 1573-0530. DOI: [10.1007/BF00417500](https://doi.org/10.1007/BF00417500).
- [24] B. S. Tsirelson. “Quantum analogues of the Bell inequalities. The case of two spatially separated domains”. In: *J. Math. Sci.* 36 (1987), pp. 557–570. DOI: [10.1007/BF01663472](https://doi.org/10.1007/BF01663472).
- [25] L. J. Landau. “Empirical two-point correlation functions”. In: *Found. Phys.* 18 (1988), pp. 449–460. DOI: [10.1007/BF00732549](https://doi.org/10.1007/BF00732549).
- [26] J. Uffink. “Quadratic Bell Inequalities as Tests for Multipartite Entanglement”. In: *Phys. Rev. Lett.* 88 (23 May 2002), p. 230406. DOI: [10.1103/PhysRevLett.88.230406](https://doi.org/10.1103/PhysRevLett.88.230406).
- [27] L. Masanes. *Necessary and sufficient condition for quantum-generated correlations*. arXiv. 2003. DOI: [10.48550/arXiv.quant-ph/0309137](https://doi.org/10.48550/arXiv.quant-ph/0309137).
- [28] A. Fine. “Hidden Variables, Joint Probability, and the Bell Inequalities”. In: *Phys. Rev. Lett.* 48 (5 Feb. 1982), pp. 291–295. DOI: [10.1103/PhysRevLett.48.291](https://doi.org/10.1103/PhysRevLett.48.291).

- [29] A. K. Ekert. “Quantum cryptography based on Bell’s theorem”. In: *Phys. Rev. Lett.* 67 (6 Aug. 1991), pp. 661–663. DOI: [10.1103/PhysRevLett.67.661](https://doi.org/10.1103/PhysRevLett.67.661).
- [30] C. H. Bennett and G. Brassard. “Quantum cryptography: Public key distribution and coin tossing”. In: *Proceedings of the International Conference on Computers, Systems and Signal Processing, Bangalore, India*. IEEE. Dec. 1984, pp. 175–179. DOI: [10.1016/j.tcs.2014.05.025](https://doi.org/10.1016/j.tcs.2014.05.025).
- [31] A. Acín et al. “Device-Independent Security of Quantum Cryptography against Collective Attacks”. In: *Phys. Rev. Lett.* 98 (23 June 2007), p. 230501. DOI: [10.1103/PhysRevLett.98.230501](https://doi.org/10.1103/PhysRevLett.98.230501).
- [32] V. Scarani. “The device-independent outlook on quantum physics”. In: *Acta Physica Slovaca* 62.4 (2012), pp. 347–409. URL: <http://www.physics.sk/aps/pubs/2012/aps-12-04/aps-12-04.pdf>.
- [33] D. Mayers and A. Yao. “Quantum cryptography with imperfect apparatus”. In: *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No.98CB36280)*. Nov. 1998, pp. 503–509. DOI: [10.1109/SFCS.1998.743501](https://doi.org/10.1109/SFCS.1998.743501).
- [34] D. Mayers and A. Yao. “Self testing quantum apparatus”. In: *Quantum Info. Comput.* 4.4 (July 2004), pp. 273–286. ISSN: 1533-7146.
- [35] M. Froissart. “Constructive generalization of Bell’s inequalities”. In: *Il Nuovo Cimento B* 64 (2 Aug. 1981), pp. 241–251. DOI: [10.1007/BF02903286](https://doi.org/10.1007/BF02903286).
- [36] A. Garg and N. D. Mermin. “Farkas’s Lemma and the nature of reality: Statistical implications of quantum correlations”. In: *Foundations of Physics* 14 (1 Jan. 1984), pp. 1–39. ISSN: 1572-9516. DOI: [10.1007/BF00741645](https://doi.org/10.1007/BF00741645).
- [37] I. Pitowsky. “The range of quantum probability”. In: *Journal of Mathematical Physics* 27.6 (June 1986), pp. 1556–1565. ISSN: 0022-2488. DOI: [10.1063/1.527066](https://doi.org/10.1063/1.527066).
- [38] I. Pitowsky. *Quantum Probability – Quantum Logic*. Springer Berlin, Heidelberg, 1989. DOI: [10.1007/BFb0021186](https://doi.org/10.1007/BFb0021186).
- [39] I. Pitowsky. “Correlation polytopes: Their geometry and complexity”. In: *Mathematical Programming* 50 (1 Mar. 1991), pp. 395–414. DOI: [10.1007/BF01594946](https://doi.org/10.1007/BF01594946).
- [40] B. S. Tsirelson. “Some results and problems on quantum Bell-type inequalities”. In: *Had. J. Suppl.* 8 (1993), p. 329. URL: <https://www.tau.ac.il/~tsirel/download/hadron.pdf>.
- [41] The conjecture was presented as a statement in Ref. [40] without proof, but it was realised as potentially wrong only in 2006 by A. Acín. URL: <https://www.tau.ac.il/~tsirel/Research/bellopalg/main.html>.
- [42] Z. Ji et al. “MIP* = RE”. In: *Commun. ACM* 64.11 (Oct. 2021), pp. 131–138. ISSN: 0001-0782. DOI: [10.1145/3485628](https://doi.org/10.1145/3485628).
- [43] T. P. Le et al. “Quantum Correlations in the Minimal Scenario”. In: *Quantum* 7 (Mar. 2023), p. 947. DOI: [10.22331/q-2023-03-16-947](https://doi.org/10.22331/q-2023-03-16-947).
- [44] G. Pereira Alves and J. Kaniewski. “Optimality of any pair of incompatible rank-one projective measurements for some nontrivial Bell inequality”. In: *Phys. Rev. A* 106 (3 Sept. 2022), p. 032219. DOI: [10.1103/PhysRevA.106.032219](https://doi.org/10.1103/PhysRevA.106.032219).

- [45] A. Tavakoli et al. “Mutually unbiased bases and symmetric informationally complete measurements in Bell experiments”. In: *Science Advances* 7.7 (2021), eabc3847. DOI: [10.1126/sciadv.abc3847](https://doi.org/10.1126/sciadv.abc3847).
- [46] S. L. Braunstein, A. Mann, and M. Revzen. “Maximal violation of Bell inequalities for mixed states”. In: *Phys. Rev. Lett.* 68 (22 June 1992), pp. 3259–3261. DOI: [10.1103/PhysRevLett.68.3259](https://doi.org/10.1103/PhysRevLett.68.3259).
- [47] S. Popescu and D. Rohrlich. “Which states violate Bell’s inequality maximally?” In: *Physics Letters A* 169.6 (Oct. 1992), pp. 411–414. ISSN: 0375-9601. DOI: [10.1016/0375-9601\(92\)90819-8](https://doi.org/10.1016/0375-9601(92)90819-8).
- [48] F. Magniez et al. “Self-testing of Quantum Circuits”. In: *Automata, Languages and Programming*. Ed. by M. Bugliesi et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 72–83. ISBN: 978-3-540-35905-0.
- [49] I. Šupic and J. Bowles. “Self-testing of quantum systems: a review”. In: *Quantum* 4 (Sept. 2020), p. 337. DOI: [10.22331/q-2020-09-30-337](https://doi.org/10.22331/q-2020-09-30-337).
- [50] G. Pereira Alves, N. Gigena, and J. Kaniewski. *Machine learning meets the CHSH scenario*. arXiv. 2024. DOI: [10.48550/arXiv.2407.14396](https://doi.org/10.48550/arXiv.2407.14396).
- [51] MOSEK ApS. *MOSEK Optimizer API for Python 10.2.2*. 2019. URL: <https://docs.mosek.com/latest/pythonapi/index.html>.
- [52] S. Diamond and S. Boyd. “CVXPY: A Python-embedded modeling language for convex optimization”. In: *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5.
- [53] A. Agrawal et al. “A rewriting system for convex optimization problems”. In: *Journal of Control and Decision* 5.1 (2018), pp. 42–60.
- [54] G. Sagnol and M Stahlberg. “PICOS: A Python interface to conic optimization solvers”. In: *Journal of Open Source Software* 7.70 (Feb. 2022), p. 3915. ISSN: 2475-9066. DOI: [10.21105/joss.03915](https://doi.org/10.21105/joss.03915).
- [55] M. Navascués, S. Pironio, and A. Acín. “Bounding the Set of Quantum Correlations”. In: *Phys. Rev. Lett.* 98 (1 Jan. 2007), p. 010401. DOI: [10.1103/PhysRevLett.98.010401](https://doi.org/10.1103/PhysRevLett.98.010401).
- [56] J. B. Lasserre. “Global Optimization with Polynomials and the Problem of Moments”. In: *SIAM Journal on Optimization* 11 (2001), pp. 796–817. DOI: [10.1137/S1052623400366802](https://doi.org/10.1137/S1052623400366802).
- [57] M. Navascués, S. Pironio, and A. Acín. “A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations”. In: *New Journal of Physics* 10.7 (July 2008), p. 073013. DOI: [10.1088/1367-2630/10/7/073013](https://doi.org/10.1088/1367-2630/10/7/073013).
- [58] A. C. Doherty et al. “The Quantum Moment Problem and Bounds on Entangled Multi-prover Games”. In: *Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity*. USA: IEEE Computer Society, 2008, pp. 199–210. ISBN: 9780769531694. DOI: [10.1109/CCC.2008.26](https://doi.org/10.1109/CCC.2008.26).
- [59] Y.-C. Liang and A. C. Doherty. “Bounds on quantum correlations in Bell-inequality experiments”. In: *Phys. Rev. A* 75 (4 Apr. 2007), p. 042103. DOI: [10.1103/PhysRevA.75.042103](https://doi.org/10.1103/PhysRevA.75.042103).
- [60] R. F. Werner and M. M. Wolf. “Bell inequalities and entanglement”. In: *Quantum Info. Comput.* 1.3 (Oct. 2001), pp. 1–25. ISSN: 1533-7146.

- [61] J. M. Donohue and E. Wolfe. “Identifying nonconvexity in the sets of limited-dimension quantum correlations”. In: *Phys. Rev. A* 92 (6 Dec. 2015), p. 062120. DOI: [10.1103/PhysRevA.92.062120](https://doi.org/10.1103/PhysRevA.92.062120).
- [62] G. Pereira Alves, N. Gigena, and J. Kaniewski. “Biased random access codes”. In: *Phys. Rev. A* 108 (4 Oct. 2023), p. 042608. DOI: [10.1103/PhysRevA.108.042608](https://doi.org/10.1103/PhysRevA.108.042608).
- [63] C. de Gois et al. “Complete Hierarchy for High-Dimensional Steering Certification”. In: *Phys. Rev. Lett.* 131 (1 July 2023), p. 010201. DOI: [10.1103/PhysRevLett.131.010201](https://doi.org/10.1103/PhysRevLett.131.010201).
- [64] R. Santos, C. Jebarathinam, and R. Augusiak. “Scalable noncontextuality inequalities and certification of multiqubit quantum systems”. In: *Phys. Rev. A* 106 (1 July 2022), p. 012431. DOI: [10.1103/PhysRevA.106.012431](https://doi.org/10.1103/PhysRevA.106.012431).
- [65] G. Drótos et al. “Towards minimal self-testing of qubit states and measurements in prepare-and-measure scenarios”. In: *New Journal of Physics* 26.6 (2024), p. 063012.
- [66] S. Wiesner. “Conjugate coding”. In: *SIGACT News* 15.1 (Jan. 1983), pp. 78–88. ISSN: 0163-5700. DOI: [10.1145/1008908.1008920](https://doi.org/10.1145/1008908.1008920).
- [67] A. Ambainis et al. “Dense quantum coding and a lower bound for 1-way quantum automata”. In: *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*. STOC '99. Atlanta, Georgia, USA: Association for Computing Machinery, 1999, pp. 376–383. ISBN: 1581130678. DOI: [10.1145/301250.301347](https://doi.org/10.1145/301250.301347).
- [68] A. Nayak. “Optimal Lower Bounds for Quantum Automata and Random Access Codes”. In: *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*. FOCS '99. USA: IEEE Computer Society, 1999, p. 369. ISBN: 0769504094.
- [69] T. Durt et al. “On mutually unbiased bases”. In: *International Journal of Quantum Information* 08.04 (2010), pp. 535–640. DOI: [10.1142/S0219749910006502](https://doi.org/10.1142/S0219749910006502).
- [70] M. Pawłowski and M. Żukowski. “Entanglement-assisted random access codes”. In: *Phys. Rev. A* 81 (4 Apr. 2010), p. 042326. DOI: [10.1103/PhysRevA.81.042326](https://doi.org/10.1103/PhysRevA.81.042326).
- [71] A. Ambainis et al. “Dense quantum coding and quantum finite automata”. In: *J. ACM* 49.4 (July 2002), pp. 496–511. ISSN: 0004-5411. DOI: [10.1145/581771.581773](https://doi.org/10.1145/581771.581773).
- [72] A. Ben-Aroya, O. Regev, and R. de Wolf. “A Hypercontractive Inequality for Matrix-Valued Functions with Applications to Quantum Computing and LDCs”. In: *2008 49th Annual IEEE Symposium on Foundations of Computer Science*. 2008, pp. 477–486. DOI: [10.1109/FOCS.2008.45](https://doi.org/10.1109/FOCS.2008.45).
- [73] A. Acín, S. Massar, and S. Pironio. “Randomness versus Nonlocality and Entanglement”. In: *Phys. Rev. Lett.* 108 (10 Mar. 2012), p. 100402. DOI: [10.1103/PhysRevLett.108.100402](https://doi.org/10.1103/PhysRevLett.108.100402).
- [74] C. Bamps and S. Pironio. “Sum-of-squares decompositions for a family of Clauser-Horne-Shimony-Holt-like inequalities and their application to self-testing”. In: *Phys. Rev. A* 91 (5 May 2015), p. 052111. DOI: [10.1103/PhysRevA.91.052111](https://doi.org/10.1103/PhysRevA.91.052111).
- [75] G. Pereira Alves, J. Kaniewski, and N. Gigena. *QRAC-tools*. GitHub. online; accessed 26-August-2024. 2022. URL: <https://github.com/jedreky/QRAC-tools>.

-
- [76] N. Miklin and M. Oszmaniec. “A universal scheme for robust self-testing in the prepare-and-measure scenario”. In: *Quantum* 5 (Apr. 2021), p. 424. ISSN: 2521-327X. DOI: [10.22331/q-2021-04-06-424](https://doi.org/10.22331/q-2021-04-06-424).
- [77] S. Sarkar, D. Saha, and R. Augusiak. “Certification of incompatible measurements using quantum steering”. In: *Phys. Rev. A* 106 (4 Oct. 2022), p. L040402. DOI: [10.1103/PhysRevA.106.L040402](https://doi.org/10.1103/PhysRevA.106.L040402).
- [78] S. Sarkar and R. Augusiak. “Self-testing of multipartite Greenberger-Horne-Zeilinger states of arbitrary local dimension with arbitrary number of measurements per party”. In: *Phys. Rev. A* 105 (3 Mar. 2022), p. 032416. DOI: [10.1103/PhysRevA.105.032416](https://doi.org/10.1103/PhysRevA.105.032416).
- [79] M. M. Wolf, D. Perez-Garcia, and C. Fernandez. “Measurements Incompatible in Quantum Theory Cannot Be Measured Jointly in Any Other No-Signaling Theory”. In: *Phys. Rev. Lett.* 103 (23 Dec. 2009), p. 230402. DOI: [10.1103/PhysRevLett.103.230402](https://doi.org/10.1103/PhysRevLett.103.230402).