Faculty of Mathematics, Physics, and Informatics
University of Gdańsk

Summary of PhD Dissertation in Physics

Máté Farkas

Certifying quantum measurements: mutually unbiased bases and measures of incompatibility

Supervisor: **dr hab. inż. Marcin Pawłowski**
Co-supervisor: **dr Jędrzej Kaniewski**
Institute of Theoretical Physics and Astrophysics
Faculty of Mathematics, Physics, and Informatics
University of Gdańsk, 80-952 Gdańsk, Poland

Gdańsk, 2019

Acknowledgement

I am immensely grateful for my supervisor, Marcin Pawłowski, for making my time as a PhD student ever so easy, providing me with constant support, let it be resources, advices, or simply having a good time. I also owe a debt of gratitude to my co-supervisor, Jędrzej Kaniewski, with whom we started off as collaborators and friends, but I have learnt so much from him during our joint projects that I only felt it adequate to ask him to be my co-supervisor, which he, to my great pleasure, accepted. I am similarly thankful to all my co-authors and colleagues with whom I had the honour to engage in fruitful and stimulating scientific discussions and projects. I am thoroughly indebted to my girlfriend, Rebeka, who accompanied me all throughout this journey, and provided me the stability, joy and care without which my work would have been unbelievably more difficult. I am truly grateful for my parents, Ágnes and György, for their continuous support, that helped me through the entirety of my education, and contributed greatly towards acquiring all the knowledge that allowed me to write the present thesis.

All throughout my PhD studies, I was funded by the Polish NCN grant Sonata UMO-2014/14/E/ST2/00020, and for a six-month period by the grant Badania Młodych Naukowców number 538-5400-B049-18, issued by the Polish Ministry of Science and Higher Education.

Szüleimnek, Ágnesnek és Györgynek

Załącznik nr 2
do Zasad przeprowadzania przewodów doktorskich
na Wydziale Matematyki, Fizyki i Informatyki UG

OŚWIADCZENIE

Ja, niżej podpisany oświadczam, iż przedłożona praca doktorska została wykonana przeze mnie samodzielnie, nie narusza praw autorskich, interesów prawnych i materialnych innych osób.

……………………………   …………………………….
data   własnoręczny podpis

Abstract

Quantum information theory is a rapidly growing field that harnesses the power of microscopic physical systems in information theoretical tasks. Some of its predictions could have a tremendous impact on near-term information technology, such as exponential speedup in computational tasks or unconditionally secure cryptographic protocols. These perspectives are highly promising, however, they call for verification schemes: one must be able to certify the correctness of quantum computations, as well as to verify the security of cryptographic devices.

While such verification schemes already exist, and are thoroughly studied, there are still a few drawbacks associated with them. The most rigorous certification scheme of "self-testing" is rather difficult to implement in the laboratory, and results in the high-dimensional setting are lacking, despite the apparent advantage of high-dimensional systems. Moreover, most verification methods focus on certifying the exact physical setup rather than some relevant properties thereof, which is impractical in some cases.

In the current thesis, I address the above shortcomings by devising experimentally friendly certification schemes of relevant properties in the high-dimensional setting. Specifically, I focus on the experimentally less demanding task of "prepare-and-measure" scenarios, in which, together with my collaborators, I introduce two methods of certifying quantum states and measurements. The first method concentrates on verifying the genuine high-dimensional nature of quantum states and measurements, a property that we refer to as 'irreducible high-dimensional systems'. Together with my collaborators, we demonstrate the applicability of our methods in a photonic experiment in dimension 1024, proving the irreducibility of the implemented quantum optical setup.

My second method uses the same prepare-and-measure protocol, however, this time I concentrate on certifying a class of measurements that has proven to be immensely useful in quantum information theory, mutually unbiased bases. Together with my collaborator, we show that these measurements can be certified in the prepare-and-measure scenario in an experimentally feasible manner. Moreover, using our results, we are able to certify two additional properties of the measurements, namely their capability of generating randomness, and their incompatibility robustness.

Finally, I focus on the above mentioned relevant property of measurements, incompatibility robustness, which measures to what extent a pair of quantum measurements is not jointly measurable. Incompatible measurements turn out to be a useful resource in various quantum information theoretic protocols, and therefore it is an important task to

quantify the extent to which a pair of measurements is incompatible. Together with my collaborators, we analyse a wide class of incompatibility robustness measures, corresponding to generic noise models. We show that some of the measures that are often used in the literature do not satisfy certain natural properties. Moreover, we show that according to one of the measures, mutually unbiased bases are among the most incompatible pairs of measurements in every dimension, but also that this is not the case for some other measures. Our results highlight that despite the significant effort dedicated to this topic, a thorough understanding of incompatibility robustness measures is still lacking in the quantum information community.

Streszczenie

Teoria informacji kwantowej jest aktywnym kierunkiem badawczym, którego celem jest wykorzystanie mikroskopowych układów fizycznych do zadań związanych z przetwarzaniem informacji. Niektóre odkrycia na tym polu mogą mieć w niedalekiej przyszłości ogromny wpływ na technologie związane z przetwarzaniem informacji, np. eksponencjalne przyspieszenie w zadaniach obliczeniowych lub bezwarunkowo bezpieczne protokoły kryptograficzne. Te perspektywy są obiecujące, ale aby je osiągnąć, potrzebne są odpowiednie procedury weryfikacji: musimy być w stanie potwierdzić, że obliczenia kwantowe dokonywane są w sposób poprawny lub że implementacja protokołu kryptograficznego jest bezpieczna.

Procedury tego typu już istnieją, ale mimo intensywnych badań nie spełniają one jeszcze wszystkich wymagań. Najbardziej rygorystyczna metoda certyfikacji, zwana "samotestowaniem", jest trudna do implementacji w eksperymencie. Ponadto istnieje niewiele wyników, które można zastosować do układów wysokowymiarowych, mimo że układy te są przydatne w wielu naturalnych zadaniach. Poza tym większość istniejących metod skupia się na certyfikacji całego układy fizycznego, a nie na konkretnych pożądanych własnościach, co w niektórych przypadkach jest niepraktyczne.

W ramach mojej pracy doktorskiej proponuję nowe i przyjazne z eksperymentalnego punktu widzenia procedury certyfikacji konkretnych istotnych cech układów wysokowymiarowych. W scenariuszu "przygotuj-i-zmierz", który jest mniej wymagający z eksperymentalnego punktu widzenia, wraz ze swoimi współpracownikami proponuję dwie metody certyfikacji stanów i pomiarów kwantowych. Pierwsza metoda pozwala weryfikować prawdziwie wysokowymiarową naturę stanu i pomiarów kwantowych, co nazywamy "nieredukowalnością" układu. Wraz ze współpracownikami zastosowaliśmy tę metodę do fotonicznego eksperymentu w wymiarze 1024, gdzie pokazaliśmy, że kwantowo-optyczny układ zaimplementowany w eksperymencie jest nieredukowalny.

Druga metoda używa tego samego protokołu w scenariuszu "przygotuj-i-zmierz", ale tym razem skupiam się na certyfikowaniu pewnej klasy pomiarów, które są niezwykle użyteczne w teorii informacji kwantowej: baz wzajemnie nieobciążonych. Wraz ze współpracownikiem pokazaliśmy, że te pomiary mogą być certyfikowane w scenariuszu "przygotuj-i-zmierz" w warunkach realistycznych z eksperymentalnego punktu widzenia. Ponadto byliśmy w stanie certyfikować dwie dodatkowe własności pomiarów: ich zdolność do generowania losowości i niekompatybilność.

W ostatniej części skupiam się na niekompatybilności pomiarów, a konkretniej na mi-

arach opartych na odporność na szum, które kwantyfikują w jakim stopniu dwa pomiary są niekompatybilne. Zrozumienie tych miar jest ważne, gdyż niekompatybilne pomiary są użytecznym zasobem w wielu kwantowych protokołach. Wraz ze współpracownikami zanalizowaliśmy szeroką gamę miar niekompatybilności, które odpowiadają naturalnym modelom szumu. Pokazaliśmy, że niektóre z miar, które są często używane w literaturze, nie spełniają pewnych naturalnych wymogów. Ponadto pokazaliśmy, że według jednej z miar bazy wzajemnie nieobciążone znajdują się wśród najbardziej niekompatybilnych pomiarów (w każdym wymiarze), ale to stwierdzenie nie jest prawdą dla innych miar. Nasze wyniki pokazują, że mimo pokaźnego wysiłku badawczego w tej tematyce, nasze zrozumienie miar niekompatybilności pomiarów wciąż jest niepełne.

Publications included in the PhD Dissertation

[A] **Certifying an irreducible 1024-dimensional photonic state using refined dimension witnesses**
E. A. Aguilar*, **M. Farkas***, D. Martínez, M. Alvarado, J. Cariñe, G. B. Xavier, J. F. Barra, G. Cañas, M. Pawłowski, G. Lima
Physical Review Letters **120**, 230503 (2018)
*contributed equally

[B] **Self-testing mutually unbiased bases in the prepare-and-measure scenario**
M. Farkas, J. Kaniewski
Physical Review A **99**, 032316 (2019)

[C] **Incompatibility robustness of quantum measurements: a unified framework**
S. Designolle*, **M. Farkas***, J. Kaniewski
New Journal of Physics **21**(11), 113053 (2019)
*contributed equally

Other published works and pre-prints

1. **Homological codes and abelian anyons**
   P. Vrana, **M. Farkas**
   Reviews in Mathematical Physics **31**(10), 1950038 (2019)

2. **Qudit homological product codes**
   M. Farkas, P. Vrana
   Quantum Information and Computation **17**(11&12), pp.0948–0958 (2017)

3. **Mutually unbiased bases and symmetric informationally complete measurements in Bell experiments: Bell inequalities, device-independent certification and applications**
   A. Tavakoli, **M. Farkas**, D. Rosset, J.-D. Bancal, J. Kaniewski
   `arXiv:`1912.03225

4. **Quantum error correction codes and absolutely maximally entangled states**
   P. Mazurek, **M. Farkas**, A. Grudka, M. Horodecki, M. Studziński
   `arXiv:`1910.07427

5. **n-fold unbiased bases: an extension of the MUB condition**
   M. Farkas
   `arXiv:`1706.04446

6. **MRCC, a quantum chemical program suite**
   M. Kállay, Z. Rolik, J. Csontos, P. Nagy, G. Samu, D. Mester, J. Csóka, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, **M. Farkas**, B. Hégely
   https://www.mrcc.hu

Contents

Part I

Summary of PhD Dissertation

I. INTRODUCTION

The main objective of physics (and of any natural science) is to *describe* and *predict* phenomena that occur in the world that surrounds us. Why does the Sun rise and set at a particular position? Why do certain objects – let them be enormous, like a star, or minuscule, like an atom – attract each other? How can we use these observations to design new tools that facilitate our everyday lives?

In the last few centuries – which have seen an unprecedented progress of human well-being – the first and foremost tool for answering such questions has been to devise *mathematical models*. Such models provide a universal language in which descriptions and predictions can be formulated, and they work incredibly well with the other cornerstone of science: *experiments*. The axiomatic structure of mathematics makes it possible for mathematical models to precisely *prescribe* an experiment, and to *describe* its outcome in an unequivocal manner. This also implies that theories based on mathematical models can be *tested*: in case experiments confirm their predictions, they are temporarily admitted. However, this status never lasts forever – whenever an experiment contradicts with its predictions, the theory should be refined, its applicability should be restricted, or, in the most extreme case, it should be completely dismissed. This tedious and never-ending trial and error process is how science has always been advancing.

About a hundred years ago, an extremely successful physical theory started to emerge: *quantum theory* aims at describing and predicting the behaviour of microscopic physical systems, and it does so with outstanding precision. As an example, one of the predictions of quantum electrodynamics is the slight deviation of the electron magnetic moment from the value $\frac{g}{2} = 1$ predicted by standard relativistic quantum mechanics. This deviation was recently measured experimentally to take the value $\frac{g}{2} = 1.00115965218085(76)$, a precision of 14 digits (11, if considering the deviation only) [OHDG06]. With experiments like this, quantum theory is arguably one of the most thoroughly tested physical theories, and it has been holding up to scrutiny incredibly well for the last century.

This extreme precision comes at a somewhat unexpected cost: in many cases, the predictions of quantum theory are in stark contrast with our everyday intuitions. Most notably, Einstein, Podolsky and Rosen pointed out in 1935 [EPR35], and later Bell formalised in 1964 [Bel64], that *the predictions of quantum theory cannot be explained by any local realistic theory*. What this means is that if the predictions of quantum theory are correct, then we cannot think of physical systems as having pre-defined (real) properties in confined (local) regions of space, and of measurements as simply revealing these properties to us.

Einstein, Podolsky, Rosen and Bell proposed an experiment in which two particles are sent to two distant laboratories. In each of the laboratories, the experimenters *simultaneously* perform a measurement on their particle, and note down the outcome of this measurement. They repeat this procedure many times with new particles, in each round possibly choosing different measurements. Since the laboratories are far away, and the measurements only take a short time, the experimenters cannot communicate during individual rounds. After many rounds, the experimenters gather their respective outcome statistics and meet to look at the statistics of the whole experiment. Astoundingly, according to quantum theory, in some cases they might find that these statistics cannot be explained by local realism. That is, if the experimenters were to assume that their respective particles had some well-defined properties (possibly different ones in each round) which the measurements could read out, they would be unable to reconstruct their experimental data. Crucially, such experiments have recently been performed in so-called *loophole-free Bell tests* [HBD$^+$15, GVW$^+$15, SMSC$^+$15]. The results confirm with very high fidelity that the predictions of quantum theory are indeed correct, and Nature cannot be modelled in a local realistic manner.

A rapidly developing field that harnesses such extraordinary feats of quantum theory is *quantum information theory*. It studies the power of microscopic particles in information theoretical tasks, such as computation or communication. Exploiting phenomena that cannot be explained in any classical (local realistic) theory, the relatively new field of quantum information theory has already led to promising, and potentially paradigm-shifting results. Some of them give us the prospect of performing computational tasks – such as factorising astronomically large numbers [Sho94], or simulating molecules consisting of a huge number of atoms [Fey82] – that are unfathomable with currently available computers. Other results open up the possibility of designing communication devices that are unhackable by any agent that is restricted by quantum theory [BB84, Eke91].

While these prognoses are immensely impressive, it is apparent that they call for *verification schemes*. How can we be sure that a quantum computation provides the right answer, if there is no way to reproduce it on a classical computer? How do we certify that the "secure" communication device – potentially obtained from an untrusted party – is not leaking any information to adversaries? Can an everyday user confirm these vital claims without having to understand all the intricacies of quantum theory and without having to track down the movement of every individual atom?

Somewhat fortunately, Bell's results also provide a way for a purely classical user to certify quantum devices in so-called *self-testing* scenarios [MY98, MY04, MMMO06]. In some Bell-type experiments, the experimenters are able to characterise their devices up to the minimal freedom that is allowed within quantum information theory. Importantly, there are also *robust* self-testing results [BLM$^+$09, MYS12, YVB$^+$14, BNS$^+$15], that allow for an approximate characterisation in imperfect experimental realisations [TWE$^+$17]. This makes making self-testing statements applicable in real-world scenarios, such as quantum computing or cryptography.

Self-testing results are extremely powerful, however, this power comes with a few drawbacks. First of all, these experiments are exceptionally difficult to implement in the laboratory – notice the time difference between Bell's theorem in 1964, and the first loophole-free Bell test in 2015: more than 50 years! Second of all, the rigid self-testing statements are not always practical: in many cases, the users might not be interested in certifying their devices up to the minimum theoretically allowed freedom, but would rather *certify certain relevant properties* of the physical systems and measurements. Lastly, deriving self-testing statements for systems with dimension (number of degrees of freedom) larger than two is also difficult theoretically, which is apparent from the lack of results in the scientific literature. On the other hand, current technology and experiments have entered a stage when they can prepare and measure high-dimensional quantum systems reliably with high precision [DLB$^+$11, FLP$^+$12, MMZ16]. Such high-dimensional systems have a provable advantage regarding noise tolerance [HP13], and their use seems inevitable if we are aiming to increase the communication capacity of existing devices, such as optical fibres [RFN13]. Therefore, in summary, new, experimentally friendly verification schemes are needed, that certify relevant properties even in the high-dimensional regime.

One example of a relevant property of quantum measurements, that also turns out to be essential for Bell-type experiments, is *measurement incompatibility* [Lud54, BLPY16]. This is yet another counter-intuitive quantum phenomenon: Some measurements cannot be performed simultaneously on a single copy of the physical state. That is, sometimes we cannot learn two different properties of a single physical system. Such measurements are called *incompatible*, and they turn out to be a useful resource in Bell-type experiments

[WPGF09], for the so-called Einstein–Podolsky–Rosen steering [QVB14, UBGP15] and state discrimination tasks [CHT19]. Therefore, it is desirable to characterise to what extent certain measurements are incompatible. Such measures of incompatibility are often studied in the literature [HMZ16], however, their properties and the relations between them are not well-understood.

In the present thesis, together with my collaborators, I address the above mentioned shortcomings in the following ways:

- I develop *experimentally friendly* certification schemes for quantum states and measurements, and I work together with an experimental team to demonstrate that these methods are applicable with currently available technologies.
- I develop methods that certify *relevant properties* of quantum states and measurements, instead of the usual rigid self-testing statements.
- My results are valid for *arbitrary dimensions*, surpassing most known results that only apply to dimension two.
- I analyse a wide class of incompatibility measures. I derive universal bounds on them and show that some widely used measures do not certify certain natural properties. I also show that what constitutes the *most incompatible measurement pair* depends on which measure we choose.

The remainder of this summary is organised as follows: In section II, I formally introduce the relevant notions, that is, the quantum formalism, certification schemes, and measurement incompatibility. Then, in section III, I summarise the findings of the papers that constitute the core material of this thesis, and that are attached to this summary in their full extent. Lastly, in section IV, I outline some potential further research directions emerging from the works that are introduced in section III and in the attachments.

II. PRELIMINARIES

A. Quantum formalism

In this section I briefly introduce a few basic notions from quantum theory, that will allow me to introduce different certification schemes and the notion of measurement incompatibility in the later sections.

1. *Quantum states and measurements*

Quantum information theory treats physical systems as *information carriers*, and measurements as means of accessing the information that is encoded in the systems. This allows for a very generic description of quantum states and measurements: If the information content of two systems are the same, then their descriptions are also the same. For example, the quantum analogue of a bit, called a *qubit*, can be realised in physically very different ways (e.g. encoded in two possible paths of a photon [CY95], or in two isolated low-energy states of a trapped ion [CZ95]). However, as long as their information content is the same, quantum information theory describes them in exactly the same mathematical manner, and it is not concerned about the details of the physical implementation.

What quantum information theory *is* concerned about, is *measurement statistics*. That is, given a physical system and a measurement, what are the probabilities of the different measurement outcomes, without referring to the physical meaning of these outcomes. Consider for example the following two experiments: (i) we prepare a photon, send it through a *beam splitter* that either transmits or reflects the photon, and place a photodetector behind the beam splitter, and (ii) we prepare a trapped ion in one of its two lowest energy states, and measure its energy. As long as we get the same probabilities for the outcomes, say, "photon detected" and "lowest energy", and also for "no photon detected" and "second lowest energy", these two experiments are described by the exact same model in quantum information theory. Therefore, in the following we adapt the most general definitions of quantum states and measurements that allow us to consistently define measurement outcome probabilities.

The *state space* describes all the possible states a physical system might occupy. Not all of these different states need to be perfectly distinguishable, and in order to be able to measure distinguishability, we identify the state space with an *inner product vector space*. The inner product of two states, $\langle\psi|\varphi\rangle$, is related to their distinguishability: if $\langle\psi|\varphi\rangle = 0$, then the states $|\psi\rangle$ and $|\varphi\rangle$ are perfectly distinguishable. On the other hand, if $|\varphi\rangle = \alpha|\psi\rangle$, that is, the state vectors are aligned, then they are indistinguishable. Therefore, we identify $|\psi\rangle$ with all states of the form $\alpha|\psi\rangle$, and pick a representative of this class such that $\langle\psi|\psi\rangle = 1$. The distinguishability $D(\psi|\varphi)$ of two states then corresponds to

$$D(\psi|\varphi) = 1 - |\langle\psi|\varphi\rangle|^2 \in [0,1]. \tag{1}$$

If $D(\psi|\varphi) = 0$, then ψ and φ are indistinguishable, whereas if $D(\psi|\varphi) = 1$, then they are perfectly distinguishable. Therefore, to describe the state space, we need an inner product vector space that has the same number of perfectly distinguishable (i.e. orthogonal) elements as the number of perfectly distinguishable possible physical states. In any quantum theory – let it be quantum electrodynamics or quantum information theory – to every physical system, we assign a *Hilbert space*:

Definition II.1. *A* ***Hilbert space*** $\mathcal{H}$ *is a linear space over the field* $\mathbb{C}$ *of complex numbers, with an inner product* $\langle . | . \rangle$*, such that all Cauchy series are convergent under the norm induced by this inner product.*

Physical states then correspond to normalised elements of this Hilbert space:

Definition II.2. *A physical* ***state*** *is described by* $|\psi\rangle \in \mathcal{H}$*, such that* $\langle\psi|\psi\rangle = 1$*.*

It is clear then that the dimension of the Hilbert space corresponds to the number of perfectly distinguishable quantum states. The historic example of this mathematical construction is Schrödinger's model of a particle moving in one dimension [Sch26]:

Example II.3. *The relevant Hilbert space for a particle moving in one dimension is* $\mathcal{H} = L^2(\mathbb{R})$*, the space of square-integrable complex functions on* $\mathbb{R}$*.*

- *The inner product of* $\psi(x) \in \mathcal{H}$ *and* $\varphi(x) \in \mathcal{H}$ *is defined as*

$$\langle\psi(x)|\varphi(x)\rangle = \int_{\mathbb{R}} \bar{\psi}(x)\varphi(x)\mathrm{d}x. \tag{2}$$

- *States correspond to unit norm elements of the Hilbert space,* $\psi(x) \in \mathcal{H}$*, such that*

$$\int_{\mathbb{R}} \bar{\psi}(x)\psi(x)\mathrm{d}x = 1. \tag{3}$$

- *The probability of finding a state* $\psi(x)$ *in the region* $S \subseteq \mathbb{R}$ *is*

$$\int_{S} \bar{\psi}(x)\psi(x)\mathrm{d}x. \tag{4}$$

Notice that the normalisation $\langle\psi|\psi\rangle = 1$ corresponds to the fact that the particle is found *somewhere* with probability 1. It is also worth noting that the choice of this Hilbert space is not arbitrary: every Hilbert space that has the same dimension as the cardinality of $\mathbb{R}$ is isometric to $L^2(\mathbb{R})$ [Con94].

In the above framework, an experimenter might be able to prepare infinitely many perfectly distinguishable states, localised at different points x on the real line $\mathbb{R}$. Setting aside the fundamental ambiguities of this possibility, the practicality of preparing infinitely many perfectly distinguishable states is severely limited. Especially if the aim – as in quantum information theory – is to encode a message in a state, or to use it to perform some computation, as practical messages and computations are always finite. The number of perfectly distinguishable states corresponds precisely to the Hilbert space dimension, and therefore in the following we will focus solely on finite-dimensional Hilbert spaces. Again, there is a canonical choice of the Hilbert space for every fixed finite dimension [Con94]:

Theorem II.4. *Every d-dimensional Hilbert space $\mathcal{H}$ is isometric to $\mathbb{C}^d$, the space of d-dimensional complex vectors, with the usual scalar product.*

Therefore, d-dimensional quantum states correspond to d-dimensional complex vectors. One might think of an abstract state $|\psi\rangle \in \mathbb{C}^d$ as an information theoretical resource, capable of encoding d perfectly distinguishable messages. Let us take the example of a photon passing through a beam splitter. A beam splitter transmits the photon with probability p (the *transmittance*), and reflects it with probability $1-p$. Let us denote the transmittance path by "T" and the reflection path by "R". Also let us denote the state of a transmitted photon by $|T\rangle$, and the state of a reflected photon by $|R\rangle$. If our beam splitter transmits every photon ($p = 1$), then the photon will occupy the state $|T\rangle$ in all cases, which can be certified by placing a detector behind the beam splitter (into the path "T"). Indeed, in an ideal experiment this detector will always detect a photon. On the other hand, if our beam splitter reflects every photon ($p = 0$), then the photon will occupy the state $|R\rangle$ in all cases, and the above detector will never detect a photon. That is, the states $|T\rangle$ and $|R\rangle$ are *perfectly distinguishable* (distinguishable with probability 1) using a detector placed behind the beam splitter. Note, however, that for any other value of the transmittance, $p \in (0, 1)$, the resulting state will not be perfectly distinguishable from either the state $|T\rangle$ or $|R\rangle$, as the detector will detect a photon with probability p. This means that the probability of distinguishing this state from, say, $|R\rangle$ is $p < 1$. Having two (and no more) possible perfectly distinguishable states in this experiment, this photonic state can be described by a two-dimensional Hilbert space:

Example II.5. *The state of a photon passing through a beam splitter can be described by a vector* $|\psi\rangle \in \mathbb{C}^2$*. If we define an orthonormal basis* $\{|T\rangle, |R\rangle\}$ *on* $\mathbb{C}^2$*, then any such state can be written as*

$$|\psi\rangle = \alpha_T|T\rangle + \alpha_R|R\rangle, \quad \langle\psi|\psi\rangle = |\alpha_T|^2 + |\alpha_R|^2 = 1. \tag{5}$$

For example, a photon passing through a beam splitter with transmittance $\frac{1}{2}$ *can be described as* $|\psi\rangle = \frac{1}{\sqrt{2}}(|T\rangle + |R\rangle)$*.*

The above construction – a two-dimensional quantum system – is called a *qubit*, and it is one of the fundamental building blocks of quantum information theory. One can think of a qubit as the quantum equivalent of a bit, which is the fundamental unit of information in classical information theory. A qubit can encode two perfectly distinguishable messages (for example, $|T\rangle$ and $|R\rangle$), just like a classical bit, but also any combination of these messages of the form in Eq. (5). On the other hand, a classical bit takes either the value "0" or "1". It is therefore apparent that the information theoretical potential of a qubit might supersede that of a classical bit.

How do we extract the encoded information from a quantum state? We have already seen an example above, using a photodetector. Putting the detector behind the beam splitter is equivalent to asking the question "is the photon transmitted?". The answer to this question is "yes" (alternatively "T") if we detect a photon, or "no" (alternatively "R", because in this case we assume that the photon is reflected) if we do not detect any photons. That is, we extract some information encoded in the path degree of freedom of the photon.

The above scheme is an example of a *quantum measurement*, which is the most general way of retrieving information from a quantum state. Physically it corresponds to measuring some property of the system (e.g. its position), and the answer carries some physical meaning (e.g. "in the transmittance path" or "in the reflection path"). In quantum information theory, on the other hand, we are only interested in the probability with which the different outcomes occur. It is therefore of little relevance how we label the outcomes, and it is convenient to call for example the outcomes "T" and "R" simply "0" and "1", analogously to a classical bit. Again, in practical scenarios, it is reasonable to assume that an experimenter has only access to measurements with finitely many possible outcomes. Therefore, mathematically a finite-outcome measurement is a linear map that takes any quantum state, and maps it to a discrete probability distribution, corresponding to the outcome probabilities. The most general way to define such a map is a *positive-operator valued measure (POVM)*:

Definition II.6. *A finite-outcome quantum measurement corresponds to a finite-outcome* ***positive-operator valued measure*** *(POVM). A POVM with n outcomes defined on the Hilbert space $\mathcal{H}$ is a set of n operators, $\{M_a\}_{a=1}^n$ from the set of bounded operators $\mathcal{B}(\mathcal{H})$ on $\mathcal{H}$, such that*

$$M_a \geq 0, \qquad \sum_{a=1}^{n} M_a = \mathbb{I}, \tag{6}$$

where $\mathbb{I}$ is the identity operator on $\mathcal{H}$ and $M_a \geq 0$ means that M_a is positive semidefinite. The operators M_a are called ***POVM elements, measurement operators*** *or* ***effects****. Given a state $|\psi\rangle \in \mathcal{H}$, the probability of outcome "a" upon measuring M on the state $|\psi\rangle$ is given by the* ***Born rule****:*

$$p(a)_\psi = \langle\psi|M_a|\psi\rangle. \tag{7}$$

It is clear that $\{p(a)_\psi\}_{a=1}^n$ is indeed a probability distribution for every $|\psi\rangle \in \mathcal{H}$, that is, from Eqs. (6) and (7) it follows that

$$p(a)_\psi \geq 0 \quad \forall a = 1, \ldots, n, \qquad \sum_{a=1}^{n} p(a)_\psi = 1 \quad \forall |\psi\rangle \in \mathcal{H}. \tag{8}$$

The above example of measuring whether a photon is transmitted through a beam splitter can also be formulated as a POVM:

Example II.7. *Measuring whether a photon is transmitted through a beam splitter corresponds to the POVM*

$$\{|T\rangle\langle T|, |R\rangle\langle R|\} \tag{9}$$

on the Hilbert space $\mathbb{C}^2$, where $|T\rangle\langle T|$ ($|R\rangle\langle R|$) is the rank-1 projection onto the vector $|T\rangle$ ($|R\rangle$). That is, given a two-path photon state

$$|\psi\rangle = \alpha_T|T\rangle + \beta|R\rangle, \qquad |\alpha_T|^2 + |\alpha_R|^2 = 1, \tag{10}$$

the probability of obtaining the answer "T" is

$$p(T)_\psi = \langle\psi|T\rangle\langle T|\psi\rangle = |\alpha_T|^2, \tag{11}$$

and the probability of obtaining the answer "R" is

$$p(R)_\psi = \langle\psi|R\rangle\langle R|\psi\rangle = |\alpha_R|^2. \tag{12}$$

The above is also an example of an important class of measurements, *projective measurements.*

Definition II.8. *A POVM* $\{P_a\}_{a=1}^n$ *is a* ***projective measurement*** *if* $P_a^2 = P_a$ *for all* $a = 1, \ldots, n$.

While POVMs are the most general definition of quantum measurements, the definition of quantum states in Definition II.2 can still be generalised. First, notice that the outcome probabilities in Eq. (7) do not change if we multiply the state $|\psi\rangle$ with a complex number of modulus 1 (a phase factor). Therefore, it is convenient to identify the state $|\psi\rangle$ with the rank-1 projector $|\psi\rangle\langle\psi|$ projecting onto $|\psi\rangle$, which is invariant under the multiplication of $|\psi\rangle$ with a phase factor. Then, the Born rule in Eq. (7) can be written as

$$p(a)_\psi = \mathrm{tr}(|\psi\rangle\langle\psi| M_a), \tag{13}$$

which is now linear in the state $|\psi\rangle\langle\psi|$. This allows for the proper treatment of the following state preparation scenario: Imagine that the experimenter has access to two devices, one of them preparing the state $|\psi_1\rangle\langle\psi_1|$, while the other one preparing $|\psi_2\rangle\langle\psi_2|$. Let us assume that the experimenter first flips a biased coin, which gives "heads" with probability q, and then, based on the outcome of the coin-flip, prepares the state $|\psi_1\rangle\langle\psi_1|$ or $|\psi_2\rangle\langle\psi_2|$. If we regard the experimenter, the two preparation devices and the coin as one big state preparation device, then the output of this device is

$$\rho = q|\psi_1\rangle\langle\psi_1| + (1-q)|\psi_2\rangle\langle\psi_2|, \quad q \in [0,1], \tag{14}$$

a *convex combination* of the states $|\psi_1\rangle\langle\psi_1|$ and $|\psi_2\rangle\langle\psi_2|$. It is easy to see that the outcome probabilities of any measurement M on this state are

$$p(a)_\rho = q \cdot p(a)_{\psi_1} + (1-q) \cdot p(a)_{\psi_2}, \tag{15}$$

which still defines a probability distribution. Since we have just given an algorithm for its preparation, ρ is still a physical state, and therefore we should allow for such convex combinations in our theory. Therefore, the most general model for a quantum state is the so-called *density operator*:

Definition II.9. *The set of quantum states on the Hilbert space* $\mathcal{H}$ *is the convex hull of states of the form* $|\psi\rangle\langle\psi|$*, that is,*

$$\mathcal{S}(\mathcal{H}) = \mathrm{Conv}\{|\psi\rangle\langle\psi| \ , \ |\psi\rangle \in \mathcal{H}, \langle\psi|\psi\rangle = 1\}. \tag{16}$$

Equivalently, a quantum state on the Hilbert space $\mathcal{H}$ *is described by a* ***density operator*** $\rho \in \mathcal{B}(\mathcal{H})$*, such that*

$$\rho \geq 0, \quad \mathrm{tr}\,\rho = 1. \tag{17}$$

The outcome probabilities of a POVM $\{M_a\}_{a=1}^{n}$ on the state ρ are given by the Born rule:

$$p(a)_\rho = \mathrm{tr}(\rho M_a). \tag{18}$$

Remark II.10. *States of the form $\rho = |\psi\rangle\langle\psi|$ are called **pure states**.*

Since positive operators can be diagonalised, every d-dimensional quantum state can be written as

$$\rho = \sum_{j=0}^{d-1} \lambda_j |\psi_j\rangle\langle\psi_j|, \tag{19}$$

where $\lambda_j \geq 0$ are the eigenvalues and $|\psi_j\rangle$ are the corresponding (orthonormal) eigenstates of ρ (in case ρ is not full-rank, some λ_j are 0, and the corresponding $|\psi_j\rangle$ can be chosen to be an orthonormal basis of the kernel of ρ), and we also have that $\mathrm{tr}\,\rho = \sum_j \lambda_j = 1$.

Similarly, we can also allow for convex combinations of classical states. For a bit, this means taking convex combinations of "0" and "1", leading to the generic two-dimensional classical state

$$c = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|, \quad p \in [0,1]. \tag{20}$$

where, for a unified description of classical and quantum states, we have identified the classical state "0" ("1") with the *fixed* projector $|0\rangle\langle 0|$ ($|1\rangle\langle 1|$) on $\mathbb{C}^2$. Similarly, we can define a generic d-dimensional classical state:

Definition II.11. *Let us fix an orthonormal basis $\{|j\rangle\}_{j=0}^{d-1}$ on $\mathbb{C}^d$. Then, every d-dimensional classical state can be written as*

$$c = \sum_{j=0}^{d-1} p_j |j\rangle\langle j|, \tag{21}$$

where $\{p_j\}$ is a probability distribution, and $|j\rangle$ is a fixed basis on $\mathbb{C}^d$.

It is clear from the Eqs. (19) and (21), that the set of d-dimensional classical states is a strict subset of the set of d-dimensional quantum states, i.e. one can think of classical states as quantum states that are diagonal in a fixed basis. In the following, we will use a canonical representation of this fixed basis, usually referred to as the *computational basis*:

Definition II.12. *The set of basis vectors $\{|j\rangle\}_{j=0}^{d-1}$ on $\mathbb{C}^d$ can be represented as the **computational basis**, with vector elements*

$$(j)_k = \delta_{j+1,k}, \quad j = 0, \ldots, d-1, \quad k = 1, \ldots, d. \tag{22}$$

As an example, the computational basis on a qubit space $\mathbb{C}^2$ can be written as

$$\{|0\rangle, |1\rangle\} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}. \tag{23}$$

So far we have only been concerned with the description of a *single* physical system. However, the natural question arises: Having a description of two systems, how to describe the *composite system* of the union of these two? The axioms of quantum theory provide a prescription for such a scenario.

Definition II.13. *Given two physical systems corresponding to the Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, the composite system corresponds to the tensor product Hilbert space, $\mathcal{H}_A \otimes \mathcal{H}_B$.*

The above definition applies to both the description of states and measurements. The composite state of the systems corresponding to ρ_A on $\mathcal{H}_A$ and ρ_B on $\mathcal{H}_B$ is $\rho = \rho_A \otimes \rho_B$. Similarly, given a measurement $\{A_a\}$ on $\mathcal{H}_A$ and $\{B_b\}$ on $\mathcal{H}_B$, we can define a measurement $\{M_{ab} = A_a \otimes B_b\}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$. As an example, the state of two independent two-path photons, one in its path "0" and the other in its path "1", is written as $|\psi\rangle\langle\psi| = |0\rangle\langle 0| \otimes |1\rangle\langle 1|$. Then, given which-path measurements $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ on both photons, we can construct the measurement $\{|0\rangle\langle 0| \otimes |0\rangle\langle 0|, |0\rangle\langle 0| \otimes |1\rangle\langle 1|, |1\rangle\langle 1| \otimes |0\rangle\langle 0|, |1\rangle\langle 1| \otimes |1\rangle\langle 1|\}$ on the two-photon system, with four different possible outcomes. In order to obtain the state of one of the subsystems from the total state, we can apply the *partial trace*, which is defined on the above simple tensors as

$$\begin{aligned} \mathrm{tr}_B(\rho_A \otimes \rho_B) &= \mathrm{tr}(\rho_B)\rho_A = \rho_A, \\ \mathrm{tr}_A(\rho_A \otimes \rho_B) &= \mathrm{tr}(\rho_A)\rho_B = \rho_B, \end{aligned} \tag{24}$$

and is extended to arbitrary tensors linearly.

Note that the probabilities on the tensor product Hilbert space are still well-defined and that in the above simple cases they factorise:

$$p(ab)_\rho = \mathrm{tr}(\rho M_{ab}) = \mathrm{tr}\left[(\rho_A \otimes \rho_B)(A_a \otimes B_b)\right] = \mathrm{tr}(\rho_A A_a)\,\mathrm{tr}(\rho_B B_b) = p(a)_{\rho_A} p(b)_{\rho_B}. \tag{25}$$

This simple fact reflects that the measurement statistics obtained from independent systems are also independent. Note, however, that on the tensor product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ much more general states than those of the form $\rho_A \otimes \rho_B$ can be defined. Naturally, one might consider convex combinations of such states:

$$\rho = \sum_k p_k \cdot \rho_A^k \otimes \rho_B^k, \quad p_k \geq 0, \quad \sum_k p_k = 1. \tag{26}$$

Having this state in mind, let us now consider a scenario in which $\mathcal{H}_A$ and $\mathcal{H}_B$ correspond to physical systems that are space-like separated, for example, two simultaneous experiments in two distant laboratories. Note that according to special relativity, no communication is allowed between the two laboratories – such experimental setups are usually referred to as *nonlocal scenarios*. Consider sets of measurements on the two subsystems $\mathcal{H}_A$ and $\mathcal{H}_B$, denoted by $\{A_a^x\}$ and $\{B_b^y\}$, where a and b are the usual outcome indices, and x and y label different measurements (*measurement settings*). Then the outcome probabilities $p(ab|xy)_\rho = \operatorname{tr}[\rho(A_a^x \otimes B_b^y)]$ should intuitively factorise to convex combinations of $p(a|x)_{\rho_A^k} \cdot p(b|y)_{\rho_B^k}$. It is easy to check that this is precisely the case, as

$$\begin{aligned} p(ab|xy)_\rho &= \operatorname{tr}\left[\left(\sum_k p_k \cdot \rho_A^k \otimes \rho_B^k\right)\left(A_a^x \otimes B_b^y\right)\right] = \sum_k p_k \operatorname{tr}(\rho_A^k A_a^x) \operatorname{tr}(\rho_B^k B_b^y) \\ &= \sum_k p_k \cdot p(a|x)_{\rho_A^k} \cdot p(b|y)_{\rho_B^k}. \end{aligned} \tag{27}$$

In the following, we will see that all outcome statistics obtained from *local realistic* models are of the above form. Local realism states that physical systems have well-defined local properties, that is, every system is in one of finitely many (as per our previous assumption) locally perfectly distinguishable states,

$$\rho = |j_A\rangle\langle j_A| \otimes |j_B\rangle\langle j_B|, \quad j_A = 0, \ldots, d_A - 1, \;\; j_B = 0, \ldots, d_B - 1, \tag{28}$$

or potentially in a convex combination of such states,

$$\rho = \sum_{\lambda \in \Lambda} p(\lambda) \cdot |j_A(\lambda)\rangle\langle j_A(\lambda)| \otimes |j_B(\lambda)\rangle\langle j_B(\lambda)|, \tag{29}$$

where $p(\lambda)$ is a probability distribution over some set Λ, and $j_{A/B}(.)$ are functions from Λ to $\{0, \ldots, d_{A/B} - 1\}$. One might think of the λ parameters as *hidden variables*. If the experimenter knew the exact value of the hidden variable, they could precisely tell which physical state the system is in. However, due to some noise or other randomness (in general, the *incomplete knowledge* of the experimenter), what the experimenter sees is just a random mixture of definite states. Intuitively, we might think that all randomness that we see in experiments is due to such incomplete knowledge, and that given a better understanding of Nature, we would be able to eliminate all randomness, and predict all experiments with certainty.

Given the full description of the physical states, measurements in the local realistic paradigm simply *read out* the pre-defined properties of the state:

$$\bar{A}_{j_A}^x \otimes \bar{B}_{j_B}^y = |j_A\rangle\langle j_A| \otimes |j_B\rangle\langle j_B|, \quad j_A = 0, \ldots, d_A - 1, \;\; j_B = 0, \ldots, d_B - 1. \tag{30}$$

That is, the outcome probabilities are deterministic:

$$p(j_A)_{|j'_A\rangle\langle j'_A|} = \delta_{j_A, j'_A}, \tag{31}$$

where $\delta_{a,b}$ is the Kronecker delta.

Clearly, the experimenter might choose to *locally* post-process the outcome of these measurements using potentially non-deterministic *response functions*

$$\begin{aligned} p_A(a|x, j_A) \geq 0, &\quad \sum_a p_A(a|x, j_A) = 1, \\ p_B(b|y, j_B) \geq 0, &\quad \sum_b p_B(b|y, j_B) = 1, \end{aligned} \tag{32}$$

that map the original outcome "j_A" ("j_B") to a new outcome "a" ("b") with probability $p_A(a|x, j_A)$ [$p_B(b|y, j_B)$], depending on the measurement setting "x" ("y"). This gives rise to the final measurements,

$$\begin{aligned} A_a^x &= \sum_{j_A} p_A(a|x, j_A) \bar{A}_{j_A}^x = \sum_{j_A} p_A(a|x, j_A) |j_A\rangle\langle j_A| \\ B_b^y &= \sum_{j_B} p_B(b|y, j_B) \bar{B}_{j_B}^y = \sum_{j_B} p_B(b|y, j_B) |j_B\rangle\langle j_B|. \end{aligned} \tag{33}$$

Finally, the outcome distribution is given by

$$\begin{aligned} p(ab|xy)_\rho = \operatorname{tr} \Bigg[&\left(\sum_{\lambda \in \Lambda} p(\lambda) |j_A(\lambda)\rangle\langle j_A(\lambda)| \otimes |j_B(\lambda)\rangle\langle j_B(\lambda)| \right) \cdot \\ &\left(\sum_{j_A} p_A(a|x, j_A) |j_A\rangle\langle j_A| \otimes \sum_{j_B} p_B(b|y, j_B) |j_B\rangle\langle j_B| \right) \Bigg] \\ = \operatorname{tr} \Bigg[&\sum_{\lambda \in \Lambda} \sum_{j_A} \sum_{j_B} p(\lambda) \cdot p_A(a|x, j_A) \cdot p_B(b|y, j_B) \cdot \delta_{j_A, j_A(\lambda)} \cdot \delta_{j_B, j_B(\lambda)} \cdot \\ &|j_A(\lambda)\rangle\langle j_A| \otimes |j_B(\lambda)\rangle\langle j_B| \Bigg] = \sum_{\lambda \in \Lambda} p(\lambda) \cdot p_A[a|x, j_A(\lambda)] \cdot p_B[b|y, j_B(\lambda)], \end{aligned} \tag{34}$$

and therefore the statistics are of the form (27). This observation justifies the following definition:

Definition II.14. *Measurement outcome statistics on the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ of the form*

$$p(ab|xy) = \sum_\lambda p(\lambda) \cdot p_A(a|x, \lambda) \cdot p_B(b|y, \lambda) \tag{35}$$

are called ***local realistic****. Statistics that cannot be written in this form are usually referred to as (Bell-)****nonlocal****.*

Crucially, not all states on $\mathcal{H}_A \otimes \mathcal{H}_B$ can be written in the form (26), and such states can potentially lead to outcome distributions that are not local realistic. Indeed, it turns out that some quantum outcome distributions are not local realistic, a phenomenon that does not occur in classical (non-quantum) theories [EPR35, Bel64]. This means that we cannot think of states as having pre-defined local properties, and measurements as simply reading out these properties. Another consequence is that some events (measurement outcomes) are *genuinely random*, that is, even if the physical state is perfectly known, the measurement outcome is impossible to predict with certainty. Therefore, whenever the violation of local realism is certified, on might also certify genuine randomness.

Let us look at a well-known example of measurement outcome statistics that violate local realism [CHSH69].

Proposition II.15. *Consider the state $\rho = |\psi\rangle\langle\psi|$ on the Hilbert space $\mathbb{C}^2 \otimes \mathbb{C}^2$, where*

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle \right). \tag{36}$$

Let us consider a pair of two-outcome POVMs $A^0 = \{A^0_+, A^0_-\}$ and $A^1 = \{A^1_+, A^1_-\}$ on the first Hilbert space, and the pair $B^0 = \{B^0_+, B^0_-\}$ and $B^1 = \{B^1_+, B^1_-\}$ on the second Hilbert space. For later convenience, we label the outcomes "+" and "-", and introduce the observables $A^x = A^x_+ - A^x_- (= 2A^x_+ - \mathbb{I})$ and $B^y = B^y_+ - B^y_- (= 2B^y_+ - \mathbb{I})$, where $x, y \in \{0, 1\}$ label the measurement settings. Note that the observables fully specify the POVMs, and let us pick the specific observables

$$A^0 = X, \quad A^1 = Z, \quad B^0 = \frac{X + Z}{\sqrt{2}}, \quad B^1 = \frac{X - Z}{\sqrt{2}}, \tag{37}$$

where

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{38}$$

are the Pauli X and Z matrices, written in the basis $\{|0\rangle, |1\rangle\}$. Then, the measurement statistics

$$p(ab|xy)_\rho = \mathrm{tr}\left[|\psi\rangle\langle\psi| \left(A^x_a \otimes B^y_b \right) \right] \tag{39}$$

do not have a local realistic description.

Proof. Let us consider the following linear functional on the measurement statistics

$$\beta = \langle A^0 B^0 \rangle + \langle A^0 B^1 \rangle + \langle A^1 B^0 \rangle - \langle A^1 B^1 \rangle \in \mathbb{R}, \tag{40}$$

where

$$\langle A^x B^y \rangle = \mathrm{tr}[|\psi\rangle\langle\psi|(A^x \otimes B^y)] = p(++|xy) - p(+-|xy) - p(-+|xy) + p(--|xy) \tag{41}$$

is the *expectation value* of the observable $A^x \otimes B^y$. Notice that for the state in Eq. (36), we have that $\mathrm{tr}[|\psi\rangle\langle\psi|(A \otimes B)] = \frac{1}{2}\mathrm{tr}(A^{\mathrm{T}}B)$, where $(.)^{\mathrm{T}}$ is the transposition in the basis $\{|0\rangle, |1\rangle\}$. It is also easy to verify that for the Pauli matrices X, Z it holds that $X^2 = Z^2 = \mathbb{I}$ and $\mathrm{tr}(XZ) = 0$. Therefore, for the state (36) and the observables (37) we have that

$$\langle A^0 B^0\rangle = \langle A^0 B^1\rangle = \langle A^1 B^0\rangle = \frac{1}{\sqrt{2}}, \quad \langle A^1 B^1\rangle = -\frac{1}{\sqrt{2}}, \tag{42}$$

and therefore the quantum value $\beta_{\mathcal{Q}}$ of the expression (40) is $\beta_{\mathcal{Q}} = 2\sqrt{2}$.

Now I will show that in any local realistic model, we obtain that the local value $\beta_{\mathcal{L}}$ satisfies $\beta_{\mathcal{L}} \leq 2$, and therefore the statistics in Eq. (42) cannot be described by any local realistic model. First, from Definition II.14 it follows that any local realistic statistics can be written as

$$\beta_{\mathcal{L}} = \sum_{\lambda} p(\lambda)\left[\langle A^0 B^0\rangle_\lambda + \langle A^0 B^1\rangle_\lambda + \langle A^1 B^0\rangle_\lambda - \langle A^1 B^1\rangle_\lambda\right] =: \sum_{\lambda} p(\lambda)\cdot\beta_{\mathcal{L}}^{\lambda}, \tag{43}$$

where

$$\begin{aligned}\langle A^x B^y\rangle_\lambda = p_A(+|x,\lambda)\cdot p_B(+|y,\lambda) - p_A(+|x,\lambda)\cdot p_B(-|y,\lambda)\\ -p_A(-|x,\lambda)\cdot p_B(+|y,\lambda) + p_A(-|x,\lambda)\cdot p_B(-|y,\lambda),\end{aligned} \tag{44}$$

since for a fixed λ, all probabilities factorise. Notice that if we aim at maximising $\beta_{\mathcal{L}}$, the hidden variable λ is not necessary, because $\beta_{\mathcal{L}}$ is linear in $p(\lambda)$. For example, if we only have two values of the hidden variable, λ_1 and λ_2 such that $\beta_{\mathcal{L}}^{\lambda_1} \geq \beta_{\mathcal{L}}^{\lambda_2}$, then it is beneficial to set $p(\lambda_1) = 1$ and $p(\lambda_2) = 0$. This argument trivially generalises to an arbitrary number of possible values of the hidden variable, that is, we can always pick the value which gives the highest value of $\beta_{\mathcal{L}}^{\lambda}$. Therefore, the optimal local realistic statistics can be written as

$$\beta_{\mathcal{L}} = \langle A^0 B^0\rangle + \langle A^0 B^1\rangle + \langle A^1 B^0\rangle - \langle A^1 B^1\rangle \tag{45}$$

where

$$\begin{aligned}\langle A^x B^y\rangle = p_A(+|x)\cdot p_B(+|y) - p_A(+|x)\cdot p_B(-|y)\\ -p_A(-|x)\cdot p_B(+|y) + p_A(-|x)\cdot p_B(-|y).\end{aligned} \tag{46}$$

Notice that this expression is linear in all the $p_A(a|x)$ and $p_B(b|y)$. Therefore, similarly to the hidden variable argument, in the optimal strategy, *deterministic* distributions suffice, that is,

$$p_A(a|x) \in \{0,1\} \quad \text{and} \quad p_B(b|y) \in \{0,1\}. \tag{47}$$

Such statistics lead to $\langle A^x B^y\rangle = a^x \cdot b^y$, where $a^x, b^y \in \{1, -1\}$. From this, it is straightforward to verify that any local realistic statistics obey $\beta_{\mathcal{L}} \leq 2$, and therefore the quantum statistics in Proposition II.15 cannot be described by any local realistic model. $\square$

The above expression, $\beta_{\mathcal{L}} \leq 2$, is known as the *CHSH inequality*, named after Clauser, Horne, Shimony and Holt [CHSH69]. Most notably, recent experiments confirm the violation of the CHSH inequality [HBD+15, GVW+15, SMSC+15], which is a solid proof that Nature itself does not behave in a local realistic manner.

The CHSH inequality is an example of *Bell inequalities*:

Definition II.16. *A **Bell inequality** is an inequality on a certain linear combination of measurement outcome probabilities,*

$$\beta = \sum_{a,b,x,y} \alpha_{abxy} \cdot p(ab|xy) \leq \beta_{\mathcal{L}}, \tag{48}$$

satisfied by every local realistic model, such that there exist quantum states and measurements that violate this inequality, that is,

$$\beta_{\mathcal{Q}} > \beta_{\mathcal{L}}. \tag{49}$$

Note that in order to violate local realism (i.e. to violate a Bell inequality), it is essential to have multiple *measurement settings* for both parties A and B. To see this, assume that B has only access to one measurement $\{B_b\}$, while A has a set of measurements $\{A_a^x\}$. Let us define the quantities

$$\begin{aligned} p(b) &= \operatorname{tr}\left[\rho\left(\mathbb{I} \otimes B_b\right)\right]. \\ p(a|x,b) &= \begin{cases} \frac{\operatorname{tr}[\rho(A_a^x \otimes B_b)]}{p(b)} & \text{if } p(b) \neq 0 \\ 0 & \text{if } p(b) = 0, \end{cases} \end{aligned} \tag{50}$$

Then the outcome statistics for an arbitrary state $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ can be written as

$$\begin{aligned} p(ab|xy)_\rho = \operatorname{tr}\left[\rho\left(A_a^x \otimes B_b\right)\right] &= p(a|x,b) \cdot p(b) = \sum_{b'} p(b') \cdot p(a|x,b') \cdot \delta_{b,b'} \\ &=: \sum_{b'} p(b') \cdot p(a|x,b') \cdot p(b|b') = \sum_{b': p(b') \neq 0} p(b') \cdot p(a|x,b') \cdot p(b|b'), \end{aligned} \tag{51}$$

where we have defined $p(b|b') = \delta_{b,b'}$, and all the objects appearing are well-defined conditional probability distributions. Therefore, we might think of b' as a hidden variable, and we see that the above statistics admit a local realistic model.

From Eq. (27), it is also clear that another prerequisite for the violation of local realism is that the state cannot be written in the form (26). Because of this fundamental importance, this characterisation of quantum states is essential in quantum theory:

Definition II.17. *A quantum state ρ on $\mathcal{H}_A \otimes \mathcal{H}_B$ is called **separable**, if it can be written as*

$$\rho = \sum_k p_k \cdot \rho_A^k \otimes \rho_B^k, \quad p_k \geq 0, \quad \sum_k p_k = 1, \tag{52}$$

for some quantum states ρ_A^k and ρ_B^k on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Otherwise, it is called ***entangled****.*

This definition naturally generalises to more than two Hilbert spaces: A quantum state ρ on $\bigotimes_j \mathcal{H}_j$ is called (fully) separable, if it can be written as

$$\rho = \sum_k p_k \bigotimes_j \rho_j^k, \quad p_k \geq 0, \quad \sum_k p_k = 1, \tag{53}$$

for some quantum states ρ_j^k on $\mathcal{H}_j$. Otherwise, it is called entangled.

Due to their importance, entangled states are studied in great detail, and apart from the violation of local realism, they are useful for quantum teleportation, superdense coding, quantum key distribution and many more tasks [HHHH09].

3. Mutually unbiased bases

In this section, I introduce a class of measurements with great information theoretical relevance, which also forms a central object of interest for this thesis. Imagine that we have access to two measurement devices with d outcomes each, that is, two POVMs $\{A_a\}_{a=1}^d$ and $\{B_b\}_{b=1}^d$ on a Hilbert space $\mathcal{H}$. Assume that for some state $|\psi\rangle \in \mathcal{H}$, we obtain a definite outcome, say "a", of the measurement A:

$$p(a)_\psi = \langle\psi|A_a|\psi\rangle = 1. \tag{54}$$

Also assume that in every such case, the outcome of the other measurement B is completely random, that is,

$$\langle\psi|B_b|\psi\rangle = \frac{1}{d} \quad \forall b = 1, \ldots, d. \tag{55}$$

Finally, let us also assume the reverse relation:

$$\langle\psi|B_b|\psi\rangle = 1 \implies \langle\psi|A_a|\psi\rangle = \frac{1}{d} \quad \forall a = 1, \ldots, d. \tag{56}$$

In words, the measurements A and B are such that if for some state $|\psi\rangle$ the outcome of one of them is certain, then the outcome of the other one is completely random. A well-known example of such measurements is called *mutually unbiased bases*:

Definition II.18. *Let $\{|\psi_a\rangle\}_{a=1}^d$ and $\{|\varphi_b\rangle\}_{b=1}^d$ be two orthonormal bases on $\mathbb{C}^d$. These bases are called* ***mutually unbiased*** *if*

$$|\langle\psi_a|\varphi_b\rangle|^2 = \frac{1}{d} \quad \forall a, b = 1, \ldots, d. \tag{57}$$

Remark II.19. *The rank-1 projective measurements* $\{|\psi_a\rangle\langle\psi_a|\}_{a=1}^{d}$ *and* $\{|\varphi_b\rangle\langle\varphi_b|\}_{b=1}^{d}$ *corresponding to mutually unbiased bases satisfy the relations*

$$\begin{aligned}\langle\xi|\psi_a\rangle\langle\psi_a|\xi\rangle = 1 &\implies |\xi\rangle = \mathrm{e}^{\mathrm{i}\alpha}|\psi_a\rangle \implies \langle\xi|\varphi_b\rangle\langle\varphi_b|\xi\rangle = |\langle\psi_a|\varphi_b\rangle|^2 = \frac{1}{d},\\ \langle\xi|\varphi_b\rangle\langle\varphi_b|\xi\rangle = 1 &\implies |\xi\rangle = \mathrm{e}^{\mathrm{i}\alpha}|\varphi_b\rangle \implies \langle\xi|\psi_a\rangle\langle\psi_a|\xi\rangle = |\langle\psi_a|\varphi_b\rangle|^2 = \frac{1}{d}\end{aligned} \tag{58}$$

for every $a, b = 1, \dots, d$.

Mutually unbiased bases (MUBs) have a wide range of applications in quantum information theory [DEBŻ10]. They are optimal for state determination [Iva81, WF89], information locking [BW07] and the so-called mean king's problem [Ara03]. They also give rise to optimal entropic uncertainty relations [MU88], and are used in cryptographic protocols [BB84]. The simplest example appears in Proposition II.15, which also shows that MUBs are useful for violating Bell inequalities:

Example II.20. *The eigenbases of the Pauli* X *and* Z *operators are mutually unbiased. Since* $Z|0\rangle = |0\rangle$ *and* $Z|1\rangle = -|1\rangle$*, the eigenbasis of the* Z *operator is simply*

$$\{|0\rangle, |1\rangle\}, \tag{59}$$

whereas since $X|0\rangle = |1\rangle$ *and* $X|1\rangle = |0\rangle$*, the eigenbasis of the* X *operator is*

$$\left\{\frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right), \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right)\right\} =: \{|+\rangle, |-\rangle\}. \tag{60}$$

It is straightforward to verify that

$$|\langle 0|+\rangle|^2 = |\langle 0|-\rangle|^2 = |\langle 1|+\rangle|^2 = |\langle 1|-\rangle|^2 = \frac{1}{2}, \tag{61}$$

and therefore these bases are mutually unbiased.

Remark II.21. *Note that the POVMs* A^0 *and* A^1 *in Proposition II.15 are*

$$A^0 = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}, \quad A^1 = \{|+\rangle\langle +|, |-\rangle\langle -|\}. \tag{62}$$

Example II.22. *The above example can be generalised to arbitrary dimensions. Consider the generalised Pauli operators in dimension* d*,*

$$X = \sum_{j=0}^{d-1} |j+1\rangle\langle j| \quad \textit{and} \quad Z = \sum_{j=0}^{d-1} \omega^j |j\rangle\langle j|, \tag{63}$$

where $\omega_d = \mathrm{e}^{\frac{2\pi\mathrm{i}}{d}}$. *The eigenbases of these two operators are mutually unbiased.*

In every prime power dimension this construction can be further generalised, and the eigenbases of the (similarly defined) operators $X, Z, XZ, XZ^2, \ldots, XZ^{d-1}$ form $d+1$ MUBs (that is, $d+1$ bases that are pairwise mutually unbiased) [BBRV02]. In fact, this number corresponds to the maximal possible number of MUBs in any given dimension d [WF89]. Therefore, the maximal number of MUBs in prime power dimensions is known exactly. However, in any composite dimension d with prime decomposition $d = \prod_j p_j^{r_j}$, it is only known that the number of MUBs is at least $\min_j\{p_j^{r_j}\} + 1$ (using the above construction). For instance, in dimension 6 the number of MUBs is $3 \leq \#\text{MUB} \leq 7$, but the exact number is unknown. Zauner conjectured in his 1991 master's thesis [Zau91] that the maximal number of MUBs in dimension 6 is 4, and the widespread belief is that this conjecture is indeed true. However, the proof has been eluding the community since almost thirty years now.

B. Certification schemes

Using the formalism of the previous section, in this section I formally introduce different certification schemes. I start with the strongest method of self-testing (also mentioned in the introduction), then I introduce the more experimentally friendly setup of prepare-and-measure scenarios and discuss how to lower bound the dimension of the physical system using them. Then, I show how to certify the quantum nature of a prepare-and-measure experiment, and how ideas from self-testing can be adapted to this scenario.

1. Self-testing

We have already seen in the previous sections that certain measurement outcome statistics in nonlocal scenarios certify that the experiment is inherently of quantum nature, that is, it does not have a local realistic description. It turns out that in some cases much stronger statements can be drawn simply from the measurement outcome statistics. The strongest such certification scheme is called *self-testing*. In self-testing, we consider a nonlocal scenario on the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, where we will often refer to the parties as Alice and Bob. We assume that there is no communication allowed between the parties (*no signalling*), that they share some state $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, and that they have access to local measurements $\{A_a^x\}$ and $\{B_b^y\}$ on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. The aim of self-testing is to characterise the physical setup (i.e. the state and the measurements) up to the minimum freedom that is theoretically possible by only looking at the measurement outcome statistics

$$p(ab|xy) = \mathrm{tr}[\rho(A_a^x \otimes B_b^y)]. \tag{64}$$

Such characterisations are usually referred to as *device-independent (DI)*, because the experimenter treats their devices as *black boxes*, and the characterisation is made solely by looking at the inputs and outputs of these boxes.

The minimum freedom for DI characterisation is defined by operations on the state and the measurements that go unnoticed when we look only at the outcome statistics (64). First, notice that we cannot make any claims on the measurements $\{A_a^x\}$ and $\{B_b^y\}$ outside of the support of the marginal states $\rho_A = \mathrm{tr}_B\,\rho$ and $\rho_B = \mathrm{tr}_A\,\rho$. Therefore, from now on we assume that the marginal states are full-rank.

Also note that the outcome statistics in Eq. (64) do not change if we apply a *local isometry*

Definition II.23. *An **isometry** on the Hilbert space $\mathcal{H}$ is a map $V : \mathcal{H} \to \mathcal{H}'$, such that $V^\dagger V = \mathbb{I}$. That is, the mapping preserves all inner products,*

$$\langle V\psi | V\varphi\rangle = \langle \psi | V^\dagger V | \varphi\rangle = \langle \psi|\varphi\rangle \quad \forall |\psi\rangle, |\varphi\rangle \in \mathcal{H}. \tag{65}$$

*The action of the isometry V on operators $A \in \mathcal{B}(\mathcal{H})$ is defined as $A \mapsto VAV^\dagger$, which preserves the **Hilbert–Schmidt inner product**:*

$$\langle VAV^\dagger, VBV^\dagger\rangle_{HS} = \mathrm{tr}(VA^\dagger V^\dagger VBV^\dagger) = \mathrm{tr}(A^\dagger B) = \langle A, B\rangle_{HS} \quad \forall A, B \in \mathcal{B}(\mathcal{H}). \tag{66}$$

Definition II.24. *A **local isometry** on the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ is a map $V : \mathcal{H}_A \otimes \mathcal{H}_B \to \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$, such that*

$$V = V_A \otimes V_B, \tag{67}$$

where V_A and V_B are isometries on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively.

Indeed, for every local isometry it holds that

$$p^V(ab|xy) = \mathrm{tr}[V\rho V^\dagger (V_A A_a^x V_A^\dagger) \otimes (V_B B_b^y V_B^\dagger)] = \mathrm{tr}[\rho(A_a^x \otimes B_b^y)] = p(ab|xy), \tag{68}$$

that is, applying a local isometry is undetectable from the outcome statistics.

Similarly, we cannot detect if an auxiliary state is appended to the system, on which the measurements act trivially, that is,

$$p^\sigma(ab|xy) = \mathrm{tr}[(\rho \otimes \sigma)(A_a^x \otimes B_b^y \otimes \mathbb{I})] = \mathrm{tr}[\rho(A_a^x \otimes B_b^y)] = p(ab|xy) \tag{69}$$

for every state $\sigma \in \mathcal{S}(\mathcal{H}_S)$ on some auxiliary Hilbert space $\mathcal{H}_S$.

Putting the above observations together, we are in a position to formally state what it means to self-test states and measurements in a nonlocal scenario [MY98, MY04].

Definition II.25. *The outcome statistics $p(ab|xy)$ **self-test** the state $\tilde{\rho} \in \mathcal{S}(\mathcal{H}_{\tilde{A}} \otimes \mathcal{H}_{\tilde{B}})$ and the measurements $\{\tilde{A}_a^x\}$ on $\mathcal{H}_{\tilde{A}}$ and $\{\tilde{B}_b^y\}$ on $\mathcal{H}_{\tilde{B}}$ if for all states and measurements $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, $\{A_a^x\}$ on $\mathcal{H}_A$ and $\{B_b^y\}$ on $\mathcal{H}_B$, such that*

$$p(ab|xy) = \mathrm{tr}[\rho(A_a^x \otimes B_b^y)], \tag{70}$$

there exists a local isometry $V = V_A \otimes V_B : \mathcal{H}_A \otimes \mathcal{H}_B \to (\mathcal{H}_{\tilde{A}} \otimes \mathcal{H}_{\tilde{B}}) \otimes (\mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$, where $V_A : \mathcal{H}_A \to \mathcal{H}_{\tilde{A}} \otimes \mathcal{H}_{A'}$ and $V_B : \mathcal{H}_B \to \mathcal{H}_{\tilde{B}} \otimes \mathcal{H}_{B'}$, such that

$$\begin{aligned} V\rho V^\dagger &= \tilde{\rho} \otimes \sigma \\ V_A A_a^x V_A^\dagger &= \tilde{A}_a^x \otimes \mathbb{I}_{A'} \quad \forall x, a \\ V_B B_b^y V_B^\dagger &= \tilde{B}_b^y \otimes \mathbb{I}_{B'} \quad \forall y, b \end{aligned} \tag{71}$$

for some $\sigma \in \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$.

That is, upon observing $p(ab|xy)$, the experimenter can be certain that there exists a local isometry mapping the physical realisation $\rho, \{A_a^x\}, \{B_b^y\}$ to the desired state and measurements $\tilde{\rho}, \{\tilde{A}_a^x\}, \{\tilde{B}_b^y\}$. More specifically, the local isometry maps the physical realisation on $\mathcal{H}_A \otimes \mathcal{H}_B$ to the tensor product of the relevant Hilbert space $\mathcal{H}_{\tilde{A}} \otimes \mathcal{H}_{\tilde{B}}$ and some "junk" Hilbert space $\mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$. The relevant Hilbert space contains the state and the measurements to be self-tested, whereas the junk Hilbert space contains an arbitrary state on which the measurements act trivially. Notice that if the parties A and B know the precise form of the isometries V_A and V_B, they can locally *extract* the state $\tilde{\rho}$ and the measurements $\{\tilde{A}_a^x\}$ and $\{\tilde{B}_b^y\}$ by applying the isometries. It is also worth noting that in general scenarios it is not known what is the largest class of operations that preserve all outcome distributions. For example, a complex conjugation leading to ρ^*, $\{(A_a^x)^*\}$ and $\{(B_b^y)^*\}$ also preserves $p(ab|xy)$, but in general this cannot be written as a local isometry. Similar observations lead to slightly different definitions of self-testing; for a recent comprehensive review see Ref. [ŠB19].

As an example, let us take another look at the CHSH inequality in Proposition II.15. We have already seen that whenever the Bell value β in Eq. (40) exceeds 2, then the experiment violates local realism. In addition to this, it turns out that the optimal Bell violation self-tests the state and the measurements from Proposition II.15 [Tsi87, SW87, PR92].

Proposition II.26. *The maximal violation of the CHSH inequality, $\beta = 2\sqrt{2}$, self-tests the state*

$$|\tilde{\psi}\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle \right), \tag{72}$$

and the measurements corresponding to the eigenbases of the observables

$$\tilde{A}^0 = X, \quad \tilde{A}^1 = Z, \quad \tilde{B}^0 = \frac{X+Z}{\sqrt{2}}, \quad \tilde{B}^1 = \frac{X-Z}{\sqrt{2}}. \tag{73}$$

That is, upon observing the maximal violation of the CHSH inequality, the experimenter can be certain that there exists a local isometry extracting the above state and measurements from the actual physically implemented setup. For the case of the CHSH inequality, this isometry is in fact constructed from the physically implemented measurements and therefore can be straightforwardly performed in the laboratory to extract the desired state and measurements.

While self-testing statements like the one above are incredibly powerful, the formulation in Definition II.25 has little relevance to actual experiments, because one never observes perfect outcome statistics in the laboratory. In fact, one never even observes any exact statistics, due to the fact that in practice only a finite number of experimental rounds can be performed, leading only to an approximation of the actual statistics. In order to overcome these limitations, the concept of *robust self-testing* has been extensively studied [BLM+09, MYS12, YVB+14, BNS+15]. Robust self-testing statements allow for an approximate characterisation of the state and the measurements for the case of imperfect statistics, and can be applied to experiments [TWE+17].

While there are plenty of possible approximate characterisation schemes for states and measurements (see Ref. [ŠB19]), let me present one example of a robust characterisation of measurements using the CHSH inequality [Kan17].

Proposition II.27. *For the observed violation of the CHSH inequality, β, it holds that*

$$\beta \leq 2\sqrt{1+t}, \tag{74}$$

where $t := \frac{1}{2}\operatorname{tr}\left(\left|[A^0, A^1]\right| \rho_A\right) \in [0,1]$ *is the* effective commutator*,* $|A| = \sqrt{A^\dagger A}$ *is the operator absolute value and* $[A^0, A^1] = A^0 A^1 - A^1 A^0$ *is the commutator.*

Note that the effective commutator is a relevant characterisation of Alice's measurements in nonlocal scenarios. It is invariant under local isometries, and it only makes statements about the measurements on the support of the marginal state. Moreover, for the ideal case, $\beta = 2\sqrt{2}$, it is equivalent to the self-testing statement in Proposition II.26, and it gives a non-trivial statement for any violation of local realism, that is, for any $\beta > 2$.

Robust self-testing statements open up the possibility of real world applications. Since these statements provide experimentally verifiable DI characterisations of states and measurements, it is natural to propose certification schemes for quantum information processing tasks based on self-testing. Accordingly, self-testing statements have been linked to device-independent randomness generation, quantum key distribution, entanglement detection and delegated quantum computing; for a thorough account, see Ref. [ŠB19].

Self-testing is an active field of research with many potential applications, however, there are a few drawbacks associated with it. Firstly, as discussed in section II A 2, in order to violate a Bell inequality, entangled states are necessary. This makes it difficult to implement self-testing protocols experimentally, as the preparation of entangled states is rather challenging, especially in high dimensions. Secondly, in high-dimensional settings deriving theoretical results is also a big challenge, partially due to the fact that the set of operations preserving all outcome distributions is unknown. Accordingly, there are only a few self-testing results in high dimensions that are not extensions of qubit results [KŠT+19, SSKA19]. This challenge is important and timely, as with current technology experimenters can already prepare and manipulate higher dimensional quantum systems with high precision [DLB+11, FLP+12, MMZ16]. Lastly, the formulation of self-testing in Definition II.25 aims at characterising the state and the measurements up to the minimal theoretically allowed freedom. While this is particularly elegant for the theory, in practice the experimenter might just want to certify some relevant property of their setup.

Therefore, in the remainder of this section I will describe relaxations of the rigid self-testing scenarios. I introduce the prepare-and-measure scenario that does not require preparing entangled states, and is therefore easier to implement experimentally. I will discuss how to bound the dimension of the physical system and certify the quantum nature of the experiment in this setup. Then I will discuss how to adapt the notion of self-testing to prepare-and-measure scenarios.

2. Prepare-and-measure scenario

Consider again the laboratories of Alice and Bob. As in the nonlocal scenario, assume that they have some *settings*, x and y, respectively. In the *prepare-and-measure scenario*, however, they do not share any physical state. Instead, Alice uses her setting x to *prepare* the quantum state ρ_x, which then she sends to Bob. Then Bob, according to his setting y, performs a measurement $\{B_b^y\}$ and announces his outcome b. The experiment is described by the outcome statistics, which in this case is written as

$$p(b|xy) = \mathrm{tr}(\rho_x B_b^y). \tag{75}$$

Our assumption is that there is no additional communication allowed between Alice and Bob, apart from the state ρ_x.

Notice that since the set of d-dimensional quantum states can be embedded in the set of $(d+1)$-dimensional quantum states, there is in principle a larger set of outcome statistics achievable with higher dimensional states. That is, if for fixed numbers of settings $|X|$ and $|Y|$ we denote the set of achievable outcome distributions in dimension d by $\mathcal{P}_d^{|X|,|Y|}$, then we have that $\mathcal{P}_d^{|X|,|Y|} \subseteq \mathcal{P}_{d+1}^{|X|,|Y|}$. This observation leads to the idea of *dimension witnesses* [GBHA10].

Definition II.28. *An inequality on a certain linear combination of outcome probabilities*

$$\beta = \sum_{bxy} \alpha_{bxy} \cdot p(b|xy) \leq Q_d \tag{76}$$

is a ***dimension witness*** *if for every set of states and measurements ρ_x, $\{B_b^y\}$ such that*

$\rho_x \in \mathcal{S}(\mathbb{C}^d)$ and $B_b^y \in \mathcal{B}(\mathbb{C}^d)$ we have that $\beta \leq Q_d$, whereas for some $\rho_x \in \mathcal{S}(\mathbb{C}^{d+1})$ and $B_b^y \in \mathcal{B}(\mathbb{C}^{d+1})$ we have that $\beta > Q_d$.

That is, violating a dimension witness for dimension d certifies that the employed quantum states and measurements are of dimension at least $d+1$. To illustrate the concept, let me introduce probably the simplest example, using only one measurement setting.

Proposition II.29. *Consider the following "compression" task: Alice has the setting $x = 0, 1, 2$, based on which she prepares the state $\rho_x \in \mathcal{S}(\mathbb{C}^d)$, which she sends to Bob. Bob has a single measurement, $\{B_b\}_{b=0}^2$, and his task is to guess Alice's input x. The* ***average success probability***

$$\bar{p} = \frac{1}{3}\sum_{x=0}^{2} p(b = x|x) \leq Q_2 \tag{77}$$

is a dimension witness for $d = 2$ with maximal quantum values $Q_2 = \frac{2}{3}$ and $Q_3 = 1$.

In order to prove the above proposition, we will make use of the notion of the operator norm:

Definition II.30. *The* ***operator norm*** *of the operator $A \in \mathcal{B}(\mathcal{H})$ is defined as*

$$\|A\| = \sup\left\{\sqrt{\langle A\psi|A\psi\rangle}\ ,\ |\psi\rangle \in \mathcal{H}, \langle\psi|\psi\rangle = 1\right\}, \tag{78}$$

For Hermitian operators, this definition is equivalent to

$$\|A\| = \sup\left\{|\langle\psi|A|\psi\rangle|\ ,\ |\psi\rangle \in \mathcal{H}, \langle\psi|\psi\rangle = 1\right\}. \tag{79}$$

Now let us turn to the proof of the above proposition.

Proof. From the above definition, it immediately follows that $p(b = x|x) = \operatorname{tr}(\rho_x B_x) \leq \|B_x\| \leq \operatorname{tr} B_x$, and we get equality if ρ_x is the eigenstate of B_x corresponding to its largest eigenvalue, and if $\|B_x\| = \operatorname{tr} B_x$ (i.e. the rank of B_x is 1). Therefore, the average success probability for 2-dimensional systems is bounded by

$$\bar{p} = \frac{1}{3}\sum_{x=0}^{2} \operatorname{tr}(\rho_x B_x) \leq \frac{1}{3}\sum_{x=0}^{2} \|B_x\| \leq \frac{1}{3}\sum_{x=0}^{2} \operatorname{tr} B_x = \frac{1}{3}\operatorname{tr}\left(\sum_{x=0}^{2} B_x\right) = \frac{1}{3}\operatorname{tr}\mathbb{I}_2 = \frac{2}{3}. \tag{80}$$

This bound is saturated e.g. for the strategy

$$\begin{aligned} &\rho_0 = |0\rangle\langle 0|, \quad \rho_1 = \rho_2 = |1\rangle\langle 1|, \\ &B_0 = |0\rangle\langle 0|, \quad B_1 = |1\rangle\langle 1|, \quad B_2 = 0, \end{aligned} \tag{81}$$

which simply corresponds to encoding $x = 1$ and 2 in the same manner, at the cost of never winning when $x = 2$, but winning all the other cases.

On the other hand, for 3-dimensional systems we obtain $\bar{p} = 1$ with the strategy

$$\begin{aligned} &\rho_0 = |0\rangle\langle 0|, \quad \rho_1 = |1\rangle\langle 1|, \quad \rho_2 = |2\rangle\langle 2| \\ &B_0 = |0\rangle\langle 0|, \quad B_1 = |1\rangle\langle 1|, \quad B_2 = |2\rangle\langle 2|. \end{aligned} \tag{82}$$

This is simply the manifestation of the fact that in a 3-dimensional state space Alice is able to encode 3 perfectly distinguishable messages (while in a 2-dimensional space she cannot). □

Therefore, if in the above task the experimenters observe an average success probability larger than 2/3, they can be certain that the dimension of the system is at least 3. Notice that in the above scenario, quantum and classical strategies achieve the same average success probability, since the optimal strategies are classical. However, in general, dimension witnesses can also be used to certify the quantum nature of some experiment. If we fix the dimension d, we can also pose a classical version of the inequality in Eq. (76),

$$\beta = \sum_{b,x,y} \alpha_{bxy} \cdot p(b|xy) \leq C_d, \tag{83}$$

where C_d is the maximum value of β achievable by d-dimensional classical states and measurements, in the sense of Definition II.11 and Eq. (33). Since the set of d-dimensional quantum states and measurements is strictly larger than the set of d-dimensional classical states and measurements, in theory it is possible to violate the inequality (83) by employing d-dimensional quantum states and measurements. Therefore, whenever some outcome statistics violate the inequality (83), the experimenter can be certain that the experiment does not have a classical description, under the assumption that the dimension does not exceed d. Let me present a simple example of such a witness, that constitutes a central object of interest of this thesis.

Example II.31. *A "$2^d \to 1$" **quantum random access code (QRAC)** is a prepare-and-measure scenario parametrised by an integer $d \geq 2$, referring to the dimension of the employed quantum system (see also Fig. 1). Alice's settings are denoted by $x = x_1x_2$, where $x_1, x_2 \in \{1, \ldots, d\} =: [d]$. According to this setting, she prepares a state $\rho_{x_1x_2} \in \mathcal{S}(\mathbb{C}^d)$, which she sends to Bob. Bob has the setting $y \in \{1, 2\}$, according to which he performs the measurement $\{B_b^y\}_{b=1}^d$. Bob's aim is to guess Alice's setting x_y. As the figure of merit, we employ the **average success probability (ASP)**:*

$$\bar{p} = \frac{1}{2d^2} \sum_{x_1,x_2=1}^{d} \sum_{y=1}^{2} p(b = x_y|x_1, x_2, y) = \frac{1}{2d^2} \sum_{x_1,x_2=1}^{d} \sum_{y=1}^{2} \mathrm{tr}(\rho_{x_1x_2} B_{x_y}^y). \tag{84}$$

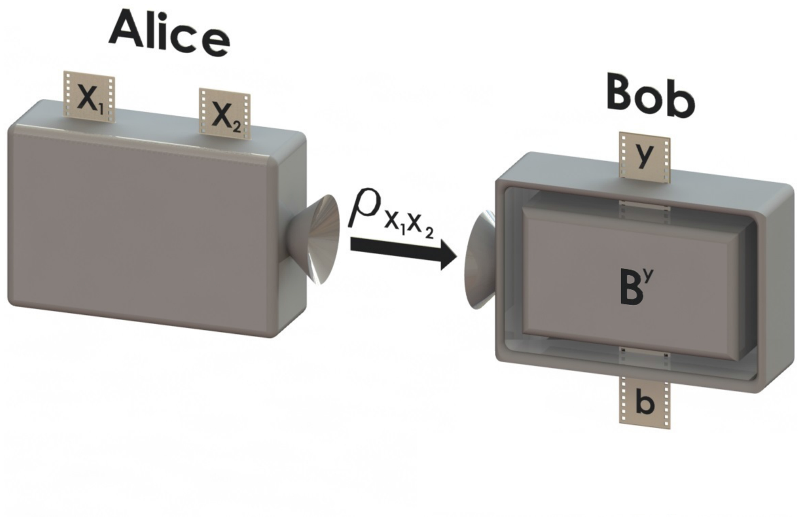


FIG. 1: A $2^d \to 1$ quantum random access code.

QRACs can also be seen as compression tasks, where Alice attempts to compress two classical dits (d-level systems) into one qudit, but this time Bob only tries to recover one of them (although Alice does not know in advance which one). As we will see in the following, quantum strategies provide an advantage in QRACs, and due to this, QRACs constitute a basic building block in many quantum information processing protocols; see e.g. [Ozo09] for a thorough account on their applications. As an example, I show that the simplest version of this task, corresponding to $d = 2$, provides a witness of quantumness:

Proposition II.32. *The $2^2 \to 1$ QRAC serves as a quantumness witness, as its classical ASP is bounded by*

$$\bar{p}_C \leq \frac{3}{4} = 0.75, \tag{85}$$

whereas quantum strategies can achieve

$$\bar{p}_Q = \frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\right) \approx 0.8536. \tag{86}$$

Proof. For the sake of this example, let us denote the settings and outcomes by $x_0, x_1, y, b \in \{0, 1\}$. Since the ASP in Eq. (84) is linear in both the state preparations and the measurements, it is sufficient to consider *extremal* states and measurements. For the classical case, this means that we fix a basis $\{|0\rangle, |1\rangle\}$ on $\mathbb{C}^2$, and each state is a pure state $\rho_x = |0\rangle\langle 0|$ or $|1\rangle\langle 1|$ (i.e. the distribution p_j in Definition II.11 is deterministic). Moreover, every measurement $\{B_0^y, B_1^y\}$ is one of the possibilities

$$\{|0\rangle\langle 0|, |1\rangle\langle 1|\}, \quad \{|1\rangle\langle 1|, |0\rangle\langle 0|\}, \quad \{\mathbb{I}, 0\}, \quad \{0, \mathbb{I}\} \tag{87}$$

(i.e. the distributions $p(b|y,j)$ in Eq. (33) are deterministic). This gives rise to finitely many possible extremal strategies, and it is straightforward to verify that $\bar{p}_C \leq \frac{3}{4}$. This bound is achieved e.g. by the strategy

$$\rho_{x_1x_2} = |x_1\rangle\langle x_1|, \quad B_b^y = |b\rangle\langle b|, \tag{88}$$

i.e. simply sending the first bit of Alice's setting, which Bob reads out. Therefore, whenever $y = 1$, they win with probability 1, and whenever $y = 2$, they win with probability $\frac{1}{2}$, resulting in the overall ASP $\bar{p}_C = \frac{3}{4}$.

For the quantum value, consider the measurements corresponding to qubit MUBs,

$$\begin{aligned} B_0^0 &= |0\rangle\langle 0|, \quad B_1^0 = |1\rangle\langle 1|, \\ B_0^1 &= |+\rangle\langle +|, \quad B_1^1 = |-\rangle\langle -| \end{aligned} \tag{89}$$

and the states $\rho_{x_0x_1}$ projecting onto the eigenvector corresponding to the largest eigenvalue of $(B_{x_0}^0 + B_{x_1}^1)$. This gives rise to

$$\begin{aligned} \bar{p} &= \frac{1}{8}\sum_{x_0,x_1=0}^{1} \mathrm{tr}\left[\rho_{x_0x_1}(B_{x_0}^0 + B_{x_1}^1)\right] = \frac{1}{8}\sum_{x_0,x_1=0}^{1} \left\|B_{x_0}^0 + B_{x_1}^1\right\| \\ &= \frac{1}{8}\left(\left\||0\rangle\langle 0| + |+\rangle\langle +|\right\| + \left\||0\rangle\langle 0| + |-\rangle\langle -|\right\| + \left\||1\rangle\langle 1| + |+\rangle\langle +|\right\| + \left\||1\rangle\langle 1| + |-\rangle\langle -|\right\|\right) \\ &= \frac{1}{8}\cdot 4\left(1 + \frac{1}{\sqrt{2}}\right) = \frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\right). \end{aligned} \tag{90}$$

□

Therefore, if the experimenters observe a $2^2 \to 1$ QRAC success probability $\bar{p} > \frac{3}{4}$, they can be certain that the systems are of quantum nature, given that the dimension is restricted to 2.

Note that in order to make any non-trivial statement about the physical setup in the prepare-and-measure scenario, it is essential to assume an upper bound on the dimension (or have some alternative assumption, see e.g. [CBB15, VHWC+17, BME+17]). Indeed, if $x = 1, \ldots, |X|$ and Alice is allowed to send any state of dimension $|X|$, then she might just send one of $|X|$ orthogonal states, which Bob is able to distinguish perfectly. This way, Bob knows exactly the setting of Alice, and they are able to reproduce every probability distribution $p(b|xy)$. This strategy works in both the quantum and the classical regime, and therefore in this case it is impossible to deduce that the experiment is of genuine quantum nature. Therefore, it is a usual assumption in prepare-and-measure scenarios to restrict the dimension to some fixed $d < |X|$, an assumption sometimes referred to as the *semi-device-independent (SDI)* assumption.

The advantage of the SDI paradigm is that it is much easier to implement experimentally than DI tasks, and it also facilitates the derivation of theoretical results. Accordingly, the SDI assumptions are used in many quantum cryptographic protocols [PB11, LPY+12, LBL+15]. SDI statements provide highly practical certification schemes, however, their potential is still not fully exploited. Note that in some scenarios, simply certifying the quantum nature of an experiment might not be completely satisfactory. Imagine that we are promised that the dimension of the quantum system is 4. However, also imagine that this 4 dimensional system is composed of the following prepare-and-measure scenario: Alice prepares a qubit, and sends it to Bob, who measures it and notes down his outcome. Then, Alice prepares another qubit, sends it again to Bob, who measures it, and his second outcome together with the first one consists his final outcome. This experiment can be written as a prepare-and-measure scenario in dimension 4, however ideally we would like to distinguish these kind of experiments from those in which Alice prepares a 4-dimensional state, sends it to Bob, who measures it and produces his outcome at once, as this latter scenario is clearly more general. Therefore, we might be able to devise a kind of *refined dimension witness*, which for $d = 4$ would read

$$\beta = \sum_{b,x,y} \alpha_{bxy} \cdot p(b|xy) \leq Q_{2\otimes 2}, \tag{91}$$

where $Q_{2\otimes 2}$ is the maximal attainable value of β with a sequential qubit strategy (mathematically, using two separable qubits and two separable measurements). These type of refined dimension witnesses are precisely what we study with my collaborators in Ref. [A], and I will discuss the results in section III A.

Furthermore, we might want to fully characterise the physical setup under the SDI assumptions. That is, we want to characterise the states ρ_x and the measurements $\{B_b^y\}$ by only looking at the outcome statistics $p(b|xy)$, assuming the dimension d. Compared to the nonlocal scenarios, in the prepare-and-measure scenario we do not allow for extra degrees of freedom, and there is no tensor product structure on the Hilbert space. Therefore, the natural class of operations preserving all outcome statistics is simply unitary transformations, since

$$p^U(b|xy) = \mathrm{tr}(U\rho_x U^\dagger U B_b^y U^\dagger) = \mathrm{tr}(\rho_x B_b^y) = p(b|xy), \tag{92}$$

whenever U is a unitary operator. Therefore, the following definition is a natural adaptation of self-testing for the prepare-and-measure scenario [TKV+18]:

Definition II.33. *The outcome statistics* $p(b|xy)$ ***self-test*** *the states* $\tilde{\rho}_x$ *and the measurements* $\{\tilde{B}_b^y\}$ *on* $\mathbb{C}^d$ ***in the prepare-and-measure scenario****, if for all states* ρ_x *and measurements* $\{B_b^y\}$ *on* $\mathbb{C}^d$ *satisfying*

$$p(b|xy) = \mathrm{tr}(\rho_x B_b^y) \tag{93}$$

there exists a unitary U *on* $\mathbb{C}^d$ *such that*

$$\begin{aligned} U\rho_x U^\dagger &= \tilde{\rho}_x \quad \forall x, \\ U B_b^y U^\dagger &= \tilde{B}_b^y \quad \forall y, b. \end{aligned} \tag{94}$$

Again, notice that in general unitaries are not the largest class of operations preserving all outcome statistics (for example, complex conjugation preserves the statistics in this scenario as well). Also notice that robust versions of this definition need to be introduced in order to make the statements experimentally relevant (which has also been studied in Ref. [TKV+18]).

While such self-testing statements in the prepare-and-measure scenario are both easy to implement experimentally and provide the theoretically most precise characterisation, they might not always be practical. In some scenarios, the experimenter might only be interested in certifying some *relevant properties* of the setup, allowing for more freedom than a unitary operation. Moreover, techniques for higher dimensional settings need to be developed in order to harness the full potential of currently available technologies. These are precisely the problems that we tackle with my collaborator in Ref. [B], and I will discuss the results in section III B.

C. Incompatible measurements

One particularly relevant property of sets of measurements in quantum information theory is that of *joint measurability*, or *compatibility*. For the sake of simplicity, I will focus on the compatibility of pairs of measurements, but this notion can be straightforwardly generalised to larger sets. This truly quantum phenomenon may occur when we have access to only a single copy of a physical state ρ. Imagine that we are interested in simultaneously measuring two distinct properties of this state, corresponding to the POVMs $\{A_a\}_{a=1}^{n_A}$ and $\{B_b\}_{b=1}^{n_B}$, which individually give rise to the outcome distributions $\{p_A(a)_\rho\}_{a=1}^{n_A} = \{\mathrm{tr}(\rho A_a)\}_{a=1}^{n_A}$ and $\{p_B(b)_\rho\}_{b=1}^{n_B} = \{\mathrm{tr}(\rho B_b)\}_{b=1}^{n_B}$. In other words, we would like to draw a variable from a joint distribution of the two measurements, $\{p(ab)_\rho\}_{a=1,b=1}^{n_A,n_B}$, that is, from a probability distribution such that $\sum_b p(ab)_\rho = p_A(a)_\rho$ and $\sum_a p(ab)_\rho = p_B(b)_\rho$ for all a, b, ρ. In quantum theory, surprisingly, there exist pairs of measurements such that it is impossible to obtain this joint distribution by measuring a single copy of the state. Such measurements are called *not jointly measurable* or sometimes *incompatible*. In order to grasp the phenomenon of incompatibility, let me start with the definition of compatible measurements, which follows rather straightforwardly from the above considerations [Lud54, BLPY16]:

Definition II.34. *Two POVMs, $\{A_a\}_{a=1}^{n_A}$ and $\{B_b\}_{b=1}^{n_B}$ on the Hilbert space $\mathbb{C}^d$ are called* ***jointly measurable*** *or* ***compatible*** *if there exists a so-called* ***parent POVM*** *$\{G_{ab}\}_{a=1,b=1}^{n_A,n_B}$ on $\mathbb{C}^d$ such that*

$$\begin{aligned} \sum_{b=1}^{n_B} G_{ab} &= A_a \quad \forall a, \\ \sum_{a=1}^{n_A} G_{ab} &= B_b \quad \forall b. \end{aligned} \tag{95}$$

Otherwise, they are called ***not jointly measurable*** *or* ***incompatible****.*

This definition captures the possibility of drawing a variable from the joint distribution of $p(a)_\rho$ and $p(b)_\rho$, via the measurement $\{G_{ab}\}$. Indeed, we get that

$$\begin{aligned} \sum_{b=1}^{n_B} p_G(ab)_\rho &= \sum_{b=1}^{n_B} \mathrm{tr}(\rho G_{ab}) = \mathrm{tr}\left(\rho \sum_{b=1}^{n_B} G_{ab}\right) = \mathrm{tr}(\rho A_a) = p_A(a)_\rho \quad \forall a, \\ \sum_{a=1}^{n_A} p_G(ab)_\rho &= \sum_{a=1}^{n_A} \mathrm{tr}(\rho G_{ab}) = \mathrm{tr}\left(\rho \sum_{a=1}^{n_A} G_{ab}\right) = \mathrm{tr}(\rho B_b) = p_B(b)_\rho \quad \forall b. \end{aligned} \tag{96}$$

In the following, we will denote the set of compatible (jointly measurable) POVM pairs in dimension d with outcome numbers n_A and n_B by $\mathbf{JM}_d^{n_A,n_B}$, and all such POVM pairs by $\mathbf{POVM}_d^{n_A,n_B}$.

To give a better idea on compatible and incompatible measurements, let me provide examples for both of them.

Example II.35. *A pair of commuting measurements,* $\{A_a\}_{a=1}^{n_A}$ *and* $\{B_b\}_{b=1}^{n_B}$ *such that*

$$[A_a, B_b] = A_a B_b - B_b A_a = 0 \quad \forall a, b \tag{97}$$

is jointly measurable with the parent POVM

$$G_{ab} = A_a B_b, \tag{98}$$

which is positive in this case, as $A_a B_b = A_a^{1/2} B_b A_a^{1/2}$*. For example, the trivial "coin toss" measurements*

$$\left\{\frac{\mathbb{I}}{n_A}\right\}_{a=1}^{n_A} \quad \text{and} \quad \left\{\frac{\mathbb{I}}{n_B}\right\}_{b=1}^{n_B} \tag{99}$$

are jointly measurable with the parent POVM

$$G_{ab} = \frac{\mathbb{I}}{n_A n_B}. \tag{100}$$

Note that for the case of projective measurements the above example provides a complete characterisation of compatible measurements [HRS08]:

Proposition II.36. *For the case of projective measurements, commutation and joint measurability are equivalent.*

This observation provides the opportunity to give an example of incompatible measurements.

Example II.37. *A pair of MUB measurements,* $\{|\psi_a\rangle\langle\psi_a|\}_{a=1}^{d}$ *and* $\{|\varphi_b\rangle\langle\varphi_b|\}_{b=1}^{d}$ *in dimension* d *are incompatible, because they are projective and*

$$\begin{aligned} \big[|\psi_a\rangle\langle\psi_a|, |\varphi_b\rangle\langle\varphi_b|\big]|\psi_{a'}\rangle &= \langle\psi_a|\varphi_b\rangle\langle\varphi_b|\psi_{a'}\rangle|\psi_a\rangle - \langle\varphi_b|\psi_a\rangle\langle\psi_a|\psi_{a'}\rangle|\varphi_b\rangle \\ &= \langle\psi_a|\varphi_b\rangle\langle\varphi_b|\psi_{a'}\rangle|\psi_a\rangle \neq 0 \end{aligned} \tag{101}$$

for $a \neq a'$*, that is, they do not commute.*

The definition of incompatibility in Definition II.34 turns out to be equivalent to the following, operationally more transparent definition [HMZ16]:

Definition II.38. *Two POVMs,* $\{A_a\}_{a=1}^{n_A}$ *and* $\{B_b\}_{b=1}^{n_B}$ *on the Hilbert space* $\mathbb{C}^d$ *are compatible if and only if there exists a parent POVM* $\{G_g\}_{g=1}^{n_G}$ *and post-processings* $p_A(.|g)$ *and* $p_B(.|g)$ *such that*

$$\begin{aligned} \sum_g p_A(a|g) G_g &= A_a \quad \forall a, \\ \sum_g p_B(b|g) G_g &= B_b \quad \forall b. \end{aligned} \tag{102}$$

This equivalent definition captures that instead of measuring both $\{A_a\}$ and $\{B_b\}$ to obtain the distributions $p(a)_\rho$ and $p(b)_\rho$, one only needs to measure $\{G_g\}$ to obtain the distribution $p(g)_\rho$, from which both $p(a)_\rho$ and $p(b)_\rho$ can be recovered via post-processing. This latter definition also makes it apparent that classical measurements in the sense of Eq. (33) are always compatible.

Furthermore, Definition II.38 makes it straightforward to argue for another classical feature of compatible measurements, namely that it is impossible to violate any Bell inequality if one of the parties has access only to compatible measurements [WPGF09]:

Proposition II.39. *In a nonlocal scenario, if one of the parties has access only to compatible measurements, then the resulting outcome statistics are local realistic.*

Proof. Without loss of generality, we can assume that the measurements of Bob, $\{B_b^y\}$, are compatible. That is, extending Definition II.38 to more than two measurements, there exists a parent POVM $\{G_g\}$ and post-processings $p(.|y,g)$ such that

$$B_b^y = \sum_g p(b|y,g) G_g \quad \forall y, b. \tag{103}$$

The outcome statistics of the experiment can then be written as

$$p(ab|xy) = \mathrm{tr}[\rho(A_a^x \otimes B_b^y)] = \mathrm{tr}\left[\rho\left(A_a^x \otimes \sum_g p(b|y,g) G_g\right)\right] = \sum_g p(b|y,g)\, \mathrm{tr}[\rho(A_a^x \otimes G_g)]. \tag{104}$$

Let us now define the quantities

$$\begin{aligned} p(g) &= \mathrm{tr}\left[\rho\left(\mathbb{I} \otimes G_g\right)\right]. \\ p(a|x,g) &= \begin{cases} \frac{\mathrm{tr}[\rho(A_a^x \otimes G_g)]}{p(g)} & \text{if } p(g) \neq 0 \\ 0 & \text{if } p(g) = 0, \end{cases} \end{aligned} \tag{105}$$

Using these, the outcome statistics read

$$p(ab|xy) = \sum_g p(g) \cdot p(a|x,g) \cdot p(b|y,g) = \sum_{g: p(g) \neq 0} p(g) \cdot p(a|x,g) \cdot p(b|y,g), \tag{106}$$

which is a well-defined local realistic model. □

Note that the argument essentially boils down to observing that if Bob's measurements are compatible, then his measurements can be simulated with just one setting corresponding to the parent measurement $\{G_g\}$. Then, for the case of one setting we have already seen in section II A 2 that any statistics have a local realistic description.

Apart from Bell nonlocality, incompatible measurements turn out to be useful in Einstein–Podolsky–Rosen steering [QVB14, UBGP15] and a particular class of state discrimination tasks [CHT19]. For this reason, it is desirable to go beyond the dichotomic characterisation of compatible/incompatible measurements, and devise measures that quantify *to what extent* a pair of measurements is incompatible [HMZ16]. Such measures then characterise the usefulness of measurement pairs in the above mentioned tasks.

One natural class of these quantifiers is measures based on *robustness to noise.* Imagine that we are given a pair of incompatible POVMs $\{A_a\}_{a=1}^{n_A}$ and $\{B_b\}_{b=1}^{n_B}$. Let us assume that due to some experimental imperfections, the actually implemented measurements are *noisy versions* of the original POVMs, that is, POVMs with elements

$$\begin{aligned} A_a^\eta &= \eta A_a + (1-\eta)\frac{\mathbb{I}}{n_A}, \\ B_b^\eta &= \eta B_b + (1-\eta)\frac{\mathbb{I}}{n_B}, \end{aligned} \tag{107}$$

or, alternatively,

$$(A^\eta, B^\eta) = \eta \cdot (A,B) + (1-\eta)\cdot\left(\frac{\mathbb{I}}{n_A}, \frac{\mathbb{I}}{n_B}\right), \tag{108}$$

adopting the notation $A = \{A_a\}_{a=1}^{n_A}$, where $\eta \in [0,1]$ is the *visibility* of the measurements. Clearly, we have that for $\eta = 1$, the measurements $\{A_a^\eta\}$ and $\{B_b^\eta\}$ are incompatible, and also that for $\eta = 0$ they are compatible. It is then apparent that there exists a *critical visibility*, η^*, at which the measurements become compatible. This critical visibility is a robustness measure of incompatibility, as it quantifies the amount of noise that needs to be added to the measurements to become compatible.

Note that in the above example, we have assumed that the noise takes the form of the trivial POVMs, $\{\mathbb{I}/n_A\}$ and $\{\mathbb{I}/n_B\}$. However, we might assume different models of noise. In full generality, we can assign any subset of the set $\mathbf{POVM}_d^{n_A,n_B}$ to be our noise model for the original POVM pair (A,B). Let us consider some subset $\mathbf{N}_{A,B}$, such that it contains at least one compatible pair. Then, the following notion is well-defined, and it is a meaningful robustness measure of incompatibility [C]:

Definition II.40. *Given two POVMs, $\{A_a\}_{a=1}^{n_A}$ and $\{B_b\}_{b=1}^{n_B}$ on $\mathbb{C}^d$, and a corresponding noise set, $\mathbf{N}_{A,B} \subseteq \mathbf{POVM}_d^{n_A,n_B}$ such that $\mathbf{N}_{A,B} \cap \mathbf{JM}_d^{n_A,n_B} \neq \emptyset$, we say that the* ***incompatibility robustness*** *$\eta^*_{A,B}$ of the pair (A,B) with respect to this noise model is*

$$\eta^*_{A,B} = \sup_{\substack{\eta\in[0,1] \\ (M,N)\in\mathbf{N}_{A,B}}} \left\{ \eta \;\middle|\; \eta\cdot(A,B) + (1-\eta)\cdot(M,N) \in \mathbf{JM}_d^{n_A,n_B} \right\}. \tag{109}$$

Note that if the noise set contains more than one pair, then we also need to optimise over this set. In other words, we have to find the noise pair (M, N) of which we need to add the least amount to (A, B) in order to make them compatible.

These measures are not only meaningful quantifiers of incompatibility, but are also relevant for experiments: They provide an error threshold for the experimenter on the amount of noise that their measurements can tolerate under a given noise model before becoming compatible, and therefore useless for certain quantum information protocols. Accordingly, several special cases of these measures have been studied in the literature, corresponding to different noise models [HMZ16, CS16]. However, despite the significant effort from the quantum information community focused on studying these measures, their properties and the relations between the different measures are still not well-understood. Moreover, the natural question whether there exists a single most-incompatible pair of measurements in a given dimension also remains unanswered. In order to fill these gaps in our general understanding of robustness based measures of incompatibility, together with my collaborators I study these measures, their properties and relations, and tackle the question of the most incompatible measurement pair in our work [C], the results of which I will discuss in section III C.

III. RESULTS

In this section, I discuss the results of the works [A], [B], [C], attached to this thesis. For each of these works, I summarise the state of the art at the time of writing the articles, the main results of the articles and the main technical details. For more details, please refer to the attached papers in part II.

A. Certifying an irreducible 1024-dimensional photonic state using refined dimension witnesses

1. *State of the art and results*

At the time of writing the article [A], it was already known that $2^d \to 1$ QRACs (see Example II.31) serve as quantumness witnesses. Specifically, it was shown in Refs. [AKR15, CSTP18] that in dimension d the maximal classical ASP is

$$\bar{p} \leq \bar{p}_C = \frac{1}{2}\left(1 + \frac{1}{d}\right), \tag{110}$$

and Ref. [THMB15] provided explicit quantum strategies employing MUB measurements reaching

$$\bar{p}_Q = \frac{1}{2}\left(1 + \frac{1}{\sqrt{d}}\right) > \bar{p}_C. \tag{111}$$

Therefore, under the assumption that the employed states and measurements in a $2^d \to 1$ QRAC are d-dimensional, if the experimenter observes an ASP larger than the value in Eq. (110), they can be certain that the experiment is of quantum nature.

However, the natural question arises: as discussed in section II B 2, it is reasonable to assume that a d-dimensional strategy that consists of lower dimensional sequential strategies will violate the inequality (110), but will not reach the value (111). That is, simply certifying the quantum nature of the experiment does not necessarily give a faithful characterisation of it. In large dimensions this becomes a crucial distinction, as one would ideally like to be able to differentiate, say, 10 qubits prepared sequentially (therefore their joint state is separable) from a full 1024-dimensional quantum state (that one might think of, mathematically, as 10 entangled qubits). The question is then: can we devise a certification method to distinguish these fundamentally different scenarios?

In our work [A], we answer this question positively, using the $2^d \to 1$ QRAC. Let us assume that the dimension d factorises as

$$d = d_1 \cdot d_2 \cdot \ldots \cdot d_r, \quad d_k \in \mathbb{N},\ d_k \geq 2, \tag{112}$$

and accordingly, the states prepared by Alice and the measurements of Bob also factorise as

$$\begin{aligned} \rho_x &= \rho_x^1 \otimes \rho_x^2 \otimes \cdots \otimes \rho_x^r, \\ M_b^y &= (M_{b^1}^y)^1 \otimes (M_{b^2}^y)^2 \otimes \cdots \otimes (M_{b^r}^y)^r, \end{aligned} \tag{113}$$

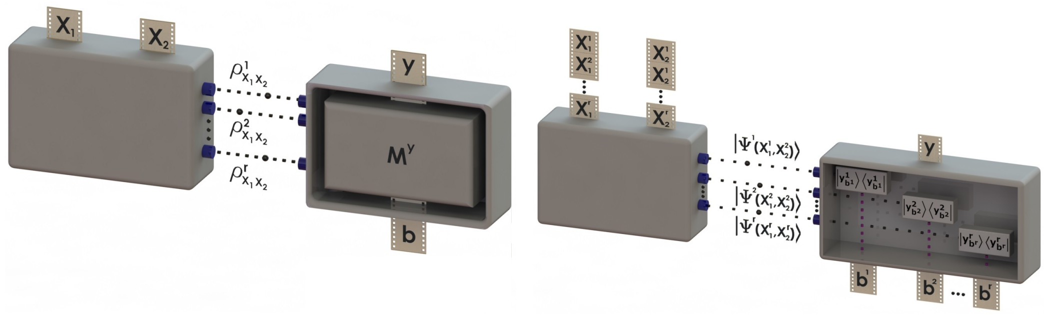
where ρ_x^k is a d_k-dimensional state and $(M_{b^k}^y)^k$ is a POVM on $\mathbb{C}^{d_k}$. We call such a construction a *product structure*, and we call it *non-trivial* if $r \geq 2$. Note that in principle we can also allow for convex combinations of states and measurements like the ones above, but due to the linearity of the ASP, these are not necessary when optimising the ASP. Also note that mathematically this formulation is equivalent to studying entanglement structures, however, in the prepare-and-measure scenario there is normally no natural tensor product structure, as we think of the systems as single d-dimensional entities. Therefore, we restrain from the terminology "entanglement", and refer to states and measurements that cannot be written in a non-trivial product form as *irreducible*. Our aim is then to certify irreducible d-dimensional states and measurements.

We achieve this certification by providing tight bounds on the achievable ASP for states and measurements of the form (113). In particular, these bounds are different for different product structures. Therefore, $2^d \to 1$ QRACs allow for certifying that the states and measurements are irreducible, if all the ASP bounds corresponding to non-trivial product structures are violated.

In order to demonstrate the applicability of our methods, we worked together with an experimental team to show that these techniques allow for certifying irreducible states and measurements of dimension 1024. The experiment performs a $2^{1024} \to 1$ QRAC using spatial degrees of freedom of a single photon. The observed ASP violates the bound for the highest non-trivial product structure, $d = 512 \cdot 2$, by more than one standard deviation, and therefore it certifies that the employed states and measurements are of irreducible dimension 1024.

2. Technical details

Our main technical tool that allows us to analyse the performance of non-trivial product structures is to show that in an optimal implementation of the QRAC game, the parties effectively play r *parallel* instances of the game, in dimensions $d_1, d_2, \ldots,$ and d_r, respectively (see Fig. 2). Accordingly, they win the full d-dimensional game if they win *all* the r sub-games. In order to reach this conclusion, first notice that the most general strategy of



(a) (b)

FIG. 2: (a) A generic QRAC with a product structure. (b) The optimal ASP can be achieved by playing r parallel QRACs, following from Lemma III.1.

Alice and Bob is in fact slightly more general than the formulation of Eq. (113) suggests. In principle, Bob might employ a so-called *sequential adaptive strategy.* Note that without loss of generality, we can assume that he applies his measurements $(M^y_{b^1})^1, (M^y_{b^2})^2, \dots, (M^y_{b^r})^r$ in a sequential fashion, since these measurements act on different systems. Then, later measurements might depend on the outcomes of former ones, i.e. Bob's measurements can in general be written as

$$M^y_b = (M^y_{b^1})^1 \otimes (M^{y,b^1}_{b^2})^2 \otimes \cdots \otimes (M^{y,b^1,b^2,\dots,b^{r-1}}_{b^r})^r. \tag{114}$$

The idea behind sequential adaptive strategies is that Bob adapts his measurements while obtaining the outcomes, therefore introducing classical correlations between the different subsystems and potentially gaining on the ASP. However, we show that these kind of strategies are not necessary.

Lemma III.1. *In a $2^d \to 1$ QRAC with non-trivial product structure, sequential adaptive strategies are not necessary to reach the optimal ASP.*

Proof. See Ref. [A]. □

What this result implies is that the optimal strategy of Alice and Bob is that they play individual QRACs on each of the r subsystems, in parallel. This is achieved by splitting up Alice's classical inputs according to the product structure, that is, $x_y = x^1_y x^2_y \dots x^r_y$, where $x^k_y \in \{1, \dots, d_k\}$ and $y \in \{1, 2\}$. Alice then encodes x^k_1 and x^k_2 into $\rho^k_{x^k_1 x^k_2}$ and sends it to Bob. Bob measures $(M^y_{b^k})^k$, and announces his outcome b^k. The final outcome is then $b = b^1 b^2 \dots b^r$, and they win the round if $b = x_y$, that is, if $b^k = x^k_y$ for all $k = 1, \dots, r$.

Let us denote by p^k_y the average probability that Bob correctly guesses x^k_y. Then, the ASP corresponding to r parallel QRACs can be written as

$$\bar{p} = \frac{1}{2}(p^1_1 \cdot p^2_1 \cdot \dots \cdot p^r_1 + p^1_2 \cdot p^2_2 \cdot \dots \cdot p^r_2). \tag{115}$$

Clearly, the quantities p_1^k and p_2^k are not independent. Our next technical result is to establish the relationship between these two. For this result, we assume that the optimal QRAC strategy is achieved by projective measurements (for a justification of this assumption, see our subsequent result in Ref. [B]).

Lemma III.2. *Let us consider a $2^d \to 1$ QRAC with projective measurements, and let us denote by p_y the average probability that Bob correctly guesses x_y. Then, the* ***quantum trade-off function*** *defined as*

$$\mathcal{M}_d^q(z) := \max\{p_2 \mid p_1 = z\}, \tag{116}$$

where the maximisation is taken over all possible d-dimensional strategies that give rise to $p_1 = z$, is given by

$$\mathcal{M}_d^q(z) = 1 - \left(\frac{d-1}{d}\right)\left(\sqrt{z} - \sqrt{\frac{1-z}{d-1}}\right)^2. \tag{117}$$

Proof. See Ref. [A]. □

Using the above trade-off functions, the ASP of a $2^d \to 1$ QRAC with non-trivial product structure can be written as

$$\bar{p} = \frac{1}{2}\left[p_1^1 \cdot p_1^2 \cdot \ldots \cdot p_1^r + \mathcal{M}_{d_1}^q(p_1^1) \cdot \mathcal{M}_{d_2}^q(p_1^2) \cdot \ldots \cdot \mathcal{M}_{d_r}^q(p_1^r)\right]. \tag{118}$$

This expression depends on r parameters, p_1^k, and can be maximised using heuristic numerical methods. This provides an effective and accurate (up to machine precision) way to obtain a tight upper bound on the QRAC ASP for any product structure.

To demonstrate the applicability of our methods, we applied our machinery to the case of $d = 1024$. Using the techniques above, we obtain bounds on the ASP for every product structure. In order to be concise, I present only a few relevant cases in Table I, but the full list can be found in Ref. [A].

Given the above values, we worked together with an experimental team to certify irreducible 1024-dimensional photonic states and measurements. While the experiment is not central to this thesis, let me briefly present a few key features of it, and its results (see Fig. 3). The quantum states are encoded in the linear transverse momentum of single photons. The single photon source is a continuous-wave laser, attenuated by an acousto-optical modulator (AOM), calibrated such that the ratio of single-photon events is 82%. The photons are sent through two spatial light modulators (SLM), which are two 32-by-32 transmissive squares, one of them modulating the amplitudes, the other one the phases.

Case	Optimal $\bar{p}$
Q_{1024}	0.515625
$Q_{512}Q_2$	0.500980
$Q_{512}C_2$	0.500973
$(Q_2)^{10}$	0.500493
Q_2C_{512}	0.500489
C_{1024}	0.500488

TABLE I: Relevant cases for a 1024-dimensional system and the respective optimal ASPs. The notation $Q_dQ_{d'}$ corresponds to a product of quantum systems of dimensions d and d'. C_d corresponds to a classical system of dimension d, and $(Q_2)^{10}$ corresponds to a product of 10 qubits.

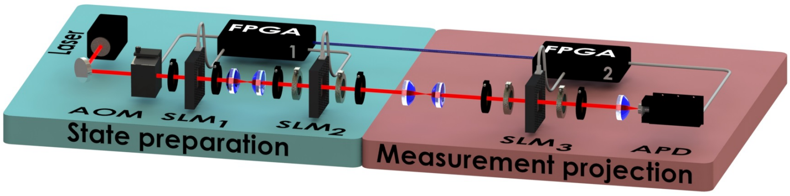


FIG. 3: Experimental setup. At the state preparation block, the spatial encoding is applied through two spatial light modulators (SLMs), and the state projection is likewise performed by an SLM combined with an avalanche single-photon detector (APD) at the measurement projection block.

The state of the photon after the two SLMs is described by

$$|\psi\rangle = \frac{1}{\sqrt{C}} \sum_{l=1}^{32} \sum_{v=1}^{32} \sqrt{t_{lv}} \mathrm{e}^{-\mathrm{i}\phi_{lv}} |c_{lv}\rangle, \tag{119}$$

where $|c_{lv}\rangle$ is the state corresponding to the square (spatial mode) (l, v), t_{lv} and ϕ_{lv} are the transmission and the phase-shift of the square (l, v), respectively, and C is a normalisation factor. Since in the experiment we have full control over the amplitudes and the phases, this state is a completely general 1024-dimensional pure quantum state.

The measurement phase ("Bob") is analogous to the state preparation. Since for the ASP in Eq. (111) we need to employ rank-1 projective measurements corresponding to two MUBs, $M_b^y = |m_b^y\rangle\langle m_b^y|$, the task is to project onto the state $|m_b^y\rangle$. We have chosen these bases carefully such that in the computational basis each vector element has the same amplitude. Therefore, in order to project onto any of these vectors, only a single SLM is needed at the measurement side, adjusting the phases. This SLM projects the state onto $|m_b^y\rangle$, and we place an avalanche single-photon detector (APD) behind it. If this detector detects a photon, we consider that it is successfully projected onto the state $|m_b^y\rangle$, and hence the outcome is "b", otherwise it is not, and the outcome is "not b".

Notice that the outcome of this measurement is binary, while in the original QRAC game the measurement needs to be d-outcome. In our case, a 1024-outcome measurement would correspond to 1024 detectors, which is certainly not feasible, and therefore this simplification was necessary. We have also adapted the ASP expression to this modified experimental setup in the following way. Let us denote by X_1 the events when we are projecting onto $|m_b^y\rangle$ and Alice's setting is such that $x_y = b$. Similarly, let us denote by X_2 the events when we are projecting onto $|m_b^y\rangle$, but $x_y \neq b$. From the experiment, we are able to count detection events in both cases, denoted by D_1 and D_2, respectively. Then, we show that the modified ASP expression

$$\bar{p} = \frac{D_1}{D_1 + D_2} \tag{120}$$

coincides with that of Eq. (84), under the assumption that $\sum_b |m_b^y\rangle\langle m_b^y| = \mathbb{I}$ for $y = 1, 2$. Using the above figure of merit, we evaluated the experiment that was running for 316 hours at an experimental round rate of 60 Hz. This frequency required the automated manipulation of the SLMs, performed by two field-programmable gate arrays (FPGA). The large sample size allowed us to evaluate the experimental data with high precision, using a Poissonian noise model on photon detection events. The results confirm an irreducible photonic state and measurements of dimension 1024, certified by the ASP $\bar{p} = 0.515 \pm 0.008$. This is more than one standard deviation larger than the second largest ASP with total dimension 1024, that is, the ASP corresponding to $Q_{512}Q_2$ (see Fig. 4).

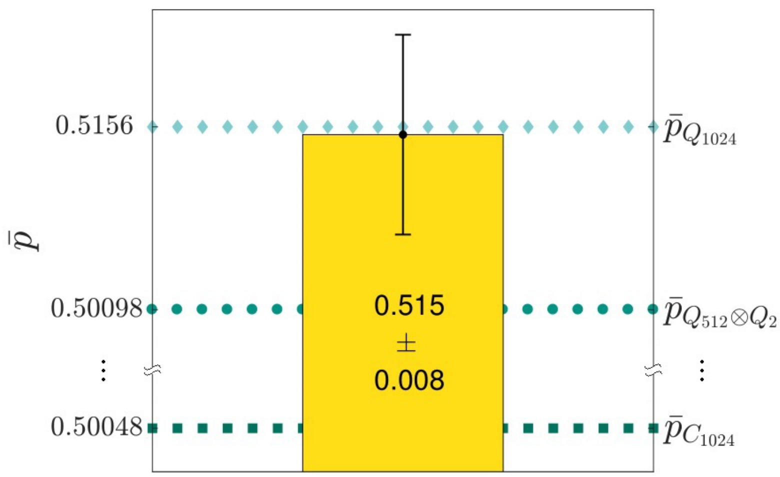


FIG. 4: Experimental results. We experimentally observe $\bar{p} = 0.515 \pm 0.008$, violating the second highest ASP bound $\bar{p}_{Q_{512}\otimes Q_2}$ (see Table I). The error bar is calculated assuming Poissonian statistics for a photon detection event.

B. Self-testing mutually unbiased bases in the prepare-and-measure scenario

1. *State of the art and results*

At the time of writing article [B], it was already known that in a $2^d \to 1$ QRAC, quantum strategies can achieve

$$\bar{p} = \frac{1}{2}\left(1 + \frac{1}{\sqrt{d}}\right) \tag{121}$$

by employing two measurements corresponding to MUBs [THMB15]. It was also a common belief that this strategy is optimal, however, it was only proven for rank-1 projective measurements [ABMP18]. It was neither clear whether this optimality holds for generic POVMs, nor whether MUBs are the unique measurements achieving this ASP. Self-testing in the prepare-and-measure scenario was also at its early stages, with only a single paper addressing this topic [TKV+18], by analysing QRACs in dimension 2. However, higher dimensional results were completely absent in the literature.

In our work [B], me and my collaborator address the above issues. We prove that indeed the ASP in Eq. (121) is the optimal quantum value using d-dimensional systems, even if allowing for POVMs. Moreover, we also show that this value can only be achieved by MUB measurements. Since different pairs of MUBs are not always equivalent up to a unitary transformation [Bri09], this is not a self-test in the sense of Definition II.33. Rather, it certifies a *relevant property* of the measurements, namely, that they correspond to MUBs. Notably, our results are essential for the methods in [ABMP18] for solving the long-standing problem of the number of MUBs in dimension 6.

Using our methods, we also provide *robust* certification schemes. Namely, we show that even by observing a sub-optimal ASP, one can lower bound the entropy of the overlaps, $\operatorname{tr}(A_i B_j)$, of Bob's two measurements $\{A_i\}_{i=1}^d$ and $\{B_j\}_{j=1}^d$, and also the sum of the operator norms, $\sum_i \|A_i\|$ and $\sum_j \|B_j\|$. The former corresponds to a sort of unbiasedness of the measurements, while the latter quantifies how close the measurements are to being rank-1 projective. When an experimenter observes the optimal ASP, both of these quantities achieve their maximal possible values, which certifies a pair of MUBs. However, even for sub-optimal ASPs, the experimenter can approximately characterise the measurements using the above quantities.

Using these quantitative characterisations, we are also able to certify two other relevant and operational properties of the measurements. Specifically, we derive a state-independent lower bound on the uncertainty generated by the two measurements, based only on the

QRAC ASP. Moreover, we provide bounds on incompatibility measures, again based only on the ASP.

2. *Technical details*

In order to prove that Eq. (121) is an upper bound for the ASP even for POVMs, our main technical tool is an operator norm inequality, proven by Kittaneh [Kit97].

Theorem III.3. *Let $A, B \geq 0$ be operators on a Hilbert space. Then* $\|A+B\| \leq \max\{\|A\|, \|B\|\} + \left\|\sqrt{A}\sqrt{B}\right\|$.

Let us denote Alice's input by $i, j \in \{1, \ldots, d\}$ and her prepared states by ρ_{ij}. Using the above theorem, and the arguments in Proposition II.32, we can bound the ASP as

$$\begin{aligned}\bar{p} &= \frac{1}{2d^2}\sum_{ij}\operatorname{tr}[\rho_{ij}(A_i+B_j)] \leq \frac{1}{2d^2}\sum_{ij}\|A_i+B_j\| \\ &\leq \frac{1}{2d^2}\sum_{ij}\left(\max\{\|A_i\|, \|B_j\|\} + \left\|\sqrt{A_i}\sqrt{B_j}\right\|\right) \leq \frac{1}{2} + \frac{1}{2d^2}\sum_{ij}\left\|\sqrt{A_i}\sqrt{B_j}\right\|,\end{aligned} \tag{122}$$

where we also used that $\|A_i\| \leq 1$ and $\|B_j\| \leq 1$ for all i, j. Then, using the fact that $\|O\| \leq \|O\|_F$, where $\|O\|_F = \sqrt{\operatorname{tr}(O^\dagger O)}$ is the Frobenius norm, we obtain

$$\bar{p} \leq \frac{1}{2} + \frac{1}{2d^2}\sum_{ij}\sqrt{\operatorname{tr}(A_iB_j)} \leq \frac{1}{2} + \frac{1}{2}\sqrt{\frac{\sum_{ij}\operatorname{tr}(A_iB_j)}{d^2}} = \frac{1}{2}\left(1+\frac{1}{\sqrt{d}}\right) =: \bar{p}_Q, \tag{123}$$

where we have used the concavity of the square-root. This concludes the proof that the value in Eq. (121) is indeed a universal quantum bound for the QRAC ASP, even for POVMs.

After this proof, we turn to the opposite question, that is, what can be said about the measurements A and B, upon observing the optimal ASP. It is clear that all the inequalities in Eqs. (122) and (123) need to be saturated. This immediately implies by the strict concavity of the square-root that all the overlaps need to be equal, that is,

$$\operatorname{tr}(A_iB_j) = \frac{1}{d} \quad \forall i, j. \tag{124}$$

It is also straightforward from the saturation of the last inequality in Eq. (122), that at least one of the measurements, say A, needs to satisfy

$$\|A_i\| = 1 \quad \forall i, \tag{125}$$

which in particular implies that A is a rank-1 projective measurement, that is, $A_i = |a_i\rangle\langle a_i|$ for some orthonormal basis $\{|a_i\rangle\}_{i=1}^d$. The last step of the certification requires a technical lemma.

Lemma III.4. *Let $A, B \geq 0$ be operators on a Hilbert space. Then, the equality $\|A+B\| = \max\{\|A\|, \|B\|\} + \left\|\sqrt{A}\sqrt{B}\right\|$ holds only if $\|A\| = \|B\|$.*

Proof. See Ref. [B]. □

This lemma in particular implies that in order to saturate Kittaneh's inequality in Eq. (122), it is required that

$$\|A_i\| = \|B_j\| \quad \forall i, j, \tag{126}$$

and we have already seen that it is necessary that $\|A_i\| = 1$ for all i. That is, both A and B need to be rank-1 projective measurements in order to achieve the optimal ASP. This together with the overlap condition Eq. (124) implies that A and B correspond to a pair of MUBs. Therefore, observing the optimal ASP certifies precisely that the measurements of Bob constitute a pair of MUBs.

In order to make this certification robust, we define an approximate characterisation of d-dimensional MUB measurements. While there is no such canonical characterisation, we choose quantities that suit our certification schemes. First, we define the *overlap entropy*

$$H_S(A, B) := H_{\frac{1}{2}}\left(\left\{\frac{1}{d}\operatorname{tr}(A_i B_j)\right\}_{ij}\right), \tag{127}$$

where $H_{\frac{1}{2}}(\{q_i\}_i) = 2\log_2\left(\sum_i \sqrt{q_i}\right)$ is the $\frac{1}{2}$-Rényi entropy of the probability distribution $\{q_i\}_i$. It is easy to see that for d-outcome measurements in dimension d,

$$H_S(A, B) \leq \log_2(d^2), \tag{128}$$

and that MUBs saturate this bound.

It is also apparent that the overlap entropy alone is not sufficient for certifying MUBs. For example, the trivial measurements $A_i = B_j = \frac{\mathbb{I}}{d}$ also saturate the bound in Eq. (128). What is missing from this characterisation is to ensure that A and B are projective, which together with the uniform overlaps implies that they are MUBs. To this end, we employ the *sum of the norms*,

$$N(A) := \sum_i \|A_i\|, \tag{129}$$

and similarly for B. It is easy to see that for d-outcome measurements in dimension d,

$$N(A) \leq d \tag{130}$$

and this bound is saturated if and only if A is rank-1 projective.

In summary, if we certify that for A and B, $H_S(A,B)$ is close to $\log_2(d^2)$ and $N(A)$ and $N(B)$ are close to d, then the measurements are close to a pair of MUBs in the sense that they are close to being rank-1 projective, and the overlaps are close to being uniform. To certify these properties, we need to derive bounds on the above quantities, as a function of the QRAC ASP. For the overlap entropy, this is a direct consequence of the first inequality in Eq. (122). It immediately follows that if we observe the ASP $\bar{p}$, then for Bob's measurements A and B it holds that

$$H_S(A,B) \geq 2\log_2[d\sqrt{d}(2\bar{p}-1)]. \tag{131}$$

This bound is non-trivial as long as $\bar{p} \geq \frac{1}{2}[1 + 1/(d\sqrt{d})]$, and the optimal ASP, $\bar{p} = \bar{p}_Q$, certifies that the overlaps are uniform. For a plot of the bound as a function of the ASP in dimension 4, see Fig. 5.

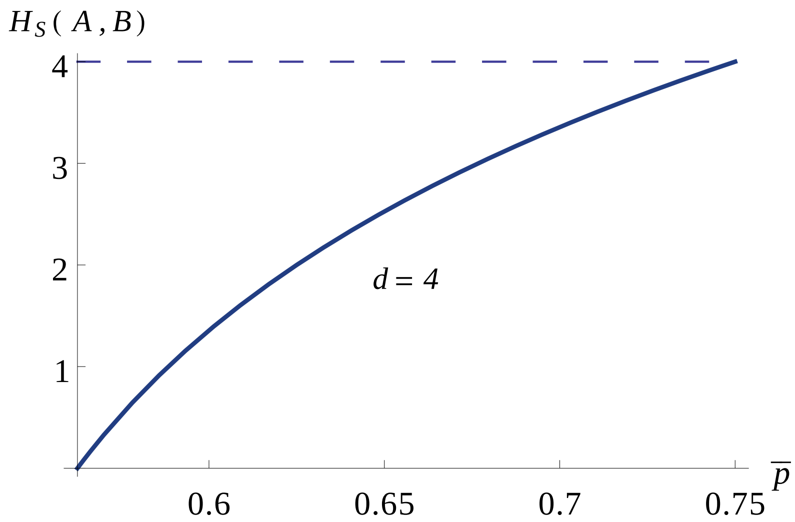


FIG. 5: Lower bound on the overlap entropy for $\bar{p} \in [\frac{1}{2} + \frac{1}{2d\sqrt{d}}, \bar{p}_Q]$ in dimension 4.

In order to devise a similar bound on the sum of the norms, our main technical tool is another operator norm inequality, proven by Kittaneh [Kit02].

Theorem III.5. *For positive semidefinite operators A and B acting on a finite-dimensional Hilbert space we have*

$$\|A+B\| \leq \frac{1}{2}\left(\|A\| + \|B\| + \sqrt{(\|A\| - \|B\|)^2 + 4\left\|\sqrt{A}\sqrt{B}\right\|^2}\right). \tag{132}$$

We introduce the quantities $n_{ij} := 1 - \frac{1}{2}(\|A_i\| + \|B_j\|)$ (norm deficiency) and $s_{ij} := \left\|\sqrt{A_i}\sqrt{B_j}\right\|$ (generalised overlap), and using the above theorem and a technical lemma, we show that

$$\bar{p} \leq \frac{1}{2} + \frac{1}{2d^2} \sum_{ij} \left[s_{ij} - (2-\sqrt{2}) s_{ij} n_{ij} \right]. \tag{133}$$

In particular, this bound serves as an alternative proof for the certification of MUBs. Note that omitting the negative term corresponds to $n_{ij} = 0$, which in turn corresponds to rank-1 projective measurements. In addition, we can bound s_{ij} by $\sqrt{\mathrm{tr}(A_i B_j)}$, which immediately gives the bound in Eq. (123).

More importantly, the bound above allows us to bound the sum of the norms in terms of the ASP. Namely, for $\bar{p} > \bar{p}_0 := \frac{1}{2} + \frac{1}{2d^2}\sqrt{(d^2-1)d}$ we can show that

$$N(A) \geq d - \frac{2+\sqrt{2}}{d}\left(1 - \sqrt{d^3(2\bar{p}-1)^2 - (d^2-1)}\right) \tag{134}$$

and by symmetry the same bound holds for $N(B)$. In particular, it is easy to see that the optimal ASP, $\bar{p} = \bar{p}_Q$, certifies $N(A) = N(B) = d$, that is, that both measurements are rank-1 projective. For a plot of the bound as a function of the ASP in dimension 4, see Fig. 6.

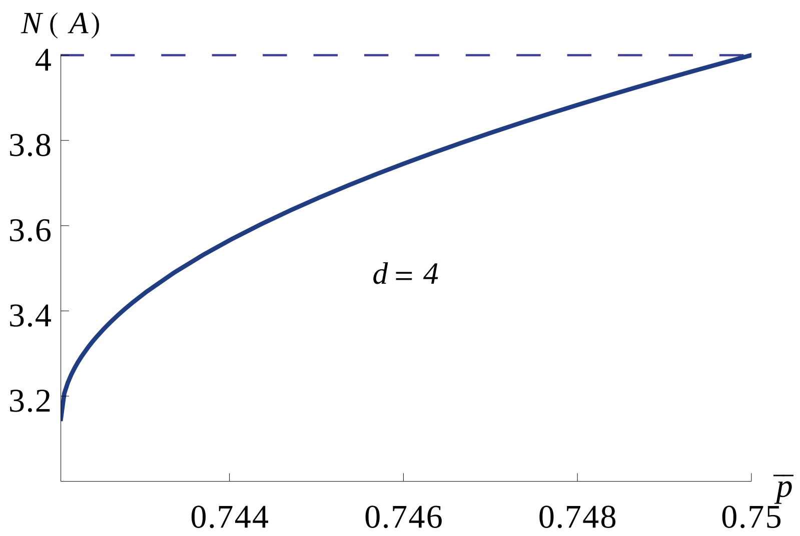


FIG. 6: Lower bound on the sum of the norms for $\bar{p} \in (\bar{p}_0, \bar{p}_Q]$ in dimension 4.

In summary, both the overlap entropy and the sum of the norms can be certified in a robust manner from the observed QRAC ASP, and these constitute a robust certification of MUB measurements.

Using these robust certificates, we are able to certify two additional relevant properties of Bob's measurements. The first such property is that of the *entropic uncertainty* of two measurements. Let us denote the Shannon entropy of the outcome distribution of the measurement A on the state ρ by $H(A)_\rho$, where the Shannon entropy of the distribution

$\{q_i\}$ is defined as $-\sum_i q_i \log_2 q_i$. Maassen and Uffink provided a state-independent lower bound on $H(A)_\rho + H(B)_\rho$ for two rank-1 projective measurements [MU88], and this lower bound is the largest for a pair of MUB measurements. That is, MUBs are optimal projective measurements for state-independent randomness extraction. The bound of Maassen and Uffink was later generalised to arbitrary POVMs in Ref. [KP02], for which it reads

$$H(A)_\rho + H(B)_\rho \geq -\log_2 c, \tag{135}$$

where $c := \max_{ij} \left\| \sqrt{A_i}\sqrt{B_j} \right\|^2$. Therefore, in order to bound the entropic uncertainty of Bob's measurements, we need an upper bound on the generalised overlap, s_{ij}. We are able to derive such a bound using our techniques, and we obtain a bound on the entropic uncertainty in terms of the QRAC ASP,

$$H(A)_\rho + H(B)_\rho \geq -2\log_2 \left(2\bar{p} - 1 + \frac{1}{d}\sqrt{d(d^2-1)[1 - d(2\bar{p}-1)^2]} \right). \tag{136}$$

The optimal ASP, $\bar{p} = \bar{p}_Q$, certifies $\log_2 d$ bits of uncertainty, which is the maximal value attainable by a pair of d-dimensional projective measurements. For a plot of the bound as a function of the ASP in dimension 4, see Fig. 7.

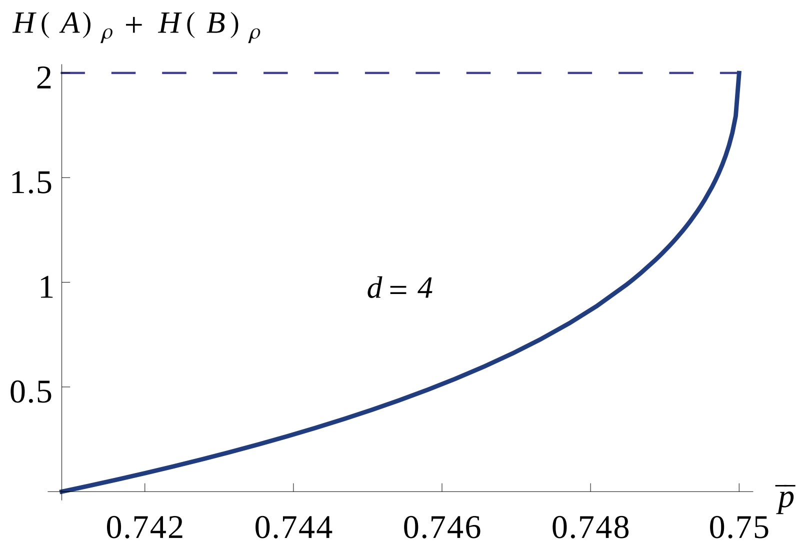


FIG. 7: Lower bound on the entropic uncertainty over the non-trivial region in dimension 4.

Finally, using our bounds, we are able to upper bound various incompatibility robustness measures of Bob's measurements in terms of the QRAC ASP. While we can obtain bounds for different measures using the upper bounds in Ref. [C], let me only present the bound for the so-called "depolarising incompatibility robustness" measure, using the upper bound from Ref. [DSFB19]. Since the analytic formula is rather complicated, let me present the bound in dimension 4 in Fig. 8, and let me remark that the optimal ASP, $\bar{p} = \bar{p}_Q$, certifies the value of the incompatibility depolarising robustness, $\eta^* = \frac{\sqrt{d}/2+1}{\sqrt{d}+1}$, which is precisely the MUB value.

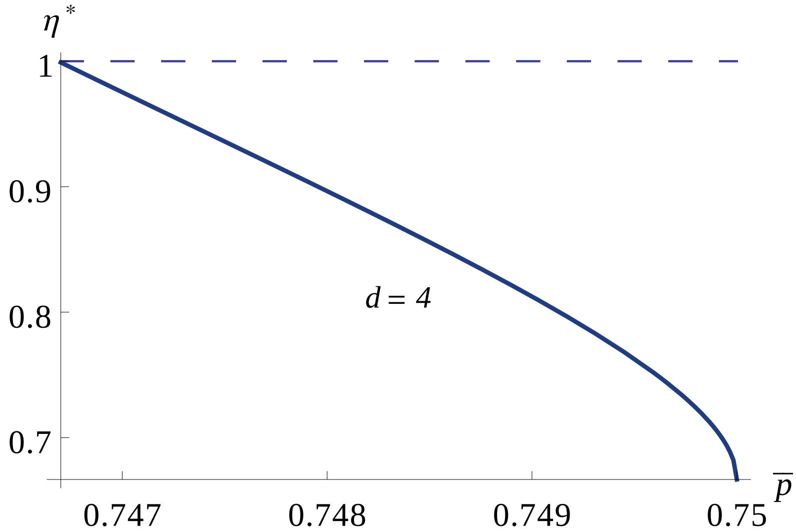


FIG. 8: Upper bound on the incompatibility robustness over the non-trivial region in dimension 4.

In summary, our techniques allow us to certify a pair of MUBs in arbitrary dimension d in a robust manner. Moreover, we are also able to robustly certify relevant properties of the measurements, namely the entropic uncertainty and different incompatibility measures.

C. Incompatibility robustness of quantum measurements: a unified framework

1. *State of the art and results*

At the time of writing article [C], robustness based measures of incompatibility have already been studied in the literature to a great extent (see Ref. [HMZ16] for an introduction). However, their properties were not systematically analysed, and the study of different measures usually appeared rather scattered in the literature. The question of which measurements are the "most incompatible" (say, in a given dimension) has not been addressed before either.

In our work [C], we address the above shortcomings by a thorough analysis of robustness based measures of incompatibility. We introduce a universal framework, that associates with every well-defined noise model a robustness measure. We make explicit connections between the properties of the noise models and the emerging properties of the corresponding incompatibility measures. Then, we turn our attention to five commonly used measures, that are all special cases of our generic framework. Using our framework, we analyse the properties of these measures, and show that some of them do not satisfy certain natural properties, and hence one should be cautious when using them. Then, using techniques from semidefinite programming, we derive universal lower bounds and measurement-dependent upper bounds on all the five measures.

We also compute the exact value of all the five measures for an arbitrary pair of rank-1 projective measurements on a qubit, and for pairs of MUBs in arbitrary dimension d. Comparing these results with our universal bounds, we deduce that for one of the measures MUBs are among the most incompatible measurement pairs in every dimension d. However, by finding explicit counterexamples, we also find that MUBs are *not* the most incompatible pairs for two other measures. Therefore, we conclude that what constitutes the most incompatible pair of measurements in general depends on the specific measure of incompatibility.

2. *Technical details*

Our universal framework for robustness based measures of incompatibility is based on Definition II.40, which I repeat here for convenience:

Definition III.6. *Given two POVMs, $\{A_a\}_{a=1}^{n_A}$ and $\{B_b\}_{b=1}^{n_B}$ on $\mathbb{C}^d$, and a corresponding noise set, $\mathbf{N}_{A,B} \subseteq \mathbf{POVM}_d^{n_A,n_B}$ such that $\mathbf{N}_{A,B} \cap \mathbf{JM}_d^{n_A,n_B} \neq \emptyset$, we say that the*

incompatibility robustness $\eta^*_{A,B}$ *of the pair* (A,B) *with respect to this noise model is*

$$\eta^*_{A,B} = \sup_{\substack{\eta\in[0,1] \\ (M,N)\in\mathbf{N}_{A,B}}} \left\{\eta \;\middle|\; \eta\cdot(A,B) + (1-\eta)\cdot(M,N) \in \mathbf{JM}_d^{n_A,n_B}\right\}. \tag{137}$$

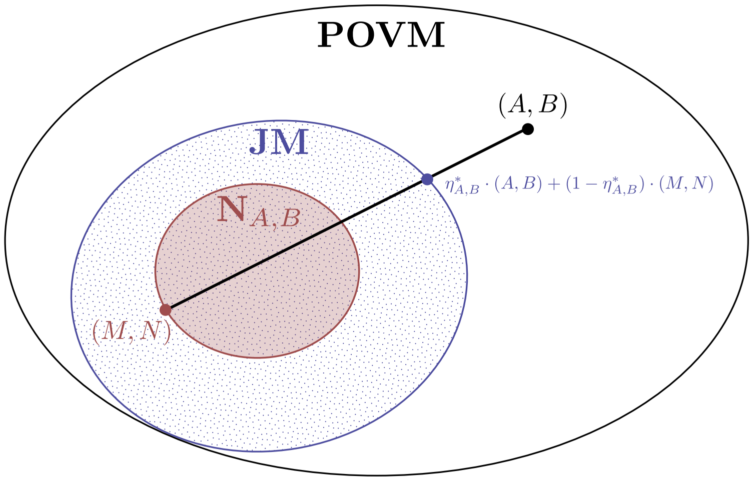


FIG. 9: Schematic representation of a generic incompatibility robustness measure for a closed and convex noise set $\mathbf{N}_{A,B}$. Note that in general the noise set need not be contained in the jointly measurable set $\mathbf{JM}$. One can also easily infer that the optimal noise pair (M,N) must lie on the boundary of $\mathbf{N}_{A,B}$ and that the optimal noisy pair $\eta^*_{A,B}\cdot(A,B)+(1-\eta^*_{A,B})\cdot(M,N)$ must lie on the boundary of $\mathbf{JM}$.

We refer to the map $\mathbf{N} : (A,B) \mapsto \mathbf{N}_{A,B}$ as the *noise model*, and the set $\mathbf{N}_{A,B}$ as the *noise set* corresponding to the pair (A,B). By noting that the set $\mathbf{JM}_d^{n_A,n_B}$ is a convex subset of the set of all POVM pairs $\mathbf{POVM}_d^{n_A,n_B}$, these robustness based measures can be interpreted geometrically, as depicted in Fig. 9. Also observe that according to Definition III.6, the lower the value $\eta^*_{A,B}$ is, the more incompatible the pair (A,B) is.

We are able to link some simple properties of the noise model to some desirable properties of the emerging incompatibility measures. In particular, whenever the noise set is closed, the supremum in Eq. (137) is always achieved. Moreover, if the noise set is covariant under unitaries, that is, $\mathbf{N}_{UAU^\dagger,UBU^\dagger} = U\,\mathbf{N}_{A,B}\,U^\dagger$, then the resulting measure is invariant under unitaries, that is, $\eta^*_{UAU^\dagger,UBU^\dagger} = \eta^*_{A,B}$.

On a more operational note, we would like our measures not to decrease under some natural operations on POVM pairs that preserve joint measurability. In other words, such "free" operations should not create more incompatible measurement pairs, a requirement motivated by resource theories [CFS16, Fri17]. We consider two such natural operations: post-processing and pre-processing.

Post-processing corresponds to stochastically relabelling measurement outcomes, just as in Eq. (32). More formally,

Definition III.7. *A post-processing β maps $\{A_a\}_{a=1}^{n_A}$ to $\{A_{a'}^{\beta}\}_{a'=1}^{n'_A}$, where*

$$A_{a'}^{\beta} = \sum_{a=1}^{n_A} \beta(a'|a) A_a, \tag{138}$$

and $\{\beta(a'|a)\}_{a'}$ is a probability distribution for every $a \in \{1, 2, \ldots, n_A\}$.

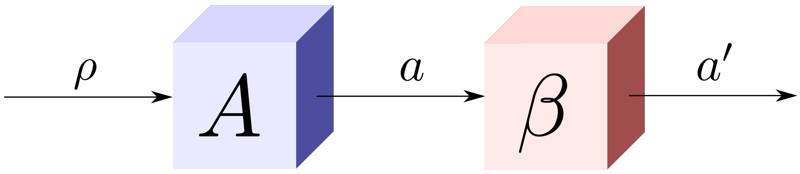


FIG. 10: Schematic representation of a post-processing of a measurement.

It is easy to verify that whenever (A, B) is jointly measurable, then so is $(A^{\beta_A}, B^{\beta_B})$, where β_A and β_B are potentially different post-processing functions.

The second class of free operations is pre-processing, which corresponds to applying a quantum channel $\Lambda^\dagger$ on the quantum state before applying the measurement.

Definition III.8. *A quantum channel is a completely positive trace preserving (CPTP) linear map $\Lambda^\dagger : \mathcal{B}(\mathbb{C}^{d'}) \to \mathcal{B}(\mathbb{C}^d)$. Complete positivity (CP) means that for every $k \in \mathbb{N}$, $k \geq 2$ we have that*

$$(\Lambda^\dagger \otimes \mathbb{I}_k) : \mathcal{B}(\mathbb{C}^{d'} \otimes \mathbb{C}^k) \to \mathcal{B}(\mathbb{C}^d \otimes \mathbb{C}^k) \tag{139}$$

preserves positivity, whereas trace preserving means that for every $\rho \in \mathcal{B}(\mathbb{C}^{d'})$, we have that $\operatorname{tr}[\Lambda^\dagger(\rho)] = \operatorname{tr}\rho$.

Formally one can think of this procedure as applying the *dual channel* Λ on the measurement. The dual of a CPTP map as the one in the above definition is a CP-unital map, that is, a CP map $\Lambda : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^{d'})$ such that $\Lambda(\mathbb{I}_d) = \mathbb{I}_{d'}$. With this definition, we can define pre-processing only in terms of the measurement.

Definition III.9. *A pre-processing Λ maps $\{A_a\}_{a=1}^{n_A}$ to $\{A_a^{\Lambda}\}_{a=1}^{n_A}$, where*

$$A_a^{\Lambda} = \Lambda(A_a), \tag{140}$$

and $\Lambda : \mathcal{B}(\mathbb{C}^d) \mapsto \mathcal{B}(\mathbb{C}^{d'})$ is a completely positive unital map.

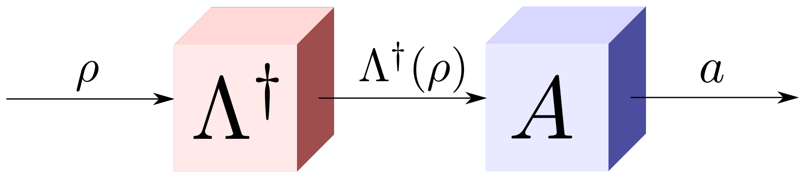


FIG. 11: Schematic representation of a pre-processing of a measurement.

It is easy to verify that whenever (A, B) is jointly measurable, then so is (A^Λ, B^Λ), where in this case we apply the same pre-processing to both measurements.

Therefore, from a meaningful measure of incompatibility, $\eta^*_{A,B}$, we expect that it does not decrease under pre- and post-processings. That is, if we denote a pre- or post-processing as a map $\Phi : (A, B) \mapsto \Phi(A, B)$, then ideally we expect that $\eta^*_{\Phi(A,B)} \geq \eta^*_{A,B}$ for all pairs (A, B). Notably, we are able to verify whether this *monotonicity* property holds for an arbitrary incompatibility robustness measure by looking only at the noise model. Specifically, whenever it holds that $\Phi(\mathbf{N}_{A,B}) \subseteq \mathbf{N}_{\Phi(A,B)}$ for all pairs (A, B), then it follows that $\eta^*_{A,B}$ is monotonic under the operation Φ.

Such monotonicity properties turn out to be crucial when looking for the *most incompatible* pairs of measurements. Specifically, if we are interested in what is the most incompatible pair of measurements in a given dimension (regardless of the number of outcomes) under a measure that is monotonic under post-processing, then the problem can be significantly simplified. Notice that every POVM pair (A, B) in dimension d can be written as a post-processing of some rank-1 POVM pair (A', B') in dimension d. One should simply consider the spectral decomposition of the POVM elements, i.e.

$$A_a = \sum_{j_a=1}^{\mathrm{rank}(A_a)} \lambda_a^{j_a} |\alpha_a^{j_a}\rangle\langle\alpha_a^{j_a}|, \tag{141}$$

and define the POVM $\tilde{A}$ with elements

$$\tilde{A}_a^{j_a} = \lambda_a^{j_a} |\alpha_a^{j_a}\rangle\langle\alpha_a^{j_a}|, \tag{142}$$

where $a = 1, \ldots, n_A$ and $j_a = 1, \ldots, \mathrm{rank}(A_a)$. The POVM $\tilde{A}$ is clearly rank-1, and using a similar construction for defining $\tilde{B}$, it is immediate to see that (A, B) can be obtained from $(\tilde{A}, \tilde{B})$ via post-processing. If our measure of incompatibility is monotonic under post-processing, then we have that $\eta^*_{A,B} \geq \eta^*_{\tilde{A},\tilde{B}}$, and therefore when looking for the most incompatible measurement pairs, it is sufficient to consider only rank-1 POVM pairs.

The last tool for finding the most incompatible measurements is to derive *universal* (measurement independent) lower bounds on incompatibility robustness measures. In case our measure is monotonic under post-processing, by the above argument it is enough to derive these bounds for rank-1 POVM pairs, as the bound will automatically apply for all pairs. Our technical tool for deriving such bounds is *semidefinite programming* [BV04]. Specifically, for a given POVM pair (A, B), the incompatibility robustness $\eta^*_{A,B}$ in Definition II.40 corresponds to the optimisation problem

$$\begin{aligned}
\eta^*_{A,B} = \sup_{\eta,\{G_{ab}\},\{M_a\},\{N_b\}} \quad & \eta \\
\text{s.t.} \quad & \eta \leq 1 \\
& G_{ab} \geq 0 \quad \forall a, b \\
& \sum_b G_{ab} = \eta A_a + (1-\eta) M_a \quad \forall a \\
& \sum_a G_{ab} = \eta B_b + (1-\eta) N_b \quad \forall b \\
& \Big(\{M_a\}_a, \{N_b\}_b\Big) \in \mathbf{N}_{A,B}\,.
\end{aligned} \tag{143}$$

For the noise models of interest, the supremum can be replaced by the maximum and the last constraint can be written as a set of linear constraints, and therefore the above optimisation problem is a semidefinite program (SDP). SDPs can be efficiently solved *numerically*, which provides a useful tool for computing the incompatibility robustness of a *specific* pair of POVMs. However, our main aim is to provide *analytic* bounds on $\eta^*_{A,B}$ for *every* pair (A, B). To this end, we can still employ the SDP formulation given in Eq. (143). Notice that any set of variables $\eta, \{G_{ab}\}, \{M_a\}, \{N_b\}$ that satisfies all the constraints will provide a lower bound on $\eta^*_{A,B}$. Therefore, in order to derive a universal lower bound on an incompatibility measure, we need to find suitable variables $\eta, \{G_{ab}\}, \{M_a\}, \{N_b\}$ such that η does not depend on the measurements (A, B), that give rise to universal bounds $\eta^*_{A,B} \geq \eta$.

The most challenging part of finding such variables turns out to be to find suitable parent POVMs $\{G_{ab}\}$. That is, a collection of positive semidefinite operators that add up to the identity, such that the marginal sums over a and b contain terms proportional to B_b and A_a, respectively. To this end, we employ a generic ansatz

$$G_{ab} \propto \{A_a, B_b\} + (\alpha_b A_a + \beta_a B_b) + \gamma_{ab}\mathbb{I} + \delta(A_a^{\frac{1}{2}} B_b A_a^{\frac{1}{2}} + B_b^{\frac{1}{2}} A_a B_b^{\frac{1}{2}}), \tag{144}$$

where α_b, β_a, γ_{ab} and δ are real parameters, and $\{A, B\} = AB+BA$ is the anticommutator and $A^{\frac{1}{2}}$ is the unique positive semidefinite operator such that $\left(A^{\frac{1}{2}}\right)^2 = A$. It is easy to see then that $\sum_{ab} G_{ab} \propto \mathbb{I}$. Notably, when both A_a and B_b are rank-1, then checking the positivity of G_{ab} is also tractable. Since for post-processing monotonic measures, bounds on rank-1 pairs of POVMs are universal, we are able to derive universal lower bounds on such measures using the above ansatz. We take a similar approach to derive measurement-dependent *upper bounds* on the robustness measures, by employing the so-called *dual* SDP, and similarly introducing ansatz solutions.

Using our techniques, we investigate five measures that are widely used in the literature. These are all special cases of Definition II.40, corresponding to different noise models $\mathbf{N}$ giving rise to the noise sets $\mathbf{N}_{A,B}$. We analyse the monotonicity of each of these measures under pre- and post-processing, derive universal lower bounds and measurement-dependent upper bounds on them, and compute the exact value for d-dimensional MUB measurements. Our findings are summarised in Table II.

	$\mathbf{N}_{A,B}$	Post	Pre	Lower	MUB value	Upper
η^{d}	$\left\{\left(\left\{\operatorname{tr} A_a \frac{\mathbb{I}}{d}\right\}_a, \left\{\operatorname{tr} B_b \frac{\mathbb{I}}{d}\right\}_b\right)\right\}$	yes	no	$\dfrac{d-2+\sqrt{d^2+4d-4}}{4(d-1)}$	$\dfrac{1}{2}\left(1+\dfrac{1}{\sqrt{d}+1}\right)$	$\dfrac{\lambda-g^{\mathrm{d}}}{f-g^{\mathrm{d}}}$
η^{r}	$\left\{\left(\left\{\frac{\mathbb{I}}{n_A}\right\}_a, \left\{\frac{\mathbb{I}}{n_B}\right\}_b\right)\right\}$	no	yes	$\dfrac{1}{2}\left(1+\dfrac{1}{\sqrt{n_A n_B}+1}\right)$		$\dfrac{\lambda-g^{\mathrm{r}}}{f-g^{\mathrm{r}}}$
η^{p}	$\left\{\left(\{p_a \mathbb{I}\}_a, \{q_b \mathbb{I}\}_b\right)\right\}$	yes		$\max\{\eta^{\mathrm{d}}, \eta^{\mathrm{r}}\}$		$\dfrac{\lambda-g^{\mathrm{p}}}{f-g^{\mathrm{p}}}$
η^{jm}	$\mathbf{JM}_d^{n_A,n_B}$	yes		$\dfrac{2\sqrt{d^2+4d-4}}{3d-2+\sqrt{d^2+4d-4}}$	$\begin{cases} 2(\sqrt{2}-1) & d=2 \\ \frac{1}{2}\left(1+\frac{1}{\sqrt{d}}\right) & d\geq 3 \end{cases}$	$\dfrac{\lambda-g^{\mathrm{jm}}}{f-g^{\mathrm{jm}}}$
η^{g}	$\mathbf{POVM}_d^{n_A,n_B}$	yes		$\dfrac{1}{2}\left(1+\dfrac{1}{\sqrt{d}}\right)$		$\dfrac{\lambda}{f}$

TABLE II: Summary of the results on the depolarising (η^{d}), random (η^{r}), probabilistic (η^{p}), jointly measurable (η^{jm}), and general (η^{g}) incompatibility robustness of pairs of POVMs. "Post" and "Pre" stand for post-processing and pre-processing monotonicity, respectively. "Lower" and "Upper" refer to lower and upper bounds on the specific measures, respectively. d is the dimension, while n_A and n_B are the outcome numbers. The quantities λ, f, g^{d}, g^{r}, g^{p} and g^{jm} are presented in Eq. (145).

The quantities λ, f, g^{d}, g^{r}, g^{p} and g^{jm} are given by

$$\begin{gathered} \lambda=\max_{a,b}\left\{\max \operatorname{Sp}\left(A_a+B_b\right)\right\}, \quad f=\sum_a \frac{\operatorname{tr} A_a^2}{d}+\sum_b \frac{\operatorname{tr} B_b^2}{d}, \\ g^{\mathrm{d}}=\sum_a\left(\frac{\operatorname{tr} A_a}{d}\right)^2+\sum_b\left(\frac{\operatorname{tr} B_b}{d}\right)^2, \quad g^{\mathrm{r}}=\frac{1}{n_A}+\frac{1}{n_B}, \\ g^{\mathrm{p}}=\min_a \frac{\operatorname{tr} A_a}{d}+\min_b \frac{\operatorname{tr} B_b}{d}, \quad \text{and} \quad g^{\mathrm{jm}}=\min_{a,b}\left\{\min \operatorname{Sp}\left(A_a+B_b\right)\right\}, \end{gathered} \tag{145}$$

where $\operatorname{Sp}(A)$ is the spectrum of the operator A.

A simple observation from Table II is that the noise sets satisfy the inclusion relations

$$\left(\mathbf{N}_{A,B}^{\mathrm{d}} \cup \mathbf{N}_{A,B}^{\mathrm{r}}\right) \subseteq \mathbf{N}_{A,B}^{\mathrm{p}} \subseteq \mathbf{N}_{A,B}^{\mathrm{jm}} \subseteq \mathbf{N}_{A,B}^{\mathrm{g}}, \tag{146}$$

which implies an ordering on the incompatibility measures

$$\max\{\eta^{\mathrm{d}}_{A,B}, \eta^{\mathrm{r}}_{A,B}\} \leq \eta^{\mathrm{p}}_{A,B} \leq \eta^{\mathrm{jm}}_{A,B} \leq \eta^{\mathrm{g}}_{A,B}. \tag{147}$$

To demonstrate our techniques and the above relations, we analytically computed the value of all five measures for a pair of rank-1 projective qubit measurements, as a function of half of the Bloch sphere angle, θ, the results of which can be seen on Fig.12. From this figure, the ordering of the measures as in Eq. (147) is apparent, as well as the observation that for $d = 2$, MUBs are the most incompatible rank-1 projective qubit measurements.

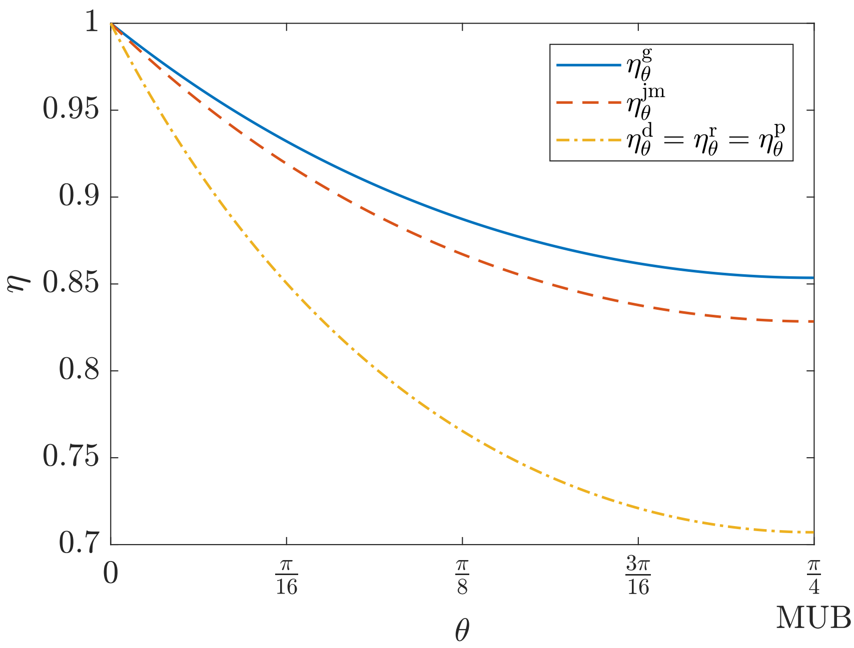


FIG. 12: The value of all the different measures for a pair of rank-one projective measurements on a qubit such that the angle between the Bloch vectors of these measurements equals 2θ. Note that the rightmost point, where $\theta = \pi/4$, corresponds to qubit MUBs, which demonstrates the fact that MUBs are the most incompatible rank-1 projective qubit measurements under all these measures. Although η^{d}, η^{r}, and η^{p} coincide in this case, this is not the case in general.

Furthermore, from Table II we see that the depolarising incompatibility robustness, η^{d}, is not monotonic under pre-processing and that the random incompatibility robustness, η^{r}, is not monotonic under post-processing. These we prove by providing explicit counterexamples in Ref. [C]. Note that the non-monotonicity of η^{r} under post-processing is essentially the reason why we cannot find a lower bound for this measure that depends only on the dimension. In particular, for this measure we show that in every dimension one can construct measurements that reach $\eta^{\mathrm{r}} = \frac{1}{2}$ by adding artificial extra outcomes that never occur (which can be considered as a post-processing).

It is also apparent from Table II that for the generalised incompatibility robustness, η^{g}, we have that MUBs are among the most incompatible pairs of d-dimensional measurements, as they saturate the universal lower bound. We also find that they are not the unique most incompatible pair. For example, if we split up one of the outcomes of a d-dimensional MUB, $A_a \to \{\frac{1}{2}A_a, \frac{1}{2}A_a\}$, leading to a measurement with an extra outcome, this will still attain the optimal value of η^{g}.

Perhaps surprisingly, we find that MUBs are *not* the most incompatible pairs of d-dimensional measurement pairs for the depolarising (η^{d}) and the probabilistic (η^{p}) incompatibility robustness measures, when the dimension is larger than 2. For the former, our best candidate for the most incompatible pair is a pair of MUBs on a 2-dimensional subspace, where the remaining subspace is irrelevant. For the latter, we find a rank-1 projective measurement pair in dimension 3 that is strictly more incompatible than both qutrit MUBs and qubit MUBs embedded in dimension 3 in the above sense. For the jointly measurable incompatibility robustness, η^{jm}, we have not found any measurement pairs in dimension 3 that are more incompatible than MUBs, but also could not prove the optimality of MUBs. To demonstrate our findings, let me present a plot of the values of the incompatibility robustness measures η^{g}, η^{jm}, η^{p} and η^{d} on a continuous path that connects 3-dimensional MUBs, $(A^{\mathrm{MUB}}, B^{\mathrm{MUB}})$, 2-dimensional MUBs embedded in 3 dimensions, $(A^{\mathrm{qMUB}}, B^{\mathrm{qMUB}})$ and the rank-1 projective pair conjectured to be optimal for η^{p}, $(A^{\mathrm{dev}}, B^{\mathrm{dev}})$, on Fig. 13.

In summary, we have thoroughly analysed the above five measures, and could verify or disprove certain natural properties. While for one of the measures we can prove that MUBs are among the most incompatible d-dimensional pairs, it is also apparent that this is not the case in general.

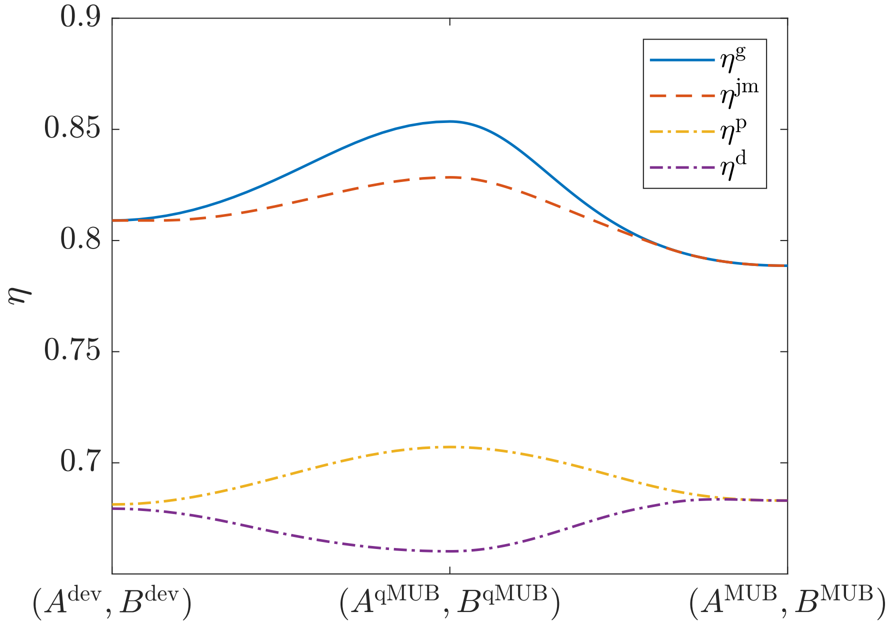


FIG. 13: The (numerical) value of the four measures along a one-parameter path of rank-one projective measurements in dimension $d = 3$. Importantly, on this path the pair $(A^{\mathrm{MUB}}, B^{\mathrm{MUB}})$ achieves the minimum value for η^{g} and η^{jm}, but it is outperformed by $(A^{\mathrm{dev}}, B^{\mathrm{dev}})$ for η^{p} and by $(A^{\mathrm{qMUB}}, B^{\mathrm{qMUB}})$ for η^{d}.

IV. OUTLOOK

While the above results fill crucial gaps in the field of semi-device-independent certification methods and measurement incompatibility, and significantly advance both of these fields, we can by no means consider these topics completely understood. In this section, I outline a few possible further research directions, stemming from, or related to the core material of this thesis.

A. Experimental self-test of MUBs in the prepare-and-measure scenario

Since the certification methods we developed in Ref. [B] are robust to noise, they are applicable to experiments. A natural continuation of this line of research is therefore to perform such an experiment. In fact, this has recently been done by my collaborators with whom I also worked together on the preparation of the article [A]. With my assistance they adapted the theory to a quantum optical setup, using multi-core optical fibres. They performed a $2^4 \to 1$ QRAC experiment with an average success probability high enough to ensure that all the quantities appearing in Ref. [B] can be certified in the non-trivial region. Together with the experimental team, we are in the process of writing up the findings of the experiment, and the work should be available on the `arXiv` repository within a few months.

B. Multi-input quantum random access codes

Both the works [A] and [B] employ $2^d \to 1$ quantum random access codes in order to certify high-dimensional quantum systems in the prepare-and-measure scenario. A natural extension of this protocol is to provide the preparation side, Alice, with more than 2 inputs, and correspondingly Bob with more than 2 measurement settings. Such QRACs are sometimes denoted as $n^d \to 1$, and have already been studied, partially by myself [Far17]. While these results are incomplete, it turns out that for $n > 2$, in general, different equivalence classes of MUBs (sets that are not related by a unitary transformation) give rise to different average success probabilities (see also Ref. [ABMP18]). Nevertheless, according to numerical evidence, our best candidates for the optimal performance in these protocols are still sets of n MUBs (whenever they exist). However, up to this date, there is no analytic proof of this, even for the case of $3^3 \to 1$ QRACs. It would be an interesting further research direction to investigate the optimal strategies in generic $n^d \to 1$ QRACs and whether it is possible to certify the optimal measurements (perhaps a specific equivalence class of MUB n-tuples) in this scenario.

C. Semi-device-independent quantum cryptography

Certification schemes in quantum theory often lead to secure quantum key distribution or random number generation protocols. In particular, the $2^2 \to 1$ QRAC was shown to give rise to secure semi-device-independent quantum key distribution [PB11]. Given our certification scheme for the general $2^d \to 1$ QRAC in Ref. [B], it is a promising future research direction to extend the methods of [PB11] and prove the semi-device-independent cryptographic security of the $2^d \to 1$ QRAC in arbitrary dimensions, potentially leading to higher key rates than that of the qubit protocol.

Another research path in this direction is to further relax the SDI assumptions, and devise certified quantum random numbers under certain plausible assumptions. This is precisely what I have been working on in the last year with my collaborators from Vienna, Brno and Bratislava. We have analysed a simple testable random number generator based on a laser, a beam splitter, a movable shutter and a photodetector. Using techniques from linear programming, we are able to bound the amount of certified randomness produced by the device, under various levels of assumptions. We do not put any constraints on the photodetector (as this is the most complex part of the setup), and provide bounds on the randomness under the assumption of a single photon source, a known photon number distribution, and a known mean value of photon numbers. We are already in the process of writing up the findings of our analysis, and are working together with an experimental team from Edinburgh to demonstrate the applicability of the device. The manuscript of this work should be available on the `arXiv` repository within a few months.

D. Device-independent certification of mutually unbiased bases

While the semi-device-independent paradigm is experimentally much more feasible, it is still of great interest whether generic d-dimensional MUBs can be self-tested in a Bell scenario. With my collaborators, we have partially solved this question in our most recent manuscript [TFR+19], by devising a family of Bell inequalities that are maximally violated by an arbitrary pair of d-dimensional MUBs. In the converse direction, the maximal violation certifies an operational definition of mutual unbiasedness, that does not refer to the Hilbert space dimension:

Definition IV.1. *We say that two d-outcome measurements $\{P_a\}_{a=1}^d$ and $\{Q_b\}_{b=1}^d$ are mutually unbiased if they are projective and the following implications hold:*

$$\begin{aligned}\langle\psi|P_a|\psi\rangle = 1 \Rightarrow \langle\psi|Q_b|\psi\rangle = \frac{1}{d}\\ \langle\psi|Q_b|\psi\rangle = 1 \Rightarrow \langle\psi|P_a|\psi\rangle = \frac{1}{d},\end{aligned} \tag{148}$$

for all a and b. That is, two projective measurements are mutually unbiased if the eigenvectors of one measurement give rise to a uniform outcome distribution for the other measurement.

It turns out that the maximal violation of the Bell inequalities introduced by us certifies precisely the above property, that can equivalently be written using the algebraic relations below.

Theorem IV.2. *Two d-outcome measurements $\{P_a\}_{a=1}^d$ and $\{Q_b\}_{b=1}^d$ are mutually unbiased if and only if*

$$P_a = dP_aQ_bP_a \qquad \textit{and} \qquad Q_b = dQ_bP_aQ_b, \tag{149}$$

for all a and b.

Naturally, any pair of d-dimensional MUBs satisfy the above criteria. However, it turns out that mutually unbiased measurements (MUMs) in the above sense are strictly more general than MUBs. In [TFR$^+$19], we prove that for $d = 2$ and 3, every MUM pair can be written as a direct sum of d-dimensional MUB pairs. However, for $d = 4$ and 5, we provide explicit examples of MUM pairs that cannot be written as a direct sum of d-dimensional MUB pairs. Lastly, we provide a protocol for device-independent quantum key distribution based on our Bell inequalities, with an optimal key rate of $\log_2 d$ bits.

E. Resource theory of incompatibility

While in our work [C] we study the monotonicity of incompatibility robustness measures under pre- and post-processing, we do not address the question of a full resource theory. That is, what are the most general, physically motivated operations that preserve joint measurability? What are the measures of incompatibility that are monotonic under such operations? Lastly, having answered these questions, can one define a resource theory of incompatibility with a single most incompatible measurement pair? These questions have been partially answered recently in Ref. [BCZ19], by considering a set of operations that allow one to freely transform between all pairs of compatible measurements. The authors also provide a complete set of incompatibility measures that are monotonic under these operations, based on certain quantum state discrimination games. However, in this theory it is unclear which incompatible measurement pairs can be transformed to which other ones, and in particular, whether there exists a single most incompatible pair of measurements. Therefore, investigating possible resource theories of incompatibility is still an open and promising future research direction.

[ABMP18] E. A. Aguilar, J. J. Borkała, P. Mironowicz, and M. Pawłowski. Connections between Mutually Unbiased Bases and Quantum Random Access Codes. *Physical Review Letters*, 121:050501, 2018.

[AKR15] A. Ambainis, D. Kravchenko, and A. Rai. Optimal Classical Random Access Codes Using Single d-level Systems. *arXiv:1510.03045*, 2015.

[Ara03] P. K. Aravind. Solution to the King's Problem in Prime Power Dimensions. *Zeitschrift für Naturforschung A*, 58(2-3):85, 2003.

[BB84] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, volume 175, page 8, 1984.

[BBRV02] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury, and F. Vatan. A new proof for the existence of mutually unbiased bases. *Algorithmica*, 34(4):512–528, 2002.

[BCZ19] F. Buscemi, E. Chitambar, and W. Zhou. A Complete Resource Theory of Quantum Incompatibility as Quantum Programmability. *arXiv:1908.11274*, 2019.

[Bel64] J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1:195–200, 1964.

[BLM+09] C.-E. Bardyn, T. C.H. Liew, S. Massar, M. McKague, and V. Scarani. Device-independent state estimation based on Bell's inequalities. *Physical Review A*, 80(6):062327, 2009.

[BLPY16] P. Busch, P. J Lahti, J.-P. Pellonpää, and K. Ylinen. *Quantum Measurement*. Springer International Publishing, 2016.

[BME+17] J. B. Brask, A. Martin, W. Esposito, R. Houlmann, J. Bowles, H. Zbinden, and N. Brunner. Megahertz-Rate Semi-Device-Independent Quantum Random Number Generators Based on Unambiguous State Discrimination. *Physical Review Applied*, 7:054018, 2017.

[BNS+15] J.-D. Bancal, M. Navascués, V. Scarani, T. Vértesi, and T. H. Yang. Physical characterization of quantum devices from nonlocal correlations. *Physical Review A*, 91:022115, 2015.

[Bri09] Brierley, S. *Mutually Unbiased Bases in Low Dimensions*. PhD thesis, University of York, 2009.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[BW07] M. A. Ballester and S. Wehner. Entropic uncertainty relations and locking: Tight bounds for mutually unbiased bases. *Physical Review A*, 75:022319, 2007.

[CBB15] R. Chaves, J. B. Brask, and N. Brunner. Device-Independent Tests of Entropy. *Physical Review Letters*, 115:110501, 2015.

[CFS16] B. Coecke, T. Fritz, and R. W. Spekkens. A mathematical theory of resources. *Information and Computation*, 250:59–86, 2016.

[CHSH69] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880, 1969.

[CHT19] C. Carmeli, T. Heinosaari, and A. Toigo. Quantum Incompatibility Witnesses. *Physical Review Letters*, 122:130402, 2019.

[Con94] J.B. Conway. *A Course in Functional Analysis*. Graduate Texts in Mathematics. Springer New York, 1994.

[CS16] D. Cavalcanti and P. Skrzypczyk. Quantitative relations between measurement incompatibility, quantum steering, and nonlocality. *Physical Review A*, 93:052112, 2016.

[CSTP18] M. Czechlewski, D. Saha, A. Tavakoli, and M. Pawłowski. Device-independent witness of arbitrary-dimensional quantum systems employing binary-outcome measurements. *Physical Review A*, 98:062305, 2018.

[CY95] I. L. Chuang and Y. Yamamoto. Simple quantum computer. *Physical Review A*, 52:3489–3496, 1995.

[CZ95] J. I. Cirac and P. Zoller. Quantum computations with cold trapped ions. *Physical Review Letters*, 74:4091–4094, 1995.

[DEBŻ10] T. Durt, B. G. Englert, I. Bengtsson, and K. Życzkowski. On mutually unbiased bases. *International Journal of Quantum Information*, 8(4):535–640, 2010.

[DLB+11] A. C. Dada, J. Leach, G. S. Buller, M. J. Padgett, and E. Andersson. Experimental high-dimensional two-photon entanglement and violations of generalized Bell inequalities. *Nature*, 7:677–680, 2011.

[DSFB19] S. Designolle, P. Skrzypczyk, F. Fröwis, and N. Brunner. Quantifying Measurement Incompatibility of Mutually Unbiased Bases. *Physical Review Letters*, 122:050402, 2019.

[Eke91] A. K. Ekert. Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67:661–663, 1991.

[EPR35] A. Einstein, B. Podolsky, and N. Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Reviews*, 47:777–780, 1935.

[Far17] M. Farkas. n-fold unbiased bases: an extension of the MUB condition. *arXiv:1706.04446*, 2017.

[Fey82] R. P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, 1982.

[FLP+12] R. Fickler, R. Lapkiewicz, W. N. Plick, M. Krenn, C. Schaeff, S. Ramelow, and A. Zeilinger. Quantum Entanglement of High Angular Momenta. *Science*, 338(6107):640–643, 2012.

[Fri17] T. Fritz. Resource convertibility and ordered commutative monoids. *Mathematical Structures in Computer Science*, 27(6):850–938, 2017.

[GBHA10] R. Gallego, N. Brunner, C. Hadley, and A. Acín. Device-independent tests of classical and quantum dimensions. *Physical Review Letters*, 105(23):230501, 2010.

[GVW+15] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-Å. 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, B. Wittmann, and A. Zeilinger. Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons. *Physical Review Letters*, 115:250401, 2015.

[HBD+15] 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. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526:682–686, 2015.

[HHHH09] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81:865–942, 2009.

[HMZ16] T. Heinosaari, T. Miyadera, and M. Ziman. An invitation to quantum incompatibility. *Journal of Physics A: Mathematical and Theoretical*, 49(12):123001, 2016.

[HP13] M. Huber and M. Pawłowski. Weak randomness in device-independent quantum key distribution and the advantage of using high-dimensional entanglement. *Physical Review A*, 88:032309, 2013.

[HRS08] T. Heinosaari, D. Reitzner, and P. Stano. Notes on Joint Measurability of Quantum Observables. *Foundations of Physics*, 38(12):1133–1147, 2008.

[Iva81] I. D. Ivanovic. Geometrical description of quantal state determination. *Journal of Physics A: Mathematical and General*, 14(12):3241, 1981.

[Kan17] J. Kaniewski. Self-testing of binary observables based on commutation. *Physical Review A*, 95:062323, 2017.

[Kit97] F. Kittaneh. Norm inequalities for certain operator sums. *Journal of Functional Analysis*, 143(2):337–348, 1997.

[Kit02] F. Kittaneh. Norm inequalities for sums of positive operators. *Journal of Operator Theory*, 48(1):95–103, 2002.

[KP02] M. Krishna and K. R. Parthasarathy. An entropic uncertainty principle for quantum measurements. *Sankhya: The Indian Journal of Statistics*, 64(3):842–851, 2002.

[KŠT+19] J. Kaniewski, I. Šupić, J. Tura, F. Baccari, A. Salavrakos, and R. Augusiak. Maximal nonlocality from maximal entanglement and mutually unbiased bases, and self-testing of two-qutrit quantum systems. *Quantum*, 3:198, 2019.

[LBL+15] T. Lunghi, J. B. Brask, C. C. W. Lim, Q. Lavigne, J. Bowles, A. Martin, H. Zbinden, and N. Brunner. Self-testing quantum random number generator. *Physical Review Letters*, 114(15):150501, 2015.

[LPY+12] H.-W. Li, M. Pawłowski, Z.-Q. Yin, G.-C. Guo, and Z.-F. Han. Semi-device-independent randomness certification using $n \to 1$ quantum random access codes. *Physical Review A*, 85(5):052308, 2012.

[Lud54] G. Ludwig. *Die Grundlagen der Quantenmechanik.* Springer Berlin Heidelberg, 1954.

[MMMO06] F. Magniez, D. Mayers, M. Mosca, and H. Ollivier. Self-testing of quantum circuits. In *Automata, languages and programming*, pages 72–83, 2006.

[MMZ16] M. Huber M. Krenn R. Fickler M. Malik, M. Erhard and A. Zeilinger. Multi-photon entanglement in high dimensions. *Nature Photonics*, 10:248–252, 2016.

[MU88] H. Maassen and J. B. M. Uffink. Generalized entropic uncertainty relations. *Physical Review Letters*, 60:1103–1106, 1988.

[MY98] D. Mayers and A. Yao. Quantum cryptography with imperfect apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, FOCS '98, page 503, 1998.

[MY04] D. Mayers and A. Yao. Self testing quantum apparatus. *Quantum Information and Computation*, 4(4):273–286, 2004.

[MYS12] M. McKague, T. H. Yang, and V. Scarani. Robust self-testing of the singlet. *Journal of Physics A: Mathematical and Theoretical*, 45(45):455304, 2012.

[OHDG06] B. Odom, D. Hanneke, B. D'Urso, and G. Gabrielse. New Measurement of the Electron Magnetic Moment Using a One-Electron Quantum Cyclotron. *Physical Review Letters*, 97:030801, 2006.

[Ozo09] Ozols, M. Quantum random access codes with shared randomness. Master's thesis, University of Waterloo, 2009.

[PB11] M. Pawłowski and N. Brunner. Semi-device-independent security of one-way quantum key distribution. *Physical Review A*, 84(1):010302, 2011.

[PR92] S. Popescu and D. Rohrlich. Which states violate Bell's inequality maximally? *Physics Letters A*, 169(6):411–414, 1992.

[QVB14] M. T. Quintino, T. Vértesi, and N. Brunner. Joint measurability, Einstein–Podolsky–Rosen steering, and Bell nonlocality. *Physical Review Letters*, 113:160402, 2014.

[RFN13] D. J. Richardson, J. M. Fini, and L. E. Nelson. Space-division multiplexing in optical fibres. *Nature Photonics*, 7:354–362, 2013.

[ŠB19] I. Šupić and J. Bowles. Self-testing of quantum systems: a review. *arXiv:1904.10042*, 2019.

[Sch26] E. Schrödinger. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review*, 28(6):1049–1070, 1926.

[Sho94] P. W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134, 1994.

[SMSC+15] 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, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam. Strong Loophole-Free Test of Local Realism. *Physical Review Letters*, 115:250402, 2015.

[SSKA19] S. Sarkar, D. Saha, J. Kaniewski, and R. Augusiak. Self-testing quantum systems of arbitrary local dimension with minimal number of measurements. *arXiv:1909.12722*, 2019.

[SW87] S. J. Summers and R. Werner. Bell's inequalities and quantum field theory. I. General setting. *Journal of Mathematical Physics*, 28(10):2440–2447, 1987.

[TFR+19] A. Tavakoli, M. Farkas, D. Rosset, J.-D. Bancal, and J. Kaniewski. Mutually unbiased bases and symmetric informationally complete measurements in Bell experiments: Bell inequalities, device-independent certification and applications. *arXiv:1912.03225*, 2019.

[THMB15] A. Tavakoli, A. Hameedi, B. Marques, and M. Bourennane. Quantum Random Access Codes Using Single d-Level Systems. *Physical Review Letters*, 114:170502, 2015.

[TKV+18] A. Tavakoli, J. Kaniewski, T. Vértesi, D. Rosset, and N. Brunner. Self-testing quantum states and measurements in the prepare-and-measure scenario. *Physical Review A*, 98:062307, 2018.

[Tsi87] B. S. Tsirelson. Quantum analogues of the Bell inequalities. The case of two spatially separated domains. *Journal of Soviet Mathematics*, 36:557, 1987.

[TWE+17] T. R. Tan, Y. Wan, S. Erickson, P. Bierhorst, D. Kienzler, S. Glancy, E. Knill, D. Leibfried, and D. J. Wineland. Chained Bell inequality experiment with high-efficiency measurements. *Physical Review Letters*, 118(13):130403, 2017.

[UBGP15] R. Uola, C. Budroni, O. Gühne, and J.-P. Pellonpää. One-to-One Mapping between Steering and Joint Measurability Problems. *Physical Review Letters*, 115:230402, 2015.

[VHWC+17] T. Van Himbeeck, E. Woodhead, N. J. Cerf, R. García-Patrón, and S. Pironio. Semi-device-independent framework based on natural physical assumptions. *Quantum*, 1:33, 2017.

[WF89] W. K. Wootters and B. D. Fields. Optimal state-determination by mutually unbiased measurements. *Annals of Physics*, 191(2):363–381, 1989.

[WPGF09] M. M. Wolf, D. Perez-Garcia, and C. Fernandez. Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory. *Physical Review Letters*, 103:230402, 2009.

[YVB+14] T. H. Yang, T. Vértesi, J.-D. Bancal, V. Scarani, and M. Navascués. Robust and versatile black-box certification of quantum devices. *Physical Review Letters*, 113(4):040401, 2014.

[Zau91] Zauner, G. Orthogonale Lateinische Quadrate und Anordnungen, Verallgemeinerte Hadamard-Matrizen und Unabhängigkeit in der Quanten-Wahrscheinlichkeitstheorie. Master's thesis, University of Vienna, 1991.

Part II

Collection of Papers

Appended to this summary is the collection of papers used in the PhD dissertation.

[A] **Certifying an irreducible 1024-dimensional photonic state using refined dimension witnesses**
E. A. Aguilar*, **M. Farkas***, D. Martínez, M. Alvarado, J. Cariñe, G. B. Xavier, J. F. Barra, G. Cañas, M. Pawłowski, G. Lima
Physical Review Letters **120**, 230503 (2018)
*contributed equally

[B] **Self-testing mutually unbiased bases in the prepare-and-measure scenario**
M. Farkas, J. Kaniewski
Physical Review A **99**, 032316 (2019)

[C] **Incompatibility robustness of quantum measurements: a unified framework**
S. Designolle*, **M. Farkas***, J. Kaniewski
New Journal of Physics **21**(11), 113053 (2019)
*contributed equally

Certifying an Irreducible 1024-Dimensional Photonic State Using Refined Dimension Witnesses

Edgar A. Aguilar,[1] Máté Farkas,[1] Daniel Martínez,[2,3] Matías Alvarado,[2,3] Jaime Cariñe,[2,3] Guilherme B. Xavier,[3,4,5] Johanna F. Barra,[2,3] Gustavo Cañas,[6] Marcin Pawłowski,[1] and Gustavo Lima[2,3]

[1]*Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics and Informatics, University of Gdansk, 80-952 Gdansk, Poland*

[2]*Departamento de Física, Universidad de Concepción, 160-C Concepción, Chile*

[3]*Millennium Institute for Research in Optics, Universidad de Concepción, 160-C Concepción, Chile*

[4]*Departamento de Ingeniería Eléctrica, Universidad de Concepción, 160-C Concepción, Chile*

[5]*Institutionen för Systemteknik, Linköpings Universitet, 581 83 Linköping, Sweden*

[6]*Departamento de Física, Universidad del Bio-Bio, Avenida Collao 1202, Concepción, Chile*

(Received 22 September 2017; revised manuscript received 23 March 2018; published 6 June 2018)

We report on a new class of dimension witnesses, based on quantum random access codes, which are a function of the recorded statistics and that have different bounds for all possible decompositions of a high-dimensional physical system. Thus, it certifies the dimension of the system and has the new distinct feature of identifying whether the high-dimensional system is decomposable in terms of lower dimensional subsystems. To demonstrate the practicability of this technique, we used it to experimentally certify the generation of an irreducible 1024-dimensional photonic quantum state. Therefore, certifying that the state is not multipartite or encoded using noncoupled different degrees of freedom of a single photon. Our protocol should find applications in a broad class of modern quantum information experiments addressing the generation of high-dimensional quantum systems, where quantum tomography may become intractable.

DOI: 10.1103/PhysRevLett.120.230503

Introduction.—The dimension d of physical systems is a fundamental property of any model, and its operational definition arguably reflects the evolution of physics itself. In quantum mechanics, it can be seen as a key resource for information processing since higher dimensional systems provide advantages in several protocols of quantum computation [1] and quantum communications [2]. In the field of quantum foundations, a recent proposal suggests that, in order to understand and create macroscopic quantum states, it will be necessary to take advantage of high-dimensional systems [3]. Therefore, it is natural to understand why there is a growing endeavor to coherently control quantum systems of large dimensions [4–16]. Nonetheless, such new technological advances require the simultaneous development of practical methods to certify that the sources are truly producing the required quantum states. In principle, one can rely on the process of quantum tomography [17–23], but this approach quickly becomes intractable in higher dimensions as at least d^2 measurements are required [24].

To address this problem, the concept of dimension witness (DW) was introduced. The original idea was based on the violation of a particular Bell inequality [25] but was then extended to the more practical prepare-and-measure scenario [26]. In general, DWs are defined as linear functions of a few measurement outcome probabilities and have classical and quantum bounds defined for each considered dimension [4,25–30]. Thus, they allow for the device-independent certification of the minimum dimension required to describe a given physical system and can also infer whether it is properly described by a coherent superposition of logical states. Nevertheless, these tests do not provide information about the composition of the system, which is crucial for high-dimensional quantum information processing. This point has been recently investigated by W. Cong *et al.* [31], where they introduced the concept of an irreducible dimension witness (IDW) to certify the presence of an irreducible four dimensional system. Specifically, their IDW distinguishes whether if the observed data are created by one pair of entangled ququarts, or two pairs of entangled qubits measured under sequential adaptive operations and classical communication.

Here, we introduce a new class of DWs, namely gamut DWs, which certifies the dimension of the system and has the new distinct feature of identifying whether any high-dimensional quantum system is irreducible. It is based on quantum random access codes (QRACs), which is a communication task defined in a prepare-and-measure scenario [32]. To demonstrate the practicability of our new technique, we experimentally certify the generation of an irreducible 1024-dimensional photonic quantum system encoded onto the transverse momentum of single photons transmitted over programmable diffractive optical devices [5,21–23,33–35]. To our knowledge, our work represents an increase of about 2 orders of magnitude to any reported experiment using path qudits. From the recorded data, one observes a violation of the bounds associated with all possible decompositions of a 1024-dimensional quantum system, thus, certifying that the generated state is not encoded using noncoupled different degrees of freedom of a photon, e.g., polarization and momentum. Nonetheless, our method is broadly relevant and should also find applications in multipartite photonic scenarios and new platforms for the fast-growing field of experimental high-dimensional quantum information processing.

Gamut dimension witness.—As stated earlier, the protocol we use in our main theorem is based on QRACs. Thus, first, we give a brief description (see, e.g., [32] for more details) of this task (see Fig. 1): one of the parties, Alice, receives two input dits: x_1 and $x_2 \in \{1, \ldots, d\}$. She is then allowed to send one d-dimensional (quantum) state, $\rho_{x_1 x_2}$ to Bob, depending on her input. Bob is then given a bit $y \in \{1, 2\}$ and his task is to guess x_y. He does so by performing a quantum measurement M^y and a classical post-processing function $\mathcal{D}^y$. As a result, he outputs $b \in \{1, \ldots, d\}$.

For a single round of the protocol, the success probability is $\mathbb{P}(b = x_y | x_1, x_2, y)$. As a figure or merit over many rounds with uniformly random inputs, we employ the average success probability (ASP): $\bar{p} = (1/2d^2)\sum_{x_1,x_2,y}\mathbb{P}(b = x_y|x_1, x_2, y)$. Thus, we are looking for the maximal value of $\bar{p}$, optimizing over all possible encoding and decoding strategies. It was proven [36] that, for classical strategies (i.e., classical states and decoding functions), the optimal ASP is $\bar{p}_{C_d} = \frac{1}{2}(1 + 1/d)$. In the quantum case, the optimal strategy is reached by using mutually unbiased bases (MUBs) for encoding and decoding [37,38], and the ASP is $\bar{p}_{Q_d} = \frac{1}{2}(1 + 1/\sqrt{d})$.

Now, we estimate the optimal ASPs for composite systems, for all possible product structures, defined as follows.

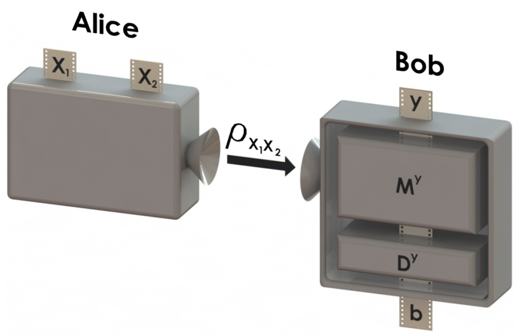


FIG. 1. Our d-dimensional QRACs scenario. Alice receives the input dits x_1 and $x_2 \in \{1, \ldots, d\}$, and prepares the state $\rho_{x_1 x_2}$ which is sent to Bob. He receives the input $y \in \{1, 2\}$, which defines the quantum measurement M^y and the classical post-processing function $\mathcal{D}^y$ to be applied to $\rho_{x_1 x_2}$. As a result, Bob outputs b.

Definition 1.—For a fixed d, we define a product structure by the set $\{r, \{d_k\}, \{\alpha_k\}\}$. For a composite system, $d = \prod_{k=1}^r d_k$, where d_k is the dimension of each subsystem and r is the number of subsystems. The state of the composite system can be written as $\rho = \rho^1_{\alpha_1} \otimes \rho^2_{\alpha_2} \otimes \cdots \otimes \rho^r_{\alpha_r}$. Here, $\alpha_k = \mathrm{c}$ and $\alpha_k = \mathrm{q}$ are used to denote the "classical" and "quantum" nature of the subsystem, respectively. Then, $\rho^k_c \in \Delta_{d_k - 1}$ is a classical state, and $\rho^k_q \in \mathcal{S}(\mathbb{C}^{d_k})$ is a quantum state.

Consider a set of measurement and state preparation settings and fix the total dimension of the physical system in question. We call a linear function on the measurement outcome probabilities a gamut dimension witness (GDW) if its extremal values for all possible product structures are different. For example, in $d = 4$, a GDW has different extremal values for a ququart, two qubits, one qubit and a bit, and a quart. The main theoretical result of this work is to demonstrate that d-dimensional QRACs can be used as GDWs for d-dimensional physical systems. To highlight this, we set it as a theorem.

Theorem 1.—d-dimensional QRACs serve as gamut dimension witnesses using the ASP function.

The proof of this theorem and all related lemmas can be found in the Supplemental Material [39]. Let us now sketch the main tools for proving the theorem. They help to understanding the problem, and can be independently used. Note that the following lemmas apply in more general QRAC scenarios as well [39].

We assume that Bob's measurements have the same product structure as the state generated by Alice. That is, we exclude that Bob's state certification would use entangling measurements. The motivation here is to rule out sequential uses of lower dimensional systems as a way to simulate higher dimensional statistics, e.g., to discriminate between n sequential uses of a d-dimensional system, and a d^n-dimensional system. A physical motivation for this assumption is to think that, if Alice cannot couple a particular set of degrees of freedom (e.g., polarization and momentum), then neither can Bob because he has access to the same equipment as Alice does [43].

Therefore, the most general strategy for decoding the d-dimensional system $\rho = \rho^1 \otimes \rho^2 \otimes \cdots \otimes \rho^r$ is as follows: Bob performs sequential adaptive measures on the subsystems in the sense of [31]. He starts by measuring subsystem ρ^1 to obtain the outcome b^1. Then, his choice of the measurement to be performed in ρ^2 may depend on b^1. Successively, each measurement on ρ^k can depend on all the measurement outcomes obtained previously. After performing all measurements, Bob feeds the obtained outcomes to a classical post-processing function and outputs his final guess on x_y, which is $b = \mathcal{D}^y(b^1 b^2, \ldots, b^r)$.

The bounds of the GDW in this general scenario are extremely hard to obtain. The following results help, making the analysis easier. First, it is argued in [32] that, in an optimal strategy, it is enough to use encoded pure states. Similarly, it has been shown that rank 1 projective measurements (explicitly: mutually unbiased bases) optimize two-input QRACs [38]. Thus, in the following, we only deal with pure states for both Alice and Bob. Additionally, we can eliminate classical post-processing functions.

Lemma 1.—In QRACs, for optimality of the ASP, there is no need for classical post-processing functions.

Last, we note that:

Lemma 2.—In QRACs, for optimality of the ASP, there is no need for sequential adaptive measurements.

Observe that the above lemmas together imply that the highest ASP for a composite system can be achieved with a strategy that consists of r QRACs in parallel, one on each subsystem ρ^k, independently. In this case, if we write Alice's inputs as dit strings $x_y = x_y^1 x_y^2, \ldots, x_y^r$, the success probability for each round is $\mathbb{P}(b = x_y | x_1, x_2, y) = \prod_{k=1}^{r} \mathbb{P}(b^k = x_y^k | x_1^k, x_2^k, y)$. The optimal $\bar{p}$ is not necessarily given by the independent optimal strategies on the individual subspaces. Therefore, in order to optimize it we introduce the trade-off function $\mathcal{M}_d(z)$ (see the Supplemental Material [39]), which provides the optimal probability of guessing dit x_2 given a fixed probability of guessing dit x_1. Let $z = \mathbb{P}(\text{Bob correctly guesses} x_1)$. Then, $\mathcal{M}_d(z)$ in dimension d is defined by $\mathcal{M}_d(z) = \max\{\mathbb{P}(\text{Bob correctly guesses} x_2)|z\}$, where the maximization is limited to all encoding-decoding strategies respecting the condition of guessing x_1 with probability z. Thus, in a general case,

$$\bar{p}_{Q_{d_1},\ldots,C_{d_r}} = \max_{z^1,\ldots,z^r} \frac{1}{2}[z^1 \cdots z^r + \mathcal{M}^q_{d_1}(z^1) \cdots \mathcal{M}^c_{d_r}(z^r)], \quad (1)$$

where we denote d-dimensional quantum and classical states by Q_d and C_d, respectively. $\mathcal{M}^q_d$ and $\mathcal{M}^c_d$ are the corresponding quantum, and classical trade-off functions [39]. Therefore, $\bar{p}$ is a function of r real variables, and its maximum can be found using standard heuristic numerical search algorithms [44]. We present the ASP optimal values for some relevant cases of a $d = 1024$ dimensional system in Table I. The full list of cases is found in the Supplemental Material [39]. Note that the gaps between the different ASP values are large enough to be experimentally observed, as we demonstrate next.

Experiment.—To demonstrate the practicability of our technique, we generate a 1024-dimensional photonic state, encoded into the linear transverse momentum of single-photons, and use the 1024-dimensional QRAC GDW to certify that it is an irreducible quantum system. To achieve this, first, we show that the ASP [Eq. (1)] can be written as a simple function of the detection events. Then, we observe that our recorded statistics violate the second highest ASP bound, $Q_{512}Q_2$, given in Table I, thus, ensuring that it is an irreducible 1024-dimensional quantum system.

In the 1024-dimensional QRAC GDW, Bob measures the elements of the two 1024-dimensional MUBs given in the Supplemental Material [39]. We denote the MUB states by $|m_j^y\rangle$, where $y = 1, 2$ defines the measuring base MUB_1 or base MUB_2, and $j = 1, \ldots, 1024$ denotes the state of a given base. Alice's state is written in terms of the two input dits x_1 and x_2 as an equal superposition of the states Bob would need to guess x_y correctly

$$|\Psi_{x_1 x_2}\rangle = \frac{1}{N}(|m^1_{x_1}\rangle + \text{sgn}(\langle m^1_{x_1}|m^2_{x_2}\rangle)|m^2_{x_2}\rangle), \quad (2)$$

where $N = \sqrt{2(1 + \frac{1}{32})}$ is a normalization factor and sign is the sign function. The optimality of the encoded states (2), and the use of MUBs is derived in the Supplemental Material [39].

For the experimental test, we resort to the setup depicted in Fig. 2. At the state preparation block, the single-photon regime is achieved by heavily attenuating optical pulses with well calibrated attenuators. An acousto-optical modulator (AOM) placed at the output of a continuous-wave laser operating at 690 nm is used to generate the optical pulses. The average number of photons per pulse is set to $\mu = 0.4$. In this case, the probability of having non-null pulses is $P(n \geq 1 | \mu = 0.4) = 33\%$. Pulses containing only one photon are the majority of the non-null pulses generated and accounts to 82% of the experimental runs. Thus, our source is a good approximation to a nondeterministic single-photon source, which is commonly adopted in quantum communications [2].

The single-photons are then sent through two spatial light modulators, SLM1 and SLM2, addressing an array of 32×32 transmissive squares. The square side is $a = 96\ \mu\text{m}$ and they are equally separated by $\delta = 160\ \mu\text{m}$ [see Fig. 2(b)], thus, effectively creating a 1024-dimensional quantum state defined in terms of the number of modes available for the photon transmission over the SLMs [5,21–23,33,34]. Specifically, the state of the transmitted photon is given by $|\Psi\rangle = (1/\sqrt{C}) \sum_{l=-l_{N_c}}^{l_{N_c}} \sum_{v=-l_{N_r}}^{l_{N_r}} \sqrt{t_{lv}}\, e^{-i\phi_{lv}} |c_{lv}\rangle$, where $|c_{lv}\rangle$ is the logical state representing the photon transmitted by the $(l,\ v)$ square. t_{lv} represents the

TABLE I. Relevant cases for a 1024-dimensional system and the respective optimal ASPs [Eq. (1)] considering each product structure. The full table can be found in the Supplemental Material [39].

Case	Optimal $\bar{p}$
Q_{1024}	0.515 625
$Q_{512}Q_2$	0.500 980
$Q_{512}C_2$	0.500 973
$Q_{32}Q_{32}$	0.500 521
$(Q_2)^{10}$	0.500 493
Q_2C_{512}	0.500 489
C_{1024}	0.500 488

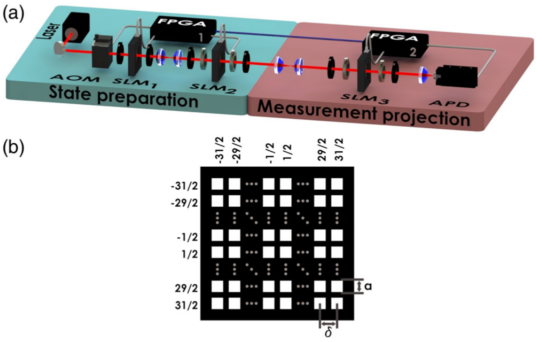


FIG. 2. (a) Experimental setup. We employ a prepare-and-measure scheme to generate and project spatial qudits, encoded into the linear transverse momentum of single-photons. At the state preparation block, the spatial encoding is applied through two spatial light modulators (SLMs), and the state projection is likewise performed by a SLM combined with a pointlike avalanche single-photon detector (APD) at the measurement projection block (see main text for details). (b) The 32×32-square mask addressed by the SLMs.

transmission and ϕ_{lv} the phase-shift given by the (l, v) square. The transmission of each square is controlled by the SLM1, which is configured for amplitude-only modulation. The phases ϕ_{lv} are controlled by SLM2 working on the configuration of phase-only modulation [22]. N_c and N_r represent the number of columns and rows, respectively. For simplicity, we define $l_{N_c} \equiv (N_c - 1)/2$, $l_{N_r} \equiv (N_r - 1)/2$, and C is the normalization factor.

At the measurement block, we use a similar scheme to the one used in the state preparation block. It consists of a SLM3, also configured for phase-modulation, and a "pointlike" avalanche single-photon detector (APD). As explained in detail in [5,22], by placing the pointlike APD at the SLM3 far-field (FF) plane, and properly adjusting the (l, v) square phase shifts, Bob can detect any state $|m_j^y\rangle$ required for the 1024-dimensional QRAC session. The pointlike APD is composed of a pinhole (aperture of 10 μm diameter) fixed at the center of the FF plane, followed by the APD module. In this case, the probability of photon detection is proportional to the overlap between the prepared and detected states. For the case of a d-dimensional QRACs implemented with a single-detector scheme, we show in the Supplemental Material (see [39] and Refs. [4,5,9,13] therein) that the ASP function can be written as

$$\bar{p} = \frac{D_1}{D_1 + D_2}. \tag{3}$$

First, we consider the events with $x_y = j$ (again, $j = 1, \ldots, 1024$ denotes the state of a given base) and define the total number of such events to be X_1. Then, we define D_1 as the number of "clicks" recorded in the experiment in those cases. Likewise, we denote X_2 to be the number of events where $x_y \neq j$ and define D_2 to be the clicks in those cases.

By means of two field-programmable gate array (FPGA) electronic modules, we are able to automate and actively control both blocks of the setup. At the state preparation block, since the state $|\Psi\rangle$ needs to be randomly selected from the set of states defined by the 1024-dimensional QRACs, a random number generator (QRNG-Quantis) is connected to FPGA1. FPGA1 controls the optical pulse production rate by the AOM, set at 60 Hz as limited by the refresh rate of the SLMs. Each attenuated optical pulse corresponds to an experimental round. At the measurement block, a second QRNG is connected to FPGA2, providing an independent and random selection for the projection $|m_j^y\rangle$ at each round. FPGA2 also records whether a detection event occurs. The overall detection efficiency is 13%. The protocol is executed as follows: In each round, FPGA1 reads the dits x_1 and x_2 produced by its QRNG. Then, FPGA1 calculates the amplitude and phase of each (l, v) square of SLM1 and SLM2 to encode the state $|\Psi_{x_1 x_2}\rangle$ onto the spatial profile of the single-photon in that experimental round. Simultaneously, FPGA2, reads from its QRNG the value of y and j. Similar to what is done in the state preparation block, FPGA2 also calculates the phase for each (l, v) square in SLM3 to implement the chosen projection $|m_j^y\rangle$. The amplitude and relative phase for each SLM was previously characterized in order to obtain the modulation curves as a function of its grey level. In this experiment, this is necessary to dynamically generate all possible states, as it would be unfeasible to prerecord predefined masks for the SLMs on the FPGAs for each one of the 1024^2 required initial states.

The experiment continuously ran over 316 hours. In this way, the statistics fluctuations observed for D_1 and D_2 were sufficiently small to unambiguously certify the generation of an irreducible 1024-dimensional quantum system. The

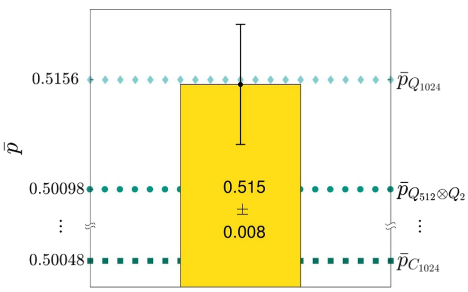


FIG. 3. Experimental results. We experimentally observe $\bar{p} = 0.515 \pm 0.008$, violating the second highest ASP bound $\bar{p}_{Q_{512}\otimes Q_2}$ (see Table I). The error bar is calculated assuming Poissonian statistics for a photon detection event.

overall visibility in our system is $97.00 \pm 0.07\%$ and the corresponding recorded average success probability is $\bar{p} = 0.515 \pm 0.008$. In Fig. 3, we compare it with the second highest ASP bound shown in Table I, associated with a composite system of the type $Q_{512}Q_2$. This certifies, only from the statistics recorded, that the generated state is not encoded using noncoupled different degrees of freedom of a photon, for instance polarization and momentum, thus, ensuring it to be an irreducible 1024-dimensional quantum system that can provide all the advantages known for high-dimensional quantum information processing, in the sense explained in [31].

Conclusion.—Dimension witnesses are practical protocols on the field of quantum information as they allow one to obtain information regarding unknown quantum states [25,26]. They are especially appealing while addressing the generation and characterization of high-dimensional quantum states, where quantum tomography demands at least d^2 measurements [24]. In general, DWs are functions of only a few measurement outcome probabilities and allow for assessments on the dimension required to describe a given quantum state in a device-independent way [4,25–30]. Here, we give a step further by introducing a new class of DW, which certifies the dimension of the system, and has the new distinct feature of allowing the identification of whether a high-dimensional system is irreducible. The application of this new feature is of broad relevance for several new architectures aiming for high-dimensional quantum information processing [4–16], and the understanding of macroscopic quantumness [3]. We demonstrate the practicability of our technique by using it to certify the generation of an irreducible 1024-dimensional photonic quantum state encoded into the linear transverse momentum of single-photons transmitted by programable diffractive apertures which have been used for several high-dimensional quantum information processing tasks [5,35,45–47].

This work was supported by First TEAM Grant No. 2016-1/5, NCN Sonata Grant No. UMO-2014/14/E/ST2/00020, Fondecyt Grants No. 1160400, No. 11150324, No. 1150101, and the Millennium Institute for Research in Optics, MIRO. E. A. A. acknowledges support from CONACyT. J. F. B. acknowledges support from Fondecyt 3170307. J. C. acknowledges support from Fondecyt 3170596. D. M. and M. A. acknowledge support from CONICYT.

E. A. A. and M. F. contributed equally to this work.

[1] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien, Nature (London) **464**, 45 (2010).
[2] Z.-S. Yuan, X.-H. Bao, C.-Y. Lu, J. Zhang, C.-Z. Peng, and J.-W. Pan, Phys. Rep. **497**, 1 (2010).
[3] N. Gisin and F. Fröwis, arXiv:1802.00736.
[4] V. D'Ambrosio, F. Bisesto, F. Sciarrino, J. F. Barra, G. Lima, and A. Cabello, Phys. Rev. Lett. **112**, 140503 (2014).
[5] S. Etcheverry, G. Cañas, E. S. Gómez, W. A. T. Nogueira, C. Saavedra, G. B. Xavier, and G. Lima, Sci. Rep. **3**, 2316 (2013).
[6] W.-B. Gao, C.-Y. Lu, X.-C. Yao, P. Xu, O. Guhne, A. Goebel, Y.-A. Chen, C.-Z. Peng, Z.-B. Chen, and J.-W. Pan, Nat. Phys. **6**, 331 (2010).
[7] A. C. Dada, J. Leach, G. S. Buller, M. J. Padgett, and E. Andersson, Nat. Phys. **7**, 677 (2011).
[8] J. T. Barreiro, N. K. Langford, N. A. Peters, and P. G. Kwiat, Phys. Rev. Lett. **95**, 260501 (2005).
[9] R. Fickler, R. Lapkiewicz, W. N. Plick, M. Krenn, C. Schaeff, S. Ramelow, and A. Zeilinger, Science **338**, 640 (2012).
[10] M. Krenn, M. Huber, R. Fickler, R. Lapkiewicz, S. Ramelow, and A. Zeilinger, Proc. Natl. Acad. Sci. U.S.A. **111**, 6243 (2014).
[11] X.-C. Yao, T.-X. Wang, P. Xu, H. Lu, G.-S. Pan, X.-H. Bao, C.-Z. Peng, C.-Y. Lu, Y.-A. Chen, and J.-W. Pan, Nat. Photonics **6**, 225 (2012).
[12] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt, Phys. Rev. Lett. **106**, 130506 (2011).
[13] X.-L. Wang, L.-K. Chen, W. Li, H.-L. Huang, C. Liu, C. Chen, Y.-H. Luo, Z.-E. Su, D. Wu, Z.-D. Li, H. Lu, Y. Hu, X. Jiang, C.-Z. Peng, L. Li, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, and J.-W. Pan, Phys. Rev. Lett. **117**, 210502 (2016).
[14] R. Barends *et al.*, Nature (London) **508**, 500 (2014).
[15] M. Malik, M. Erhard, M. Huber, M. Krenn, R. Fickler, and A. Zeilinger, Nat. Photonics **10**, 248 (2016).
[16] R. Fickler, G. Campbell, B. Buchler, P. K. Lam, and A. Zeilinger, Proc. Natl. Acad. Sci. U.S.A. **113**, 13642 (2016).
[17] G. M. D'Ariano and P. Lo Presti, Phys. Rev. Lett. **86**, 4195 (2001).
[18] G. M. D'Ariano, L. Maccone, and P. Lo Presti, Phys. Rev. Lett. **93**, 250407 (2004).
[19] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A **64**, 052312 (2001).
[20] R. T. Thew, K. Nemoto, A. G. White, and W. J. Munro, Phys. Rev. A **66**, 012303 (2002).
[21] G. Lima, E. S. Gómez, A. Vargas, R. O. Vianna, and C. Saavedra, Phys. Rev. A **82**, 012302 (2010).
[22] G. Lima, L. Neves, R. Guzmán, E. S. Gómez, W. A. T. Nogueira, A. Delgado, A. Vargas, and C. Saavedra, Opt. Express **19**, 3542 (2011).
[23] D. Goyeneche, G. Cañas, S. Etcheverry, E. S. Gómez, G. B. Xavier, G. Lima, and A. Delgado, Phys. Rev. Lett. **115**, 090401 (2015).
[24] W. K. Wootters and B. D. Fields, Ann. Phys. (N.Y.) **191**, 363 (1989).
[25] N. Brunner, S. Pironio, A. Acin, N. Gisin, A. A. Méthot, and V. Scarani, Phys. Rev. Lett. **100**, 210503 (2008).
[26] R. Gallego, N. Brunner, C. Hadley, and A. Acín, Phys. Rev. Lett. **105**, 230501 (2010).
[27] J. Ahrens, P. Badziag, A. Cabello, and M. Bourennane, Nat. Phys. **8**, 592 (2012).
[28] M. Hendrych, R. Gallego, M. Micuda, N. Brunner, A. Acin, and J. P. Torres, Nat. Phys. **8**, 588 (2012).
[29] N. Brunner, M. Navascués, and T. Vértesi, Phys. Rev. Lett. **110**, 150501 (2013).

[30] J. Bowles, M. T. Quintino, and N. Brunner, Phys. Rev. Lett. **112**, 140407 (2014).

[31] W. Cong, Y. Cai, J.-D. Bancal, and V. Scarani, Phys. Rev. Lett. **119**, 080401 (2017).

[32] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, in *Proceedings of the Thirty-first Annual ACM Symposium on Theory of Computing, STOC '99* (ACM, New York, 1999) pp. 376–383.

[33] L. Neves, G. Lima, J. G. Aguirre Gómez, C. H. Monken, C. Saavedra, and S. Pádua, Phys. Rev. Lett. **94**, 100501 (2005).

[34] G. Lima, A. Vargas, L. Neves, R. Guzmán, and C. Saavedra, Opt. Express **17**, 10688 (2009).

[35] G. Cañas, N. Vera, J. Cariñe, P. González, J. Cardenas, P. W. R. Connolly, A. Przysiezna, E. S. Gómez, M. Figueroa, G. Vallone, P. Villoresi, T. F. da Silva, G. B. Xavier, and G. Lima, Phys. Rev. A **96**, 022317 (2017).

[36] M. Czechlewski, D. Saha, and M. Pawłowski, arXiv:1803.05245.

[37] E. Aguilar, J. Borkała, P. Mironowicz, and M. Pawłowski, arXiv:1709.04898.

[38] M. Farkas, arXiv:1803.00363.

[39] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.120.230503, which includes Refs. [40–42] for previous results on QRACs and majorization.

[40] A. Ambainis, D. Leung, L. Mancinska, and M. Ozols, Quantum Random Access Codes with Shared Randomness, Master's thesis, University of Waterloo (2009).

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

[42] A. Marshall, I. Olkin, and B. Arnold, *Inequalities: Theory of Majorization and Its Applications*, Springer Series in Statistics (Springer, New York, 2010).

[43] We also note that numerical evidence on small dimensional examples suggests that, if the quantum states are in a product form, then entangling measurements do not improve the ASP.

[44] W. H. Press, S. A. Teukolski, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes* (Cambridge University Press, Cambridge, 1992).

[45] B. Marques, A. A. Matoso, W. M. Pimenta, A. J. Gutiérrez-Esparza, M. F. Santos, and S. Pádua, Sci. Rep. **5**, 16049 (2015).

[46] G. Cañas, S. Etcheverry, E. S. Gómez, C. Saavedra, G. B. Xavier, G. Lima, and A. Cabello, Phys. Rev. A **90**, 012119 (2014).

[47] M. A. Solís-Prosser, M. F. Fernandes, O. Jiménez, A. Delgado, and L. Neves, Phys. Rev. Lett. **118**, 100501 (2017).

Supplemental Material: Certifying an irreducible 1024-dimensional photonic state using refined dimension witnesses

Edgar A. Aguilar,[1,*] Máté Farkas,[1,*] Daniel Martínez,[2,3,4] Matías Alvarado,[2,3,4] Jaime Cariñe,[2,3,4] Guilherme B. Xavier,[3,4,5,6] Johanna F. Barra,[2,3,4] Gustavo Cañas,[2,3,4] Marcin Pawłowski,[1] and Gustavo Lima[2,3,4]

[1]*Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics and Informatics, University of Gdansk, 80-952 Gdansk, Poland*
[2]*Departamento de Física, Universidad de Concepción, 160-C Concepción, Chile*
[3]*Center for Optics and Photonics, Universidad de Concepción, Concepción, Chile*
[4]*MSI-Nucleus for Advanced Optics, Universidad de Concepción, Concepción, Chile*
[5]*Departamento de Ingeniería Eléctrica, Universidad de Concepción, 160-C Concepción, Chile*
[6]*Institutionen för Systemteknik, Linköpings Universitet, 581 83 Linköping, Sweden*

(Dated: May 4, 2018)

The supplemental material is organized into two sections: Theory (S1), and Experimental Considerations (S2).

The theoretical section makes all the formal definitions and provides the proofs of Theorem 1, Lemma 1, and Lemma 2 of the main text. We further clarify Equation (1) of the main text, as well as showing the explicit form of the trade-off functions. The theoretical section ends with two examples. In particular we calculate a table of all of the possible quantum partitions for $d = 1024$ as direct proof that indeed: $Q_{1024} > Q_{512}Q_2 >$ "all other partitions". (Table S2)

The experimental section explicitly show the representation of the MUBs that were used in the experiment. We also formalize the single-detector scheme, and explain how the figure of merit (Equation (3) of the main text) is derived. Finally, we show how this figure of merit depends on the overall detection efficiency ν and average photon number per pulse μ.

S1. THEORY

S1.A. Formal Definitions and Problem Formulation

We begin by defining $n^d \to 1$ *Random Access Codes* (RACs) rigorously. RACs is a strategy in which Alice tries to compress an n-dit string into 1 dit, such that Bob can recover any of the n dits with high probability [1]. Specifically, Alice receives an input string $x = x_1x_2\dots x_n$ drawn from a uniform distribution, where $x_i \in [d]$, with $[d] = \{1, 2, \dots, d\}$. Note that in the special case of the main manuscript, we always use $x = x_1x_2$. She then uses an encoding function $\mathcal{E} : [d]^n \to [d]$, and is allowed to send one dit $a_x = \mathcal{E}(x)$ to Bob. On the other side, Bob receives an input $y \in [n]$ (also uniformly distributed), and together with Alice's message a_x uses one of n decoding functions $\mathcal{D}^y : [d] \to [d]$, to output $b = \mathcal{D}^y(a_x)$ as a guess for x_y. If Bob's guess is correct (i.e. $b = x_y$) then we say that they *win*, otherwise we say that they *lose*. We can then quantify their probability of success $\mathrm{P}(\mathcal{D}^y(\mathcal{E}(x)) = x_y)$, which in general depends on their inputs and on the chosen *strategy* $(\mathcal{E}, \mathcal{D})$, where $\mathcal{D} = \{\mathcal{D}^y\}_{y=1}^n$.

Similarly, one defines the d-dimensional $n^d \to 1$ Quantum Random Access Codes (QRACs) with the only change being that Alice tries to compress her input string into a d-dimensional quantum system (see Fig.S1). Alice encodes her n-dit string via $\mathcal{E} : [d]^n \to \mathcal{S}(\mathrm{C}^d)$, and sends the d-dimensional system $\rho_x = \mathcal{E}(x)$ to Bob. He then performs some decoding to output his guess $b \in [d]$ for x_y. The decoding function is a quantum measurement followed by classical post-processing, as we clarify next.

Definition S1.1
*A **quantum decoding strategy** is $\mathcal{D} = \left\{\{M_l^y\}_l, \mathcal{D}^y\right\}_{y=1}^n$, i.e. n pairs of measurement operators $\{M_l^y\}_l$ (normalized $\sum_l M_l^y = 1\ \forall y$, and positive semi-definite $M_l^y \geq 0\ \forall l, y$), and classical post-processing functions $\mathcal{D}^y : [d] \to [d]$, such that if Bob receives as input ρ_x and y, he outputs $b = \mathcal{D}^y(l)$ with probability tr$[\rho_x M_l^y]$.*

To quantify the performance of a given encoding-decoding strategy, we shall employ the *average success probability* (ASP) $\bar{p}$ as our figure of merit.

Definition S1.2
*The **Average Success Probability** of a given encoding-decoding strategy $(\mathcal{E}, \mathcal{D})$ is:*

$$\bar{p} = \frac{1}{nd^n} \sum_{x,y} \mathrm{P}(B = x_y | X = x, Y = y), \tag{S1}$$

* These authors contributed equally to this work.

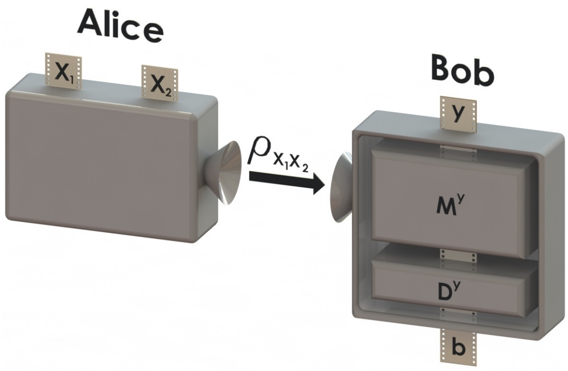


FIG. S1: d-dimensional $2^d \to 1$ QRACs scenario, which is the one considered in the main manuscript. Alice receives the input dits x_1 and $x_2 \in \{1, \dots, d\}$, and prepares the state $\rho_{x_1 x_2}$ which is sent to Bob. He receives the input $y \in \{1, 2\}$, which defines the quantum measurement M^y and the classical post-processing function $\mathcal{D}^y$ to be applied to $\rho_{x_1 x_2}$. As a result, Bob outputs b.

where uppercase letters X, Y, B denote random variables, while the corresponding lowercase letters represent the events (i.e. the values the random variables can take). Another useful way of understanding the ASP is by viewing the whole QRAC protocol as a game and thinking of the ASP as the probability that Alice and Bob win any given round. Loosely speaking:

$$\bar{p} = \mathrm{P}(B = \text{correct}). \tag{S2}$$

Nonetheless, the real object of interest is the *optimal average success probability*, which corresponds to the maximal value of $\bar{p}$ taken over all possible encoding-decoding strategies. Explicitly:

$$\bar{p}_{(C,Q)_d} = \max_{\{\mathcal{E},\mathcal{D}\}} \frac{1}{nd^n} \sum_{x,y} \mathrm{P}(B = x_y | X = x, Y = y), \tag{S3}$$

with C and Q respectively representing the classical and quantum scenarios.

Definition S1.3
For a fixed d*, we define a product structure by the set* $\{r, \{d_k\}, \{\alpha_k\}\}$*. For a composite system,* $d = \prod_{k=1}^{r} d_k$*, where* d_k *is the dimension of each subsystem and* r *is the number of subsystems. The state of the composite system can be written as* $\rho = \rho^1_{\alpha_1} \otimes \rho^2_{\alpha_2} \otimes \cdots \otimes \rho^r_{\alpha_r}$*. Here,* $\alpha_k = c$ *and* $\alpha_k = q$*, are used to denote the "classical" and "quantum" nature of the subsystem, respectively. Then,* $\rho^k_c \in \Delta_{d_k - 1}$ *is a classical state, and* $\rho^k_q \in \mathcal{S}(\mathbb{C}^{d_k})$ *is a quantum state.*

We are now in a position to formally pose the central question of this paper. Suppose Alice creates states of dimension d with a certain *product structure*, i.e. she creates the state $\rho = \rho^1_{\alpha_1} \otimes \rho^2_{\alpha_2} \otimes \cdots \otimes \rho^r_{\alpha_r}$. When dealing with separable states, it is easier to speak as if the information was encoded into distinct non-interacting physical systems. Of course it could equivalently be the case that there is only one physical system with non-interacting degrees of freedom creating the abstract separable structure, but for the sake of clarity we will keep the first picture in mind. This may be viewed as adding constraints to Alice's possible encoding functions $\mathcal{E}$.

We must further assume the same constraints on Bob's measurements. This might seem arbitrary, as we are only interested in the nature of the prepared state. Nevertheless, one can argue that if e.g. Bob is allowed to perform "entangling" measurements, this device might as well be located in Alice's lab, allowing her to prepare an arbitrary entangled state which does not respect the original constraints. That is, we are interested in the scenario where both Alice and Bob have the same technological equipment at their disposal, as is the case in experiments [2]. We remark that this assumption was also used to prove robustness in [3]. Table S1 gives an example of different product structures if $r \leq 2$.

Our main theorem states that the optimal ASPs of QRACs serve as a tool to differentiate these product structures. For convenience we also restate it here.

Theorem 1 (Main theorem) *d-dimensional* $2^d \to 1$ *QRACs serve as gamut dimension witnesses using the ASP function.*

The rest of this section is dedicated to proving Theorem 1.

Case	Constraints on $\mathcal{E}$, and $\mathcal{D}$
$Q_{d_1 d_2}$	Fully Quantum (No Constraints) $\rho \in \mathcal{S}(\mathrm{C}^d)$
$Q_{d_1} Q_{d_2}$	Separable Quantum States $\rho = \rho_q^1 \otimes \rho_q^2$ $\rho_q^1 \in \mathcal{S}(\mathrm{C}^{d_1})$, $\rho_q^2 \in \mathcal{S}(\mathrm{C}^{d_2})$
$Q_{d_1} C_{d_2}$	Classical Quantum $\rho = \rho_q^1 \otimes \rho_c^2$ $\rho_q^1 \in \mathcal{S}(\mathrm{C}^{d_1})$, $\rho_c^2 \in \Delta_{d_2-1}$
$C_{d_1} Q_{d_2}$	Classical Quantum $\rho = \rho_c^1 \otimes \rho_q^2$ $\rho_c^1 \in \Delta_{d_1-1}$, $\rho_q^2 \in \mathcal{S}(\mathrm{C}^{d_2})$
$C_{d_1 d_2}$	Classical $\rho \in \Delta_{d_1 d_2 - 1}$

TABLE S1: Example of Alice's possible product structures, if the dimension $d = d_1 d_2$ factorizes and $r \leq 2$. We assume that the measurement $\mathcal{D}$ has the same product structure as the encoding $\mathcal{E}$.

S1.B. Proofs of Lemmas 1 & 2

We will show how to transform from the most general setup from Fig. S2(a), into the setup of Fig. S2(b). In order to do this, we restrict the encoding function to only pure states (the optimality of which is demonstrated in Ref.[1]), the measurements to be projectives (shown optimal for our case in [4]), and prove two lemmas that show that both (1) classical post-processing functions, and (2) sequential adaptive strategies, are all unnecessary on Bob's side. Note that these lemmas apply in the general $n^d \to 1$ case.

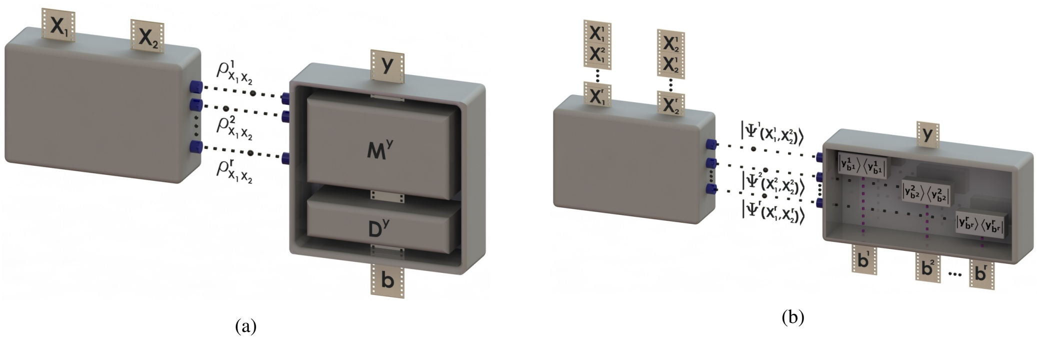


FIG. S2: (a) A generic QRAC with a product structure. (b) A simplified version using Lemmas 1,2.

The first simplification we make is to show that the optimal quantum strategy does not require classical post-processing functions $\mathcal{D}^y$. That is, Bob's output b can simply be read out from his quantum measurements. This is typically assumed in all QRAC papers (e.g. [1, 5]) but without proof.

Lemma 1
Given a quantum decoding strategy $(\{M_l^y\}_l, \mathcal{D}^y)$ with average success probability $\bar{p}$, there exists another quantum decoding strategy $(\{\tilde{M}_l^y\}_l, \tilde{\mathcal{D}}_y)$ with average success probability $\tilde{p} \geq \bar{p}$ and with trivial classical post processing $\tilde{\mathcal{D}}_y = id$.

Proof of Lemma 1 *Let $\rho_x = \mathcal{E}(x)$ be the states which achieve the optimal average success probability $\bar{p}$. Then Eq (S1) can be expressed as:*

$$\bar{p} = \frac{1}{nd^n} \sum_{x,y} tr \left[\rho_x \sum_{k: D_y(k) = x_y} M_k^y \right]. \tag{S4}$$

Now, let us define new operators:

$$\tilde{M}_k^y = \sum_{j:D_y(j)=k} M_j^y. \tag{S5}$$

We can now use the same encoding states $\rho_{x_1,x_2,\ldots,x_n}$ and write the original average success probability in terms of the new operators:

$$\bar{p} = \frac{1}{nd^n} \sum_{x,y} tr\left[\rho_x \tilde{M}_{x_y}^y\right]. \tag{S6}$$

Since we used a fixed encoding strategy and have a new decoding strategy, in principle we could have $\tilde{p} \geq \bar{p}$ after further optimization. Also, we see in Eq (S6) that there is no need for explicit classical post-processing (i.e. $\tilde{D}_y(k) = k$). Thus, hereafter, quantum decoding strategies will simply be written as $\{M_b^y\}_b$, since they will directly output the guess b.

Therefore, the most general allowed measurement strategy is:

Definition S1.4
*Assume that Bob receives r states from Alice: $\rho = \rho_{\alpha_1}^1 \otimes \rho_{\alpha_2}^2 \otimes \cdots \otimes \rho_{\alpha_r}^r$ (in fact, by [1] these could be assumed to be pure states), where each $\rho_{\alpha_i}^i \in \mathcal{S}(\mathbb{C}^{d_i})$ and $d = d_1 d_2 \cdots d_r$. By Lemma 1, let the measurement outcome of $\rho_{\alpha_i}^i$ be $b^i \in \{1, 2, \ldots, d_i\}$. We call a **sequential adaptive strategy** any scheme where Bob uses previous measurement outputs to determine the measurement basis of future states. That is, when measuring the state $\rho_{\alpha_j}^j$, the basis $\{M_l^{y,b^1,b^2,\ldots,b^{j-1}}\}_{l=1}^{d_j}$ could depend on the previously measured systems.*

This scenario is problematic, since optimizing sequential adaptive quantum strategies turns out to be extremely complicated in general. One of our main technical contributions is to show that they are not necessary for optimality.

Lemma 2
There exists an optimal strategy that does not use sequential adaptive measurements.

Proof of Lemma 2 *Let's assume we have a strategy that uses sequential adaptive measurements. Fix the choice of all encoded states and measurements. Then, we show that there exists a strategy without sequential adaptive measurements, that gives at least as high average success probability, as the original one. To show this, let us write the average success probability for the mentioned sequential adaptive strategy as:*

$$\begin{aligned}
\bar{p} &= \frac{1}{2d^2} \sum_{x,y} \mathbb{P}(B^1 = x_y^1, B^2 = x_y^2, \cdots, B^r = x_y^r \mid X = x, Y = y) \\
&= \frac{1}{2d^2} \sum_{x,y} \mathbb{P}(B^1 = correct, B^2 = correct, \cdots, B^r = correct \mid X = x, Y = y) \\
&= \mathbb{P}(B^1 = correct, B^2 = correct, \cdots, B^r = correct) \\
&= \mathbb{P}(B^r = correct \mid B^{r-1} = correct, \cdots, B^1 = correct)\mathbb{P}(B^{r-1} = correct, \cdots, B^1 = correct) \\
&= \ldots = \prod_{k=r}^{1} \mathbb{P}(B^k = correct \mid B^{k-1} = correct, \cdots, B^1 = correct),
\end{aligned} \tag{S7}$$

where we used the definition of conditional probability multiple times. By construction, B^k can only depend on such B^js that $j < k$. Now, we can use the fact, that the conditional probability is again a valid probability measure, thus we can apply completeness of probabilities. Let us denote $\prod_{k=r}^{m} \mathbb{P}(B^k = correct \mid B^{k-1} = correct, \cdots, B^1 = correct) \equiv \mathcal{P}^m$. Then

$$\begin{aligned}
\bar{p} &= \mathcal{P}^3 \cdot \mathbb{P}(B^2 = correct \mid B^1 = correct)\mathbb{P}(B^1 = correct) \\
&= \mathcal{P}^3 \Big(\sum_{s=1}^{d_1} \mathbb{P}(B^2 = correct \mid B^1 = correct, B^1 = s)\mathbb{P}(B^1 = s \mid B^1 = correct) \Big) \mathbb{P}(B^1 = correct) \\
&= \mathcal{P}^3 \Big(\sum_{s=1}^{d_1} \mathbb{P}(B^2 = correct \mid B^1 = correct, B^1 = s)\mathbb{P}(B^1 = s, B^1 = correct) \Big).
\end{aligned} \tag{S8}$$

We see that the events $(B^1 = \textit{correct})$ *and* $(B^2 = \textit{correct})$ *are independent when conditioning on the value of* B^1*, i.e.*

$$\begin{aligned} &\mathbb{P}(B^2 = \textit{correct}, B^1 = \textit{correct} \mid B^1 = s) \\ &= \mathbb{P}(B^1 = \textit{correct} \mid B^1 = s)\mathbb{P}(B^2 = \textit{correct} \mid B^1 = s), \end{aligned} \tag{S9}$$

for any $s \in \{1, \dots, d_1\}$*. This is because if we condition on the value of* B^1*, we fix all the states and measurements (remember that the strategy is fixed, and the only freedom is in the choice of measurement basis on qudit 2 (see Fig. S2(b))). Then, since our qudits are in a product state, their outcomes are independent.*

From equation (S9) it follows that

$$\mathbb{P}(B^2 = \textit{correct} \mid B^1 = \textit{correct}, B^1 = s) = \mathbb{P}(B^2 = \textit{correct} \mid B^1 = s), \tag{S10}$$

and thus

$$\begin{aligned} \bar{p} &= \mathcal{P}^3 \Big(\sum_{s=1}^{d_1} \mathbb{P}(B^2 = \textit{correct} \mid B^1 = s)\mathbb{P}(B^1 = s, B^1 = \textit{correct}) \Big) \\ &\leq \mathcal{P}^3 \Big(\sum_{s=1}^{d_1} \mathbb{P}(B^2 = \textit{correct})\mathbb{P}(B^1 = s, B^1 = \textit{correct}) \Big) = \mathcal{P}^3 \cdot \mathbb{P}(B^2 = \textit{correct})\mathbb{P}(B^1 = \textit{correct}), \end{aligned} \tag{S11}$$

where $\mathbb{P}(B^2 = \textit{correct}) = \max_{s \in \{1,\dots,d_1\}} \mathbb{P}(B^2 = \textit{correct} \mid B^1 = s)$*, i.e. we choose the measurement basis which gives the greatest success probability for qudit 2, hence eliminating adaptiveness on this qudit. Now, we use the same reasoning in order to get rid of adaptiveness on consequtive qudits. We show that this indeed works on qudit 3, and then the idea generalizes trivially. At this point, we have that*

$$\begin{aligned} \bar{p} &= \mathcal{P}^4 \cdot \mathbb{P}(B^3 = \textit{correct} \mid B^2 = \textit{correct}, B^1 = \textit{correct})\mathbb{P}(B^2 = \textit{correct})\mathbb{P}(B^1 = \textit{correct}) \\ &= \mathcal{P}^4 \Big(\sum_{s=1}^{d_2} \sum_{t=1}^{d_1} \mathbb{P}(B^3 = \textit{correct} \mid B^2 = \textit{correct}, B^1 = \textit{correct}, B^2 = s, B^1 = t) \\ &\times \mathbb{P}(B^2 = s \mid B^2 = \textit{correct})\mathbb{P}(B^1 = t \mid B^1 = \textit{correct}) \Big) \mathbb{P}(B^2 = \textit{correct})\mathbb{P}(B^1 = \textit{correct}) \\ &= \mathcal{P}^4 \Big(\sum_{s=1}^{d_2} \sum_{t=1}^{d_1} \mathbb{P}(B^3 = \textit{correct} \mid B^2 = \textit{correct}, B^1 = \textit{correct}, B^2 = s, B^1 = t) \\ &\times \mathbb{P}(B^2 = s, B^2 = \textit{correct})\mathbb{P}(B^1 = t, B^1 = \textit{correct}) \Big) \end{aligned} \tag{S12}$$

(here, we implicitly used the already proven fact that qudits 1 and 2 are independent of each other). Now the conditional independence goes as

$$\begin{aligned} &\mathbb{P}(B^3 = \textit{correct}, B^2 = \textit{correct}, B^1 = \textit{correct} \mid B^2 = s, B^1 = t) \\ &= \mathbb{P}(B^3 = \textit{correct} \mid B^2 = s, B^1 = t)\mathbb{P}(B^2 = \textit{correct}, B^1 = \textit{correct} \mid B^2 = s, B^1 = t), \end{aligned} \tag{S13}$$

since fixing all measurement bases yields independent outcomes. From this it follows that

$$\mathbb{P}(B^3 = \textit{correct} \mid B^2 = \textit{correct}, B^1 = \textit{correct}, B^2 = s, B^1 = t) = \mathbb{P}(B^3 = \textit{correct} \mid B^2 = s, B^1 = t), \tag{S14}$$

and thus

$$\begin{aligned} \bar{p} &= \mathcal{P}^4 \Big(\sum_{s=1}^{d_2} \sum_{t=1}^{d_1} \mathbb{P}(B^3 = \textit{correct} \mid B^2 = s, B^1 = t)\mathbb{P}(B^2 = s, B^2 = \textit{correct})\mathbb{P}(B^1 = t, B^1 = \textit{correct}) \\ &\leq \mathcal{P}^4 \Big(\sum_{s=1}^{d_2} \sum_{t=1}^{d_1} \mathbb{P}(B^3 = \textit{correct})\mathbb{P}(B^2 = s, B^2 = \textit{correct})\mathbb{P}(B^1 = t, B^1 = \textit{correct}) \Big) \\ &= \mathcal{P}^4 \cdot \mathbb{P}(B^3 = \textit{correct})\mathbb{P}(B^2 = \textit{correct})\mathbb{P}(B^1 = \textit{correct}), \end{aligned} \tag{S15}$$

where $\mathbb{P}(B^3 = \textit{correct}) = \max_{\substack{s \in \{1,\dots,d_2\} \\ t \in \{1,\dots,d_1\}}} \mathbb{P}(B^3 = \textit{correct} \mid B^2 = s, B^1 = t)$*, meaning that we choose the measurement basis that gives the greatest success probability on qudit 3. It is clear now that this reasoning applies for all qudits and thus*

$$\bar{p} = \prod_{k=r}^{1} \mathbb{P}(B^k = \textit{correct}), \tag{S16}$$

and it is a non-adaptive strategy.

S1.C. Trade-Off Functions

The usefulness of non-adaptive strategies is that in essence, Alice and Bob are playing r QRACs in parallel (see Fig. S2(b)). However, the optimal average success probability is not necessarily given by the independent optimal strategies on the individual subspaces. This is easily understood when one remembers that the winning condition is that $b = x_y$ as a whole, and no "partial points" are awarded if only a part of the string is guessed correctly. Before proceeding, it is illustrative to look at the ASP once again, but written in the following way:

$$\begin{aligned}\bar{p} &= \frac{1}{2}\left[\frac{1}{d^2}\left(\sum_{x_1,x_2} \mathrm{P}(B = x_1|X = x_1x_2, Y = 1)\right) + \frac{1}{d^2}\left(\sum_{x_1,x_2} \mathrm{P}(B = x_2|X = x_1x_2, Y = 2)\right)\right] \\ &= \frac{1}{2}\left[\mathrm{P}(\text{Bob correctly guesses } x_1) + \mathrm{P}(\text{Bob correctly guesses } x_2)\right],\end{aligned} \tag{S17}$$

where we have defined $\mathrm{P}(\text{Bob correctly guesses } x_y)$ as the average probability of success, if the y-th dit is asked. Let us remark that these probabilities are not independent and are clearly strategy dependent. It is this first dependency that will be our object of study:

Definition S1.5
*Let $z = \mathrm{P}(\text{Bob correctly guesses } x_1)$. Then we define the **quantum trade-off function** $\mathcal{M}_d^q(z)$ in dimension d as:*

$$\mathcal{M}_d^q(z) = \max_{(\mathcal{E},\{M_l^y\}_l)} \{\mathrm{P}(\textit{Bob correctly guesses } x_2)|\mathrm{P}(\textit{Bob correctly guesses } x_1) = z\}, \tag{S18}$$

where the maximization is limited to all quantum encoding-decoding strategies which respect the condition of guessing x_1.

In fact, one could formally write the optimal ASP in terms of the trade-off function as:

$$\bar{p}_{Q_d} = \max_{z\in[\frac{1}{d},1]} \frac{1}{2}\left[z + \mathcal{M}_d^q(z)\right]. \tag{S19}$$

We will devote a later Lemma (3) to investigating the functional form of the quantum $\mathcal{M}_d^q$. For now, we return to the problem of the r QRACs in parallel. When writing out the average success probability, we have to calculate the probability that Alice and Bob win given inputs x_1, x_2, y. That is,

$$\begin{aligned}\mathrm{P}(B = x_y|X = x_1x_2, Y = y) &= \mathrm{P}(B^1 = x_y^1, B^2 = x_y^2, \dots, B^r = x_y^r|X = x_1x_2, Y = y) \\ &= \prod_{k=1}^{r} \mathrm{P}(B^k = x_y^k|X = x_1x_2, Y = y).\end{aligned} \tag{S20}$$

The first equality is just expanding the dits into r substrings ($B = B^1B^2\dots B^r$ and $x_y = x_y^1x_y^2\dots x_y^r$). To obtain the second equality, we use the fact that the QRACs are independent. According to Lemmas 1 and 2, Bob will use identity decoding on each measurement and output $b = b^1b^2\dots b^r$ as a guess for x_y. This in turn implies that the kth information carrier only has information about x_1^k and x_2^k, i.e. $\mathrm{P}(B^k = x_y^k|X = x_1x_2, Y = y)$ only depends on x_1^k and x_2^k.

Hence, let us define

$$\mathrm{P}(\text{Bob correctly guesses } x_y^k) = \frac{1}{(d_k)^2} \sum_{x_1^k,x_2^k\in[d_k]} \mathrm{P}(B^k = x_y^k|X^k = x_1^kx_2^k, Y = y). \tag{S21}$$

Then, Alice and Bob are trying to maximize the following global expression:

$$\bar{p}_{Q_{d_1}Q_{d_2}\dots Q_{d_r}} = \max_{z^1\in[\frac{1}{d_1},1],z^2\in[\frac{1}{d_2},1],\dots,z^r\in[\frac{1}{d_r},1]} \frac{1}{2}\left[z^1z^2\dots z^r + \mathcal{M}_{d_1}^q(z^1)\mathcal{M}_{d_2}^q(z^2)\dots\mathcal{M}_{d_r}^q(z^r)\right]. \tag{S22}$$

By optimizing (S22) , we are able to calculate the average success probability for separable states, and compare it to the optimal average success probability of (S19). We now turn to showing the form of $\mathcal{M}_d^q(z)$.

Lemma 3
The following are equivalent forms of $\mathcal{M}_d^q(z)$:

$$\mathcal{M}_d^q(z) = 1 - \left(\frac{d-1}{d}\right)\left(\sqrt{z} - \sqrt{\frac{1-z}{d-1}}\right)^2, \tag{S23}$$

$$\mathcal{M}_d^q(z) = \cos^2\left(\cos^{-1}\left(\frac{1}{\sqrt{d}}\right) - \cos^{-1}\left(\sqrt{z}\right)\right). \tag{S24}$$

Furthermore, they are achieved when Bob's measurement bases are mutually unbiased.

Proof of Lemma 3 *Let Bob's decoding bases be $\{|\psi_k\rangle\}_k$, and $\{|\phi_k\rangle\}_k$, corresponding to $y = 1$ and 2, respectively. Given inputs x_1, x_2, Alice's best strategy is to encode a superposition of $|\psi_{x_1}\rangle$ and $|\phi_{x_2}\rangle$. Having any orthogonal components to these states will drop her average success probability and hence those strategies will not appear in the maximization performed for the trade-off function. Explicitly:*

$$\mathcal{E}(x) = |x\rangle = \frac{1}{\sqrt{N}}\left(t|\psi_{x_1}\rangle + e^{\mathbf{i}\zeta}(1-t)|\phi_{x_2}\rangle\right), \tag{S25}$$

where $N = 1 + 2t(1-t)\left(\Re[e^{\mathbf{i}\zeta}\langle\psi_{x_1}|\phi_{x_2}\rangle] - 1\right)$ is a normalization factor, $t \in [0,1]$ is a parameter that will vary to change Bob's probability of correctly guessing the first dit, and $\zeta \in [0, 2\pi)$ is a phase. It can be verified that $\zeta = -Arg(\langle\psi_{x_1}|\phi_{x_2}\rangle)$, i.e. $e^{\mathbf{i}\zeta}\langle\psi_{x_1}|\phi_{x_2}\rangle \in \mathrm{R}^+$ simultaneously maximizes both $|\langle\psi_{x_1}|x\rangle|^2$ and $|\langle\phi_{x_2}|x\rangle|^2$, for all $t \in [0,1]$. These are the probabilities of Bob correctly guessing x_1 and x_2, respectively. With this choice of ζ then:

$$z_x \equiv |\langle\psi_{x_1}|x\rangle|^2 = \frac{\left(t + \sqrt{s_x}(1-t)\right)^2}{N}, \tag{S26}$$

$$|\langle\phi_{x_2}|x\rangle|^2 = \frac{\left(t\sqrt{s_x} + (1-t)\right)^2}{N}, \tag{S27}$$

where $s_x = |\langle\psi_{x_1}|\phi_{x_2}\rangle|^2$. Inverting equation (S26) to have $t = t(z_x, s_x)$:

$$t = \frac{-z_x + \sqrt{s_x}(\sqrt{s_x} + z_x - 1) \pm \sqrt{(s_x-1)z_x(z_x-1)}}{(\sqrt{s_x}-1)(\sqrt{s_x} - 1 + 2z_x)}. \tag{S28}$$

Then, inserting it into (S27) we obtain the probability of correctly guessing the second dit, as a function of the probability of correctly guessing the first (z_x).

$$|\langle\phi_{x_2}|x\rangle|^2 = (1-z_x) + s_x(2z_x - 1) \pm 2\sqrt{s_x(s_x-1)z_x(z_x-1)}. \tag{S29}$$

We take the positive sign, since we want to maximize the average success probability. Hence, we are trying to maximize the expression:

$$\bar{p} = \max_{\{|\psi_k\rangle\},\{|\phi_k\rangle\}} \frac{1}{2d^2}\sum_x \left(1 + s_x(2z_x-1) + 2\sqrt{s_x(s_x-1)z_x(z_x-1)}\right), \tag{S30}$$

subject to the conditions $0 \le s_x, z_x \le 1$, $\sum_x s_x = d$, and $\sum_x z_x = zd^2$, where $z = \mathrm{P}$(Bob correctly guesses x_1).

The non-constant part of the above expression can be written as $\sum_x f(s_x, z_x)$, where $f(s_x, z_x) = s_x z_x + \sqrt{s_x(1-s_x)z_x(1-z_x)}$. This sum is a function of the 2-by-d^2 matrix $S = \binom{\vec{s}^T}{\vec{z}^T}$, where the x-th element of the vector $\vec{s}$ ($\vec{z}$) is s_x (z_x). Note that for any matrix S satisfying the constraints on the s_x and z_x,

$$S^* \equiv \begin{pmatrix} \frac{1}{d} & \frac{1}{d} & \cdots & \frac{1}{d} \\ z & z & \cdots & z \end{pmatrix} = S \begin{pmatrix} \frac{1}{d^2} & \cdots & \frac{1}{d^2} \\ \vdots & & \vdots \\ \frac{1}{d^2} & \cdots & \frac{1}{d^2} \end{pmatrix}. \tag{S31}$$

Here, the last matrix is doubly stochastic, and hence we say that any matrix S satisfying the constraints on the s_x and z_x majorizes S^* (see [6, Definition 15.A.2]). But this is equivalent ([6, Proposition 15.A.4]) to the statement that $\sum_x \phi(s_x, z_x) \le$

$\sum_x \phi(\frac{1}{d}, z)$ *for all continuous concave functions* $\phi : \mathbb{R}^2 \to \mathbb{R}$*. It is straightforward to show that the function* $f(s_x, z_x)$ *is concave (i.e. its Hessian is negative semi-definite) on the domain* $[0,1] \times [0,1]$*, and hence, considering the above, the ASP (Eq. (S30)) is maximized by* $s_x = \frac{1}{d}$ *and* $z_x = z$ *for all* x*. Substituting these into Eq. (S29) we get the form of the trade-off function:*

$$M_d^q(z) = 1 - z + \frac{2z-1}{d} + 2\frac{\sqrt{(d-1)z(1-z)}}{d}, \tag{S32}$$

which can be furthered simplified into (S23).

To obtain the other form of $\mathcal{M}_d^q(z)$ *we can visualize the problem geometrically, by regarding the angle* θ *between two state vectors* $|\xi\rangle$ *and* $|\chi\rangle$ *to be* $\theta = \cos^{-1}(|\langle\xi|\chi\rangle|)$*. We have shown that the trade-off function is obtained when Bob uses two mutually unbiased bases, hence the measurement vectors* $|\psi_{x_1}\rangle$ *and* $|\phi_{x_2}\rangle$ *have an angle of* $\cos^{-1}\left(d^{-1/2}\right)$ *between them. Alice's encoded state* $|x\rangle$ *must lie on the plane of the measurement vectors and the angle between* $|x\rangle$ *and* $|\psi_{x_1}\rangle$ *is* $\cos^{-1}(\sqrt{z})$*. The trade-off function (S24) is then obtained when we see that the angle between* $|x\rangle$ *and* $|\phi_{x_2}\rangle$ *is the difference of the two angles described above.*

Notice that in the discussion following (S31) it was shown that $s_x = |\langle\psi_{x_1}|\phi_{x_2}\rangle|^2 = 1/d$ for all x. This is precisely the MUB condition on Bob's measurements. To arrive at Alice's optimal strategy we need to maximize (S22), using the derived representation (S24) of $\mathcal{M}_d^q(z)$. The maximization can easily done by setting $\frac{d\bar{p}_{Q_d}}{dz} = 0$, to find $z_{\max}$. Explicitly:

$$z_{\max} = \mathcal{M}_d^q(z_{\max}) = \frac{1}{2}\left(1 + \frac{1}{\sqrt{d}}\right). \tag{S33}$$

This means that the best strategy for Alice is to encode every state $|x\rangle$ into an equal superposition of $|\psi_{x_1}\rangle$ and $|\phi_{x_2}\rangle$ in order for the success probability to be the same, no matter which basis Bob chooses to do a measurement in. We put this into a corollary:

Corollary 1 *For* $2^d \to 1$ *QRACs, the optimal average success probability is achieved when Bob uses two mutually unbiased bases* $(\{|\psi_{x_1}\rangle\}_{x_1}, \{|\phi_{x_2}\rangle\}_{x_2})$*, and Alice encodes her inputs into states* $|x_1x_2\rangle$ *which are equal superpositions of* $|\psi_{x_1}\rangle$ *and* $|\phi_{x_2}\rangle$*.*

Note that this optimal quantum strategy for d-dimensional $2^d \to 1$ QRACs has been discussed in [7]. The optimal encoding strategy for Alice involves encoding her state into the eigenvector corresponding to the highest eigenvalue of the operator $(|\psi_{x_1}\rangle\langle\psi_{x_1}| + |\phi_{x_2}\rangle\langle\phi_{x_2}|)$. This is the state given in Equation (2) of the main text.

For completeness, we also define the *classical trade-off function* $\mathcal{M}_d^c(z)$ in an analogous way to Definition S1.5, except that the maximization is done over classical encoding-decoding strategies. In fact, this function is linear:

$$\mathcal{M}_d^c(z) = \frac{d+1}{d} - z. \tag{S34}$$

This can easily be checked, since the optimal success probability for $2^d \to 1$ RACs is known to be $\bar{p}_{C_d} = (d+1)/2d$ [8]. This success probability can be obtained by the pure coding schemes of just sending the first or second dit, and all convex combinations of these strategies would give the same maximum. See Fig. S3 for a visualization of the trade-off functions with varying dimensions. Note, however, that classical strategies factorize, so that we never use the trade-off functions in this setting alone, but only in conjunction with the quantum functions, e.g. if Alice is able to encode her input dits into quantum systems of dimensions $d_1, d_2, \ldots, d_{r-1}$ and the rest of the information of dimension d_r classically, we would have to maximize:

$$\bar{p}_{Q_{d_1}Q_{d_2}\ldots Q_{d_{r-1}}C_{d_r}} = \max_{z^1\in[\frac{1}{d_1},1],\ldots,z^r\in[\frac{1}{d_r},1]} \frac{1}{2}\left[z^1z^2\cdots z^r + \mathcal{M}_{d_1}^q(z^1)\cdots\mathcal{M}_{d_{r-1}}^q(z^{r-1})\mathcal{M}_{d_r}^c(z^r)\right]. \tag{S35}$$

S1.D. Two Examples

S1.D.I. d=39

Here, we take the case $d = 39$ into consideration, which will highlight the necessity of the trade-off functions. We have that $\bar{p}_{Q_{39}} = \frac{1}{2}\left(1 + \frac{1}{\sqrt{39}}\right) \approx 0.5801$. Now, we wish to know the optimal ASP if the preparation and measurement are split in terms of two systems with dimensions $d_1 = 13$ and $d_2 = 3$. Numerically we optimize (S22):

$$\bar{p}_{Q_{13}Q_3} = \max_{z^1\in[\frac{1}{13},1],z^2\in[\frac{1}{3},1]} \frac{1}{2}\left[z^1z^2 + \mathcal{M}_{13}^q(z^1)\mathcal{M}_3^q(z^2)\right] \approx 0.5217. \tag{S36}$$

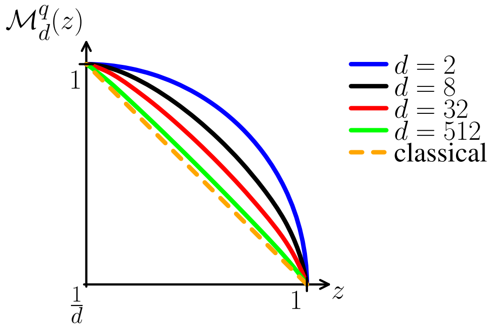


FIG. S3: Visualization of the quantum trade-off functions $\mathcal{M}_d^q(z)$, with varying dimensions.

A contour plot of of the function being maximized (S36) with the maxima highlighted can be seen in Fig. S4. In fact, the maximum is obtained in two different points. Let $(z^1, z^2) = (0.1944, 0.4302)$ be the first point, then in fact $(\mathcal{M}_{13}^q(0.1944), \mathcal{M}_3^q(0.4302)) = (0.9695, 0.9900)$ is the other point which achieves the maximum. The first point, where both z^1 and z^2 are relatively small, the strategy gives a strong bias to guessing the second dit x_2 at the expense of lowering the probability of correctly guessing the first input x_1. Explicitly for the first point; $\mathrm{P}(\text{Bob correctly guesses } x_1) = z^1 z^2 \approx 0.0836$, whereas $\mathrm{P}(\text{Bob correctly guesses } x_2) = \mathcal{M}_{13}^q(z^1)\mathcal{M}_3^q(z^2) \approx 0.9598$. It is clear then, that the second point which achieves the maximum is just a reflection of this strategy, now giving a positive bias towards guessing the first dit.

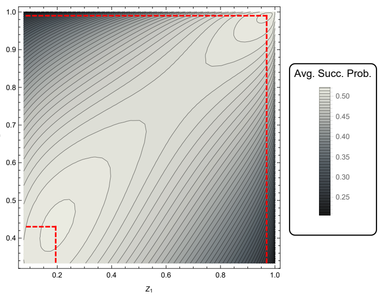


FIG. S4: Contour plot of (S36), for the example $d = 39$. See text for details.

To conclude, we explicitly see that $\bar{p}_{Q_{13}Q_3} > \bar{p}_{Q_{13}}\bar{p}_{Q_3} \approx 0.5037$. That is, even though Alice and Bob are using two non-interacting Hilbert spaces, the optimal strategy is a global one, instead of playing strictly independent QRACs.

S1.D.II. d=1024

Now, we look at the case $d = 1024$, the dimension we certify in our experiment. We compute the optimal success probabilities for all possible quantum partitions of a 1024-dimensional quantum system. The values were calculated using Eq. (S35). The aim here is to show that $Q_{512}Q_2$ was the relevant bound for the experiment, and not e.g. $Q_{32}Q_{32}$ or any other partition. See Table S2.

Notice that, since $\mathcal{M}_d^q(z) > \mathcal{M}_d^c(z)$, there is no need to calculate the classical-quantum partitions, as they would clearly be worse than the equivalent fully quantum partition. However, it is interesting to note that $Q_{512}C_2 > Q_{256}Q_4$.

S2. EXPERIMENTAL CONSIDERATIONS

In this section, we deal with the analysis supporting our photonic experiment in dimension $d = 1024$.

S2.A. Useful Representation of the MUBs

From a theoretical point of view, any two mutually unbiased bases in dimension $d = 1024$ would yield the optimal average success probability. However, in our optical setup, for simplicity it is better to consider a representation of the two MUBs which have only matrix elements given by ± 1. Thus requiring only phase-modulations of 0 or π to be addressed by the SLMs to encode and decode the required states. To construct such MUBs in dimension 1024, we first consider two MUBs in dimension 4:

$$MUB_1^{d=4} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}, \tag{S37}$$

$$MUB_2^{d=4} = \frac{1}{2}\begin{pmatrix} 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 \\ -1 & -1 & 1 & -1 \end{pmatrix}. \tag{S38}$$

Now, if we consider the following tensor products:

$$MUB_1 = (MUB_1^{d=4})^{\otimes 5}, \qquad MUB_2 = (MUB_2^{d=4})^{\otimes 5}, \tag{S39}$$

we end up with two MUBs in dimension 1024, where the columns represent the basis states.

S2.B. Single Detector Scheme

In our photonic experiment, we are dealing with a very large dimension ($d = 1024$). The protocol requires Bob to perform a full von Neumann projective measurement on one of two bases before outputting his guess b. In the laboratory this would translate to having 1024 different photo-detectors associated to each of the eigenvalues of the measurement performed, which is practically impossible. However, one can simulate a full d-outcome projective measurement to overcome this limitation, as it has been commonly done in the field of high-dimensional quantum information processing [9–12]. The basic idea is that Bob uses a flexible detector scheme, which can project the incoming state to each one of the MUBs states required in the protocol. Thus, estimating the probability for each basis state collapse individually with only one detector.

In this case, one uses an extra randomly uniform input $j \in [d]$ on Bob's side. Depending on his inputs y, j Bob will measure the operators $\{|m_j^y\rangle\langle m_j^y|, 1 - |m_j^y\rangle\langle m_j^y|\}$. If Alice's state collapses on $|m_j^y\rangle\langle m_j^y|$, i.e. a photon is recorded by Bob while the scheme is set to make the projection $|m_j^y\rangle\langle m_j^y|$, he will claim that $x_y = j$. Otherwise, he will assume that $x_y \neq j$. A full von Neumann measurement is simulated in the case that

$$\sum_{j\in[d]} |m_j^y\rangle\langle m_j^y| = 1, \;\; \forall y \in [n]. \tag{S40}$$

Let us consider the events where $x_y = j$ and define the total number of such events X_1. Let us also define D_1 as the number of "clicks" from the experiment in those cases. Likewise, let X_2 denote the number of events where $x_y \neq j$, and D_2 the clicks in those cases. Assuming uniform sampling, $(d-1)X_1 \approx X_2$.

To get an appropriate figure of merit for the experiment in this scenario, consider first the *total experimental efficiency*:

$$\nu := \frac{\#\text{ real clicks}}{\#\text{ theoretically expected clicks}}. \tag{S41}$$

Note that this efficiency does not assume anything about the inner-workings of the actual experimental setup, making it still compatible with the device independent approach. Let q be the average success probability of a given strategy, then:

$$\nu = \frac{D_1 + D_2}{qX_1 + \left(\frac{1-q}{d-1}\right)X_2} = \frac{D_1 + D_2}{X_1}. \tag{S42}$$

To calculate the number of theoretically expected clicks, we use the average failure probability $\left(\frac{1-q}{d-1}\right)$ for simplicity, but without loss of generality. Furthermore, note that the average success probability is the ratio of the times Bob correctly guessed $x_y = j$, to the number of times he should have guessed it to be $x_y = j$:

$$\frac{D_1}{X_1} = \nu q, \tag{S43}$$

Then, by combining equations (S42) and (S43), we obtain:

$$q = \frac{D_1}{D_1 + D_2}, \tag{S44}$$

which will be our main experimental figure of merit to calculate the average success probability q of the strategy. There are several benefits of using (S44) : (1) It has an easy operational interpretation as "fraction of times Bob clicks correctly, compared to the total number of clicks", (2) since it only uses the data from the clicks, it is more experimentally friendly, not lowering the statistics due to detector malfunction or lossy channels, (3) from how it was derived, it does not assume the inner workings of the experiment, making it quite general, and most importantly (4) with the assumption of Eq. (S40), it is equivalent to the standard form of the ASP, i.e., Eq. (S6).

S2.C. Robustness of the ASP to Detection Efficiency and Poissonian Source

In the previous section, we arrived at (S43) by assuming that there was only one photon present in each experimental round. However, in our experimental setup we do not have a perfect single photon source, and multi-photon events can occur. The problem with having more than one photon in the system, is that our detector does not resolve the number of detected photons (otherwise this would not be an issue, and we would simply discard events with more than one photon). The nature of our detection event, the so-called "click", is in fact the probabilistic event "*at least 1 photon detected*". Of course this event can be understood as the complement of the event "*no photon detected*". If we assume for a brief moment that $\nu = 1$, and that there is a n-photon event, the probability of having a "click"-event would be:

$$\mathrm{P}(\text{detecting at least 1 photon } | n\text{-photon event}) = 1 - (1-q)^k. \tag{S45}$$

Due to the nature of laser light formation, we consider a Poisson distribution for our photon production, with mean μ which can be experimentally tuned. Now, we return to the case of having experimental efficiency ν. Imagine that there are n photons with Alice's state $|\Psi\rangle$ present, out of which only k collapse onto the correct state $|\Phi\rangle$ during the measurement process, and then each of the k photons have a ν probability of being detected. Hence, the probability of at least one click would be:

$$\frac{D_1}{X_1} = \sum_{n=1}^{\infty} \mathbb{P}(n \text{ photons produced}) \sum_{k=1}^{n} \mathbb{P}(k \text{ of the } n \text{ photons collapsing on } |\Phi\rangle |\; n\text{-photon event})) \mathbb{P}(\text{at aeast 1 detected}). \tag{S46}$$

This expression is fully general. We now explicitly introduce the Poissonian distribution:

$$\frac{D_1}{X_1} = \sum_{n=1}^{\infty} \left(\frac{\mu^n e^{-\mu}}{n!} \right) \sum_{k=1}^{n} \binom{n}{k} q^k (1-q)^{n-k} \left(1 - (1-\nu)^k\right). \tag{S47}$$

To simplify matters, we look just at the inner summation to get:

$$\sum_{k=1}^{n} \binom{n}{k} q^k (1-q)^{n-k} \left(1 - (1-\nu)^k\right) = 1 - (1-\nu q)^n. \tag{S48}$$

Which is what we could have intuitively guessed since the beginning. If there are k photons present, then the probability to detect at least 1 photon with a ν experimental efficiency is just $1-(1-\nu q)^n$. Then, putting (S48) into (S47) and carrying out the sum we obtain:

$$\frac{D_1}{X_1} = 1 - e^{-\nu\mu q}. \tag{S49}$$

We note that while deriving this, we have been assuming the optimal QRAC strategy for the encoded states and measurement operators. In particular, q does not depend on the inputs of Alice and Bob, (as shown in lemma 3), i.e. every round performs the same as the average. In the same way, the average failing probability $\left(\frac{1-q}{d-1}\right)$ will be modified as:

$$\frac{D_2}{X_2} = 1 - e^{-\nu\mu\left(\frac{1-q}{d-1}\right)}. \tag{S50}$$

Then, if we divide the rhs of (S44) by X_1, and we use (S49) and (S50), we obtain:

$$\frac{D_1}{D_1+D_2}=\frac{1-e^{-\nu\mu q}}{1-e^{-\nu\mu q}+(d-1)\left(1-e^{-\nu\mu\left(\frac{1-q}{d-1}\right)}\right)}, \tag{S51}$$

which relates the theoretical average success probability of the strategy q, to our experimental figure of merit. We interpret this as follows: suppose Alice and Bob's strategy predicts an average success probability of q, and we experimentally know the value $\nu\mu$. Then, equation (S51) gives the maximally allowed value of the figure of merit, assuming no other experimental errors. Experimentally, this allows us to fine-tune the μ parameter, to be sure the $Q_{512}Q_2$ value can be violated.

The first order term of (S51) in the small parameter $\nu\mu$ (0.052 in our setup) is:

$$\frac{D_1}{D_1+D_2}=q-\frac{1}{2}\left(\frac{1-q}{d-1}\right)q(dq-1)\nu\mu+O\left((\nu\mu)^2\right). \tag{S52}$$

[1] Ozols, Maris, *Quantum Random Access Codes with Shared Randomness*, Master's thesis (2009).

[2] We also note that numerical evidence on small dimensional examples suggests that if the quantum states are in a product form, then entangling measurements do not improve the ASP.

[3] W. Cong, Y. Cai, J.-D. Bancal, and V. Scarani, Phys. Rev. Lett. **119**, 080401 (2017).

[4] M. Farkas, ArXiv e-prints (2018), arXiv:1803.00363 [quant-ph].

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

[6] A. Marshall, I. Olkin, and B. Arnold, *Inequalities: Theory of Majorization and Its Applications*, Springer Series in Statistics (Springer New York, 2010).

[7] E. Aguilar, J. Borkała, P. Mironowicz, and M. Pawłowski, (2017), arXiv:1709.04898 [quant-ph].

[8] M. Czechlewski, D. Saha, and M. Pawłowski, ArXiv e-prints (2018), arXiv:1803.05245 [quant-ph].

[9] V. D'Ambrosio, F. Bisesto, F. Sciarrino, J. F. Barra, G. Lima, and A. Cabello, Phys. Rev. Lett. **112**, 140503 (2014).

[10] S. Etcheverry, G. Cañas, E. S. Gómez, W. A. T. Nogueira, C. Saavedra, G. B. Xavier, and G. Lima, Sci. Rep. **3**, 2316 (2013).

[11] X.-L. Wang, L.-K. Chen, W. Li, H.-L. Huang, C. Liu, C. Chen, Y.-H. Luo, Z.-E. Su, D. Wu, Z.-D. Li, H. Lu, Y. Hu, X. Jiang, C.-Z. Peng, L. Li, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, and J.-W. Pan, Phys. Rev. Lett. **117**, 210502 (2016).

[12] R. Fickler, R. Lapkiewicz, W. N. Plick, M. Krenn, C. Schaeff, S. Ramelow, and A. Zeilinger, Science **338**, 640 (2012).

Case	Optimal $\bar{p}$
Q_{1024}	0.515625
$Q_{512}Q_2$	0.500980
$Q_{256}Q_4$	0.500654
$Q_{256}Q_2Q_2$	0.500653
$Q_{128}Q_8$	0.500563
$Q_{128}Q_4Q_2$	0.500561
$Q_{128}Q_2Q_2Q_2$	0.500560
$Q_{64}Q_{16}$	0.500530
$Q_{64}Q_8Q_2$	0.500525
$Q_{64}Q_4Q_4$	0.500524
$Q_{64}Q_4Q_2Q_2$	0.500523
$Q_{64}Q_2Q_2Q_2Q_2$	0.500523
$Q_{32}Q_{32}$	0.500521
$Q_{32}Q_{16}Q_2$	0.500512
$Q_{32}Q_8Q_4$	0.500509
$Q_{32}Q_8Q_2Q_2$	0.500508
$Q_{32}Q_4Q_4Q_2$	0.500507
$Q_{32}Q_4Q_2Q_2Q_2$	0.500507
$Q_{32}Q_2Q_2Q_2Q_2Q_2$	0.500506
$Q_{16}Q_{16}Q_4$	0.500505
$Q_{16}Q_{16}Q_2Q_2$	0.500504
$Q_{16}Q_8Q_8$	0.500503
$Q_{16}Q_8Q_4Q_2$	0.500501
$Q_{16}Q_8Q_2Q_2Q_2$	0.500501
$Q_{16}Q_4Q_4Q_4$	0.500500
$Q_{16}Q_4Q_4Q_2Q_2$	0.500500
$Q_{16}Q_4Q_2Q_2Q_2Q_2$	0.500499
$Q_{16}Q_2Q_2Q_2Q_2Q_2Q_2$	0.500499
$Q_8Q_8Q_8Q_2$	0.500499
$Q_8Q_8Q_4Q_4$	0.500498
$Q_8Q_8Q_4Q_2Q_2$	0.500498
$Q_8Q_8Q_2Q_2Q_2Q_2$	0.500497
$Q_8Q_4Q_4Q_4Q_2$	0.500497
$Q_8Q_4Q_4Q_2Q_2Q_2$	0.500496
$Q_8Q_4Q_2Q_2Q_2Q_2Q_2$	0.500496
$Q_8Q_2Q_2Q_2Q_2Q_2Q_2Q_2$	0.500495
$Q_4Q_4Q_4Q_4Q_4$	0.500496
$Q_4Q_4Q_4Q_4Q_2Q_2$	0.500495
$Q_4Q_4Q_4Q_2Q_2Q_2Q_2$	0.500495
$Q_4Q_4Q_2Q_2Q_2Q_2Q_2Q_2$	0.500494
$Q_4Q_2Q_2Q_2Q_2Q_2Q_2Q_2Q_2$	0.500494
$Q_2Q_2Q_2Q_2Q_2Q_2Q_2Q_2Q_2Q_2$	0.500493

TABLE S2: All quantum cases for a 1024-dimensional system and the respective optimal ASPs considering each product structure.

Self-testing mutually unbiased bases in the prepare-and-measure scenario

Máté Farkas[1,*] and Jędrzej Kaniewski[2]

[1]*Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics and Informatics, University of Gdansk, 80-952 Gdansk, Poland*

[2]*Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland*

(Received 10 April 2018; revised manuscript received 31 August 2018; published 12 March 2019)

Mutually unbiased bases (MUBs) constitute the canonical example of incompatible quantum measurements. One standard application of MUBs is the task known as quantum random access code (QRAC), in which classical information is encoded in a quantum system, and later part of it is recovered by performing a quantum measurement. We analyze a specific class of QRACs, known as the $2^d \to 1$ QRAC, in which two classical dits are encoded in a d-dimensional quantum system. It is known that among rank-1 projective measurements MUBs give the best performance. We show (for every d) that this cannot be improved by employing nonprojective measurements. Moreover, we show that the optimal performance can only be achieved by measurements which are rank-1 projective and mutually unbiased. In other words, the $2^d \to 1$ QRAC is a self-test for a pair of MUBs in the prepare-and-measure scenario. To make the self-testing statement robust we propose measures which characterize how well a pair of (not necessarily projective) measurements satisfies the MUB conditions and show how to estimate these measures from the observed performance. Similarly, we derive explicit bounds on operational quantities like the incompatibility robustness or the amount of uncertainty generated by the uncharacterized measurements. For low dimensions the robustness of our bounds is comparable to that of currently available technology, which makes them relevant for existing experiments. Last, our results provide essential support for a recently proposed method for solving the long-standing existence problem of MUBs.

DOI: 10.1103/PhysRevA.99.032316

I. INTRODUCTION

Mutually unbiased bases (MUBs) play an important role in many quantum information processing tasks. They are optimal for quantum state determination [1,2], information locking [3,4], and the mean king's problem [5,6]. Moreover, they give rise to the strongest entropic uncertainty relations (among projective measurements) [7–9]. One intuitive way to look at them is the following: Imagine that we encode a classical message in a pure state corresponding to an element of a basis. If we measure this state in a basis unbiased to the initial one, then each measurement outcome occurs with the same probability. That is, we do not learn anything about the originally encoded message. Formally, two bases $\{|a_i\rangle\}_{i=1}^d$ and $\{|b_j\rangle\}_{j=1}^d$ in $\mathbb{C}^d$ are mutually unbiased if

$$|\langle a_i|b_j\rangle|^2 = \frac{1}{d} \quad \forall i, j \in [d] := \{1, 2, \dots, d\}. \tag{1}$$

Due to their importance, significant effort has been dedicated to investigating their structure (see Ref. [10] for a survey and Ref. [11] for a classification in dimensions 2–5). It is known that in dimension d, there are at least 3 and at most $d+1$ MUBs and the upper bound is saturated in prime power dimensions. The maximal number of MUBs in composite dimensions is a long-standing open problem (see Refs. [12–17] for the case of dimension 6).

Another scenario in which MUBs perform well is the so-called $2^d \to 1$ quantum random access code (QRAC) [18,19]. In this setup, two classical dits are encoded into a qudit, and the aim is to recover one of them chosen uniformly at random. It is well-known that sending a quantum system gives an advantage over sending a classical system (of the same dimension) [20] and this fact is used in many quantum information protocols [21–25]. It is commonly believed that the optimal performance of the $2^d \to 1$ QRAC is achieved when the measurements correspond to a pair of MUBs in dimension d, but this claim has only been proven for a restricted class of measurements [26].

The observation that quantum systems can give rise to stronger-than-classical correlations was first made by Bell [27] (although in a slightly different setup). Moreover, it turns out that some of these strongly nonclassical correlations can be achieved in an essentially *unique* manner. That is, the observed statistics allow us to identify the employed states and measurements (up to local isometries and extra degrees of freedom). The most prominent example of this kind is the well-known CHSH inequality [28], which is maximally violated by a pair of MUBs in dimension 2 on both sides [29–32]. Whenever such an inference—characterizing the state and/or measurements based solely on the observed statistics—can be made, it is referred to as *self-testing* [33–35]. Self-testing is closely related to the concept of device-independent (DI) quantum information processing, in which the devices used in the protocol are a priori untrusted [36–40]. It is clear that what makes DI cryptography possible is precisely the self-testing character of the correlations observed during the

*mate.farkas@phdstud.ug.edu.pl

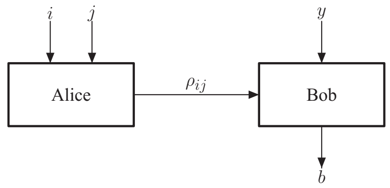


FIG. 1. Schematic representation of the $2^d \to 1$ QRAC protocol.

protocol. By now self-testing is a well-developed field [41–48] and includes results which are robust to noise [49–55]. Such statements are of particular interest, as they can be directly applied to experiments [56].

Recently the notion of self-testing has been extended to prepare-and-measure scenarios [57]. In this setup, a preparation device creates one of many possible quantum states and then sends it to a measurement device. The latter performs one of many possible measurements on the state, and then produces a classical output. This scenario encompasses many important quantum communication protocols, e.g., the BB84 and B92 quantum cryptography protocols [58,59] and the aforementioned QRACs.

In the prepare-and-measure scenario one cannot distinguish between classical and quantum systems, unless additional restrictions are imposed. The standard choice is to place an upper bound on the dimension of the system transmitted between the devices [60–62]. This is often referred to as the semi-device-independent (SDI) model for which several cryptographic protocols have been proposed [63–65]. In analogy to the DI model, it is clear that the security of SDI protocols is related to self-testing results in the prepare-and-measure scenario.

In this paper, we investigate the self-testing properties of the $2^d \to 1$ QRAC. In Ref. [57], the authors derive robust self-testing results for $d = 2$ and ask whether similar statements hold for larger d. We resolve this question by deriving a robust self-testing statement for arbitrary d. We show that the optimal performance in the $2^d \to 1$ QRAC certifies that the two measurements correspond to MUBs. To make the statement robust we propose measures that characterize how close a pair of POVMs is to the MUB arrangement and derive explicit bounds on those in terms of the QRAC performance. Finally, we use this characterization to obtain explicit bounds on operationally relevant quantities like the incompatibility robustness [66] or the amount of uncertainty produced.

II. SETUP

In the $2^d \to 1$ QRAC scenario (see Fig. 1), on the preparation side Alice gets two uniformly random inputs, $i, j \in [d]$. Based on these inputs she prepares a d-dimensional state ρ_{ij}, and sends it to Bob who is on the measurement side. He gets a uniformly random input $y \in \{1, 2\}$, which tells him which of Alice's inputs he is supposed to guess. If $y = 1$, he aims to guess i, otherwise j. This is performed by a measurement on ρ_{ij}, which we describe by the operators $\{A_i\}_i$ for $y = 1$, and $\{B_j\}_j$ for $y = 2$, where $A_i, B_j \geqslant 0$, $\sum_i A_i = \sum_j B_j = \mathbb{I}$ and $i, j \in [d]$. The outcome of the measurement determines Bob's guess and the figure of merit is the *average success probability* (ASP), which can be written, using the above notation, as

$$\bar{p} = \frac{1}{2d^2} \sum_{ij} \mathrm{tr}[\rho_{ij}(A_i + B_j)]. \tag{2}$$

III. IDEAL SELF-TEST

To obtain the ideal self-testing statement we derive an achievable upper bound on the ASP and identify situations in which all the steps in the derivation are tight. Note that $\mathrm{tr}\,[\rho_{ij}(A_i + B_j)] \leqslant \|A_i + B_j\|$, where $\|.\|$ is the operator norm, and since $(A_i + B_j) \geqslant 0$, one can always find a state ρ_{ij} such that this inequality is saturated. Let us from now on assume that the preparations are always chosen optimally, which allows us to focus solely on the measurements. Finding the maximal ASP can be performed using operator norm inequalities and other tools from matrix analysis and yields the following theorem.

Theorem 1. The average success probability of the $2^d \to 1$ QRAC is upper bounded by

$$\bar{p} \leqslant \frac{1}{2}\left(1 + \frac{1}{\sqrt{d}}\right) =: \bar{p}_Q, \tag{3}$$

and this bound can only be attained if Bob's measurements are rank-1 projective and mutually unbiased. Moreover, in the optimal case the prepared states are the unique eigenstates of $A_i + B_j$, corresponding to the highest eigenvalue.

It was previously known that this upper bound holds if we restrict ourselves to rank-1 projective measurements and that among these measurements only MUBs can actually achieve it [26]. What we show is that the QRAC performance cannot be improved by employing nonprojective measurements and that the optimal performance indeed requires MUBs, even if we allow for generic measurements. Note that this does not follow from any extremality argument, as in general projective measurements are not the only extremal d-outcome measurements [67].

For a complete proof, we refer the reader to Appendix A. Here, we state that the crucial step is to use operator norm inequalities to show that the ASP is bounded by

$$\bar{p} \leqslant \frac{1}{2} + \frac{1}{2d^2} \sum_{ij} \sqrt{t_{ij}}, \tag{4}$$

where $t_{ij} := \mathrm{tr}(A_i B_j) \geqslant 0$, and therefore $\sum_{ij} t_{ij} = d$. The right-hand side is strictly Schur-concave in $\{t_{ij}\}_{ij}$, and hence is uniquely maximized by the uniform distribution, $t_{ij} = \frac{1}{d}$ [68], which yields $\bar{p}_Q$. A separate argument implies that to reach $\bar{p}_Q$ both measurements must be rank-1 projective and combining these two facts leads to the conclusion that the two measurements correspond to MUBs.

Theorem 1 implies that the $2^d \to 1$ QRAC is an SDI self-test for a pair of MUBs in dimension d: observing the optimal ASP implies that the two measurements constitute a pair of MUBs. One might wonder whether the self-testing statement can be made even stronger, in the sense of providing more details about the measurements, but this is not possible. It is easy to check that every pair of MUBs is capable of producing

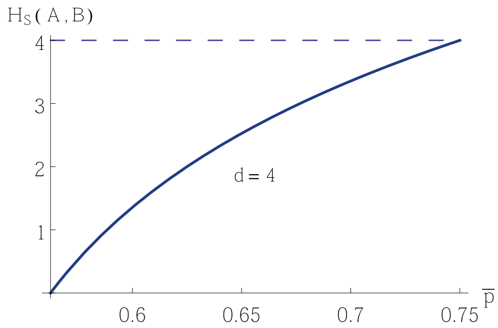


FIG. 2. Lower bound on the overlap entropy for $\bar{p} \in [\frac{1}{2} + \frac{1}{2d\sqrt{d}}, \bar{p}_Q]$ in dimension 4.

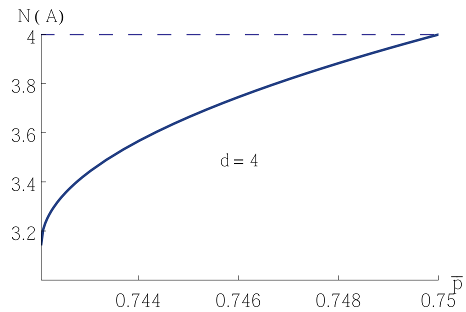


FIG. 3. Lower bound on the sum of the norms for $\bar{p} \in (\bar{p}_0, \bar{p}_Q]$ in dimension 4.

the optimal ASP. This ideal self-test is crucial for the success of the methods described in Ref. [26], as there it is essential that the optimal QRAC ASP can only be obtained with an arbitrary pair of MUBs.

IV. ROBUST SELF-TEST

Since in a real experiment one never observes the optimal performance, the ideal self-testing result is not sufficient. Instead, we need a *robust* self-testing statement, which tells us what can be certified in the case of sub-optimal performance.

Inequality Eq. (4) implies that observing the optimal ASP forces the distribution $\{t_{ij}\}_{ij}$ to be uniform. For suboptimal performance we immediately get a bound on the $\frac{1}{2}$-Rényi entropy ($H_{\frac{1}{2}}(\{q_i\}) = 2\log_2[\sum_i \sqrt{q_i}]$) of the distribution $\{t_{ij}/d\}_{ij}$, which we call the *overlap entropy* $H_S(A,B) := H_{\frac{1}{2}}(\{t_{ij}/d\}_{ij})$. More concretely, from Eq. (4) we deduce that

$$H_S(A,B) \geqslant 2\log_2[d\sqrt{d}(2\bar{p}-1)]. \tag{5}$$

This bound is nontrivial as long as $\bar{p} > \frac{1}{2} + \frac{1}{2d\sqrt{d}}$ and observing $\bar{p} = \bar{p}_Q$ implies $H_S(A,B) = \log_2(d^2)$, which is the maximal value of the overlap entropy for a pair of POVMs. For $d = 4$ the lower bound is plotted in Fig. 2.

Looking at the overlap entropy is not sufficient, because the maximal value can be achieved by measurements which are not MUBs, for instance, the trivial measurements corresponding to $A_i = B_j = \mathbb{I}/d$. The missing part is an argument showing that the measurements are close to being rank-1 projective. For a d-outcome measurement $\{A_i\}_i$ acting on $\mathbb{C}^d$ this property can be assessed by looking at the sum of the norms, $N(A) := \sum_i \|A_i\|$, since for all measurements $N(A) \leqslant d$ and the maximal value is attained if and only if the measurement is rank-1 projective. Therefore, saturating $N(A) = N(B) = d$ and $H_S(A,B) = \log_2(d^2)$ certifies the MUB arrangement.

To obtain a bound on $N(A)$ we need a stronger version of Eq. (4). In Appendix B we show that

$$\bar{p} \leqslant \frac{1}{2} + \frac{1}{2d^2}\sum_{ij}[s_{ij} - (2-\sqrt{2})s_{ij}n_{ij}], \tag{6}$$

where $n_{ij} := 1 - \frac{1}{2}(\|A_i\| + \|B_j\|)$ and $s_{ij} := \|\sqrt{A_i}\sqrt{B_j}\|$. This bound reduces to Eq. (4) if we omit the negative term and bound s_{ij} by $\sqrt{t_{ij}}$, which constitutes an alternative derivation of Theorem 1 (as $n_{ij} = 0$ for all i, j implies that both measurements are rank-1 projective).

The important feature of Eq. (6) is that it allows us to lower bound the sum of the norms. In Appendix B we show that for $\bar{p} > \bar{p}_0 := \frac{1}{2} + \frac{1}{2d^2}\sqrt{(d^2-1)d}$ we have

$$N(A) \geqslant d - \frac{2+\sqrt{2}}{d}[1 - \sqrt{d^3(2\bar{p}-1)^2 - (d^2-1)}], \tag{7}$$

and by symmetry the same bound holds for $N(B)$. It is easy to check that for $\bar{p} = \bar{p}_Q$, the right-hand side evaluates to d, i.e., the optimal performance certifies that both measurements are rank-1 projective. The lower bound given in Eq. (7) is plotted for $d = 4$ in Fig. 3.

Since Eqs. (7) and (5) allow us to robustly certify the two defining properties of MUBs (rank-1 projectivity and uniformity of overlaps, respectively), combining them yields a robust self-test for MUBs. Note that the robustness is limited by Eq. (7) which requires that $\bar{p} > \bar{p}_0$.

V. OPERATIONAL BOUNDS

In the previous paragraph we have focused on quantities tailored to certify closeness to the MUB arrangement. Let us now show that a similar approach can be used to derive bounds on quantities which have an immediate operational meaning.

We begin with the incompatibility robustness. We say that two POVMs $\{A_i\}_i$ and $\{B_j\}_j$ are compatible (or jointly measurable) if there exists a parent POVM $\{M_{ij}\}_{ij}$, such that $\sum_j M_{ij} = A_i$ and $\sum_i M_{ij} = B_j$ for all i, j. Otherwise, they are incompatible, which is often taken as the definition of nonclassicality. To quantify incompatibility beyond this binary characterization, the notion of incompatibility robustness has been introduced [66]. Consider the noisy POVMs, $A_i^\eta = \eta A_i + (1-\eta)\,\mathrm{tr}\,A_i\,\mathbb{I}/d$, and similarly B_j^η. The incompatibility robustness η^* of A and B is defined as the largest η such that $\{A_i^\eta\}_i$ and $\{B_j^\eta\}_j$ are compatible. According to this measure MUBs are highly incompatible, but, perhaps surprisingly, they are not the most incompatible among rank-1 projective measurements in dimension d [69].

Recently an analytic upper bound on η^* has been derived for an arbitrary set of POVMs [70]. For a pair of POVMs the

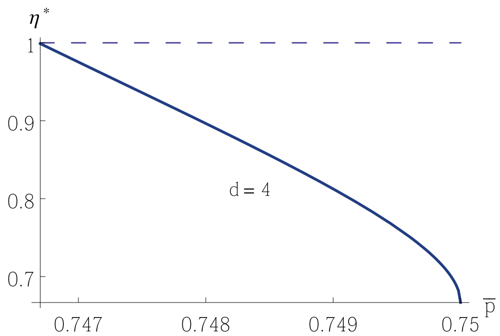


FIG. 4. Upper bound on the incompatibility robustness over the nontrivial region in dimension 4.

bound reads

$$\eta^* \leqslant \frac{d^2 \max_{ij} \|A_i + B_j\| - \sum_i (\operatorname{tr} A_i)^2 - \sum_j (\operatorname{tr} B_j)^2}{d \sum_i \operatorname{tr} A_i^2 + d \sum_j \operatorname{tr} B_j^2 - \sum_i (\operatorname{tr} A_i)^2 - \sum_j (\operatorname{tr} B_j)^2}. \tag{8}$$

All the quantities appearing in this expression can be bounded using the previously developed methods, which leads to a bound which depends only on the observed performance $\bar{p}$. Since the final bound is rather complex, we do not present it here and refer the interested reader to Appendix C. The important feature of the bound is that for the optimal performance $\bar{p} = \bar{p}_Q$ we recover the correct value of the incompatibility robustness for a pair of MUBs, i.e., $\eta^* = \frac{\sqrt{d}/2+1}{\sqrt{d}+1}$. In Fig. 4 we plot the bound for $d = 4$ over the region where it is nontrivial.

We note here that similar bounds can be derived for other measures of incompatibility robustness using the same techniques. Among these is a measure that uses arbitrary POVMs as noise [71], for which MUBs are the most incompatible pair of POVMs (of any number of outcomes) in dimension d [72]. This can also be certified by observing $\bar{p} = \bar{p}_Q$.

The second operational quantity we consider is the amount of randomness produced by the uncharacterized measurements. For a POVM A, let $H(A)_\rho := H(\{\operatorname{tr}(A_i\rho)\}_i)$ be the Shannon entropy of the outcome statistics of A on the state ρ. Maassen and Uffink derived a state-independent lower bound on $H(A)_\rho + H(B)_\rho$ for rank-1 projective measurements [7]. For our purposes we need a more general statement which covers nonprojective measurements. Such a bound has been derived in Ref. [73] and reads

$$H(A)_\rho + H(B)_\rho \geqslant -\log_2 c, \tag{9}$$

where $c := \max_{ij} \|\sqrt{A_i}\sqrt{B_j}\|^2$. Therefore, we need an upper bound on s_{ij} and such a bound has already been derived in Appendix B. The final statement reads

$$\begin{aligned} &H(A)_\rho + H(B)_\rho \\ &\geqslant -2\log_2\left(2\bar{p} - 1 + \frac{1}{d}\sqrt{d(d^2-1)[1-d(2\bar{p}-1)^2]}\right). \end{aligned} \tag{10}$$

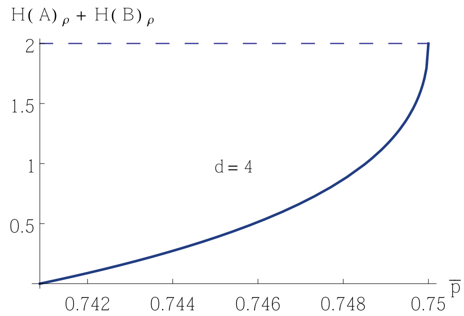


FIG. 5. Lower bound on the entropic uncertainty over the nontrivial region in dimension 4.

The optimal performance certifies $\log_2 d$ bits of randomness, which is the maximal value for a pair of projective measurements. We plot the above bound for $d = 4$ over the region where it is nontrivial in Fig. 5.

We note that a similar bound can be derived for the one-shot analog of the Shannon entropy, the min-entropy $H_{\min}$ (which coincides with the ∞-Rényi entropy), which is often preferred in cryptographic scenarios. It was shown in Ref. [74] that for a pair of POVMs, $H_{\min}(A)_\rho + H_{\min}(B)_\rho \geqslant -\log_2(\frac{1+\sqrt{c}}{2})$, for which we can derive a similar bound to that of Eq. (10).

VI. SUMMARY AND OUTLOOK

We have shown that the $2^d \to 1$ QRAC constitutes a robust self-test for MUBs in arbitrary dimension. Observing sufficiently high ASP allows us to deduce that the employed measurements are close to being rank-1 projective and that their overlaps are close to being uniform. The same approach can be used to bound operationally relevant quantities like the incompatibility robustness or the amount of randomness produced. For low dimensions the robustness of our bounds makes them interesting from the experimental point of view.

The most obvious direction for further research is to use our self-testing results to prove SDI security of prepare-and-measure quantum key distribution using high-dimensional systems. One of the main components of the SDI security proof given in Ref. [63] is the relation between the observed QRAC performance and the randomness produced for $d = 2$ (qubits). In this work we derive precisely such relations for arbitrary d and we believe that one can use them directly in security proofs.

There is an important difference between SDI self-testing and DI self-testing. In the usual DI self-testing we certify systems up to local isometries and extra degrees of freedom. Since the second equivalence is not relevant in the SDI setup (the dimension of the system is fixed), one might expect that SDI self-testing should characterize the measurements up to a unitary transformation. However, this is generally not the case: While in some dimensions all pairs of MUBs are equivalent up to unitaries (and possibly complex conjugation), e.g., $d = 2, 3, 5$, there are dimensions where this is not the case, e.g., $d = 4$ [11]. It is natural to ask whether these inequivalent classes of MUBs can be distinguished by considering more complex QRACs. In fact, a related version of this question appears readily if we consider $n^d \to 1$ QRACs with $n > 2$. In this case it is known that different classes of n-tuples of MUBs perform differently [26,75]. Numerical evidence for $n = 3$ and low d suggests that the optimal performance is achieved by one of these classes, so one might conjecture that such QRACs self-test this particular class. Again, it is not clear how to certify the remaining classes.

The $2^d \to 1$ QRAC analyzed in this paper is closely related, at least in spirit, to the family of Bell inequalities proposed by Bechmann-Pasquinucci and Gisin [76]. We hope that the understanding gained in this work will help us to prove self-testing statements for those inequalities. It would be particularly interesting to see whether the need for "more-than-unitary" freedom can also appear in the standard nonlocality-based self-testing.

ACKNOWLEDGMENTS

We thank Michał Oszmaniec for fruitful discussions. M.F. acknowledges support from the Polish NCN Grant Sonata No. UMO-2014/14/E/ST2/00020. J.K. acknowledges support under POLONEZ programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 665778.

APPENDIX A: IDEAL SELF-TEST

In the main text, we establish that the QRAC ASP can be upper bounded by

$$\bar{p} \leqslant \frac{1}{2d^2} \sum_{ij} \|A_i + B_j\|, \tag{A1}$$

and this can always be saturated by suitable states ρ_{ij} on the preparation side. To bound the above quantity, we use a special case of a matrix norm inequality derived by Kittaneh [77], applied to the square-root function and the operator norm. For further purposes, we briefly reproduce the proof here as well. We will make use of the fact that for operators A, B on a Hilbert space, $\|A \oplus B\| = \max\{\|A\|, \|B\|\}$ [78].

Theorem 2. Let $A, B \geqslant 0$ be operators on a Hilbert space. Then $\|A + B\| \leqslant \max\{\|A\|, \|B\|\} + \|\sqrt{A}\sqrt{B}\|$.

Proof. Consider the block-operator

$$X = \begin{pmatrix} \sqrt{A} \\ \sqrt{B} \end{pmatrix}, \quad \text{and thus } X^\dagger X = A + B. \tag{A2}$$

Therefore,

$$\begin{aligned} \|A + B\| &= \|X^\dagger X\| = \|XX^\dagger\| = \left\| \begin{pmatrix} A & \sqrt{A}\sqrt{B} \\ \sqrt{B}\sqrt{A} & B \end{pmatrix} \right\| \\ &= \left\| \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} + \begin{pmatrix} 0 & \sqrt{A}\sqrt{B} \\ \sqrt{B}\sqrt{A} & 0 \end{pmatrix} \right\| \\ &\leqslant \left\| \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \right\| + \left\| \begin{pmatrix} 0 & \sqrt{A}\sqrt{B} \\ \sqrt{B}\sqrt{A} & 0 \end{pmatrix} \right\| \\ &= \max\{\|A\|, \|B\|\} + \|\sqrt{A}\sqrt{B}\|, \end{aligned} \tag{A3}$$

where we used some basic properties of the operator norm (see, e.g., Ref. [78]; or Ref. [77] for a more detailed and general version of the proof).

Using the above theorem, we get

$$\bar{p} \leqslant \frac{1}{2d^2} \sum_{ij} (\max\{\|A_i\|, \|B_j\|\} + \|\sqrt{A_i}\sqrt{B_j}\|). \tag{A4}$$

From $\sum_i A_i = \sum_j B_j = \mathbb{I}$ it follows that $A_i, B_j \leqslant \mathbb{I}$, and thus $\|A_i\|, \|B_j\| \leqslant 1$. Then

$$\bar{p} \leqslant \frac{1}{2d^2} \sum_{ij} (1 + \|\sqrt{A_i}\sqrt{B_j}\|) = \frac{1}{2} + \frac{1}{2d^2} \sum_{ij} \|\sqrt{A_i}\sqrt{B_j}\|. \tag{A5}$$

Now we use the fact that for any operator O, $\|O\| \leqslant \|O\|_F$, where $\|O\|_F := \sqrt{\operatorname{tr}(O^\dagger O)}$ is the Frobenius norm [78]. Therefore,

$$\bar{p} \leqslant \frac{1}{2} + \frac{1}{2d^2} \sum_{ij} \|\sqrt{A_i}\sqrt{B_j}\|_F = \frac{1}{2} + \frac{1}{2d^2} \sum_{ij} \sqrt{\operatorname{tr}(A_i B_j)}. \tag{A6}$$

Recall that $t_{ij} := \operatorname{tr}(A_i B_j)$ and, therefore, $t_{ij} \geqslant 0$ and $\sum_{ij} t_{ij} = d$. The right-hand side of Eq. (A6) is a symmetric and strictly concave function of the t_{ij}, and as such, it is strictly Schur-concave (see, e.g., Ref. [68]). Therefore, it is maximized *uniquely* by setting all the t_{ij} uniform, $t_{ij} = \frac{1}{d}$ for all $i, j \in [d]$. The upper bound on the ASP set by such t_{ij} is then

$$\bar{p} \leqslant \frac{1}{2} + \frac{1}{2d^2} \sum_{ij} \frac{1}{\sqrt{d}} = \frac{1}{2}\left(1 + \frac{1}{\sqrt{d}}\right). \tag{A7}$$

Note that this bound is saturated by measuring in MUBs (see also Ref. [26]).

Now, let us turn our attention to necessary conditions for saturating the above bound. We first show that at least one of the measurements must be rank-1 projective to reach the optimal ASP. Saturating the upper bound requires $\operatorname{tr}(A_i B_j) = \frac{1}{d}$ for all $i, j \in [d]$ and by summing over one of the indices, we see that $\operatorname{tr} A_i = \operatorname{tr} B_j = 1$ for all i, j. Investigating the chain of inequalities obtained above, it is necessary for optimality that $\max\{\|A_i\|, \|B_j\|\} = 1$ for all $i, j \in [d]$; otherwise, $\bar{p} < \frac{1}{2d^2} \sum_{ij} (1 + \|\sqrt{A_i}\sqrt{B_j}\|) \leqslant \frac{1}{2}(1 + \frac{1}{\sqrt{d}})$. Assume that there exists a j^* such that $\|B_{j^*}\| < 1$. Then to fulfill $\max\{\|A_i\|, \|B_{j^*}\|\} = 1$ for all $i \in [d]$, it is necessary that $\|A_i\| = 1$ for all $i \in [d]$. Since these operators must all be trace-1 and positive semidefinite, it follows that $A_i = |a_i\rangle\langle a_i|$ for all $i \in [d]$. If there is no such j^*, then $\|B_j\| = 1$ for all $j \in [d]$, and we arrive at an analogous condition for B_j. Thus, without loss of generality we can assume that $A_i = |a_i\rangle\langle a_i|$ for all $i \in [d]$.

The rest of this Appendix is dedicated to showing that the other measurement must also be rank-1 projective. Let us analyze the inequality derived by Kittaneh and to do so, we first recall a few definitions from matrix analysis. We denote by $\mathcal{L}(\mathcal{H})$ the algebra of linear operators on the Hilbert space $\mathcal{H}$, and by $\|.\|_{\mathcal{H}}$ the Hilbert space norm. The numerical range of an operator O is $W(O) := \{\langle x|Ox\rangle | \|x\|_{\mathcal{H}} = 1\}$, while the numerical radius is $w(O) := \sup_{\|x\|_{\mathcal{H}}=1} |\langle x|Ox\rangle|$.

By construction every complex number $c \in W(O)$ satisfies $|c| \leqslant w(O)$ and we always have $w(O) \leqslant \|O\|$ [78].

In Theorem 2, the inequality comes from the triangle inequality and to investigate when this holds as an equality we use a result by Barraa and Boumazgour [79].

Theorem 3. Let $S, T \in \mathcal{L}(\mathcal{H})$ be nonzero. Then the equation $\|S + T\| = \|S\| + \|T\|$ holds if and only if $\|S\|\|T\| \in \overline{W(S^\dagger T)}$.

For a finite-dimensional Hilbert space the numerical range is always closed [78], thus in our case the closure in the theorem is redundant. It is immediate to see that a necessary condition for the operators S and T to saturate the triangle inequality is that $\|S\|\|T\| \leqslant w(S^\dagger T)$. However, from the submultiplicativity of the operator norm, we know that $w(S^\dagger T) \leqslant \|S^\dagger T\| \leqslant \|S^\dagger\|\|T\| = \|S\|\|T\|$, and hence this condition is equivalent to $w(S^\dagger T) = \|S\|\|T\|$.

We will also use the following bound on the numerical radius, obtained by Kittaneh [80].

Theorem 4. If $O \in \mathcal{L}(\mathcal{H})$, then

$$[w(O)]^2 \leqslant \tfrac{1}{2}\|O^\dagger O + OO^\dagger\|. \tag{A8}$$

We are now ready to derive a necessary condition to saturate Kittaneh's inequality in Theorem 2.

Lemma 5. Let $A, B \geqslant 0$ be operators on a Hilbert space. Then, the equality $\|A + B\| = \max\{\|A\|, \|B\|\} + \|\sqrt{A}\sqrt{B}\|$ holds only if $\|A\| = \|B\|$.

Proof. Let us denote the block-operators appearing in the proof of Theorem 2 by

$$S = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = S^\dagger, \quad T = \begin{pmatrix} 0 & \sqrt{A}\sqrt{B} \\ \sqrt{B}\sqrt{A} & 0 \end{pmatrix} = T^\dagger. \tag{A9}$$

Then, following from Theorem 3 and the discussion below it, a necessary condition for $A, B \geqslant 0$ to saturate Kittaneh's inequality is that $w(ST) = \|S\|\|T\| = \max\{\|A\|, \|B\|\}\|\sqrt{A}\sqrt{B}\|$.

Applying Theorem 4 to ST, we get that

$$\begin{aligned} (ST)^\dagger ST &= \begin{pmatrix} \sqrt{A}B^3\sqrt{A} & 0 \\ 0 & \sqrt{B}A^3\sqrt{B} \end{pmatrix}, \\ ST(ST)^\dagger &= \begin{pmatrix} A^{\frac{3}{2}}BA^{\frac{3}{2}} & 0 \\ 0 & B^{\frac{3}{2}}AB^{\frac{3}{2}} \end{pmatrix}, \end{aligned} \tag{A10}$$

and hence

$$\begin{aligned} (w(ST))^2 &\leqslant \frac{1}{2}\max\left\{\left\|\sqrt{A}B^3\sqrt{A} + A^{\frac{3}{2}}BA^{\frac{3}{2}}\right\|, \left\|\sqrt{B}A^3\sqrt{B} + B^{\frac{3}{2}}AB^{\frac{3}{2}}\right\|\right\} \\ &\leqslant \frac{1}{2}\max\left\{\|\sqrt{A}B^3\sqrt{A}\| + \left\|A^{\frac{3}{2}}BA^{\frac{3}{2}}\right\|, \|\sqrt{B}A^3\sqrt{B}\| + \left\|B^{\frac{3}{2}}AB^{\frac{3}{2}}\right\|\right\} \\ &= \frac{1}{2}\max\left\{\left\|\sqrt{A}B^{\frac{3}{2}}\right\|^2 + \left\|A^{\frac{3}{2}}\sqrt{B}\right\|^2, \left\|A^{\frac{3}{2}}\sqrt{B}\right\|^2 + \left\|\sqrt{A}B^{\frac{3}{2}}\right\|^2\right\} \\ &= \frac{1}{2}\left(\left\|A^{\frac{3}{2}}\sqrt{B}\right\|^2 + \left\|\sqrt{A}B^{\frac{3}{2}}\right\|^2\right) \leqslant \frac{1}{2}(\|A\|^2 + \|B\|^2)\|\sqrt{A}\sqrt{B}\|^2 \\ &\leqslant \max\{\|A\|^2, \|B\|^2\}\|\sqrt{A}\sqrt{B}\|^2. \end{aligned} \tag{A11}$$

Here, in the second line, we used the triangle inequality, in the third line the identity $\|O\|^2 = \|O^\dagger O\|$ and in the fourth line submultiplicativity. The last inequality is trivial, and is only saturated if $\|A\| = \|B\|$. Therefore, $\|A + B\| = \max\{\|A\|, \|B\|\} + \|\sqrt{A}\sqrt{B}\|$ only if $\|A\| = \|B\|$.

This lemma shows that saturating the upper bound on the ASP implies that $\|B_j\| = \|A_i\| = 1$ for all $i, j \in [d]$. It was also necessary that $\operatorname{tr} B_j = 1$, and therefore (similarly to the A_i), $B_j = |b_j\rangle\langle b_j|$ for all $j \in [d]$, and both measurements must be rank-1 projective. From here, it follows immediately from the condition $\operatorname{tr}(A_i B_j) = \frac{1}{d}$, that the bases defining the measurements must be mutually unbiased.

APPENDIX B: ROBUST SELF-TEST

While it is clear what it means for two measurements to be *exactly* mutually unbiased, there are multiple ways of turning this definition into an approximate statement (particularly if we allow for nonprojective measurements). For our purposes it is natural to split the definition of MUBs into two standalone conditions and consider them separately.

The first condition, which is usually implicit in the definition of MUBs, is that both measurements are projective and that the measurement operators are rank-1. Let $\{A_i\}_i$ be a d-outcome measurement on a d-dimensional system and let us consider the sum of the norms, $N(A) := \sum_i \|A_i\|$. This is a suitable quantity, because

$$N(A) = \sum_i \|A_i\| \leqslant \sum_i \operatorname{tr} A_i = d,$$

and since $\|A_i\| \leqslant 1$, the maximum is achieved iff every measurement operator is a rank-1 projector. Therefore, the difference between $\sum_i \|A_i\|$ and the maximal value d tells us how much $\{A_i\}_i$ deviates from being rank-1 projective.

The second condition, often referred to as *the* MUB condition, requires that the overlap between every pair of measurement operators is the same. The question here is how to generalize the overlap to nonprojective measurements. The quantity $\sqrt{\operatorname{tr}(A_i B_j)}$ discussed in the main text is a valid generalisation of the overlap in the sense that it reduces to the overlap for rank-1 projective measurements. However, the argument given below naturally leads to a different quantity, namely $\|\sqrt{A_i}\sqrt{B_j}\|$. Note that this is a commonly used definition of the overlap, e.g., in the context of uncertainty relations.

The main purpose of this Appendix is to derive a lower bound on $N(A)$ as a function of the observed performance.

However, to do that, we must first derive explicit bounds on the range of $\|\sqrt{A_i}\sqrt{B_j}\|$.

In our argument we use the following technical lemma.

Lemma 6. The function

$$h(x, y) := x + y - \alpha xy - \sqrt{x^2 + y^2}$$

for $\alpha := 2 - \sqrt{2}$ satisfies $h(x, y) \geqslant 0$ for $x, y \in [0, 1]$.

Proof. If we express x and y in terms of the polar coordinates

$$x = r\cos(\theta - \pi/4),$$
$$y = r\sin(\theta - \pi/4),$$

the function becomes

$$\begin{aligned} h(r, \theta) &= r[\cos(\theta - \pi/4) + \sin(\theta - \pi/4) - 1] \\ &\quad - \frac{\alpha r^2}{2}\sin[2(\theta - \pi/4)] = r(\sqrt{2}\sin\theta - 1) \\ &\quad + \frac{\alpha r^2}{2}\cos 2\theta. \end{aligned}$$

To cover the square $x, y \in [0, 1]$ we prove the statement for $r \in [0, \sqrt{2}]$ and $\theta \in [\pi/4, 3\pi/4]$. For fixed θ the function $h(r, \theta)$ is a quadratic function of r and the coefficient of the quadratic term is nonpositive. This means that to determine the minimum value, it suffices to consider the extreme points, i.e., $r = 0$ and $r = \sqrt{2}$. Since $h(0, \theta) = 0$, we only have to look at the latter. We have

$$\begin{aligned} h(\sqrt{2}, \theta) &= 2\sin\theta - \sqrt{2} + \alpha\cos 2\theta \\ &= -2\alpha\sin^2\theta + 2\sin\theta + 2 - 2\sqrt{2} \\ &= 2\alpha(1 - \sin\theta)\left(\sin\theta - \frac{1}{\sqrt{2}}\right), \end{aligned}$$

and it is easy to see that for $\theta \in [\pi/4, 3\pi/4]$ each term is nonnegative.

Moreover, we use the following operator norm inequality derived by Kittaneh [81].

Theorem 7. For positive semidefinite operators A and B acting on a finite-dimensional Hilbert space we have

$$\|A + B\| \leqslant \frac{1}{2}\left(\|A\| + \|B\| + \sqrt{(\|A\| - \|B\|)^2 + 4\|\sqrt{A}\sqrt{B}\|^2}\right). \tag{B1}$$

In our argument A and B will be particular measurement operators from the two measurements. We define the *generalized overlap* between A_i and B_j as

$$s_{ij} := \|\sqrt{A_i}\sqrt{B_j}\| \in [0, 1].$$

Another relevant quantity of a pair of measurement operators is the *norm deficiency* defined as

$$n_{ij} := 1 - (\|A_i\| + \|B_j\|)/2 \in [0, 1].$$

It is easy to see that if $n_{ij} = 0$ for all i, j, we have

$$\sum_i \|A_i\| = \sum_j \|B_j\| = d,$$

i.e., both measurements are rank-1 projective. Our goal now is to relate the right-hand side of Eq. (B1) to s_{ij} and n_{ij}. First, note that

$$\begin{aligned} \|A_i\| - \|B_j\| &= 2\|A_i\| - (\|A_i\| + \|B_j\|) \\ &\leqslant 2 - 2(1 - n_{ij}) = 2n_{ij} \end{aligned}$$

and similarly

$$\|B_j\| - \|A_i\| \leqslant 2n_{ij}.$$

These two inequalities imply that

$$(\|A_i\| - \|B_j\|)^2 \leqslant 4n_{ij}^2,$$

and plugging this back into Eq. (B1) gives

$$\|A_i + B_j\| \leqslant 1 - n_{ij} + \sqrt{n_{ij}^2 + s_{ij}^2}.$$

Applying the inequality derived in Lemma 6 to s_{ij} and n_{ij} gives

$$\|A_i + B_j\| \leqslant 1 + s_{ij} - \alpha s_{ij} n_{ij},$$

where $\alpha = 2 - \sqrt{2}$. Applying this upper bound to Eq. (A1) immediately yields

$$\begin{aligned} \bar{p} &\leqslant \frac{1}{2d^2}\sum_{ij}(1 + s_{ij} - \alpha s_{ij} n_{ij}) \\ &= \frac{1}{2} + \frac{1}{2d^2}\sum_{ij} s_{ij} - \frac{\alpha}{2d^2}\sum_{ij} s_{ij} n_{ij}. \end{aligned} \tag{B2}$$

Let us first bound the range of s_{ij}, i.e., find explicit functions of $\bar{p}$ denoted by $s_{\min}$ and $s_{\max}$ such that

$$s_{ij} \in [s_{\min}, s_{\max}]$$

for all i, j. To do this we drop the last term in Eq. (B2) to obtain

$$\bar{p} \leqslant \frac{1}{2} + \frac{1}{2d^2}\sum_{ij} s_{ij}.$$

To bound the sum of s_{ij} we bound the operator norm by the Frobenius norm:

$$s_{ij} = \|\sqrt{A_i}\sqrt{B_j}\| \leqslant \|\sqrt{A_i}\sqrt{B_j}\|_F = \sqrt{\operatorname{tr}(A_i B_j)} = \sqrt{t_{ij}}$$

and finally use the normalization condition $\sum_{ij} t_{ij} = d$. Let us now separate one term from the rest of the sum. For simplicity we choose the first term, i.e., s_{11}, but by symmetry the same argument applies to every s_{ij}. We obtain

$$\begin{aligned} \bar{p} &\leqslant \frac{1}{2} + \frac{1}{2d^2}\left(s_{11} + \sum_{ij\neq 11} s_{ij}\right) \\ &\leqslant \frac{1}{2} + \frac{1}{2d^2}\left(s_{11} + \sum_{ij\neq 11}\sqrt{t_{ij}}\right). \end{aligned} \tag{B3}$$

Since the remaining sum contains $d^2 - 1$ terms, concavity of the square root implies that

$$\sum_{ij\neq 11}\frac{1}{d^2 - 1}\sqrt{t_{ij}} \leqslant \sqrt{\frac{\sum_{ij\neq 11} t_{ij}}{d^2 - 1}} = \sqrt{\frac{d - t_{11}}{d^2 - 1}} \leqslant \sqrt{\frac{d - s_{11}^2}{d^2 - 1}},$$

where in the last step we used the fact that $s_{11} \leqslant \sqrt{t_{11}}$. Plugging this bound into Eq. (B3) gives

$$\bar{p} \leqslant \frac{1}{2} + \frac{1}{2d^2}\big[s_{11} + \sqrt{(d^2-1)\big(d - s_{11}^2\big)}\big] =: f(s_{11}).$$

Computing the derivative of f shows that f is increasing for $s_{11} < 1/\sqrt{d}$ and decreasing for $s_{11} > 1/\sqrt{d}$. The maximum achieved for $s_{11} = 1/\sqrt{d}$ corresponds to the optimal ASP. This implies that the lowest and highest values of s_{11} compatible with the observed $\bar{p}$ can be determined by computing the two solutions of the equality

$$\bar{p} = \frac{1}{2} + \frac{1}{2d^2}\big[s_{11} + \sqrt{(d^2-1)\big(d - s_{11}^2\big)}\big].$$

This reduces to solving a quadratic equation and finally we deduce that $s_{11} \in [s_{\min}, s_{\max}]$, where

$$s_{\min} := 2\bar{p} - 1 - \frac{1}{d}\sqrt{d(d^2-1)[1 - d(2\bar{p}-1)^2]}, \quad \text{(B4)}$$

$$s_{\max} := 2\bar{p} - 1 + \frac{1}{d}\sqrt{d(d^2-1)[1 - d(2\bar{p}-1)^2]}. \quad \text{(B5)}$$

The optimal performance, i.e. $\bar{p} = \frac{1}{2} + \frac{1}{2\sqrt{d}}$, implies that $s_{\min} = s_{\max} = \frac{1}{\sqrt{d}}$. Moreover, since both functions are continuous in $\bar{p}$, for sufficiently good performance we obtain bounds stronger than the trivial $s_{11} \geqslant 0$ and $s_{11} \leqslant 1$. This concludes the first part of the argument, i.e., providing explicit bounds on the range of the generalized overlaps.

For the second part of the argument, in which we show that the measurements are close to being rank-1 projective, we need all the overlaps to be bounded away from 0, i.e., $s_{\min} > 0$. According to Eq. (B4) this is guaranteed as long as $\bar{p} > \bar{p}_0$ for

$$\bar{p}_0 := \frac{1}{2} + \frac{1}{2d^2}\sqrt{(d^2-1)d}.$$

Using the concavity result while keeping the negative term in Eq. (B2) leads to

$$\bar{p} \leqslant \frac{1}{2} + \frac{1}{2d^2}\big(s_{11} + \sqrt{(d^2-1)\big(d - s_{11}^2\big)}\big) - \frac{\alpha}{2d^2}\sum_{ij} s_{ij} n_{ij}.$$

Without loss of generality we can assume that s_{11} is the smallest overlap and then

$$\bar{p} \leqslant \frac{1}{2} + \frac{1}{2d^2}\big(s_{11} + \sqrt{(d^2-1)\big(d - s_{11}^2\big)}\big) - \frac{\alpha s_{11}}{2d^2}\sum_{ij} n_{ij},$$

which is equivalent to

$$\sum_{ij} n_{ij} \leqslant \frac{1}{\alpha s_{11}}\big[s_{11} + \sqrt{(d^2-1)\big(d - s_{11}^2\big)} - d^2(2\bar{p}-1)\big]. \quad \text{(B6)}$$

To analyze the right-hand side, we define

$$g(x) := 1 + \sqrt{(d^2-1)\left(\frac{d}{x^2} - 1\right)} - \frac{d^2(2\bar{p}-1)}{x},$$

and now our goal is to maximize $g(x)$ over $x \in [0, 1/\sqrt{d}]$, as $s_{\min} \leqslant 1/\sqrt{d}$. Recall that we work under the assumption that $\bar{p} > \bar{p}_0$ and therefore $2\bar{p} - 1 > 0$. We can analytically compute the derivative dg/dx and set it to 0 to conclude that the only stationary point corresponds to

$$x^* := \frac{\sqrt{d^3(2\bar{p}-1)^2 - (d^2-1)}}{d(2\bar{p}-1)} = \sqrt{d - \frac{d^2-1}{d^2(2\bar{p}-1)^2}}.$$

Evaluating the second derivative d^2g/dx^2 at x^* tells us that this is a maximum and since this is the only stationary point, it must be the unique maximizer in the interval $[0, 1/\sqrt{d}]$. Therefore, in Eq. (B6) we can set $s_{11} = x^*$ to obtain

$$\sum_{ij} n_{ij} \leqslant \frac{1}{\alpha}\big[1 - \sqrt{d^3(2\bar{p}-1)^2 - (d^2-1)}\big].$$

Finally, we can use this bound to obtain lower bounds on the sums of the norms $\sum_i \|A_i\|$ and $\sum_j \|B_j\|$ for the individual measurements. Since

$$\sum_{ij} n_{ij} = d^2 - \frac{d}{2}\left(\sum_i \|A_i\| + \sum_j \|B_j\|\right),$$

we can use the trivial bound $N(B) = \sum_j \|B_j\| \leqslant d$ to obtain

$$\begin{aligned} N(A) = \sum_i \|A_i\| &\geqslant d - \frac{2}{d}\sum_{ij} n_{ij} \\ &\geqslant d - \frac{2}{\alpha d}\big[1 - \sqrt{d^3(2\bar{p}-1)^2 - (d^2-1)}\big]. \end{aligned} \quad \text{(B7)}$$

Clearly, the same lower bound holds for $N(B)$.

APPENDIX C: INCOMPATIBILITY ROBUSTNESS

In this Appendix we derive an analytic upper bound on the incompatibility robustness as a function of the observed ASP. We start with a bound derived recently in Ref. [70]:

$$\eta^* \leqslant \frac{d^2 \max_{ij}\|A_i + B_j\| - \sum_i (\operatorname{tr} A_i)^2 - \sum_j (\operatorname{tr} B_j)^2}{d\sum_i \operatorname{tr} A_i^2 + d\sum_j \operatorname{tr} B_j^2 - \sum_i (\operatorname{tr} A_i)^2 - \sum_j (\operatorname{tr} B_j)^2}. \quad \text{(C1)}$$

The aim is to bound all the terms appearing in this formula by quantities which we have already bounded in Appendix B.

Let us start with the numerator. The first term is easy to bound since

$$\|A_i + B_j\| \leqslant 1 + s_{ij},$$

and $\max_{ij} s_{ij} \leqslant s_{\max}$ given in Eq. (B5).

To bound the second term we use the fact that for positive semidefinite operators $(\operatorname{tr} A)^2 \geqslant \operatorname{tr} A^2$ and then bound the Frobenius norm by the operator norm:

$$(\operatorname{tr} A_i)^2 \geqslant \operatorname{tr} A_i^2 = \|A_i\|_F^2 \geqslant \|A_i\|^2.$$

To bound the sum of the squares $\sum_i \|A_i\|^2$ we use a standard inequality for vector p-norms which for d-dimensional vectors reads $\|x\|_2 \geqslant \frac{1}{\sqrt{d}}\|x\|_1$. Applying this to the real vector whose components are given by $x_i = \|A_i\|$ yields

$$\sum_i \|A_i\|^2 \geqslant \frac{1}{d}\left(\sum_i \|A_i\|\right)^2.$$

Putting the two inequalities together gives

$$\sum_i (\operatorname{tr} A_i)^2 \geqslant \frac{1}{d}\left(\sum_i \|A_i\|\right)^2,$$

which can be bounded using Eq. (B7).

The first term in the denominator we have already bounded: From the previous argument we see that

$$\sum_i \operatorname{tr} A_i^2 \geqslant \frac{1}{d}\left(\sum_i \|A_i\|\right)^2.$$

Bounding the last term turns out to be slightly more involved, so we state it as a separate lemma.

Lemma 8. Let $\{A_i\}_i$ be a d-outcome measurement acting on $\mathbb{C}^d$. If

$$\sum_i \|A_i\| \geqslant q,$$

then

$$\sum_i (\operatorname{tr} A_i)^2 \leqslant d + (d-q)(d-q+1).$$

Proof. Before proceeding to the technical details, let us briefly explain the idea behind the proof. Suppose we are given a partition of the d measurement outcomes into two disjoint sets. Moreover, we are promised that the trace of the measurement operators corresponding to the outcomes in the first (second) set belongs to the interval $[0, 1]$ ($[1, d]$). It turns out that an upper bound on the desired quantity can be derived in terms of simple properties of this partition. Maximising this bound over all valid partitions leads to the main result of the lemma.

Formally, we are given two sets X and Y such that $X \cup Y = [d]$ and $X \cap Y = \emptyset$. Moreover, we have

$$i \in X \Rightarrow \operatorname{tr} A_i \in [0, 1],$$
$$i \in Y \Rightarrow \operatorname{tr} A_i \in [1, d].$$

Define $n := |X|$, $\gamma := \sum_{i\in X} \operatorname{tr} A_i$ and clearly

$$n - \gamma \geqslant 0. \tag{C2}$$

Moreover, the assumption of the lemma implies

$$\begin{aligned} q \leqslant \sum_i \|A_i\| = \sum_{i\in X} \|A_i\| + \sum_{i\in Y} \|A_i\| \leqslant \sum_{i\in X} \operatorname{tr} A_i + |Y| \\ = \gamma + d - n, \end{aligned}$$

and therefore

$$n - \gamma \leqslant d - q. \tag{C3}$$

For the rest of the argument let us think of n and γ as some fixed values. Once we derive the final upper bound in terms of these two variables, we will maximize it over the allowed pairs of n and γ.

For $i \in X$ we have $(\operatorname{tr} A_i)^2 \leqslant \operatorname{tr} A_i$, and therefore

$$\sum_{i\in X} (\operatorname{tr} A_i)^2 \leqslant \sum_{i\in X} \operatorname{tr} A_i = \gamma.$$

To bound the second term we must explicitly determine the allowed combinations of $\{\operatorname{tr} A_i\}_{i\in Y}$. Since $\{\operatorname{tr} A_i\}_{i\in Y} \in [1, d]^{|Y|}$ and

$$\sum_{i\in Y} \operatorname{tr} A_i = d - \gamma,$$

the valid choices of $\{\operatorname{tr} A_i\}_{i\in Y}$ form a polytope. It is easy to see that all the vertices of this polytope correspond to setting $|Y| - 1$ values to 1 and the last value to $[d - \gamma - (|Y| - 1)]$. Since $\sum_{i\in Y} (\operatorname{tr} A_i)^2$ is a convex function of the traces, the maximal value is achieved at a vertex, and therefore

$$\sum_{i\in Y} (\operatorname{tr} A_i)^2 \leqslant (|Y| - 1) + [d - \gamma - (|Y| - 1)]^2.$$

Plugging in $|Y| = d - n$ gives

$$\begin{aligned} \sum_{i\in Y} (\operatorname{tr} A_i)^2 &\leqslant d - n - 1 + (n - \gamma + 1)^2 \\ &= d + (n-\gamma)(n-\gamma+1) - \gamma. \end{aligned}$$

Putting the two bounds together leads to

$$\begin{aligned} \sum_i (\operatorname{tr} A_i)^2 &= \sum_{i\in X} (\operatorname{tr} A_i)^2 + \sum_{i\in Y} (\operatorname{tr} A_i)^2 \\ &\leqslant d + (n-\gamma)(n-\gamma+1). \end{aligned}$$

Now we must maximize the right-hand side subject to the constraints given in Eqs. (C2) and (C3). The maximum is achieved when the latter is saturated, which leads to the final result of the lemma.

The final bound reads

$$\eta^* \leqslant \frac{\frac{1}{2}d^2(1 + s_{\max}) - \frac{q^2}{d}}{q^2 - d - (d-q)(d-q+1)}, \tag{C4}$$

where $s_{\max}$ is the quantity defined in Eq. (B5), while q is the right-hand side of Eq. (B7).

[1] I. D. Ivonovic, J. Phys. A: Math. Gen. **14**, 3241 (1981).
[2] W. K. Wootters and B. D. Fields, Ann. Phys. **191**, 363 (1989).
[3] M. A. Ballester and S. Wehner, Phys. Rev. A **75**, 022319 (2007).
[4] D. P. DiVincenzo, M. Horodecki, D. W. Leung, J. A. Smolin, and B. M. Terhal, Phys. Rev. Lett. **92**, 067902 (2004).
[5] P. K. Aravind, Zeitschrift für Naturforschung A **58**, 85 (2003).
[6] B.-G. Englert and Y. Aharonov, Phys. Lett. A **284**, 1 (2001).
[7] H. Maassen and J. B. M. Uffink, Phys. Rev. Lett. **60**, 1103 (1988).
[8] S. Wehner and A. Winter, New J. Phys. **12**, 025009 (2010).
[9] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner, Rev. Mod. Phys. **89**, 015002 (2017).
[10] T. Durt, B. Englert, I. Bengtsson, and K. Życzkowski, Int. J. Quantum Inf. **8**, 535 (2010).

[11] S. Brierley, S. Weigert, and I. Bengtsson, Quantum Inf. Comput. **10**, 803 (2010).
[12] G. Zauner, Int. J. Quantum Inf. **09**, 445 (2011).
[13] M. Grassl, arXiv:quant-ph/0406175 (2004).
[14] P. Jaming, M. Matolcsi, P. Móra, F. Szöllősi, and M. Weiner, J. Phys. A: Math. Theor. **42**, 245305 (2009).
[15] F. Szöllősi, J. London Math. Soc. **85**, 616 (2012).
[16] S. Brierley and S. Weigert, J. Phys.: Conf. Ser. **254**, 012008 (2010).
[17] P. Butterley and W. Hall, Phys. Lett. A **369**, 5 (2007).
[18] M. Ozols, Quantum random access codes with shared randomness, Master's thesis, University of Waterloo (2009).
[19] A. Tavakoli, A. Hameedi, B. Marques, and M. Bourennane, Phys. Rev. Lett. **114**, 170502 (2015).
[20] A. Ambainis, D. Kravchenko, and A. Rai, arXiv:1510.03045 (2015).
[21] S. Wiesner, SIGACT News **15**, 78 (1983).
[22] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, J. ACM **49**, 496 (2002).
[23] E. F. Galvão, Foundations of quantum theory and quantum information applications, Ph.D. thesis, University of Oxford (2002).
[24] M. Hayashi, K. Iwama, H. Nishimura, R. Raymond, and S. Yamashita, in *Proceedings of the 2006 IEEE International Symposium on Information Theory* (IEEE, Piscataway, NJ, 2006), pp. 446–450.
[25] I. Kerenidis, Quantum encodings and applications to locally decodable codes and communication complexity, Ph.D. thesis, University of California at Berkeley (2004).
[26] E. A. Aguilar, J. J. Borkała, P. Mironowicz, and M. Pawłowski, Phys. Rev. Lett. **121**, 050501 (2018).
[27] J. S. Bell, in *John S. Bell on the Foundations of Quantum Mechanics* (World Scientific, Singapore, 2001), pp. 74–83.
[28] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[29] B. S. Tsirelson, J. Sov. Math. **36**, 557 (1987).
[30] S. J. Summers and R. Werner, J. Math. Phys. **28**, 2440 (1987).
[31] S. Popescu and D. Rohrlich, Phys. Lett. A **169**, 411 (1992).
[32] B. S. Tsirelson, Hadron. J. Suppl. **8**, 329 (1993).
[33] D. Mayers and A. Yao, in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science, FOCS'98* (IEEE Computer Society, Washington, DC, 1998), p. 503.
[34] D. Mayers and A. Yao, Quantum Inf. Comput. **4**, 273 (2004).
[35] F. Magniez, D. Mayers, M. Mosca, and H. Ollivier, in *Automata, Languages, and Programming* (Springer, Berlin, 2006), pp. 72–83.
[36] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).
[37] A. Acín, N. Gisin, and L. Masanes, Phys. Rev. Lett. **97**, 120405 (2006).
[38] R. Colbeck, Quantum and relativistic protocols for secure multi-party computation, Ph.D. thesis, University of Cambridge (2006).
[39] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).
[40] S. Pironio, A. Acín, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, Nature **464**, 1021 (2010).
[41] M. McKague, in *Theory of Quantum Computation, Communication, and Cryptography* (Springer, Berlin, 2014), pp. 104–120.
[42] M. McKague, T. H. Yang, and V. Scarani, J. Phys. A: Math. Theor. **45**, 455304 (2012).
[43] T. H. Yang and M. Navascués, Phys. Rev. A **87**, 050102 (2013).
[44] C. Bamps and S. Pironio, Phys. Rev. A **91**, 052111 (2015).
[45] Y. Wang, X. Wu, and V. Scarani, New J. Phys. **18**, 025021 (2016).
[46] I. Šupić, R. Augusiak, A. Salavrakos, and A. Acín, New J. Phys. **18**, 035013 (2016).
[47] A. Coladangelo, K. T. Goh, and V. Scarani, Nat. Commun. **8**, 15485 (2017).
[48] J. Bowles, I. Šupić, D. Cavalcanti, and A. Acín, Phys. Rev. Lett. **121**, 180503 (2018).
[49] C.-E. Bardyn, T. C. H. Liew, S. Massar, M. McKague, and V. Scarani, Phys. Rev. A **80**, 062327 (2009).
[50] J.-D. Bancal, M. Navascués, V. Scarani, T. Vértesi, and T. H. Yang, Phys. Rev. A **91**, 022115 (2015).
[51] T. H. Yang, T. Vértesi, J.-D. Bancal, V. Scarani, and M. Navascués, Phys. Rev. Lett. **113**, 040401 (2014).
[52] K. F. Pál, T. Vértesi, and M. Navascués, Phys. Rev. A **90**, 042340 (2014).
[53] X. Wu, Y. Cai, T. H. Yang, H. N. Le, J.-D. Bancal, and V. Scarani, Phys. Rev. A **90**, 042339 (2014).
[54] J. Kaniewski, Phys. Rev. Lett. **117**, 070402 (2016).
[55] J. Kaniewski, Phys. Rev. A **95**, 062323 (2017).
[56] T. R. Tan, Y. Wan, S. Erickson, P. Bierhorst, D. Kienzler, S. Glancy, E. Knill, D. Leibfried, and D. J. Wineland, Phys. Rev. Lett. **118**, 130403 (2017).
[57] A. Tavakoli, J. Kaniewski, T. Vértesi, D. Rosset, and N. Brunner, Phys. Rev. A **98**, 062307 (2018).
[58] C. H. Bennett and G. Brassard, Theor. Comput. Sci. **560**, 7 (2014).
[59] C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992).
[60] R. Gallego, N. Brunner, C. Hadley, and A. Acín, Phys. Rev. Lett. **105**, 230501 (2010).
[61] M. Hendrych, R. Gallego, M. Mičuda, N. Brunner, A. Acín, and J. P. Torres, Nat. Phys. **8**, 588 (2012).
[62] J. Ahrens, P. Badziag, A. Cabello, and M. Bourennane, Nat. Phys. **8**, 592 (2012).
[63] M. Pawłowski and N. Brunner, Phys. Rev. A **84**, 010302 (2011).
[64] H.-W. Li, M. Pawłowski, Z.-Q. Yin, G.-C. Guo, and Z.-F. Han, Phys. Rev. A **85**, 052308 (2012).
[65] T. Lunghi, J. B. Brask, C. C. W. Lim, Q. Lavigne, J. Bowles, A. Martin, H. Zbinden, and N. Brunner, Phys. Rev. Lett. **114**, 150501 (2015).
[66] T. Heinosaari, J. Kiukas, and D. Reitzner, Phys. Rev. A **92**, 022115 (2015).
[67] M. Oszmaniec, L. Guerini, P. Wittek, and A. Acín, Phys. Rev. Lett. **119**, 190501 (2017).
[68] A. Marshall, I. Olkin, and B. Arnold, *Inequalities: Theory of Majorization and Its Applications*, Springer Series in Statistics (Springer, New York, 2010).
[69] J. Bavaresco, M. T. Quintino, L. Guerini, T. O. Maciel, D. Cavalcanti, and M. T. Cunha, Phys. Rev. A **96**, 022110 (2017).
[70] S. Designolle, P. Skrzypczyk, F. Fröwis, and N. Brunner, Phys. Rev. Lett. **122**, 050402 (2019).
[71] E. Haapasalo, J. Phys. A: Math. Theore. **48**, 255303 (2015).
[72] S. Designolle, M. Farkas, and J. Kaniewski (unpublished).
[73] M. Krishna and K. R. Parthasarathy, Sankhya: Indian J. Stat. **64**, 842 (2002).

[74] M. Tomamichel, S. Fehr, J. Kaniewski, and S. Wehner, New J. Phys. **15**, 103002 (2013).
[75] M. Farkas, arXiv:1706.04446 (2017).
[76] H. Bechmann-Pasquinucci and N. Gisin, Quantum Inf. Comput. **3**, 157 (2003).
[77] F. Kittaneh, J. Funct. Anal. **143**, 337 (1997).
[78] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics (Springer, New York, 1996).
[79] M. Barraa and M. Boumazgour, Proc. Am. Math. Soc. **130**, 471 (2002).
[80] F. Kittaneh, Studia Mathematica **168**, 73 (2005).
[81] F. Kittaneh, J. Operator Theory **48**, 95 (2002).

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG | IOP Institute of Physics

PAPER • OPEN ACCESS

Incompatibility robustness of quantum measurements: a unified framework

To cite this article: Sébastien Designolle *et al* 2019 *New J. Phys.* **21** 113053

New Journal of Physics
The open access journal at the forefront of physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
21 June 2019

REVISED
16 September 2019

ACCEPTED FOR PUBLICATION
22 October 2019

PUBLISHED
26 November 2019

PAPER

Incompatibility robustness of quantum measurements: a unified framework

Sébastien Designolle[1,5,6], Máté Farkas[2,5] and Jędrzej Kaniewski[3,4]

1 Group of Applied Physics, University of Geneva, 1211 Geneva, Switzerland
2 Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics and Informatics, University of Gdansk, 80-952 Gdansk, Poland
3 Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland
4 Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland
5 These authors contributed equally to this work.
6 Author to whom any correspondence should be addressed.

E-mail: sebastien.designolle@unige.ch, mate.frks@gmail.com and jkaniewski@fuw.edu.pl

Keywords: joint measurability, incompatibility robustness, mutually unbiased bases

Abstract

In quantum mechanics performing a measurement is an invasive process which generally disturbs the system. Due to this phenomenon, there exist incompatible quantum measurements, i.e. measurements that cannot be simultaneously performed on a single copy of the system. It is then natural to ask what the most incompatible quantum measurements are. To answer this question, several measures have been proposed to quantify how incompatible a set of measurements is, however their properties are not well-understood. In this work, we develop a general framework that encompasses all the commonly used measures of incompatibility based on robustness to noise. Moreover, we propose several conditions that a measure of incompatibility should satisfy, and investigate whether the existing measures comply with them. We find that some of the widely used measures do not fulfil these basic requirements. We also show that when looking for the most incompatible pairs of measurements, we obtain different answers depending on the exact measure. For one of the measures, we analytically prove that projective measurements onto two mutually unbiased bases are among the most incompatible pairs in every dimension. However, for some of the remaining measures we find that some peculiar measurements turn out to be even more incompatible.

1. Introduction

It is well-known that the concept of a measurement in quantum physics challenges our everyday intuition. In a classical theory objects have properties, whether we look at them or not, and a measurement simply reveals to us their pre-existing values. In quantum mechanics, on the other hand, performing a measurement is an invasive process, which necessarily disturbs the state (except for some special cases). Moreover, even if we have complete knowledge about the system, we can only predict the probabilities of different outcomes, which can be computed using the Born rule. An intriguing consequence of the quantum formalism is the existence of measurements that are *incompatible*, i.e. that cannot be measured simultaneously given only one copy of the system. The best known example consists of the position and momentum of a quantum mechanical particle, which cannot be measured simultaneously with arbitrary precision.

In this work we study the incompatibility of measurements with a finite number of outcomes. These measurements assign to each physical state ρ a discrete probability distribution $\{p_a(\rho)\}_a$, whose elements we interpret as the probability of outcome a on the state ρ. We say that two measurements are *compatible* (or *jointly measurable*) if there exists a single measurement, referred to as the *parent measurement*, that is able to universally replace the two [1, 2]. More specifically, on any state the outcome probabilities of both measurements can be recovered from the outcome probabilities of the parent measurement. Therefore, the two measurements can be performed simultaneously by performing the parent measurement. If such a parent measurement does not exist, we say that the measurements are *incompatible* (or *not jointly measurable*). We remark here that other notions of compatibility, such as commutativity, non-disturbance and coexistence, are also used in the literature [1, 3]; let us for completeness briefly explain how they are related. Commutativity of a measurement pair implies non-disturbance, which in turn implies joint measurability, which then implies coexistence. Moreover, it is known that none of the converse implications hold in general, therefore these notions are strictly distinct [4]. In this work we focus solely on the notion of joint measurability, because the existence (or not) of a parent measurement has a clear operational meaning. Therefore, throughout the present paper we use the terms '(in) compatibility' and '(non-)joint measurability' interchangeably. It is important to notice that whenever two measurements are compatible, they cannot be used to produce quantum advantage in tasks like Bell nonlocality [5] or Einstein–Podolsky–Rosen steering [6, 7]. Moreover, it was recently shown that joint measurability is equivalent to a specific notion of classicality, namely, preparation non-contextuality [8, 9]. Hence, one may think of compatible measurements as 'classical' , and incompatible measurements as a resource for the above tasks. Therefore, it is of fundamental importance to characterise and understand the structure of incompatible measurements.

What is particularly important is to go beyond the dichotomy of compatible and incompatible measurements, and quantify *to what extent* a pair of measurements is incompatible. A natural framework for this quantification, often used in the literature, is to define measures based on robustness to noise. Briefly speaking, robustness-based measures of incompatibility quantify the minimal amount of noise that needs to be added to a pair of measurements to make them compatible. The more noise is required, the more incompatible the measurements are. Note that measures of this type are directly relevant to experiments, because in real-world implementations measurements are always noisy, due to inevitable experimental imperfections.

Robustness-based measures are also natural measures of incompatibility in the context of resource theories [10, 11]. Here one considers a set of 'free' objects (compatible measurements) and quantify the usefulness of 'resource' objects (incompatible measurements) by so-called resource monotones. While in this work we do not develop a full resource theory of incompatibility, we note that robustness-based measures are good candidates for resource monotones if they satisfy certain natural properties [12–15]. In resource theories one defines 'free operations' that do not create resource (that is, do not map compatible measurements to incompatible ones). Properly defined resource monotones should then be monotonic under such free operations. Once measures with the desired properties are found, the question 'what are the most incompatible pairs of measurements?' is well-defined with respect to each of these measures.

Several robustness-based measures have been proposed in the literature (see [16] for an introduction), the essential difference between them being the assumed noise model. Nevertheless, some basic properties of these measures have not been determined and little effort has been dedicated to understanding the similarities and differences among them. In this work we make the following contributions to fill this gap.

- We develop a framework in which a robustness-based measure can be defined with respect to an arbitrary noise model. We identify the minimum assumptions on the noise model that ensure that the resulting measure satisfies some basic requirements, i.e. we provide an explicit connection between the properties of the noise model and the desired properties of the measure.
- We apply our framework to study five measures already introduced in the literature in a unified fashion. By giving explicit counterexamples we show that some widely used measures do not satisfy certain natural properties motivated by resource theories.
- We show that when looking for the most incompatible pairs, we obtain different answers depending on the specific measure of incompatibility. For one of the measures we analytically prove that mutually unbiased bases (MUBs) are among the most incompatible pairs of measurements in every dimension. For three other measures we can explicitly show that, for dimensions larger than two, MUBs are *not* among the most incompatible pairs. Our study for the last measure is inconclusive.

In section 2 we define incompatibility robustness in a fashion that is independent of the specific noise model, introduce the natural properties that the measures should desirably satisfy and relate them to the properties of the noise model, formulate the notion of most incompatible measurement pairs, and discuss the measures' semidefinite program (SDP) formulation and how to use this formulation to derive bounds on them. Then in section 3 we introduce the five measures already used in the literature, illustrate them on a simple example, analyse their relevant properties, and derive new bounds on each of them. At the end of this section we discuss the relations between the measures, apply our results to compute all the different measures for MUBs, then summarise the main results in a compact form. In section 4 we address the question of the most incompatible pairs of measurements under the five measures. Finally, in section 5 we summarise the new findings and pose some important open questions arising from our work.

We note here that the notion of incompatibility naturally generalises to more than two measurements, but for simplicity in the main text we restrict ourselves to pairs of measurements. For a formal treatment of larger sets of measurements, and results regarding them, we refer the interested reader to appendix E.

2. Definitions and basic properties

In this section we formalise the main definitions and concepts outlined in the introduction. We give a mathematically precise definition of (in)compatibility and of robustness-based measures for an arbitrary noise model. Then we specify a few natural properties the measures should satisfy, and give concrete conditions on the noise model under which these are automatically fulfilled. We also rigorously formulate the notion of 'most incompatible measurements', and discuss how to efficiently search for them. Finally, we introduce the notion of SDP, and how to use it to derive bounds on robustness-based measures.

2.1. Incompatible measurements

Throughout this paper we analyse the most general model of quantum measurements, positive operator valued measures (POVMs). For this model, we establish that the physical system lives on a d-dimensional Hilbert space, $\mathcal{H} \simeq \mathbb{C}^d$. The relevant objects are all elements of the set of linear operators on this space, $\mathcal{B}(\mathbb{C}^d)$. The state of the system is described by a positive semidefinite operator with unit trace, denoted by ρ. A POVM with n outcomes is a set of n positive semidefinite operators, $\{A_a\}_{a=1}^n$, such that $\sum_{a=1}^n A_a = \mathbb{1}$, where $\mathbb{1}$ is the identity operator. The probability of observing outcome a is given by the Born rule, $p_a(\rho) = \mathrm{tr}(A_a \rho)$. In the following, we will use the terms 'measurement' and 'POVM' interchangeably.

We will often refer to the following three important classes of POVMs. *Rank-one* POVMs are measurements whose elements are rank-one operators, $A_a \propto |\varphi_a\rangle\langle\varphi_a|$, where $|\varphi_a\rangle\langle\varphi_a|$ is the projector onto $|\varphi_a\rangle \in \mathbb{C}^d$. Note that such measurements cannot have fewer elements than the dimension of the Hilbert space, that is, $n \geqslant d$ with the above notation. *Projective* measurements are POVMs whose elements are projectors. Note that such measurements cannot have more non-zero elements than the dimension of the Hilbert space. Since the set of measurements with n outcomes acting on dimension d is a convex set, we will talk about *extremal* POVMs (in the convex geometry sense). Recall that every POVM can be written as a convex combination of extremal POVMs and these have been extensively studied in [17].

The ability to recover the outcome probabilities of two POVMs on any state from the statistics of a single measurement is referred to as *joint measurability* and can be formulated in the following way.

Definition 1. Given two POVMs, $\{A_a\}_{a=1}^{n_A}$ and $\{B_b\}_{b=1}^{n_B}$, we say that they are *jointly measurable* (or *compatible*) if there exists a POVM $\{G_{ab}\}_{a=1,b=1}^{n_A,n_B}$ such that $\sum_{b=1}^{n_B} G_{ab} = A_a$ for all a, and $\sum_{a=1}^{n_A} G_{ab} = B_b$ for all b. We call such a POVM a *parent* measurement of $\{A_a\}_{a=1}^{n_A}$ and $\{B_b\}_{b=1}^{n_B}$.

This definition captures the idea that the parent measurement provides a joint outcome distribution of the two initial measurements on every state. It is worth pointing out that the notion of joint measurability in which the parent POVM is allowed an arbitrary (finite) outcome set and arbitrary classical post-processing turns out to be equivalent to the one above (see e.g. [16], section 3.1).

We note that a parent POVM is not necessarily unique for a fixed pair of measurements [18, 19]. It is clear that in order to recover the outcome probabilities of A and B, one only needs to measure G and add up the relevant probabilities (in the following we sometimes drop the outcome indices to refer to the POVMs, when it does not lead to confusion; this notation is to be understood as $A = \{A_a\}_{a=1}^{n_A}$). A simple example of a jointly measurable pair is the trivial measurement pair, $\left\{\frac{1}{n_A}\right\}_{a=1}^{n_A}$ and $\left\{\frac{1}{n_B}\right\}_{b=1}^{n_B}$ with the parent POVM $\left\{\frac{1}{n_A n_B}\right\}_{a=1,b=1}^{n_A,n_B}$. In fact any POVM pair with pairwise commuting measurement operators, $[A_a, B_b] = 0$ for all a and b, is jointly measurable. This can be seen by employing the parent POVM G with elements $G_{ab} = A_a B_b$, which is guaranteed to be positive in this case. Note that commutativity becomes necessary and sufficient if one of the two measurements is projective, see [18], proposition 8 for a proof.

If a parent POVM does not exist, we say that A and B are *not jointly measurable* (or *incompatible*). A standard example of incompatible d-outcome measurement pairs in dimension $d \geqslant 2$ is a pair of projective measurements onto two MUBs [20]. These consist of rank-one projectors $A^{\mathrm{MUB}} = \{|\varphi_a\rangle\langle\varphi_a|\}_{a=1}^{d}$ and $B^{\mathrm{MUB}} = \{|\psi_b\rangle\langle\psi_b|\}_{b=1}^{d}$ onto the orthonormal bases $\{|\varphi_a\rangle\}_{a=1}^{d}$ and $\{|\psi_b\rangle\}_{b=1}^{d}$, such that all the pairwise overlaps (moduli of inner products) are uniform: $|\langle\varphi_a|\psi_b\rangle| = 1/\sqrt{d}$ for all a, b. As these measurements are projective and non-commuting, they are incompatible.

In the following we will denote the set of POVM pairs with outcome numbers n_A and n_B in dimension d by $\mathbf{POVM}_d^{n_A,n_B}$, and its elements by (A, B). Note that POVM pairs inherit the convex structure of POVMs (denoted by $\mathbf{POVM}_d^n$), therefore convex combinations of them are well-defined. For the subset corresponding to jointly measurable pairs, we will use the notation $\mathbf{JM}_d^{n_A,n_B}$, but drop the indices whenever it does not lead to confusion. Note that the set $\mathbf{JM}_d^{n_A,n_B}$ is a convex subset of $\mathbf{POVM}_d^{n_A,n_B}$: it is straightforward to verify that if $(A^0, B^0) \in \mathbf{JM}_d^{n_A,n_B}$ with parent POVM G^0, and $(A^1, B^1) \in \mathbf{JM}_d^{n_A,n_B}$ with parent POVM G^1, then $(1-p)(A^0, B^0) + p(A^1, B^1) \in \mathbf{JM}_d^{n_A,n_B}$ with parent POVM $(1-p)G^0 + pG^1$ for all $p \in [0, 1]$. That is, taking convex combinations preserves joint measurability.

2.2. Incompatibility robustness

In order to talk about noisy measurements, we define what we mean by a *noise model*.

Definition 2. A *noise model* $\mathbf{N}$ is a map $\mathbf{N}\colon \mathbf{POVM}_d^n \to \mathbb{P}(\mathbf{POVM}_d^n)$, where $\mathbb{P}$ is the set of all subsets, that maps every POVM $A \in \mathbf{POVM}_d^n$ to a subset of all n-outcome POVMs in dimension d, that is, $\mathbf{N}\colon A \mapsto \mathbf{N}_A \subseteq \mathbf{POVM}_d^n$. We will refer to $\mathbf{N}_A$ as the *noise set* of A under this noise model.

Given a noise model, we can define *noisy versions* of POVMs as convex combinations of POVMs with elements of their corresponding noise sets. Specifically, if $M \in \mathbf{N}_A$ and $\eta \in [0, 1]$, then a noisy version of A with *visibility* η is the POVM

$$\eta A + (1-\eta)M \in \mathbf{POVM}_d^n. \tag{1}$$

Noise models will be crucial for our analysis, as different noise models give rise to different measures of incompatibility. Initially, for a unified treatment of robustness based measures, we will discuss properties that do not depend on the precise choice of the noise model, and only introduce explicit choices in section 3, where we analyse the five specific measures.

In order to apply it to incompatibility, we extend the concept of a noise model to pairs of measurements: in this case, the noise model $\mathbf{N}$ is a map $\mathbf{N}\colon \mathbf{POVM}_d^{n_A,n_B} \to \mathbb{P}(\mathbf{POVM}_d^{n_A,n_B})$ that maps every pair $(A, B) \in \mathbf{POVM}_d^{n_A,n_B}$ to its corresponding noise set, $\mathbf{N}\colon (A, B) \mapsto \mathbf{N}_{A,B} \subseteq \mathbf{POVM}_d^{n_A,n_B}$. Note that the set $\mathbf{N}_{A,B}$ may actually depend on the measurements A and B, and not simply on their dimension or number of outcomes (whenever the map $\mathbf{N}$ is not constant). The simplest example of a noise model is $\mathbf{N}_{A,B} = \left\{\left(\{\frac{1}{n_A}\}_{a=1}^{n_A}, \{\frac{1}{n_B}\}_{b=1}^{n_B}\right)\right\}$, that maps every POVM pair to the one-element set containing only the trivial measurement pair. On the other end of the spectrum, the largest possible choice of the noise model is $\mathbf{N}_{A,B} = \mathbf{POVM}_d^{n_A,n_B}$, mapping every POVM pair to the set of all POVM pairs.

We will now define a measure of incompatibility corresponding to an arbitrary noise model. To ensure that the measure is well-defined, we require that the map $\mathbf{N}$ is such that for every pair (A, B) the noise set $\mathbf{N}_{A,B}$ contains at least one jointly measurable pair. For any such noise model, one can define an incompatibility robustness measure for pairs of POVMs, i.e. the maximal visibility at which the noisy pair is still compatible.

Definition 3. Given two POVMs, $\{A_a\}_{a=1}^{n_A}$ and $\{B_b\}_{b=1}^{n_B}$ on $\mathbb{C}^d$, and a noise model $\mathbf{N}$, we say that the *incompatibility robustness* $\eta_{A,B}^*$ of the pair (A, B) with respect to this noise model is

$$\eta_{A,B}^* = \sup_{\substack{\eta\in[0,1]\\(M,N)\in\mathbf{N}_{A,B}}} \{\eta \mid \eta\cdot(A, B) + (1-\eta)\cdot(M, N) \in \mathbf{JM}_d^{n_A,n_B}\}. \tag{2}$$

This definition has a clear geometric interpretation, see figure 1. Note that regardless of the choice of the noise model, $\eta_{A,B}^* = 1$ if and only if A and B are jointly measurable, and that under this definition the lower $\eta_{A,B}^*$ is, the more incompatible the measurements are.

There are several other requirements one might impose on the noise model. Let us briefly discuss some of these and explain what their consequences are.

- If we assume that for every pair (A, B), the noise set $\mathbf{N}_{A,B}$ is *closed*, we are guaranteed that the supremum is achieved, i.e. there exists an optimal noise pair. In this case the supremum in equation (2) can be replaced by a maximum. Note that since we are dealing with finite-dimensional objects, it is irrelevant which topology we choose to define the notion of closedness.
- If we assume that for every pair (A, B), the noise set $\mathbf{N}_{A,B}$ is *convex*, we are guaranteed to find a decomposition of the form given in equation (2) for any $\eta \in [0, \eta_{A,B}^*)$. It suffices to find a noise pair (M', N') and a visibility

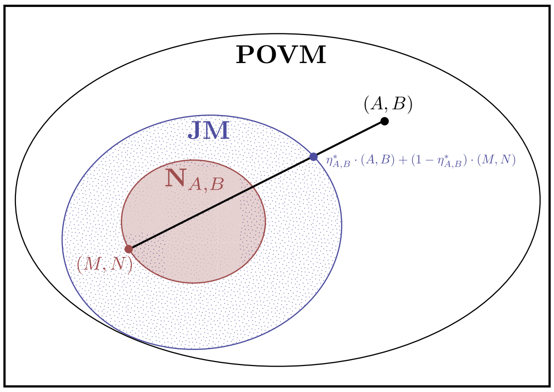


Figure 1. Schematic representation of a generic incompatibility robustness measure for a noise model which maps to closed and convex sets. Note that in general the noise set $\mathbf{N}_{A,B}$ need not be contained in the jointly measurable set **JM**. One can also easily infer that the optimal noise pair (M, N) must lie on the boundary of $\mathbf{N}_{A,B}$ and that the optimal noisy pair $\eta^*_{A,B} \cdot (A, B) + (1 - \eta^*_{A,B}) \cdot (M, N)$ must lie on the boundary of **JM**.

$\eta' \geqslant \eta$ such that

$$\eta' \cdot (A, B) + (1 - \eta') \cdot (M', N') \in \mathbf{JM}. \tag{3}$$

Such (M', N') and η' are guaranteed to exist, since $\eta < \eta^*_{A,B}$. Then pick $(M^{\mathbf{JM}}, N^{\mathbf{JM}}) \in \mathbf{N}_{A,B}$ such that

$$(M^{\mathbf{JM}}, N^{\mathbf{JM}}) \in \mathbf{JM}, \tag{4}$$

which is again guaranteed to exist by our fundamental assumption on the noise model. From the convexity of **JM** it follows that taking the convex combination of equation (3) with weight η/η' and equation (4) with weight $(1 - \eta/\eta')$ leads to $\eta \cdot (A, B) + (1 - \eta) \cdot (M, N) \in \mathbf{JM}$, where

$$(M, N) = \frac{\eta}{1-\eta} \cdot \frac{1-\eta'}{\eta'} \cdot (M', N') + \left(1 - \frac{\eta}{1-\eta} \cdot \frac{1-\eta'}{\eta'}\right) \cdot (M^{\mathbf{JM}}, N^{\mathbf{JM}}), \tag{5}$$

and the convexity of $\mathbf{N}_{A,B}$ ensures that $(M, N) \in \mathbf{N}_{A,B}$. Note that a looser constraint, namely that $\mathbf{N}_{A,B}$ is a radial set at $(M^{\mathbf{JM}}, N^{\mathbf{JM}})$ (the line segments between $(M^{\mathbf{JM}}, N^{\mathbf{JM}})$ and all other elements of $\mathbf{N}_{A,B}$ are contained in $\mathbf{N}_{A,B}$) is sufficient for this property.

- Another property one might require from the noise set is *covariance with respect to unitaries*. Intuitively, this means that if two pairs of measurements are related by a unitary, then so should be their respective noise sets. More specifically, if (A, B) and (A', B') satisfy

$$A'_a = UA_aU^\dagger \quad \text{and} \quad B'_b = UB_bU^\dagger \tag{6}$$

for all outcomes *a* and *b* and for some fixed unitary *U*, then

$$(M, N) \in \mathbf{N}_{A,B} \Longleftrightarrow (UMU^\dagger, UNU^\dagger) \in \mathbf{N}_{A',B'}. \tag{7}$$

This property is sufficient to ensure that the resulting incompatibility robustness measure is unitarily invariant, i.e. $\eta^*_{A,B} = \eta^*_{A',B'}$.

- Finally, one might require that for every choice of (A, B) the corresponding noise set $\mathbf{N}_{A,B}$ is *invariant under unitaries*, i.e.

$$(M, N) \in \mathbf{N}_{A,B} \Longrightarrow (UMU^\dagger, UNU^\dagger) \in \mathbf{N}_{A,B} \tag{8}$$

for every unitary *U*. An advantage of this property is that if we assume that the noise set is convex, then we can average over the Haar measure on unitary matrices, which leads to a noise pair whose every element is proportional to the identity operator. We will use this property in section 2.4 to derive non-trivial lower bounds on the resulting incompatibility measure.

The last two properties are clearly related. Indeed, if the noise set does not depend on the pair (A, B) beyond the dimension and the outcome numbers (the map $\mathbf{N}$ is constant), they turn out to be equivalent. However, in full generality these two properties are independent, i.e. we can have one without the other. To conclude let us simply state that *all the measures considered in this work satisfy all the requirements stated above*.

In section 3, we will replace the star in $\eta^*_{A,B}$ with a reference to the specific noise model in order to make clear which measure we use. In general we are looking for noise models that give rise to measures of incompatibility that satisfy certain natural properties motivated by resource theories.

2.3. Monotonicity

The natural properties we consider capture the intuition that measures of incompatibility should not decrease under operations that do not create incompatibility. In other words, measurements should not become more incompatible under such operations. This is well-motivated from the resource theoretic point of view, allowing for a partial order of measurement pairs based on their incompatibility robustness.

Consider an operation $\Phi: (A, B) \mapsto \Phi(A, B)$, that maps every POVM pair to another POVM pair, not necessarily preserving the dimension or the outcome numbers. We say that this operation is *joint measurability-preserving* if for all $(A, B) \in \mathbf{JM}$ we have that $\Phi(A, B) \in \mathbf{JM}$. It is desirable that our measures are non-decreasing under such operations, that is, $\eta^*_{\Phi(A,B)} \geqslant \eta^*_{A,B}$ for every joint measurability-preserving operation Φ. If this inequality holds for every (A, B) we say that η^* is *monotonic under* Φ.

Whenever the joint measurability-preserving operation Φ is linear, a simple property of the noise model $\mathbf{N}$ implies monotonicity, namely, $\Phi(\mathbf{N}_{A,B}) \subseteq \mathbf{N}_{\Phi(A,B)}$ for all (A, B). To see this, consider a measurement pair (A, B) and its corresponding noise set $\mathbf{N}_{A,B}$. Following from definition 3, we have that

$$\eta^*_{A,B} \cdot (A, B) + (1 - \eta^*_{A,B}) \cdot (M, N) \in \mathbf{JM} \tag{9}$$

for some $(M, N) \in \mathbf{N}_{A,B}$. Applying Φ to the left-hand side, we obtain

$$\eta^*_{A,B} \cdot \Phi(A, B) + (1 - \eta^*_{A,B}) \cdot \Phi(M, N) \in \mathbf{JM}, \tag{10}$$

as Φ is linear and joint measurability-preserving. Whenever $\Phi(\mathbf{N}_{A,B}) \subseteq \mathbf{N}_{\Phi(A,B)}$, the left-hand side of equation (10) is a noisy version of $\Phi(A, B)$ with visibility $\eta^*_{A,B}$, which implies that $\eta^*_{\Phi(A,B)} \geqslant \eta^*_{A,B}$. Therefore, if the image of the noise set under Φ is contained in the noise set of the image for every measurement pair, then η^* based on this noise model is monotonic under Φ. In many cases, the stronger property $\Phi(\mathbf{N}_{A,B}) = \mathbf{N}_{\Phi(A,B)}$ holds for all (A, B), and then we say that the noise model is *invariant* under Φ.

In this paper we will consider two natural classes of joint measurability-preserving operations, which are transformations of the measurement outputs and inputs. The first class acts on the outputs of the measurements and is therefore called *post-processings*. The second class, on the other hand, acts on the inputs (quantum states) of the measurements, and is accordingly called *pre-processings* (see figures 2 and 3, respectively). Post-processings amount to recording the outcome of the measurement and then applying a response function to it. It can therefore be formulated in the following way.

Definition 4. A post-processing β maps $\{A_a\}_{a=1}^{n_A}$ to $\{A^\beta_{a'}\}_{a'=1}^{n'_A}$, where

$$A^\beta_{a'} = \sum_{a=1}^{n_A} \beta(a'|a) A_a, \tag{11}$$

and $\{\beta(a'|a)\}_{a'}$ is a probability distribution for every $a \in \{1, 2,\ldots,n_A\}$.

A post-processing is called deterministic if the probability distribution $\{\beta(a'|a)\}_{a'}$ is deterministic for all $a \in \{1, 2,\ldots,n_A\}$, that is, $\beta(a'|a) \in \{0, 1\}$. If such a post-processing decreases the number of outcomes, it is referred to as *coarse-graining* or *binning*, e.g. the operation mapping the POVM $\{A_1, A_2, A_3\}$ to $\{A_1, A_2 + A_3\}$. What is important is that every POVM can be obtained by coarse-graining a rank-one POVM with potentially more outcomes.

Note that post-processings preserve the dimension but might change the outcome number. For pairs (A, B) the operation $\Phi^\beta: (A, B) \mapsto (A^{\beta_A}, B^{\beta_B})$ is joint measurability-preserving (note that the post-processings applied to A and B are independent): assume that $(A, B) \in \mathbf{JM}$ with parent POVM G. Then it is straightforward to verify that $(A^{\beta_A}, B^{\beta_B}) \in \mathbf{JM}$ with parent POVM G^β, where $G^\beta_{a'b'} = \sum_{ab} \beta_A(a'|a)\beta_B(b'|b)G_{ab}$.

The second class, pre-processings, amounts to first applying a quantum channel to the measured state and then performing the measurement. Denoting the channel acting on the state by $\Lambda^\dagger$ (the dual of the map Λ), we arrive at the following definition.

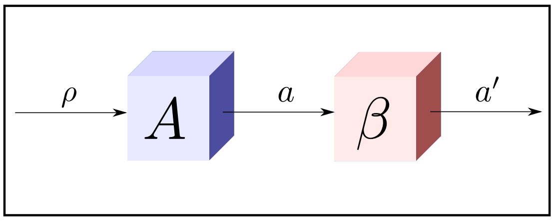


Figure 2. Schematic representation of a post-processing of a measurement.

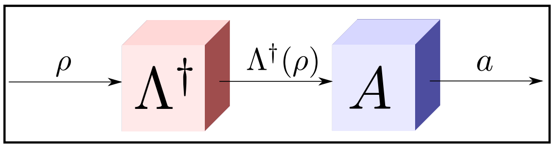


Figure 3. Schematic representation of a pre-processing of a measurement.

Definition 5. A pre-processing Λ maps $\{A_a\}_{a=1}^{n_A}$ to $\{A_a^{\Lambda}\}_{a=1}^{n_A}$, where

$$A_a^{\Lambda} = \Lambda(A_a), \tag{12}$$

and $\Lambda: \mathcal{B}(\mathbb{C}^d) \mapsto \mathcal{B}(\mathbb{C}^{d'})$ is a completely positive unital map.

Note that, for our formal treatment the unital map Λ does only need to be positive (and not necessarily completely positive), although all the positive unital maps appearing in this work are also completely positive.

A well-known example of pre-processings is the one in Naimark's dilation theorem. This states that for every POVM A on $\mathbb{C}^d$, there exists $d' \in \mathbb{N}$, an isometry $V: \mathbb{C}^d \to \mathbb{C}^{d'}$, and a projective measurement P on $\mathbb{C}^{d'}$ such that $A_a = V^{\dagger} P_a V$ for all a, that is, $A = P^{\Lambda}$, where $\Lambda(.) = V^{\dagger}(.)V$ is a (completely) positive unital map. That is, every POVM can be obtained by pre-processing a projective measurement acting on a potentially higher dimensional Hilbert space.

Note that pre-processings preserve the outcome number but might change the dimension. For pairs (A, B) the operation $\Phi^{\Lambda}: (A, B) \mapsto (A^{\Lambda}, B^{\Lambda})$ is joint measurability-preserving (in contrast to the case of post-processing, here there is just a single pre-processing applied to both A and B): assume that $(A, B) \in \mathbf{JM}$ with parent POVM G. Then it is straightforward to verify that $(A^{\Lambda}, B^{\Lambda}) \in \mathbf{JM}$ with parent POVM G^{Λ}. Note also that an incompatibility measure that is monotonic under pre-processings necessarily satisfies unitary invariance, as already mentioned in [12], section C.

Finally, let us consider another natural operation that preserves joint-measurability, although it is of a different flavour than pre- and post-processings. Namely, recall that taking convex combinations preserves joint measurability, that is, for any $(A^0, B^0) \in \mathbf{JM}$ and $(A^1, B^1) \in \mathbf{JM}$ we have that $(A^p, B^p) = (1 - p)(A^0, B^0) + p(A^1, B^1) \in \mathbf{JM}$ for all $p \in [0, 1]$ (see section 2.1). For this reason, it is desirable that our measures do not decrease under taking convex combinations, that is, $\eta^*_{A^p,B^p} \geqslant \min\{\eta^*_{A^0,B^0}, \eta^*_{A^1,B^1}\}$ for all $p \in [0, 1]$, a property sometimes referred to as *quasi-concavity*.

It is easy to see that this condition holds whenever the noise model satisfies the simple property that, using the above notation, for any $(M^0, N^0) \in \mathbf{N}_{A^0,B^0}$ and $(M^1, N^1) \in \mathbf{N}_{A^1,B^1}$, we have $(M^p, N^p) = (1 - p)(M^0, N^0) + p(M^1, N^1) \in \mathbf{N}_{A^p,B^p}$. To see this, let us define $\eta^*_{\min} = \min\{\eta^*_{A^0,B^0}, \eta^*_{A^1,B^1}\}$. From the convexity of the noise set, there exist $(M^0, N^0) \in \mathbf{N}_{A^0,B^0}$ and $(M^1, N^1) \in \mathbf{N}_{A^1,B^1}$ such that $\eta^*_{\min} \cdot (A^0, B^0) + (1 - \eta^*_{\min}) \cdot (M^0, N^0) \in \mathbf{JM}$ and $\eta^*_{\min} \cdot (A^1, B^1) + (1 - \eta^*_{\min}) \cdot (M^1, N^1) \in \mathbf{JM}$ (see section 2.2). Taking a convex combination of these two relations with coefficients $1 - p$ and p, respectively, results in $\eta^*_{\min} \cdot (A^p, B^p) + (1 - \eta^*_{\min}) \cdot (M^p, N^p) \in \mathbf{JM}$, that is, $\eta^*_{A^p,B^p} \geqslant \min\{\eta^*_{A^0,B^0}, \eta^*_{A^1,B^1}\}$. All the noise models considered in this paper satisfy the requirement stated above and therefore the corresponding measures are non-decreasing under convex combinations.

A stronger property that is often desired is *joint concavity*, which using the above notation reads $\eta^*_{A^p,B^p} \geqslant p\eta^*_{A^0,B^0} + (1 - p)\eta^*_{A^1,B^1}$ (note that throughout this paper we will write 'concavity' and 'convexity' instead of 'joint concavity' and 'joint convexity', for simplicity). However, what one naturally deduces by looking at the noise model turns out to be slightly different. More specifically, if the noise set is convex for every pair and the noise model is a constant map we may conclude that the inverse of the measure is convex, i.e. $1/\eta^*_{A^p,B^p} \leqslant (1-p)/\eta^*_{A^0,B^0} + p/\eta^*_{A^1,B^1}$, similarly to the proof in [21], proposition 2. It is easy to see that the concavity of η^* implies that $1/\eta^*$ is convex [22], equation (3.11), but the converse does not hold in general. In fact, in appendix A, using an explicit counterexample, we show that none of the measures studied in this paper are concave. It is common to use the measure $t^* = 1/\eta^* - 1$ instead of η^* because it is easy to prove its convexity, and it also has the appealing property that it vanishes for every $(A, B) \in \mathbf{JM}$ (a property referred to as *faithfulness* in [23]—also note that in [24], faithfulness, post-processing monotonicity and convexity were postulated as natural properties of any measure of incompatibility). Moreover, whenever η^* is monotonic under pre- or post-processings, then so is t^* (with opposite relation in the inequality defining monotonicity). Nevertheless, in the following we will study η^* since it suits our purposes better and it is easily interconvertible with t^*.

In section 3, we will investigate the properties introduced above for each specific measure. As all these measures are quasi-concave and none of them are concave, we will only explicitly address pre- and post-processing monotonicity of η^*, and convexity of the corresponding inverse measure, t^*.

2.4. Most incompatible measurements

For any given measure of incompatibility, one can ask what the most incompatible pairs of POVMs are. To make this question well-defined, we introduce the following quantity.

Definition 6. Given a measure of incompatibility, η^*, we define $\chi^*(d;\, n_A,\, n_B)$ to be its lowest possible value for dimension d and outcome numbers n_A and n_B.

$$\chi^*(d;\, n_A,\, n_B) = \min\,\{\eta^*_{A,B} \mid (A,\, B) \in \mathbf{POVM}_d^{n_A,n_B}\}. \tag{13}$$

The minimum in this definition is justified, as the set $\mathbf{POVM}_d^{n_A,n_B}$ is closed and bounded. For a fixed measure this definition yields a real number from the range $[0,\, 1]$ for all positive integers $d,\, n_A,\, n_B$. Sometimes, however, we might be interested in less detailed information. We might just ask the question 'what are the most incompatible measurement pairs in dimension d?', regardless of the outcome numbers, leading to the quantity

$$\chi^*(d) = \inf_{n_A,n_B} \chi^*(d;\, n_A,\, n_B), \tag{14}$$

where the infimum is taken over positive integers and it is not clear whether $\chi^*(d)$ is achieved for any finite n_A and n_B. Alternatively, we might only fix the outcome numbers, leading to $\chi^*(n_A,\, n_B)$, or fix neither the dimension nor the outcome numbers, leading to χ^*.

One might wonder whether a non-trivial lower bound on χ^* can be derived based only on the previously assumed property of the noise model, namely, that for every POVM pair the corresponding noise set contains at least one jointly measurable pair, but this turns out not to be the case. For every pair of incompatible measurements $(A,\, B)$ we can choose the noise set to contain a single jointly measurable pair with the property that the interior of the line segment connecting $(A,\, B)$ and the noise pair lies outside the jointly measurable set. Clearly, in this case $\eta^*_{A,B} = 0$ for all incompatible pairs $(A,\, B)$, and η^* defined through this construction is just the indicator function of joint measurability.

However, a mild additional assumption on the noise model allows us to get a non-trivial lower bound on χ^*. Suppose that for every incompatible pair $(A,\, B)$ there exists a valid noise pair $(M,\, N)$ such that the measurement operators of A commute with those of N and similarly the measurement operators of B commute with those of M. Then, the POVM given by

$$G_{ab} = \frac{1}{2}(A_a N_b + M_a B_b) \tag{15}$$

is a valid parent POVM for $\frac{1}{2}(A + M)$ and $\frac{1}{2}(B + N)$, therefore it ensures that $\eta^*_{A,B} \geqslant \frac{1}{2}$, and we conclude that $\chi^* \geqslant \frac{1}{2}$. Clearly, the above condition is fulfilled whenever we are guaranteed to find a noise pair where all the elements are proportional to the identity (a direct consequence of the unitary invariance property discussed in section 2.2). This is the case for all the measures that we study.

To make the search for the most incompatible pairs of measurements efficient, it is crucial to identify operations under which the measure is monotonic, as it significantly shrinks the set over which we need to optimise. Specifically, if we want to compute $\chi^*(d;\, n_A,\, n_B)$ and we deal with a measure that is non-decreasing under convex combinations, we only need to consider pairs of extremal measurements. If our goal is to compute $\chi^*(d)$, i.e. we do not care about the number of outcomes, and our measure is monotonic under post-processings, we do not need to consider measurement pairs that are post-processings of another pair. Since every POVM can be written as a post-processing (coarse-graining) of some rank-one POVM with possibly more outcomes, for post-processing monotonic measures the value $\chi^*(d)$ can be found by searching only over rank-one measurements. Similarly, if we aim to compute $\chi^*(n_A, n_B)$, i.e. we do not care about the dimension, and our measure is monotonic under pre-processings, we do not need to consider measurement pairs that are pre-processings of another pair. Due to Naimark's dilation theorem, every POVM can be obtained by pre-processing a projective measurement that possibly acts on a higher dimensional space, therefore projective measurements achieve $\chi^*(n_A, n_B)$ for pre-processing monotonic measures.

2.5. Semidefinite programming

It is clear from equation (2) that incompatibility robustness measures are defined through an optimisation problem. The class of optimisation problems that arises in our case is called SDP and can be seen as a generalisation of linear programming [22]. An SDP is an optimisation problem whose optimisation variables are matrices, and whose objective function and constraints are linear functions of these variables. The constraints can be either matrix equalities or matrix inequalities (recall that for matrices the inequality $A \geqslant B$ is equivalent to $A - B$ being a positive semidefinite matrix). For every SDP, later referred to as the *primal*, another SDP, called the *dual*, can be defined such that its solution bounds the primal one. In this paper the primal SDP is a maximisation problem and the dual SDP is a minimisation problem whose solution upper bounds the primal solution. In all the examples that we study in this work, the solutions of these two SDPs in fact coincide, as we will see in section 3.1.1. Thanks to this feature, it is possible to efficiently solve such SDPs on a computer, which gives us a tool to study incompatibility robustness measures *numerically*. This tool we often employed using the MATLAB computing environment together with the YALMIP [25], SDPT3 [26] and MOSEK [27] optimisation toolboxes. However, the main objective of our work is to study these measures *analytically*. In order to do so, we find *feasible points* for the SDPs, that is, assignments of variables that satisfy all the constraints, but that are not necessarily optimal. By finding feasible points for the primal and dual problems, we obtain lower and upper bounds, respectively, on the value of the optimisation problem. In the next two sections we introduce objects that will come in useful for finding such feasible points.

2.5.1. Lower bounds

Feasible points for the primal SDP lead to lower bounds on the incompatibility robustness. For a fixed pair (A, B) feasible points correspond to a noise pair (M, N), a visibility η, and a parent POVM G for $\eta \cdot (A, B) + (1 - \eta) \cdot (M, N)$, all of these satisfying the constraints of the SDP. That is, the noise pair should satisfy $(M, N) \in \mathbf{N}_{A,B}$, and the visibility must be in the range $\eta \in [0, 1]$. Crucially, the parent POVM G should give $\eta A + (1 - \eta)M$ and $\eta B + (1 - \eta)N$ as marginals (which also guarantees its proper normalisation), and all its measurement operators should be positive semidefinite. In order to find feasible parent POVMs satisfying these properties, we introduce an ansatz solution. This ansatz encompasses all possible choices of the parent POVM elements that are linear combinations of the elements of A and B, their square-roots, and products thereof, such that the normalisation of the parent POVM is ensured. Namely, let

$$G_{ab} \propto \{A_a, B_b\} + (\alpha_b A_a + \beta_a B_b) + \gamma_{ab}\mathbb{1} + \delta(A_a^{\frac{1}{2}} B_b A_a^{\frac{1}{2}} + B_b^{\frac{1}{2}} A_a B_b^{\frac{1}{2}}) \tag{16}$$

for some real parameters $\alpha_b, \beta_b, \gamma_{ab}$ and δ. It is clear then that $\sum_{ab} G_{ab} \propto \mathbb{1}$.

In this construction the anticommutator term plays a crucial role. When the measurement operators of the two POVMs commute, i.e. we have $A_a B_b = B_b A_a$ for all a and b, the anticommutator is guaranteed to be positive semidefinite. We can therefore set $G_{ab} = \frac{1}{2}\{A_a, B_b\}$, which is a valid parent POVM for A and B. For non-commuting measurement operators, however, the anticommutator might have some negative eigenvalues for which the remaining terms are supposed to compensate. Note that the same construction for parent POVMs has recently been used in [28].

For a pair of rank-one POVMs checking the positivity of equation (16) becomes analytically tractable: in this case we can write the operator as a direct sum of an operator acting on the two-dimensional subspace spanned by the eigenvectors of A_a and B_b, and a multiple of the identity on the orthogonal subspace (which is non-trivial for $d \geqslant 3$). This allows us to explicitly compute the eigenvalues and check positivity. For this reason, for our methods to work efficiently and provide tight bounds, it is extremely important that the measure we study is monotonic under post-processings. This is because in this case it is enough to look at rank-one POVMs in order to find the most incompatible pairs, and the robustness of any POVM pair can be bounded by the robustness of their rank-one decompositions.

Note that computing the marginals of the POVM in equation (16) is also easy in general, except for the terms multiplying the parameter δ. However, for most constructions we will choose $\delta = 0$, and only include this term in a special (albeit very important) case.

As an example, let us present a known result initially presented for pairs in [29] and then generalised to arbitrary number of measurements [16, 30, 31]. The idea is to try to perform two measurements simultaneously by duplicating the input state and then feeding each measurement with one of the copies. By virtue of the no-cloning theorem, the duplication process cannot be perfect. Thanks to a duality between noiseless measurements acting on noisy states and noisy measurements acting on noiseless states, one can obtain a parent POVM from this procedure

$$G_{ab} = \frac{1}{2(d+1)}[\{A_a, B_b\} + \mathrm{tr}(B_b)A_a + \mathrm{tr}(A_a)B_b], \tag{17}$$

which is indeed of the form (16). The positivity of G_{ab} defined in this way follows straightforwardly from the fact that $[A_a/\mathrm{tr}(A_a) + B_b/\mathrm{tr}(B_b)]^2 \geqslant 0$ (we assume that $\mathrm{tr}A_a\mathrm{tr}B_b > 0$; the other cases are trivial). This parent POVM gives rise to a universal lower bound on some measures, see equation (26).

2.5.2. Upper bounds

In order to derive upper bounds on incompatibility robustness measures, we need to find feasible points for the dual SDPs. These SDPs have a similar structure for all the different measures that we study in this work, and therefore some quantities will often appear in the upper bounds. For this reason, we define them here:

$$f = \sum_a \frac{\mathrm{tr}A_a^2}{d} + \sum_b \frac{\mathrm{tr}B_b^2}{d} \quad \text{and} \quad \lambda = \max_{a,b}\{\max \mathrm{Sp}(A_a + B_b)\}, \tag{18}$$

where $\mathrm{Sp}(M)$ is the spectrum of the operator M (note that $A_a + B_b$ is always positive semidefinite). It is easy to see that $f \leqslant 2$ and the inequality is saturated if and only if both measurements are projective. We will also need the following four quantities:

$$g^{\mathrm{d}} = \sum_a \left(\frac{\mathrm{tr}A_a}{d}\right)^2 + \sum_b \left(\frac{\mathrm{tr}B_b}{d}\right)^2, \quad g^{\mathrm{r}} = \frac{1}{n_A} + \frac{1}{n_B},$$
$$g^{\mathrm{p}} = \min_a \frac{\mathrm{tr}A_a}{d} + \min_b \frac{\mathrm{tr}B_b}{d}, \quad \text{and} \quad g^{\mathrm{jm}} = \min_{a,b}\{\min \mathrm{Sp}(A_a + B_b)\}. \tag{19}$$

Note that $g^{\mathrm{d}} = g^{\mathrm{r}} = g^{\mathrm{p}} = 2/d$ whenever both measurements are rank-one projective.

2.6. Example

We will compute all the studied incompatibility robustness measures for a pair of rank-one projective qubit measurements parametrised as

$$A_a(\theta) = \frac{1}{2}[\mathbb{1} + (-1)^a(\cos\theta\,\sigma_z + \sin\theta\,\sigma_x)] \quad \text{and} \quad B_b(\theta) = \frac{1}{2}[\mathbb{1} + (-1)^b(\cos\theta\,\sigma_z - \sin\theta\,\sigma_x)], \tag{20}$$

where σ_z and σ_x are the Pauli Z and X matrices, $\theta \in [0, \pi/4]$ and $a, b = 1, 2$. Note that we choose the angle θ to be half of the angle between the Bloch vectors of the two measurements. For this pair of rank-one projective measurements, we can compute the different parameters defined in equations (18) and (19), namely, $f = 2$, $\lambda = 1 + \cos\theta$, $g^{\mathrm{d}} = g^{\mathrm{r}} = g^{\mathrm{p}} = 1$, and $g^{\mathrm{jm}} = 1 - \cos\theta$. In the following, when discussing any measure of incompatibility for this pair, we will use η_θ^* as a shorthand for $\eta_{A(\theta),B(\theta)}^*$. We will also make use of the following compact notation to write down the primal and dual variables:

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \quad \text{and} \quad (X, Y) = \left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}\right), \tag{21}$$

where the elements G_{ab}, X_a and Y_b are 2×2 Hermitian matrices.

3. Five relevant measures

In this section we introduce five different explicit noise models, which give rise to five different robustness-based measures of incompatibility that are commonly used in the literature. For each measure we write down both the primal and the dual SDPs, analyse their desired properties, illustrate their computation on a pair of rank-one projective qubit measurements, and derive explicit lower and upper bounds on them. A compact summary of the main results can be found at the end of this section in table 1.

3.1. Incompatibility depolarising robustness

3.1.1. Definition and properties

In this case the noise model is defined by the map

$$\mathbf{N}_{A,B}^{\mathrm{d}} = \left\{\left(\left\{\mathrm{tr}(A_a)\frac{\mathbb{1}_d}{d}\right\}_{a=1}^{n_A}, \left\{\mathrm{tr}(B_b)\frac{\mathbb{1}_d}{d}\right\}_{b=1}^{n_B}\right)\right\}. \tag{22}$$

The noise set depends on the specific measurements, which makes this measure different than all the other measures considered in this work. It has been investigated in many works [12, 28, 30, 32–36], often in relation with Einstein–Podolsky–Rosen steering. This specific type of noise has also been considered in scenarios different from incompatibility [37]. The physical motivation is as follows: take a depolarising quantum channel $\Lambda_\eta^\dagger(.)$, which acts on states as $\Lambda_\eta^\dagger(\rho) = \eta\rho + (1-\eta)\mathrm{tr}(\rho)\mathbb{1}/d$, that is, by mixing them with white noise. If we measure a system that has undergone such an evolution, we obtain the outcome probabilities $p(a) = \mathrm{tr}[A_a\Lambda_\eta^\dagger(\rho)] = \mathrm{tr}[\Lambda_\eta(A_a)\rho]$, where $\Lambda_\eta(A_a) = \eta A_a + (1-\eta)\mathrm{tr}(A_a)\mathbb{1}/d$ is the dual of the depolarising channel, which leads precisely to the type of noise set defined in equation (22).

The corresponding incompatibility robustness, as introduced in definition 3, can be computed via the SDPs

$$\eta^{\mathrm{d}}_{A,B} = \begin{cases} \max\limits_{\eta,\{G_{ab}\}_{ab}} & \eta \\ \text{s.t.} & G_{ab} \geqslant 0, \quad \eta \leqslant 1 \\ & \sum\limits_b G_{ab} = \eta A_a + (1-\eta)\mathrm{tr}A_a\dfrac{\mathbb{1}}{d} \\ & \sum\limits_a G_{ab} = \eta B_b + (1-\eta)\mathrm{tr}B_b\dfrac{\mathbb{1}}{d} \end{cases} = \begin{cases} \min\limits_{\substack{\{X_a\}_a\\ \{Y_b\}_b}} & 1 + \sum\limits_a \mathrm{tr}(X_a A_a) + \sum\limits_b \mathrm{tr}(Y_b B_b) \\ \text{s.t.} & X_a = X_a^\dagger, \quad Y_b = Y_b^\dagger, \quad X_a + Y_b \geqslant 0 \\ & 1 + \sum\limits_a \mathrm{tr}(X_a A_a) + \sum\limits_b \mathrm{tr}(Y_b B_b) \\ & \geqslant \sum\limits_a \dfrac{\mathrm{tr}A_a}{d}\mathrm{tr}X_a + \sum\limits_b \dfrac{\mathrm{tr}B_b}{d}\mathrm{tr}Y_b \end{cases}, \tag{23}$$

where in the following the first formulation will be referred to as the primal, and the second as the dual. The primal variables G_{ab} and η are simply the measurement operators of the parent POVM and the visibility, respectively. The dual variables X_a and Y_b are Lagrange multipliers corresponding to the primal equality constraints. Note that the normalisation of G is not enforced as it follows from the other constraints. For an explicit derivation of the dual problem, see [36], appendix A. Slater's theorem states that whenever a strictly feasible point (a point satisfying all the constraints strictly) exists for either the primal or the dual, the duality gap is zero, thus the primal and dual solutions coincide [22]. In this case, we can take $X_a = Y_b = \delta\,\mathbb{1}$, which is a strictly feasible point of the dual for sufficiently large δ. Thus, the theorem applies and justifies the equality between the two problems in equation (23). Similar arguments apply to all pairs of primal-dual SDPs that we discuss in this work.

As the noise set $\mathbf{N}^{\mathrm{d}}_{A,B}$ defined in equation (22) is invariant under post-processings by linearity of the trace, it follows from section 2.3 that η^{d} is monotonic under post-processings. It turns out, however, that η^{d} does not satisfy the other two natural properties introduced in section 2.3, namely monotonicity under non trace-preserving pre-processings and convexity of the inverse; see appendix A for counterexamples. Note that the monotonicity under pre-processings was incorrectly claimed in [12], proposition 2.

3.1.2. Example

From a result by Busch [38], theorem 4.5 on the joint measurability of pairs of two-outcome qubit measurements, also rephrased by Uola *et al* more recently [39], section III C, we get

$$\eta^{\mathrm{d}}_\theta = \frac{1}{\cos\theta + \sin\theta}. \tag{24}$$

This value is plotted in figure 4 together with the other measures. For completeness and later reference, we give optimal solutions to both the primal and the dual stated in equation (23)

$$\begin{aligned} G &= \frac{1}{\cos\theta + \sin\theta}\begin{bmatrix} \cos\theta\,\dfrac{\mathbb{1}-\sigma_z}{2} & \sin\theta\,\dfrac{\mathbb{1}-\sigma_x}{2} \\ \sin\theta\,\dfrac{\mathbb{1}+\sigma_x}{2} & \cos\theta\,\dfrac{\mathbb{1}+\sigma_z}{2} \end{bmatrix}, \\ (X, Y) &= \frac{1}{4(\cos\theta + \sin\theta)}\left(\begin{bmatrix} \mathbb{1} + (\sigma_z + \sigma_x) \\ \mathbb{1} - (\sigma_z + \sigma_x) \end{bmatrix}, \begin{bmatrix} \mathbb{1} + (\sigma_z - \sigma_x) \\ \mathbb{1} - (\sigma_z - \sigma_x) \end{bmatrix}\right), \end{aligned} \tag{25}$$

where we have used the notation introduced in equation (21).

3.1.3. Lower bound

As mentioned before, a lower bound on η^{d} is already known [16, 29–31]. The parent POVM given in equation (17) is indeed a feasible point for the primal in equation (23) together with

$$\eta = \frac{1}{2}\left(1 + \frac{1}{d+1}\right). \tag{26}$$

For a pair (A, B) of rank-one measurements in dimension $d \geqslant 2$, this bound can be improved. Let us introduce a feasible point for the primal in equation (23) with G of the form (16), where

$$\begin{pmatrix}\alpha_b\\ \beta_a\end{pmatrix} = \frac{-2+\sqrt{d^2+4d-4}}{d}\begin{pmatrix}\mathrm{tr}B_b\\ \mathrm{tr}A_a\end{pmatrix}, \quad \gamma_{ab} = \left(\frac{d+2-\sqrt{d^2+4d-4}}{2d}\right)^2 \mathrm{tr}A_a\mathrm{tr}B_b, \quad \text{and} \quad \delta = 0. \tag{27}$$

For a proof that this leads to valid measurement operators G_{ab} and for a measurement-dependent refinement we refer the reader to appendix C.1.1. This construction gives a lower bound on η^{d} for all pairs of rank-one measurements. However, since the measure is monotonic under post-processings, the bound is actually universal, i.e. for an arbitrary pair $(A,\ B)$ of measurements in dimension d we have

$$\eta^{\mathrm{d}}_{A,B} \geqslant \frac{d-2+\sqrt{d^2+4d-4}}{4(d-1)}. \tag{28}$$

Importantly, this bound turns out to be strictly better than equation (26), which was the best lower bound known so far.

3.1.4. Upper bound

Following the idea used in [36], we provide a valid assignment of the dual variables X_a and Y_b for the dual problem given in equation (23) to get an upper bound on η^{d}, namely,

$$X_a = \frac{\frac{\lambda}{2}\mathbb{1} - A_a}{(f-g^{\mathrm{d}})d} \quad \text{and} \quad Y_b = \frac{\frac{\lambda}{2}\mathbb{1} - B_b}{(f-g^{\mathrm{d}})d}, \tag{29}$$

where f and λ are defined in equation (18) and g^{d} in equation (19). Here we implicitly assume that $f \neq g^{\mathrm{d}}$, but one can show that the equality $f = g^{\mathrm{d}}$ holds if and only if all POVM elements of A and B are proportional to $\mathbb{1}$, in which case the pair is trivially compatible (see appendix E.3.1). The resulting upper bound is given by

$$\eta^{\mathrm{d}}_{A,B} \leqslant \frac{\lambda - g^{\mathrm{d}}}{f - g^{\mathrm{d}}} = 1 - \frac{f-\lambda}{f-g^{\mathrm{d}}}, \tag{30}$$

where the last equality makes clear that this upper bound is non-trivial whenever $f > \lambda$ (since $f > g^{\mathrm{d}}$ from appendix E.3.1). In the following we always implicitly assume that this condition is satisfied when we discuss the various upper bounds.

3.2. Incompatibility random robustness

3.2.1. Definition and properties

In this case the noise model is defined by the map

$$\mathbf{N}^{\mathrm{r}}_{A,B} = \left\{\left(\left\{\frac{\mathbb{1}_d}{n_A}\right\}_{a=1}^{n_A}, \left\{\frac{\mathbb{1}_d}{n_B}\right\}_{b=1}^{n_B}\right)\right\}, \tag{31}$$

a single element containing the trivial measurement, i.e. the measurement generating a uniform distribution of outcomes regardless of the state. It has been investigated in many works [7, 28, 31, 34, 35, 40], and also in the framework of general probabilistic theories [41, 42].

The corresponding incompatibility robustness, as introduced in definition 3, can be computed via the SDPs [40]

$$\eta^{\mathrm{r}}_{A,B} = \left\{\begin{array}{ll} \max\limits_{\eta,\{G_{ab}\}_{ab}} & \eta \\ \text{s.t.} & G_{ab} \geqslant 0, \quad \eta \leqslant 1 \\ & \sum\limits_b G_{ab} = \eta A_a + (1-\eta)\dfrac{\mathbb{1}}{n_A} \\ & \sum\limits_a G_{ab} = \eta B_b + (1-\eta)\dfrac{\mathbb{1}}{n_B} \end{array}\right. = \left\{\begin{array}{ll} \min\limits_{\substack{\{X_a\}_a\\ \{Y_b\}_b}} & 1 + \sum\limits_a \mathrm{tr}(X_a A_a) + \sum\limits_b \mathrm{tr}(Y_b B_b) \\ \text{s.t.} & X_a = X_a^\dagger, \quad Y_b = Y_b^\dagger, \quad X_a + Y_b \geqslant 0 \\ & 1 + \sum\limits_a \mathrm{tr}(X_a A_a) + \sum\limits_b \mathrm{tr}(Y_b B_b) \\ & \geqslant \sum\limits_a \dfrac{1}{n_A}\mathrm{tr}X_a + \sum\limits_b \dfrac{1}{n_B}\mathrm{tr}Y_b \end{array}\right. . \tag{32}$$

Note that the normalisation of G is not enforced as it follows from the other constraints.

As the noise set $\mathbf{N}^{\mathrm{r}}_{A,B}$ defined in equation (31) is invariant under pre-processings (recall that pre-processings are unital), it follows from section 2.3 that η^{r} is monotonic under pre-processings. Moreover, as this set is also convex and independent of the specific form of A and B (the map $\mathbf{N}^{\mathrm{r}}$ is constant), we know from section 2.3 that $1/\eta^{\mathrm{r}}$ is convex. However, this measure is not monotonic under non outcome number-preserving post-processings, see appendix A for a counterexample.

3.2.2. Example

For rank-one projective measurements η^{d} and η^{r} coincide, therefore

$$\eta_\theta^{\mathrm{r}} = \eta_\theta^{\mathrm{d}} = \frac{1}{\cos\theta + \sin\theta}. \tag{33}$$

3.2.3. Lower bound

As η^{r} is not monotonic under post-processings, we cannot use a solution for rank-one measurements as in section 3.1.3 to deduce a general lower bound. Thus, we consider an arbitrary pair (A, B) of measurements in dimension d and we introduce a feasible point for the primal in equation (32) with G of the form (16), where

$$\alpha_b = \sqrt{\frac{n_A}{n_B}}, \quad \beta_a = \sqrt{\frac{n_B}{n_A}}, \quad \gamma_{ab} = 0, \quad \text{and} \quad \delta = 0 \tag{34}$$

from which we obtain the bound

$$\eta_{A,B}^{\mathrm{r}} \geqslant \frac{1}{2}\left(1 + \frac{1}{\sqrt{n_A n_B} + 1}\right). \tag{35}$$

The positivity of this parent POVM follows from

$$0 \leqslant \sqrt{n_A n_B}\left(\frac{A_a}{\sqrt{n_B}} + \frac{B_b}{\sqrt{n_A}}\right)^2 = \{A_a, B_b\} + \sqrt{\frac{n_A}{n_B}} A_a^2 + \sqrt{\frac{n_B}{n_A}} B_b^2 \leqslant \{A_a, B_b\} + \sqrt{\frac{n_A}{n_B}} A_a + \sqrt{\frac{n_B}{n_A}} B_b, \tag{36}$$

where the last inequality is due to $A_a^2 \leqslant A_a$ and $B_b^2 \leqslant B_b$.

3.2.4. Upper bound

In the case of η^{r} we choose the dual variables as

$$X_a = \frac{\frac{\lambda}{2}\mathbb{1} - A_a}{(f - g^{\mathrm{r}})d} \quad \text{and} \quad Y_b = \frac{\frac{\lambda}{2}\mathbb{1} - B_b}{(f - g^{\mathrm{r}})d}, \tag{37}$$

where f and λ are defined in equation (18) and g^{r} in equation (19). Here we implicitly assume that $f \neq g^{\mathrm{r}}$, but one can show that the equality $f = g^{\mathrm{r}}$ holds if and only if all POVM elements of A and B are proportional to $\mathbb{1}$, in which case the pair is trivially compatible (see appendix E.3.1). The resulting upper bound is given by

$$\eta_{A,B}^{\mathrm{r}} \leqslant \frac{\lambda - g^{\mathrm{r}}}{f - g^{\mathrm{r}}}. \tag{38}$$

3.3. Incompatibility probabilistic robustness

3.3.1. Definition and properties

In this case the noise model is defined by the map

$$\mathbf{N}_{A,B}^{\mathrm{p}} = \left\{(\{p_a\,\mathbb{1}_d\}_{a=1}^{n_A}, \{q_b\,\mathbb{1}_d\}_{b=1}^{n_B}) \mid p_a \geqslant 0, q_b \geqslant 0, \sum_a p_a = 1 = \sum_b q_b\right\}, \tag{39}$$

where $\{p_a\}_a$ and $\{q_b\}_b$ are probability distributions. This measure has been investigated in many works [12, 16, 29, 31, 34, 35, 42–45], and also in the framework of general probabilistic theories [46, 47].

The corresponding incompatibility robustness, as introduced in definition 3, can be computed via the SDPs

$$\eta_{A,B}^{\mathrm{p}} = \left\{\begin{array}{ll} \max\limits_{\substack{\eta,\{G_{ab}\}_{ab}\\ \{\tilde{p}_a\}_a,\{\tilde{q}_b\}_b}} & \eta \\ \text{s.t.} & G_{ab} \geqslant 0, \quad \tilde{p}_a \geqslant 0, \quad \tilde{q}_b \geqslant 0 \\ & \sum_a \tilde{p}_a = 1 - \eta = \sum_b \tilde{q}_b \\ & \sum_b G_{ab} = \eta A_a + \tilde{p}_a \mathbb{1} \\ & \sum_a G_{ab} = \eta B_b + \tilde{q}_b \mathbb{1} \end{array}\right. = \left\{\begin{array}{ll} \min\limits_{\substack{\{X_a\}_a,\xi\\ \{Y_b\}_b,\upsilon}} & 1 + \sum_a \mathrm{tr}(X_a A_a) + \sum_b \mathrm{tr}(Y_b B_b) \\ \text{s.t.} & X_a = X_a^\dagger, \quad Y_b = Y_b^\dagger, \quad X_a + Y_b \geqslant 0 \\ & 1 + \sum_a \mathrm{tr}(X_a A_a) + \sum_b \mathrm{tr}(Y_b B_b) \geqslant \xi + \upsilon \\ & \xi \geqslant \mathrm{tr}X_a, \quad \upsilon \geqslant \mathrm{tr}Y_b \end{array}\right. . \tag{40}$$

Note that, in order to make the problem linear in its variables, we have introduced sub-normalised probability distributions $\tilde{p}_a = (1 - \eta)p_a$ and $\tilde{q}_b = (1 - \eta)q_b$. Note also that the normalisation of G and the constraint $\eta \leqslant 1$ are not enforced as they follow from the other constraints. As the noise set $\mathbf{N}_{A,B}^{\mathrm{p}}$ defined in equation (39) contains both $\mathbf{N}_{A,B}^{\mathrm{d}}$ of equation (22) and $\mathbf{N}_{A,B}^{\mathrm{r}}$ of equation (31), the constraints of the primal in equation (40) are looser than the ones in equations (23) and (32). By duality, the constraints of the dual in equation (40) are then tighter than the ones in equations (23) and (32), which can indeed be seen by plugging suitable convex combinations of the constraints $\xi \geqslant \mathrm{tr}X_a$ and $\upsilon \geqslant \mathrm{tr}Y_b$ into $1 + \sum_a \mathrm{tr}(X_a A_a) + \sum_b \mathrm{tr}(Y_b B_b) \geqslant \xi + \upsilon$.

As the noise set $\mathbf{N}^{\mathrm{p}}_{A,B}$ defined in equation (39) is invariant under pre- and post-processings (by unitality and linearity, respectively), it follows from section 2.3 that η^{p} is monotonic under pre- and post-processings. Moreover, as this set is also convex and independent of the specific form of A and B (the map $\mathbf{N}^{\mathrm{p}}$ is constant), we know from section 2.3 that $1/\eta^{\mathrm{p}}$ is convex. Thus, η^{p} is the first measure that satisfies all the properties introduced in section 2 except for concavity.

3.3.2. Example

The dual feasible points from section 3.1.2 satisfy the additional trace constraints of the dual given in equation (40). Thus, the measures η^{d} and η^{p} coincide on this family of measurements:

$$\eta^{\mathrm{p}}_{\theta} = \eta^{\mathrm{d}}_{\theta} = \frac{1}{\cos\theta + \sin\theta}. \tag{41}$$

Note, however, that η^{d} and η^{p} differ in general, even for rank-one projective measurement pairs (see section 4.3 for an explicit example).

3.3.3. Lower bound

Since the noise set $\mathbf{N}^{\mathrm{p}}_{A,B}$ contains both $\mathbf{N}^{\mathrm{d}}_{A,B}$ and $\mathbf{N}^{\mathrm{r}}_{A,B}$ for all $(A,\ B)$, lower bounds on η^{d} and η^{r} immediately apply to η^{p}.

3.3.4. Upper bound

In the case of η^{p} we choose the dual variables as

$$X_a = \frac{\frac{\lambda}{2}\mathbb{1} - A_a}{(f - g^{\mathrm{p}})d}, \quad Y_b = \frac{\frac{\lambda}{2}\mathbb{1} - B_b}{(f - g^{\mathrm{p}})d}, \quad \xi = \max_a \mathrm{tr}X_a, \quad \text{and} \quad \upsilon = \max_b \mathrm{tr}Y_b, \tag{42}$$

where f and λ are defined in equation (18) and g^{p} in equation (19). Here we implicitly assume that $f \neq g^{\mathrm{p}}$, but one can show that the equality $f = g^{\mathrm{p}}$ holds if and only if all POVM elements of A and B are proportional to $\mathbb{1}$, in which case the pair is trivially compatible (see appendix E.3.1). The resulting upper bound is given by

$$\eta^{\mathrm{p}}_{A,B} \leqslant \frac{\lambda - g^{\mathrm{p}}}{f - g^{\mathrm{p}}}. \tag{43}$$

3.4. Incompatibility jointly measurable robustness

3.4.1. Definition and properties

In this case the noise model is defined by the map

$$\mathbf{N}^{\mathrm{jm}}_{A,B} = \mathbf{JM}^{n_A,n_B}_{d}, \tag{44}$$

the set of jointly measurable pairs of POVMs with n_A and n_B outcomes in dimension d. To the best of our knowledge, this measure has only been considered in [40], section II C.

The corresponding incompatibility robustness, as introduced in definition 3, can be computed via the SDPs

$$\eta^{\mathrm{jm}}_{A,B} = \left\{\begin{array}{ll} \max\limits_{\eta,\ \{G_{ab}\}_{ab},\ \{\tilde{H}_{ab}\}_{ab}} & \eta \\ \text{s.t.} & G_{ab} \geqslant 0, \quad \sum\limits_{ab} G_{ab} = \mathbb{1}, \quad \tilde{H}_{ab} \geqslant 0 \\ & \sum\limits_{b}(G_{ab} - \tilde{H}_{ab}) = \eta A_a \\ & \sum\limits_{a}(G_{ab} - \tilde{H}_{ab}) = \eta B_b \end{array}\right. = \left\{\begin{array}{ll} \min\limits_{N,\ \{X_a\}_a,\ \{Y_b\}_b} & \mathrm{tr}N \\ \text{s.t.} & N = N^{\dagger}, \quad X_a = X_a^{\dagger}, \quad Y_b = Y_b^{\dagger} \\ & N \geqslant X_a + Y_b \geqslant 0 \\ & \sum\limits_{a}\mathrm{tr}(X_a A_a) + \sum\limits_{b}\mathrm{tr}(Y_b B_b) \geqslant 1 \end{array}\right. . \tag{45}$$

Note that the noise POVMs do not explicitly appear in the primal problem, since optimising over jointly measurable pairs is equivalent to optimising over the parent measurement, here denoted by H. To make the problem linear in its variables, we have introduced a sub-normalised parent POVM of the noise, $\tilde{H} = (1 - \eta)H$. Note also that the constraint $\eta \leqslant 1$ is not enforced as it follows from summing up one of the marginal constraints.

In analogy with η^{p}, the measure η^{jm} also satisfies the properties introduced in section 2, namely monotonicity under pre- and post-processings, and convexity of the inverse.

3.4.2. Example

The value of this measure for a pair of rank-one projective qubit measurements is strictly higher than for the previous measures, whenever the pair is incompatible. Specifically,

$$\eta_\theta^{\mathrm{jm}} = \frac{2}{1+\cos\theta+\sin\theta}. \tag{46}$$

This value is plotted in figure 4 together with the other measures. Interestingly, even for such a simple example the primal problem given in equation (45) admits multiple optimal solutions. More specifically, we obtain a continuous one-parameter family, which reads

$$G = \begin{bmatrix} r\,\dfrac{\mathbb{1}-\sigma_z}{2} & (1-r)\dfrac{\mathbb{1}-\sigma_x}{2} \\ (1-r)\dfrac{\mathbb{1}+\sigma_x}{2} & r\,\dfrac{\mathbb{1}+\sigma_z}{2} \end{bmatrix},$$

$$\tilde{H} = (1-\eta_\theta^{\mathrm{jm}})\begin{bmatrix} s\,\dfrac{\mathbb{1}+\sigma_z}{2} & (1-s)\dfrac{\mathbb{1}+\sigma_x}{2} \\ (1-s)\dfrac{\mathbb{1}-\sigma_x}{2} & s\,\dfrac{\mathbb{1}-\sigma_z}{2} \end{bmatrix}, \quad \text{where} \quad r = \eta_\theta^{\mathrm{jm}}(s+\cos\theta) - s \tag{47}$$

and s is a free parameter taken from the interval $[0,\ 1]$ to ensure the positivity of the elements of H. Different values of s correspond to applying noise along different axes: for $s=0$ the noise only affects the X direction, while for $s=1$ it only affects the Z direction. A feasible optimal point for the dual given in equation (45) reads

$$(X,\ Y) = \frac{1}{4(1+\cos\theta+\sin\theta)}\left(\begin{bmatrix}\mathbb{1}-(\sigma_z+\sigma_x)\\ \mathbb{1}+(\sigma_z+\sigma_x)\end{bmatrix}, \begin{bmatrix}\mathbb{1}-(\sigma_z-\sigma_x)\\ \mathbb{1}+(\sigma_z-\sigma_x)\end{bmatrix}\right), \quad \text{and} \quad N = \frac{1}{1+\cos\theta+\sin\theta}\mathbb{1}. \tag{48}$$

3.4.3. Lower bound

Let us consider a pair $(A,\ B)$ of rank-one measurements in dimension d. Finding a feasible point for the primal in equation (45) is not an easy task, as we have to find two parent POVMs at once. For G_{ab}, we make the same choice as for η^{d}, i.e. equation (27) in section 3.1.3. We choose the subnormalised noise POVM $\tilde{H}$ to be of the form (16) with

$$\begin{pmatrix}\alpha_b\\ \beta_a\end{pmatrix} = \frac{-2-\sqrt{d^2+4d-4}}{d}\begin{pmatrix}\mathrm{tr}B_b\\ \mathrm{tr}A_a\end{pmatrix}, \quad \gamma_{ab} = \left(\frac{d+2+\sqrt{d^2+4d-4}}{2d}\right)^2 \mathrm{tr}A_a\mathrm{tr}B_b, \quad \text{and} \quad \delta = 0, \tag{49}$$

which leads to

$$\eta = \frac{2\sqrt{d^2+4d-4}}{3d-2+\sqrt{d^2+4d-4}} \leqslant \eta_{A,B}^{\mathrm{jm}}. \tag{50}$$

Details about this specific point can be found in appendix C.4 together with a measurement-dependent refinement. As η^{jm} is monotonic under post-processings, this bound on pairs of rank-one measurements extends to all pairs of measurements in dimension d.

3.4.4. Upper bound

Consider the following feasible point for the dual given in equation (45):

$$X_a = \frac{A_a - \frac{g^{\mathrm{jm}}}{2}\mathbb{1}}{(f-g^{\mathrm{jm}})d}, \quad Y_b = \frac{B_b - \frac{g^{\mathrm{jm}}}{2}\mathbb{1}}{(f-g^{\mathrm{jm}})d}, \quad \text{and} \quad N = \frac{\lambda - g^{\mathrm{jm}}}{f-g^{\mathrm{jm}}}\cdot\frac{\mathbb{1}}{d}, \tag{51}$$

where f and λ are defined in equation (18) and g^{jm} in equation (19). Here we implicitly assume that $f \neq g^{\mathrm{jm}}$, but one can show that the equality $f = g^{\mathrm{jm}}$ holds if and only if all POVM elements of A and B are proportional to $\mathbb{1}$, in which case the pair is trivially compatible (see appendix E.3.1). The above feasible point immediately implies that

$$\eta_{A,B}^{\mathrm{jm}} \leqslant \frac{\lambda - g^{\mathrm{jm}}}{f - g^{\mathrm{jm}}}. \tag{52}$$

3.5. Incompatibility generalised robustness

3.5.1. Definition and properties

In this case the noise model is defined by the map

$$\mathbf{N}^{\mathrm{g}}_{A,B} = \mathbf{POVM}_d^{n_A,n_B}, \tag{53}$$

the set of all POVM pairs with n_A and n_B outcomes, respectively, in dimension d. To the best of our knowledge, this measure was first introduced in [21] and studied further in [7, 33, 40, 48]. Recently, it was given an operational meaning through state discrimination tasks [13, 49, 50].

The corresponding incompatibility robustness, as introduced in definition 3, can be computed via the SDPs

$$\eta^{\mathrm{g}}_{A,B} = \begin{cases} \max\limits_{\eta,\{G_{ab}\}_{ab}} & \eta \\ \text{s.t.} & G_{ab} \geqslant 0, \quad \sum\limits_{ab} G_{ab} = \mathbb{1} \\ & \sum\limits_{b} G_{ab} \geqslant \eta A_a \\ & \sum\limits_{a} G_{ab} \geqslant \eta B_b \end{cases} = \begin{cases} \min\limits_{N,\{X_a\}_a} & \mathrm{tr}N \\ \text{s.t.} & N = N^\dagger, \quad N \geqslant X_a + Y_b \\ & X_a \geqslant 0, \quad Y_b \geqslant 0 \\ & \sum\limits_{a} \mathrm{tr}(X_a A_a) + \sum\limits_{b} \mathrm{tr}(Y_b B_b) \geqslant 1 \end{cases}. \tag{54}$$

Note that in the primal, the noise POVMs do not appear, because we can explicitly solve for these variables, which gives rise to matrix inequalities instead of equalities for the marginals. These looser constraints give us additional freedom and allow us to employ operator inequalities. Note also that the constraint $\eta \leqslant 1$ is not enforced as it follows from summing up one of the marginal constraints. The constraints in the primal in equation (54) are looser than in the primal in equation (45), because the noise set is larger for all measurement pairs. In turn, the feasible set of the dual problem shrinks, as the dual constraints $X_a \geqslant 0$ and $Y_b \geqslant 0$ are tighter than $X_a + Y_b \geqslant 0$.

In analogy with η^{p} and η^{jm}, the measure η^{g} also satisfies the properties we introduced in section 2, namely monotonicity under pre- and post-processings, and convexity of the inverse.

3.5.2. Example

The value of this measure for the running example is even higher than for the previous measures, specifically

$$\eta^{\mathrm{g}}_{\theta} = \frac{\sqrt{2}+1}{\sqrt{2}+\cos\theta+\sin\theta}. \tag{55}$$

This value is plotted in figure 4 together with the other measures. A feasible point for the primal in equation (54) reads

$$G = \begin{bmatrix} r\dfrac{\mathbb{1}-\sigma_z}{2} & (1-r)\dfrac{\mathbb{1}-\sigma_x}{2} \\ (1-r)\dfrac{\mathbb{1}+\sigma_x}{2} & r\dfrac{\mathbb{1}+\sigma_z}{2} \end{bmatrix}, \quad \text{where} \quad r = \frac{1-\sin\theta+(\sqrt{2}+1)\cos\theta}{\sqrt{2}(\sqrt{2}+\cos\theta+\sin\theta)}, \tag{56}$$

and for the dual,

$$(X, Y) = \frac{\sqrt{2}}{4(\sqrt{2}+\cos\theta+\sin\theta)}\left(\begin{bmatrix} \mathbb{1}-\dfrac{\sigma_z+\sigma_x}{\sqrt{2}} \\ \mathbb{1}+\dfrac{\sigma_z+\sigma_x}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \mathbb{1}-\dfrac{\sigma_z-\sigma_x}{\sqrt{2}} \\ \mathbb{1}+\dfrac{\sigma_z-\sigma_x}{\sqrt{2}} \end{bmatrix}\right), \quad \text{and} \quad N = \frac{\sqrt{2}+1}{2(\sqrt{2}+\cos\theta+\sin\theta)}\mathbb{1}. \tag{57}$$

3.5.3. Lower bound

For a pair (A, B) of rank-one measurements in dimension d, let us introduce a feasible point for the primal in equation (54) with G of the form (16), where

$$\begin{pmatrix}\alpha_b \\ \beta_a\end{pmatrix} = \frac{1}{2\sqrt{d}}\begin{pmatrix}\mathrm{tr}B_b \\ \mathrm{tr}A_a\end{pmatrix}, \quad \gamma_{ab} = 0, \quad \text{and} \quad \delta = \frac{\sqrt{d}}{2}, \tag{58}$$

so that we obtain the bound

$$\eta = \frac{1}{2}\left(1+\frac{1}{\sqrt{d}}\right) \leqslant \eta^{\mathrm{g}}_{A,B}. \tag{59}$$

A proof of feasibility of this specific point is given below. For more details, see appendix C.5 which also contains a measurement-dependent refinement. As η^{g} is monotonic under post-processings, this bound on pairs of rank-one measurements extends to all pairs of measurements in dimension d.

The novelty in equation (58), as compared to the parent POVMs used for the other measures, is the fact that δ is non-zero. What enables us to introduce this term is the extra freedom in the primal in equation (54), namely, the inequalities in the marginal constraints instead of equalities, which allows us to analyse the marginals for non-zero δ.

For the proof of feasibility, we write the parent POVM defined by the coefficients in equation (58) as

$$G_{ab} = \frac{1}{4(d+\sqrt{d})}[\mathrm{tr}(B_b)A_a + \mathrm{tr}(A_a)B_b + 2\sqrt{d}\{A_a, B_b\} + d(A_a^{\frac{1}{2}}B_bA_a^{\frac{1}{2}} + B_b^{\frac{1}{2}}A_aB_b^{\frac{1}{2}})]. \tag{60}$$

Since A_a and B_b are rank-one, we can write $A_a = \mathrm{tr}(A_a)P_a$ and $B_b = \mathrm{tr}(B_b)Q_b$ for some $P_a = |\varphi_a\rangle\langle\varphi_a|$ and $Q_b = |\psi_b\rangle\langle\psi_b|$. Therefore, we can rewrite (60) as

$$G_{ab} = \frac{\mathrm{tr}(A_a)\mathrm{tr}(B_b)}{4(d+\sqrt{d})}[(P_a + \sqrt{d}P_aQ_b)^\dagger(P_a + \sqrt{d}P_aQ_b) + (Q_b + \sqrt{d}Q_bP_a)^\dagger(Q_b + \sqrt{d}Q_bP_a)] \geqslant 0, \tag{61}$$

which shows that G is a valid POVM.

Next we should compute its marginals. The first one reads

$$\sum_b G_{ab} = \frac{1}{4(d+\sqrt{d})}\left[dA_a + \mathrm{tr}(A_a)\mathbb{1} + 4\sqrt{d}A_a + d\left(A_a + \sum_b B_b^{\frac{1}{2}}A_aB_b^{\frac{1}{2}}\right)\right], \tag{62}$$

where the terms are ordered as in equation (60) for clarity. Moreover, we have that for every $|\xi\rangle$,

$$\begin{aligned} d\,\langle\xi|\sum_b B_b^{\frac{1}{2}}A_aB_b^{\frac{1}{2}}|\xi\rangle &= \sum_{b'}\mathrm{tr}(B_{b'})\sum_b \mathrm{tr}(B_b)\mathrm{tr}(A_a)\langle\xi|\psi_b\rangle\langle\psi_b|\varphi_a\rangle\langle\varphi_a|\psi_b\rangle\langle\psi_b|\xi\rangle \\ &= \mathrm{tr}(A_a)\sum_{b'}|\sqrt{\mathrm{tr}(B_{b'})}|^2\sum_b|\sqrt{\mathrm{tr}(B_b)}\langle\xi|\psi_b\rangle\langle\psi_b|\varphi_a\rangle|^2 \\ &\geqslant \mathrm{tr}(A_a)\left|\sum_b \mathrm{tr}(B_b)\langle\xi|\psi_b\rangle\langle\psi_b|\varphi_a\rangle\right|^2 = \mathrm{tr}(A_a)|\langle\xi|\varphi_a\rangle|^2 = \langle\xi|A_a|\xi\rangle, \end{aligned} \tag{63}$$

where we used the Cauchy–Schwarz inequality. Therefore, $d\sum_b B_b^{1/2}A_aB_b^{1/2} \geqslant A_a$, which together with $\mathrm{tr}(A_a)\mathbb{1} \geqslant A_a$ enables us to lower bound the marginal (62), namely,

$$\sum_b G_{ab} \geqslant \frac{2(d+2\sqrt{d}+1)}{4(d+\sqrt{d})}A_a = \frac{1}{2}\left(1+\frac{1}{\sqrt{d}}\right)A_a. \tag{64}$$

By symmetry of equation (60) the same conclusion holds for the second marginal, which shows that the point defined in equations (58) and (59) is indeed feasible.

3.5.4. Upper bound

Consider the following feasible point for the dual given in equation (54):

$$X_a = \frac{A_a}{fd}, \quad Y_b = \frac{B_b}{fd}, \quad \text{and} \quad N = \frac{\lambda}{f}\cdot\frac{\mathbb{1}}{d}, \tag{65}$$

where f and λ are defined in equation (18). This immediately implies that

$$\eta^{\mathrm{g}}_{A,B} \leqslant \frac{\lambda}{f}. \tag{66}$$

3.6. Relations between the measures

Certain inclusions between the noise sets defined in equations (22), (31), (39), (44), and (53), imply an ordering of the measures. More specifically, from

$$(\mathbf{N}^{\mathrm{d}}_{A,B} \bigcup \mathbf{N}^{\mathrm{r}}_{A,B}) \subseteq \mathbf{N}^{\mathrm{p}}_{A,B} \subseteq \mathbf{N}^{\mathrm{jm}}_{A,B} \subseteq \mathbf{N}^{\mathrm{g}}_{A,B}, \tag{67}$$

we conclude that

$$\max\{\eta^{\mathrm{d}}_{A,B}, \eta^{\mathrm{r}}_{A,B}\} \leqslant \eta^{\mathrm{p}}_{A,B} \leqslant \eta^{\mathrm{jm}}_{A,B} \leqslant \eta^{\mathrm{g}}_{A,B} \tag{68}$$

for every pair (A, B). It turns out that η^{d} and η^{r} are incomparable (see appendix A for an example). A more detailed analysis allows us to prove that some of the inequalities given in equation (68) are in fact strict. Specifically, in appendix B we derive improved relations between η^{d} and η^{jm}, η^{d} and η^{g}, and η^{r} and η^{g}, which imply that for a pair of incompatible measurements (A, B) the separations between these measures are strict, i.e. $\eta^{\mathrm{d}}_{A,B} < \eta^{\mathrm{jm}}_{A,B}$, $\eta^{\mathrm{d}}_{A,B} < \eta^{\mathrm{g}}_{A,B}$, and $\eta^{\mathrm{r}}_{A,B} < \eta^{\mathrm{g}}_{A,B}$. Moreover, the examples given in section 3.7 show that in some cases η^{d} coincides with η^{p}, as well as η^{r} with η^{p} and η^{jm} with η^{g}. The question whether the separation between η^{p} and η^{jm} is strict or not is left open.

3.7. Mutually unbiased bases

We have mentioned earlier that MUBs constitute a standard example of a pair of incompatible measurements on a d-dimensional system. Indeed, they might seem like natural candidates for the most incompatible pair of measurements in dimension d. In this section we show that for a pair of MUBs all the previously introduced measures can be computed analytically. The specific values we obtain will be compared against the findings of section 4, in which we look for the most incompatible pairs of measurements.

For a pair $(A^{\mathrm{MUB}}, B^{\mathrm{MUB}})$ of projective measurements onto two MUBs in dimension d (see section 2.1), we will use $\eta^*_{\mathrm{MUB}}(d)$ as a shorthand for $\eta^*_{A^{\mathrm{MUB}},B^{\mathrm{MUB}}}$. Note that although in higher dimensions not all pairs of MUBs are unitarily equivalent, they nevertheless give the same value for all the measures studied in this work. Hence, for these measures the quantity $\eta^*_{\mathrm{MUB}}(d)$ turns out to be well-defined.

In dimension $d = 2$ a pair of MUB measurements is a special case of the example introduced in section 2.6, corresponding to $\theta = \pi/4$. Therefore equations (24), (46), and (55) imply that

$$\eta^{\mathrm{d}}_{\mathrm{MUB}}(2) = \eta^{\mathrm{r}}_{\mathrm{MUB}}(2) = \eta^{\mathrm{p}}_{\mathrm{MUB}}(2) = \frac{1}{\sqrt{2}}, \quad \eta^{\mathrm{jm}}_{\mathrm{MUB}}(2) = 2(\sqrt{2} - 1), \quad \text{and} \quad \eta^{\mathrm{g}}_{\mathrm{MUB}}(2) = \frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\right). \tag{69}$$

For a pair of projective measurements onto two MUBs in dimension $d \geqslant 3$, the parameters given in equations (18) and (19) equal $f = 2$, $\lambda = 1 + 1/\sqrt{d}$, $g^{\mathrm{d}} = g^{\mathrm{r}} = g^{\mathrm{p}} = 2/d$, and $g^{\mathrm{jm}} = 0$. It turns out that for MUBs the upper bounds given in equations (30), (52), and (66) are actually tight. Therefore, the only missing component is a feasible point for the primal.

For η^{d} and η^{r} our feasible solution consists of

$$\eta = \frac{1}{2}\left(1 + \frac{1}{\sqrt{d} + 1}\right) \tag{70}$$

and

$$G_{ab} = \frac{1}{2(\sqrt{d} + 1)}\left(\{A_a, B_b\} + \frac{1}{\sqrt{d}}A_a + \frac{1}{\sqrt{d}}B_b\right). \tag{71}$$

This parent POVM, inspired by [39], section IV, is of the form of equation (16). The positivity of these operators can be confirmed using the techniques presented in appendix C and let us stress that the proof crucially relies on the fact that the bases are mutually unbiased. For η^{p} we must explicitly include the weights and we choose them to be uniform $p_a = q_b = 1/d$ for all a, b. This assignment saturates the upper bound given in equation (30), which implies that

$$\eta^{\mathrm{d}}_{\mathrm{MUB}}(d) = \eta^{\mathrm{r}}_{\mathrm{MUB}}(d) = \eta^{\mathrm{p}}_{\mathrm{MUB}}(d) = \frac{1}{2}\left(1 + \frac{1}{\sqrt{d} + 1}\right). \tag{72}$$

For η^{g} we use the same parent POVM, but the more flexible form of noise allows for higher visibility:

$$\eta = \frac{1}{2}\left(1 + \frac{1}{\sqrt{d}}\right). \tag{73}$$

For η^{jm} we must supplement our solution with a sub-normalised parent POVM of the noise pair

$$\tilde{H}_{ab} = \frac{1 - \eta^{\mathrm{jm}}_{\mathrm{MUB}}(d)}{d(d-2)}\left[\mathbb{1} + \frac{d}{d-1}(\{A_a, B_b\} - A_a - B_b)\right], \tag{74}$$

which has already been used in [49], and is of the form of equation (16). This construction is only valid for $d \geqslant 3$, because for $d = 2$ the corresponding noise pair $\{(\mathbb{1} - A_a)/(d-1)\}_a$ and $\{(\mathbb{1} - B_b)/(d-1)\}_b$ is not jointly measurable (see equation (47) for a family of optimal feasible points for the primal). In both cases the visibility given in equation (73) saturates the upper bounds (55) and (46), respectively, which implies that for all $d \geqslant 3$, we have

$$\eta^{\mathrm{jm}}_{\mathrm{MUB}}(d) = \eta^{\mathrm{g}}_{\mathrm{MUB}}(d) = \frac{1}{2}\left(1 + \frac{1}{\sqrt{d}}\right). \tag{75}$$

Note that the value $\eta^{\mathrm{g}}_{\mathrm{MUB}}(d)$ was already derived in [21]. Also notice that equation (75) together with equation (59) implies that MUBs are among the most incompatible measurement pairs with respect to η^{g} in every dimension.

3.8. Summary

In table 1 we give a compact summary of the results for the differents robustness-based measures of incompatibility: definition of the noise sets, properties introduced in section 2.3, lower and upper bounds, and

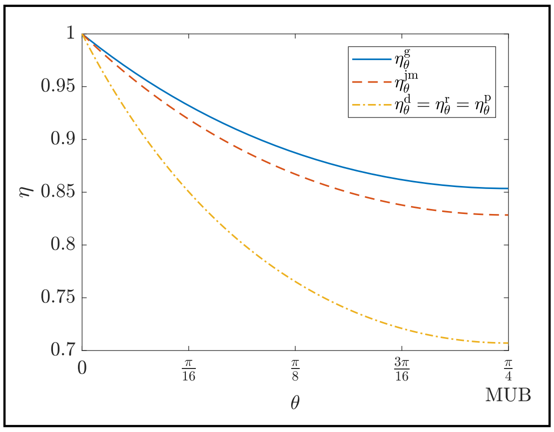


Figure 4. The value of all the different measures (see table 1) for a pair of rank-one projective measurements on a qubit such that the angle between the Bloch vectors of these measurements equals 2θ; see equation (20). Note that the rightmost point where $\theta = \pi/4$ corresponds to qubit MUBs, which demonstrates the fact that MUBs are the most incompatible rank-one projective qubit measurements under all these measures. From bottom to top, the curves are $\eta^{\mathrm{d}} = \eta^{\mathrm{r}} = \eta^{\mathrm{p}}$ from equation (24), then η^{jm} from equation (46), and finally η^{g} from equation (55). Although η^{d}, η^{r}, and η^{p} coincide in this case, this is not the case in general.

Table 1. Summary of the results on the depolarising, random, probabilistic, jointly measurable, and general incompatibility robustness of pairs of POVMs. Recall that d is the dimension, while n_A and n_B are the outcome numbers. 'Post' and 'Pre' stand for post-processing and pre-processing monotonicity, respectively, see section 2.3. 'Cvx' stands for the convexity of the inverse of the measure, see section 2.3. For a pair of rank-one projective measurements $(A,\ B)$, the quantities appearing in the upper bounds are $f = 2$, $\lambda = \max_{a,b}\{\max \mathrm{Sp}(A_a + B_b)\}$, $g^{\mathrm{jm}} = \min_{a,b}\{\min \mathrm{Sp}(A_a + B_b)\}$, and $g^{\mathrm{d}} = g^{\mathrm{r}} = g^{\mathrm{p}} = 2/d$; see equations (18) and (19) for definitions.

	Form of the noise	Post	Pre	Cvx	Lower bound	MUB value	Upper bound
η^{d}	$\left\{\left(\left\{\mathrm{tr}\, A_a \frac{\mathbb{1}}{d}\right\}_a, \left\{\mathrm{tr}\, B_b \frac{\mathbb{1}}{d}\right\}_b\right)\right\}$	yes	no	no	$\frac{d-2+\sqrt{d^2+4d-4}}{4(d-1)}$		$\frac{\lambda - g^{\mathrm{d}}}{f - g^{\mathrm{d}}}$
η^{r}	$\left\{\left(\left\{\frac{\mathbb{1}}{n_A}\right\}_a, \left\{\frac{\mathbb{1}}{n_B}\right\}_b\right)\right\}$	no	yes	yes	$\frac{1}{2}\left(1+\frac{1}{\sqrt{n_A n_B}+1}\right)$	$\frac{1}{2}\left(1+\frac{1}{\sqrt{d}+1}\right)$	$\frac{\lambda - g^{\mathrm{r}}}{f - g^{\mathrm{r}}}$
η^{p}	$\left\{\left(\left\{p_a \mathbb{1}\right\}_a, \left\{q_b \mathbb{1}\right\}_b\right)\right\}$		yes		$\max\{\eta^{\mathrm{d}}, \eta^{\mathrm{r}}\}$		$\frac{\lambda - g^{\mathrm{p}}}{f - g^{\mathrm{p}}}$
η^{jm}	$\mathbf{JM}_d^{n_A, n_B}$		yes		$\frac{2\sqrt{d^2+4d-4}}{3d-2+\sqrt{d^2+4d-4}}$	$\begin{cases} 2(\sqrt{2}-1) & d=2 \\ \frac{1}{2}\left(1+\frac{1}{\sqrt{d}}\right) & d \geqslant 3 \end{cases}$	$\frac{\lambda - g^{\mathrm{jm}}}{f - g^{\mathrm{jm}}}$
η^{g}	$\mathbf{POVM}_d^{n_A, n_B}$		yes		$\frac{1}{2}\left(1+\frac{1}{\sqrt{d}}\right)$		$\frac{\lambda}{f}$

value for a specific example of two projective measurements onto MUBs (see section 3.7). In figure 4 we plot the values of η_θ^* achieved by a pair of rank-one projective measurements acting on a qubit.

4. Most incompatible pairs of measurements

In this section, we address the question of the most incompatible measurement pairs in dimension d, for all the measures introduced in section 3. This question has already been raised and partially answered in previous works: in infinite dimension for η^{p} in [29] and numerically for η^{d} and η^{g} in [33]. Perhaps surprisingly, we find that the answer depends on which incompatibility measure we consider. We have already seen that projective measurements onto a pair of MUBs are among the most incompatible pairs under η^{g} in every dimension. On the other hand, for the measures η^{d} and η^{p} we give explicit constructions of pairs which are more incompatible than MUBs for any dimension $d \geqslant 3$. For η^{jm}, our study is inconclusive, and we do not find measurements that are more incompatible than MUBs in any dimension. First we discuss the special case of η^{r}, then we solve the qubit case for all the measures, and finally we discuss higher dimensions.

4.1. Incompatibility random robustness

Recall that in order to find the most incompatible measurement pair in dimension d regardless of the outcome numbers, it is enough to consider rank-one POVMs if the measure in consideration is monotonic under post-processings. As we see from table 1, this is not the case for η^{r}, which, at first glance, makes this problem hard to tackle. However, what turns out is that for this measure the answer is trivial. Consider a pair of measurements (A, B) and increase artificially the number of outcomes by adding zero POVM elements to both measurements. Let us add these elements one-by-one, and denote the POVM pair at step i by (A^i, B^i). In appendix C.2.2 we show that if $\lambda < 2$ and $2(\lambda - 1) < f$, we have

$$\lim_{i\to\infty} \eta^{\mathrm{r}}_{A^i,B^i} \leqslant \frac{2-\lambda}{f-2(\lambda-1)}, \tag{76}$$

where f and λ are defined in equation (18). It is then clear that whenever $f = 2$ and $\lambda < 2$ (e.g. any pair of rank-one projective measurements onto two bases that do not have any eigenvectors in common), this limit reaches $\frac{1}{2}$. As it coincides with the trivial lower bound mentioned in section 2.4, this shows that $\chi^{\mathrm{r}}(d) = \frac{1}{2}$ for $d \geqslant 2$. In the rest of this section, we will not discuss this measure anymore. However, recall that for pairs of rank-one projective measurements η^{r} coincides with η^{d}, and therefore some of the results later in this section also apply to this measure.

4.2. Qubit case

In section 3.7 we have shown that for a pair of MUBs all the incompatibility measures can be computed analytically. What is special in the case of $d = 2$ is that these values coincide with the universal lower bounds (see table 1). This means that pairs of projective measurements onto MUBs are among the most incompatible pairs under η^{d}, η^{p}, η^{jm}, and η^{g} in dimension $d = 2$. Formally, using the notation introduced in section 2.4, we have that

$$\chi^{\mathrm{d}}(2) = \chi^{\mathrm{p}}(2) = \frac{1}{\sqrt{2}}, \quad \chi^{\mathrm{jm}}(2) = 2(\sqrt{2}-1), \quad \text{and} \quad \chi^{\mathrm{g}}(2) = \frac{1}{2}\left(1+\frac{1}{\sqrt{2}}\right). \tag{77}$$

For η^{d}, this was known for pairs of two-outcome POVMs [36], appendix G.

It is important to point out that there exist other pairs of measurements reaching these minimal values: from the upper bounds given in appendix E.3.2, it is clear that any rank-one POVM pair such that $A_a = |a\rangle\langle a|$ and the Bloch vectors of B lie in the xy-plane of the Bloch sphere gives rise to the same value as MUBs. As an example, one might choose $A_a = |a\rangle\langle a|$ and B as a trine measurement in the xy-plane.

In appendix E.4, we extend this result to triplets of qubit measurements. In this case, we show that triplets of projective measurements onto MUBs are among the most incompatible measurements under η^{d}, η^{p}, η^{jm}, and η^{g} in dimension $d = 2$.

Also note that the value of $\chi^{\mathrm{d}}(2)$ (respectively its equivalent for three measurements) has interesting consequences for Einstein–Podolsky–Rosen steering. This is because joint measurability is intimately linked to this notion [6, 7], as the depolarising map in η^{d} can be equivalently applied to the state we wish to steer, due to its self-duality. We refer to [36], appendix F for details on this connection and only mention here that our results show that in a steering scenario with two (respectively three) measurements and an isotropic state of local dimension two, POVMs do not provide any advantage over projective measurements.

4.3. Higher dimensions

4.3.1. Dimension $d = 3$

In the previous section we have seen that in dimension $d = 2$ pairs of projective measurements onto two MUBs are among the most incompatible pairs of measurements under η^{d}, η^{p}, η^{jm}, and η^{g}. Starting from dimension $d = 3$, the picture changes dramatically. To show this, we plot the (numerical) value of these four measures for a particular one-parameter path of rank-one projective measurements in dimension three, see figure 5. It is evident from this plot that, contrary to the qubit case, MUBs do not achieve the lowest value of the incompatibility robustness under η^{d} and η^{p}. Instead, the lowest value among rank-one projective measurements is reached by other bases, which we have found through an extensive numerical search among pairs of rank-one projective measurements, using a parametrisation of unitary matrices in dimension three [51].

In this section we only look at rank-one projective measurements. Due to the unitary invariance of all the measures we assume without loss of generality that the first measurement corresponds to the computational basis $A_a = |a\rangle\langle a|$, so that we only need to specify the second measurement B.

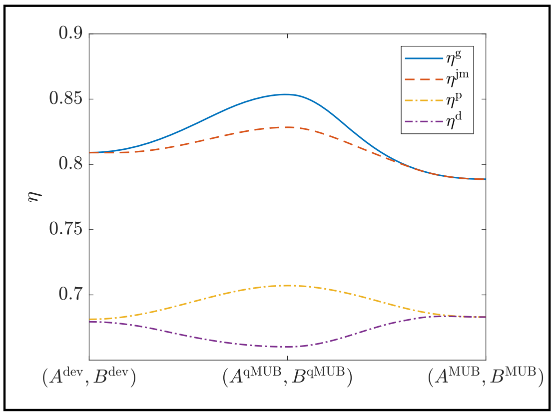


Figure 5. The (numerical) value of the four measures along a one-parameter path of rank-one projective measurements in dimension $d = 3$. The pair $(A^{\mathrm{dev}}, B^{\mathrm{dev}})$ is defined in equation (79), $(A^{\mathrm{qMUB}}, B^{\mathrm{qMUB}})$ in equation (78), and $(A^{\mathrm{MUB}}, B^{\mathrm{MUB}})$ at the beginning of this section. Details about the specific path used can be found in appendix D. Importantly, on this path the pair $(A^{\mathrm{MUB}}, B^{\mathrm{MUB}})$ achieves the minimum value with respect to η^{g} and η^{jm}, but it is outperformed by $(A^{\mathrm{dev}}, B^{\mathrm{dev}})$ with respect to η^{p} and by $(A^{\mathrm{qMUB}}, B^{\mathrm{qMUB}})$ with respect to η^{d}.

For η^{d}, the optimum is reached, among others, by

$$B_b^{\mathrm{qMUB}} = U|b\rangle\langle b|U^\dagger, \quad \text{where} \quad U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{78}$$

Note that it is simply a pair of *qubit MUBs* on a two-dimensional subspace together with a trivial third outcome on the orthogonal subspace. The incompatibility depolarising robustness of this pair, $\eta^{\mathrm{d}}_{\mathrm{qMUB}}(3) \approx 0.6602$ (see equation (80) below for an analytical value) outperforms substantially not only $\eta^{\mathrm{d}}_{\mathrm{MUB}}(3) \approx 0.6830$, but also the minimal value 0.6794 found numerically in [33], table 4.

For η^{p}, the optimum is reached, among others, by

$$B_b^{\mathrm{dev}} = U|b\rangle\langle b|U^\dagger, \quad \text{where} \quad U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}, \tag{79}$$

which gives $\eta^{\mathrm{p}}_{\mathrm{dev}} \approx 0.6813$, showing a slight *deviation* from $\eta^{\mathrm{p}}_{\mathrm{MUB}}(3) \approx 0.6830$.

For η^{jm}, the numerical search did not yield an improvement on the MUB value, and for η^{g} we already have an analytical proof that MUBs are among the most incompatible pairs in every dimension.

4.3.2. Dimension $d \geqslant 4$

For η^{d}, the qubit MUB structure found in dimension $d = 3$ has several natural generalisations in higher dimensions. The general idea is to divide the Hilbert space into orthogonal subspaces of various dimensions, and define the measurements as either MUBs or trivial measurements on the different subspaces. Among these, we found numerically that the most incompatible construction is to define a pair of qubit MUBs on a two-dimensional subspace, while on the orthogonal subspace the remaining measurement operators turn out to be irrelevant. For simplicity, we choose trivial measurements on the orthogonal subspace, that is, $A_a = |a\rangle\langle a|$ and $B_b = |b\rangle\langle b|$ for $a, b \geqslant 3$, while $\{A_1, A_2\}$ and $\{B_1, B_2\}$ is a pair of MUBs on the qubit subspace. For this construction, we get a lower bound in equation (C13) and an upper bound in equation (C23), which give the

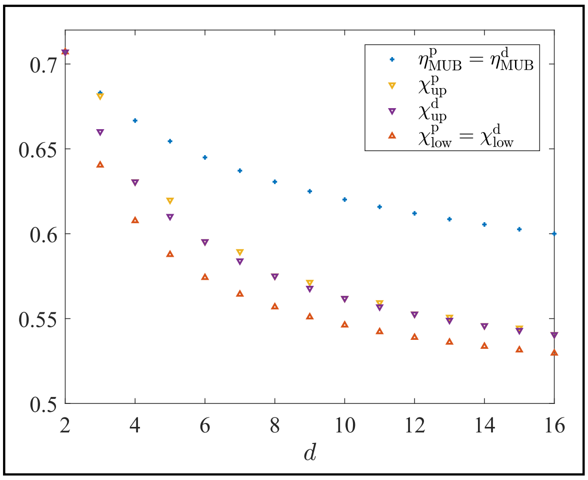


Figure 6. Illustration of the improvement over MUBs for η^{d} and η^{p} when the dimension d ranges from 2 to 16. From top to bottom are depicted the MUB value (equations (72) and (75)), the lowest value we found for η^{p} (that is, equation (80) for even dimensions and numerical results based on an analytical construction described in the main text for odd dimensions), the lowest value we found for η^{d} (equation (80)), and the lower bound (28).

same value and therefore the incompatibility depolarising robustness of this pair is

$$\eta^{\mathrm{d}}_{\mathrm{qMUB}}(d) = \frac{1}{2}\left(1 + \frac{\sqrt{2}}{d + \sqrt{2}}\right) < \eta^{\mathrm{d}}_{\mathrm{MUB}}(d). \tag{80}$$

In figure 6 we plot the improvement over MUBs that this construction achieves. In particular, it is worth stressing that, in contrast to a pair of MUBs, this construction exhibits the same asymptotic scaling as the lower bound derived in section 3.1.3. More specifically, expanding the right-hand side of equation (28) gives

$$\frac{1}{2} + \frac{1}{2d} + O(d^{-2}), \tag{81}$$

whereas

$$\eta^{\mathrm{d}}_{\mathrm{qMUB}}(d) = \frac{1}{2} + \frac{1}{\sqrt{2}\,d} + O(d^{-2}), \tag{82}$$

$$\eta^{\mathrm{d}}_{\mathrm{MUB}}(d) = \frac{1}{2} + \frac{1}{2\sqrt{d}} + O(d^{-1}). \tag{83}$$

The reason why this pair performs so well is the fact that the two measurements are highly incompatible on the qubit subspace, while the noise is spread uniformly over the entire space. Note that an analogous structure has been found while searching for the quantum state whose nonlocal statistics are the most robust to white noise [52]. Supported by the optimisation in dimension $d = 3$ together with one billion random instances in dimensions $d = 4$ and $d = 5$, and the asymptotic scalings, we conjecture that this pair is among the most incompatible pairs of rank-one projective measurements under η^{d} for all dimensions. For general pairs of measurements we leave the question open.

For η^{p}, fixing MUBs on a qubit subspace no longer determines the incompatibility robustness any more, as the noise can now be adjusted to have different weights on the different subspaces. In fact the construction that uses trivial measurements on the orthogonal subspace does not surpass the d-dimensional MUB value any more. However, employing some other rank-one projective measurements on the orthogonal subspace gives rise to measurements that outperform MUBs. In even dimensions, by decomposing the space into many qubit subspaces and by having MUBs on each of them, we can reach again the value of equation (80). For instance in dimension $d = 4$ this means

$$B_b = U|b\rangle\langle b|U^\dagger, \quad \text{where} \quad U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}. \tag{84}$$

The parent POVM is then the same as for η^{d} whereas the construction of the dual variables is explained in appendix C.1.2. Our conjecture on η^{d} then translates straightforwardly to η^{p} in even dimensions as $\eta^{\mathrm{d}} \leqslant \eta^{\mathrm{p}}$. In odd dimensions, this construction is not applicable. We conjecture that in dimension $d = 3$ the pair defined in equation (79) is among the most incompatible pairs of projective measurements under η^{p}. In higher odd dimensions, taking this pair on a qutrit subspace together with MUBs on all remaining qubit subspaces always outperforms MUBs (see figure 6). As there might be some more involved construction giving a lower value, we leave the question of the lowest value of η^{p} open for odd dimensions higher than $d = 5$. Note nonetheless that with one billion random pairs of rank-one measurements in dimension $d = 5$ we were not able to surpass it.

For η^{jm}, encouraged by the optimisation in dimension $d = 3$ and the one billion random sampling in dimensions $d = 4$ and $d = 5$, we conjecture that pairs of MUBs in any dimension cannot be outperformed by any pair of rank-one projective measurements.

Regarding η^{g}, the incompatibility generalised robustness of a pair of MUBs is precisely the universal lower bound that we derived in equation (59). This means that MUBs are among the most incompatible pairs among all pairs of measurements in dimension d, regardless of the number of outcomes. Formally, using the notation introduced in section 2.4, this means that

$$\chi^{\mathrm{g}}(d) = \frac{1}{2}\left(1 + \frac{1}{\sqrt{d}}\right). \tag{85}$$

5. Conclusions

In this work we develop a unified framework to study various robustness-based measures of incompatibility of quantum measurements. We find that some of the widely used measures do not satisfy some natural properties, which means that one should be cautious when dealing with them. In particular, they are not suitable for constructing a resource theory of incompatibility. Moreover, we find that the most incompatible measurement pair depends on the exact measure that we use, even when all the addressed natural properties are satisfied. We are able to show that for one of the measures a pair of rank-one projective measurements onto mutually unbased bases is among the most incompatible pairs, but also that this is not the case for some other measures. Our work shows that the different measures exhibit genuinely different properties and we conclude that despite a substantial effort dedicated to the topic, our understanding is still rather limited.

One natural future direction arising from our work would be to obtain a complete characterisation of the most incompatible measurement pairs in all scenarios for all the measures. We expect, however, that this might be rather difficult, so one might start by restricting the task to natural scenarios, e.g. $d = n_A = n_B$ or even just searching over rank-one projective measurements.

Many results in this paper can be straightforwardly extended to the case of more than two measurements. We refer to appendix E for the SDP formulations of the various measures, the upper bounds and a few lower bounds. This could serve as a good starting point for future research.

A last promising research direction arising from our work concerns the possibility of constructing a resource theory of incompatibility. Are some of the existing measures suitable as resource monotones? Are there some additional conditions that one should require? What is the most general class of operations that preserves joint measurability? Answering these questions will help us to understand how to quantify and classify incompatibility in a meaningful and operational manner.

Acknowledgments

The authors thank Nicolas Brunner, Bartosz Reguła, René Schwonnek, Marco Tomamichel and Roope Uola for useful discussions. Financial support by the Swiss National Science Foundation (Starting grant DIAQ, NCCR-QSIT) is gratefully acknowledged. MF acknowledges support from the Polish NCN grant Sonata UMO-2014/14/E/ST2/00020, and the grant Badania Młodych Naukowców number 538-5400-B049-18, issued by the

Polish Ministry of Science and Higher Education. JK acknowledges support from POLONEZ programme which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 665778.

Appendix A. Counterexamples

In this appendix, we prove some claims made in the main text through explicit examples. Note that some of the values in this section are obtained via numerics, but as these values are solutions of SDPs, they are exact up to machine precision.

Counterexample 1. The measure η^{d} is not monotonic under pre-processings.

Note that an incorrect proof of this statement appeared in [12], proposition 2. The issue with the argument is that it implicitly assumes pre-processings to be trace-preserving. The following counterexample exploits this loophole. Consider a pair of qubit MUBs measurements:

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},\quad A_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},\quad B_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix},\quad B_2 = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{A1}$$

For these the value $\eta^{\mathrm{d}}_{A,B} = 1/\sqrt{2}$ is well-known. See for example [39], section III A or the example in the main text (section 3.1.2). Let us create new qutrit measurements A^{Λ} and B^{Λ} by pre-processing, specifically, by applying the map $\Lambda(.) = K_1(.)K_1^{\dagger} + K_2(.)K_2^{\dagger}$, where

$$K_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } K_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ so that } \Lambda\left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right] = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & d \end{pmatrix}. \tag{A2}$$

Crucially, $\mathrm{tr}A_2 = 1 \neq 2 = \mathrm{tr}A_2^{\Lambda}$ and similarly for B. From the following feasible point for the dual in (23):

$$X_1 = \begin{pmatrix} \frac{9}{4} & 0 & 0 \\ 0 & \frac{27}{20} & 0 \\ 0 & 0 & \frac{3}{4} \end{pmatrix},\ X_2 = \begin{pmatrix} \frac{27}{10} & 0 & 0 \\ 0 & \frac{3}{4} & 0 \\ 0 & 0 & \frac{3}{4} \end{pmatrix},$$
$$Y_1 = \begin{pmatrix} \frac{2\sqrt{39}-99}{40} & -\frac{1}{4} & 0 \\ -\frac{1}{4} & \frac{4\sqrt{39}-63}{60} & 0 \\ 0 & 0 & -\frac{3}{4} \end{pmatrix}, Y_2 = \begin{pmatrix} \frac{2\sqrt{39}-99}{40} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{4\sqrt{39}-63}{60} & 0 \\ 0 & 0 & -\frac{3}{4} \end{pmatrix}, \tag{A3}$$

we get the bound

$$\eta^{\mathrm{d}}_{A^{\Lambda},B^{\Lambda}} \leqslant \frac{14\sqrt{39}-3}{120} \approx 0.7036 < 0.7071 \approx \frac{1}{\sqrt{2}} = \eta^{\mathrm{d}}_{A,B}. \tag{A4}$$

Counterexample 2. The measure $1/\eta^{\mathrm{d}}$ is not convex.

Consider the following pairs (A^0, B^0) and (A^1, B^1) of qubit measurements

$$A_1^0 = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix},\quad A_2^0 = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{pmatrix},\quad B_1^0 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix},\quad B_2^0 = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix},\quad A_1^1 = \mathbb{1},\quad A_2^1 = 0,\quad B^1 = B^0. \tag{A5}$$

In [53], jointly measurable pairs of two-outcome qubit measurements are fully characterised. From this, we can compute

$$\eta^{\mathrm{d}}_{A^0,B^0} = \sqrt{\frac{5+\sqrt{5}}{10}}, \quad \eta^{\mathrm{d}}_{A^2,B^2} = 1, \quad \text{and} \quad \eta^{\mathrm{d}}_{\frac{A^0+A^1}{2},\frac{B^0+B^1}{2}} = \sqrt{\frac{25+\sqrt{13}}{34}}, \tag{A6}$$

from which the convexity of $1/\eta^{\mathrm{d}}$ is immediately negated as

$$\frac{1}{2}\left(\frac{1}{\eta^{\mathrm{d}}_{A^0,B^0}} + \frac{1}{\eta^{\mathrm{d}}_{A^1,B^1}}\right) \approx 1.0878 < 1.0902 \approx \frac{1}{\eta^{\mathrm{d}}_{\frac{A^0+A^1}{2},\frac{B^0+B^1}{2}}}. \tag{A7}$$

Note that the non-concavity of η^{d} follows from the non-convexity of $1/\eta^{\mathrm{d}}$.

Counterexample 3. The measure η^{r} is not monotonic under post-processings.

Consider a pair (A, B) of qubit MUBs measurements, as given in equation (A1). Let us create a new three-outcome measurement A^β by the post-processing

$$\beta(1|1) = \beta(2|1) = \frac{1}{2}, \quad \beta(3|1) = \beta(1|2) = \beta(2|2) = 0,$$
$$\text{and } \beta(3|1) = 1 \text{ so that } A_1^\beta = A_2^\beta = \frac{A_1}{2} \text{ and } A_3^\beta = A_2. \tag{A8}$$

The incompatibility random robustness of A^β and B is lower than $1/\sqrt{2}$, which can be seen by the feasible point for the dual in (32)

$$X_1 = X_2 = \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & \frac{27}{10} \end{pmatrix}, X_3 = \begin{pmatrix} \frac{27}{20} & 0 \\ 0 & \frac{9}{4} \end{pmatrix},$$
$$Y_1 = \begin{pmatrix} \frac{4\sqrt{39}-63}{60} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{2\sqrt{39}-99}{40} \end{pmatrix}, Y_2 = \begin{pmatrix} \frac{4\sqrt{39}-63}{60} & \frac{1}{4} \\ \frac{1}{4} & \frac{2\sqrt{39}-99}{40} \end{pmatrix}, \tag{A9}$$

which gives rise to

$$\eta^{\mathrm{r}}_{A^\beta,B} \leqslant \frac{14\sqrt{39}-3}{120} \approx 0.7036 < 0.7071 \approx \frac{1}{\sqrt{2}} = \eta^{\mathrm{r}}_{A,B}. \tag{A10}$$

Counterexample 4. The measures η^{d} and η^{r} are incomparable.

Using [53] and the pair of measurements (A^0, B^0) defined in equation (A5), one gets

$$\eta^{\mathrm{d}}_{A^0,B^0} = \sqrt{\frac{5+\sqrt{5}}{10}} \approx 0.8507 < 0.8660 \approx \frac{\sqrt{3}}{2} = \eta^{\mathrm{r}}_{A^0,B^0}. \tag{A11}$$

To get the other direction, we consider a pair of two-outcome measurements in dimension $d = 3$, namely,

$$A_1^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_2^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B_1^2 = \begin{pmatrix} \frac{1}{32} & \frac{1}{8} & -\frac{1}{8} \\ \frac{1}{8} & \frac{3}{4} & -\frac{1}{8} \\ -\frac{1}{8} & -\frac{1}{8} & \frac{3}{4} \end{pmatrix}, \quad B_2^2 = \begin{pmatrix} \frac{31}{32} & -\frac{1}{8} & \frac{1}{8} \\ -\frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{4} \end{pmatrix}, \tag{A12}$$

which gives $\eta^{\mathrm{r}}_{A^2,B^2} \approx 0.8799 < 0.8816 \approx \eta^{\mathrm{d}}_{A^2,B^2}$.

Counterexample 5. None of the measures defined in the main text is concave.

Consider the following pairs (A^0, B^0) and (A^1, B^1) of qubit measurements

$$A_a^0 = |a\rangle\langle a|, \quad B_1^0 = \begin{pmatrix} \frac{1}{20} & \frac{1}{20} \\ \frac{1}{20} & \frac{19}{20} \end{pmatrix}, \quad B_2^0 = \begin{pmatrix} \frac{19}{20} & -\frac{1}{20} \\ -\frac{1}{20} & \frac{1}{20} \end{pmatrix}, \quad A_a^1 = U_A|a\rangle\langle a|U_A^\dagger, \quad B_b^1 = U_B|b\rangle\langle b|U_B^\dagger, \tag{A13}$$

where

$$U_A = \begin{pmatrix} \sqrt{\frac{19}{20}} & \sqrt{\frac{1}{20}} \\ \sqrt{\frac{1}{20}} & -\sqrt{\frac{19}{20}} \end{pmatrix} \quad \text{and} \quad U_B = \begin{pmatrix} \sqrt{\frac{1}{5}} & \sqrt{\frac{1}{5}} \\ \sqrt{\frac{1}{5}} & -\sqrt{\frac{4}{5}} \end{pmatrix}. \tag{A14}$$

With this example, the concavity of all five measures studied in the main text is negated, that is,

$$\eta^*_{\frac{A^0+A^1}{2},\frac{B^0+B^1}{2}} < \frac{\eta^*_{A^0,B^0} + \eta^*_{A^1,B^1}}{2}, \tag{A15}$$

as one can confirm by solving the respective SDPs up to machine precision.

Appendix B. Relations between the measures

In the main text we observed that inclusions between the different noise sets immediately imply certain inequalities between the measures. More specifically, equation (68) states that

$$\max\{\eta^{\mathrm{d}}_{A,B}, \eta^{\mathrm{r}}_{A,B}\} \leqslant \eta^{\mathrm{p}}_{A,B} \leqslant \eta^{\mathrm{jm}}_{A,B} \leqslant \eta^{\mathrm{g}}_{A,B}. \tag{B1}$$

In this appendix we show that these relations can be strengthened, which leads to strict separations between some of the measures.

In order to tighten the inequality between η^{d} and η^{jm}, we take the optimal point for the primal for η^{d} in equation (23), and construct from it a feasible point for the primal for η^{jm} in equation (45). Specifically, for a pair of measurements (A, B) we subtract some fraction of the original POVM element from the noise reaching the optimum in the primal for η^{d} in equation (23), such that the remaining noise is jointly measurable and can thus serve as a feasible point for the primal for η^{jm} in equation (45):

$$\eta^{\mathrm{d}}_{A,B}A_a + (1-\eta^{\mathrm{d}}_{A,B})\mathrm{tr}A_a\frac{\mathbb{1}}{d} = \left(\eta^{\mathrm{d}}_{A,B} + \frac{1-\eta^{\mathrm{d}}_{A,B}}{d}\epsilon\right)A_a + \left(1-\eta^{\mathrm{d}}_{A,B} - \frac{1-\eta^{\mathrm{d}}_{A,B}}{d}\epsilon\right)\frac{\mathrm{tr}(A_a)\mathbb{1} - \epsilon A_a}{d-\epsilon}, \tag{B2}$$

and similarly for B_b. The challenge now is to determine the largest value of ϵ for which the noise pair

$$\left(\left\{\frac{\mathrm{tr}(A_a)\mathbb{1} - \epsilon A_a}{d-\epsilon}\right\}_a, \left\{\frac{\mathrm{tr}(B_b)\mathbb{1} - \epsilon B_b}{d-\epsilon}\right\}_b\right) \tag{B3}$$

is jointly measurable. This can be done by finding rank-one POVMs which can be post-processed to give A and B, respectively. Let $\{R_r\}_r$ be a rank-one POVM which under post-processing β_R gives $\{A_a\}_a$ and similarly let $\{S_s\}_s$ give $\{B_b\}_b$ under β_S. The parent POVM given in equation (49) implies that the noise pair

$$\left(\left\{\frac{\mathrm{tr}(R_r)\mathbb{1} - \epsilon R_r}{d-\epsilon}\right\}_r, \left\{\frac{\mathrm{tr}(S_s)\mathbb{1} - \epsilon S_s}{d-\epsilon}\right\}_s\right) \tag{B4}$$

is jointly measurable for

$$\epsilon = \frac{2d}{d+\sqrt{d^2+4d-4}}. \tag{B5}$$

Now note that $A_a = \sum_r \beta_R(a|r)R_r$ implies

$$\sum_r \beta_R(a|r)\frac{\mathrm{tr}(R_r)\mathbb{1} - \epsilon R_r}{d-\epsilon} = \frac{\mathrm{tr}(A_a)\mathbb{1} - \epsilon A_a}{d-\epsilon}. \tag{B6}$$

Clearly, if we apply the post-processings β_R and β_S to the noise pair given in equation (B4), we will obtain the noise pair given in equation (B3) for the same value of ϵ. Since post-processings preserve joint measurability we deduce that

$$\eta^{\mathrm{d}}_{A,B} + \frac{2(1-\eta^{\mathrm{d}}_{A,B})}{d+\sqrt{d^2+4d-4}} \leqslant \eta^{\mathrm{jm}}_{A,B}. \tag{B7}$$

In order to tighten the inequality (B1) between η^{d} and η^{g}, we take the optimal point for the primal for η^{d} in equation (23), and construct from it a feasible point for the primal for η^{g} in equation (54). Specifically, we use $\mathrm{tr}(A_a)\mathbb{1} \geqslant A_a$ and $\mathrm{tr}(B_b)\mathbb{1} \geqslant B_b$ to obtain

$$\eta^{\mathrm{d}}_{A,B}A_a + (1 - \eta^{\mathrm{d}}_{A,B})\mathrm{tr}A_a\frac{\mathbb{1}}{d} \geqslant \left(\eta^{\mathrm{d}}_{A,B} + \frac{1 - \eta^{\mathrm{d}}_{A,B}}{d}\right)A_a \tag{B8}$$

and a similar relation for B_b. These together imply that

$$\eta^{\mathrm{d}}_{A,B} + \frac{1 - \eta^{\mathrm{d}}_{A,B}}{d} \leqslant \eta^{\mathrm{g}}_{A,B}. \tag{B9}$$

In order to tighten the inequality (B1) between η^{r} and η^{g}, we take the optimal point for the primal for η^{r} in equation (32), and construct from it a feasible point for the primal for η^{g} in equation (54). Specifically, we use $\mathbb{1} \geqslant A_a$ and $\mathbb{1} \geqslant B_b$ to obtain

$$\eta^{\mathrm{r}}_{A,B}A_a + (1 - \eta^{\mathrm{r}}_{A,B})\frac{\mathbb{1}}{n_A} \geqslant \left(\eta^{\mathrm{r}}_{A,B} + \frac{1 - \eta^{\mathrm{r}}_{A,B}}{n_A}\right)A_a \tag{B10}$$

and a similar relation for B_b. These together imply that

$$\eta^{\mathrm{r}}_{A,B} + \frac{1 - \eta^{\mathrm{r}}_{A,B}}{\max\{n_A,\, n_B\}} \leqslant \eta^{\mathrm{g}}_{A,B}. \tag{B11}$$

Note that all the above improved relations are saturated by pairs of MUBs in dimension two, see section 3.7.

Appendix C. Bounds on the different measures

In this appendix we provide details about various bounds that we introduce in the main text, namely equations (28), (30), (50), and (59). Moreover, we provide measurement-dependent refinements of the lower bounds and we generalise the upper bound on η^{d}, η^{r}, and η^{p} for certain classes of measurements with some specific structures.

We will use the ansatz defined in equation (16), but only for the case of rank-one measurements A and B. Note that in this case $A_a^{1/2}B_bA_a^{1/2} = \mathrm{tr}(A_aB_b)A_a/\mathrm{tr}A_a \propto A_a$, and similarly, $B_b^{1/2}A_aB_b^{1/2} \propto B_b$. Therefore, we can write equation (16) as

$$G_{ab} \propto \{A_a,\, B_b\} + (\tilde{\alpha}_{ab}A_a + \tilde{\beta}_{ab}B_b) + \gamma_{ab}\mathbb{1}, \tag{C1}$$

where the proportionality constant is fixed by the normalisation, and we introduced the new parameters $\tilde{\alpha}_{ab}$ and $\tilde{\beta}_{ab}$ that now depend on both indices. Clearly, the operator is non-trivial only on the subspace spanned by the eigenvectors of A_a and B_b, which allows us to compute its spectrum. The eigenvalues of (C1) are then

$$\frac{1}{2}(\tilde{\alpha}_{ab}\mathrm{tr}A_a + \tilde{\beta}_{ab}\mathrm{tr}B_b + 2\mathrm{tr}(A_aB_b) \pm \sqrt{(\tilde{\alpha}_{ab}\mathrm{tr}A_a - \tilde{\beta}_{ab}\mathrm{tr}B_b)^2 + 4\mathrm{tr}(A_aB_b)(\tilde{\alpha}_{ab} + \mathrm{tr}B_b)(\tilde{\beta}_{ab} + \mathrm{tr}A_a)}) + \gamma_{ab}, \tag{C2}$$

together with γ_{ab} when $d \geqslant 3$.

C.1. Incompatibility depolarising robustness

C.1.1. Lower bound. For a pair $(A,\, B)$ of rank-one measurements, an ansatz of the form (C1) that is easy to analyse is defined by $\tilde{\alpha}_{ab} = x\mathrm{tr}B_b$, $\tilde{\beta}_{ab} = x\mathrm{tr}A_a$, and $\gamma_{ab} = y\mathrm{tr}A_a\mathrm{tr}B_b$, so that

$$G_{ab} = \frac{1}{2(1 + dx) + d^2y}(\{A_a,\, B_b\} + x(A_a\mathrm{tr}B_b + B_b\mathrm{tr}A_a) + y\mathrm{tr}A_a\mathrm{tr}B_b\mathbb{1}). \tag{C3}$$

Clearly if either $A_a = 0$ or $B_b = 0$, we have $G_{ab} = 0$, so in the following we restrict ourselves to the case $\mathrm{tr}A_a\mathrm{tr}B_b > 0$. From equation (C2) we deduce that in order to have $G_{ab} \geqslant 0$, we should have

$$y \geqslant 0 \quad \text{and} \quad x + c_{ab}^2 \pm (1 + x)c_{ab} + y \geqslant 0, \quad \text{where} \quad c_{ab} = \sqrt{\frac{\mathrm{tr}(A_aB_b)}{\mathrm{tr}A_a\mathrm{tr}B_b}}. \tag{C4}$$

For $x \geqslant -1$ the second constraint is tighter with the minus sign which gives

$$y \geqslant -x(1 - c_{ab}) + c_{ab}(1 - c_{ab}). \tag{C5}$$

For a fixed c_{ab} this defines a half-plane in the $(x,\, y)$ plane. Taking the intersection of all the half-planes corresponding to $c_{ab} \in [0,\, 1]$ yields the region of $(x,\, y)$ for $x \geqslant -1$ which is allowed for all possible measurements. To explicitly characterise the region we maximise the right-hand side of equation (C5) over $c_{ab} \in [0,\, 1]$ for every fixed value of $x \geqslant -1$. Since the expression is a quadratic function of c_{ab} the maximum is achieved at $c_{ab} = (1 + x)/2$ if this value lies in the range $[0,\, 1]$ or at one of the endpoints $c_{ab} = 0$, $c_{ab} = 1$. A straightforward case-by-case analysis yields the allowed region for $x \geqslant -1$.

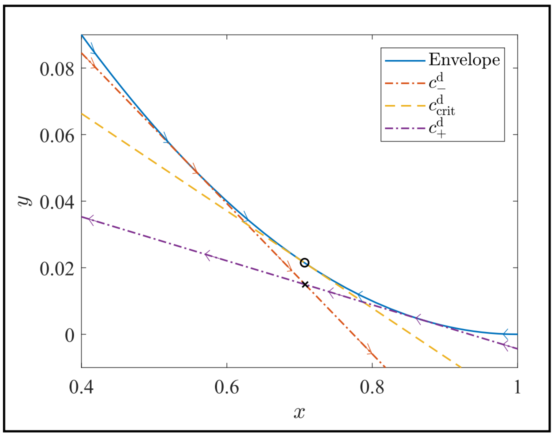


Figure C1. Illustration of the measurement-dependent refinement of the lower bound on η^d. The upper curve is the boundary of the allowed region defined in equation (C7), while the three lines are the boundaries of half-planes defined in equation (C5) for three different overlaps $c_-^d < c_{crit}^d < c_+^d$. Note that the arrows indicate the gradient of the objective function defined in equation (C8) along the different curves. In particular, the objective function is constant on the dashed line corresponding to c_{crit}^d. The black circle corresponds to the point given in equation (C9) while the black cross to the one given in equation (C11). In this figure we have used $d = 3$ and $c_\pm^d = c_{crit}^d \pm 0.08$.

For $x \leqslant -1$ the tighter constraint reads

$$y \geqslant -x(1 + c_{ab}) - c_{ab}(1 + c_{ab}) \tag{C6}$$

and the same procedure leads to the allowed region for $x \leqslant -1$. Combining the two results gives the overall allowed region:

$$y \geqslant \begin{cases} -2(1+x) & \text{if } x \leqslant -3 \\ \dfrac{(1-x)^2}{4} & \text{if } -3 \leqslant x \leqslant 1, \\ 0 & \text{if } 1 \leqslant x \end{cases} \tag{C7}$$

over which we want to maximise the objective function of the primal in equation (23), that is,

$$\eta = \frac{2 + dx}{2(1 + dx) + d^2 y}. \tag{C8}$$

Since the right-hand side increases as y decreases, the maximum is reached on the boundary of the allowed region. Then we can plug y with equality in equation (C7) into the function (C8) and differentiate the resulting single variable function with respect to x to obtain the following optimal assignment:

$$x = \frac{-2 + \sqrt{d^2 + 4d - 4}}{d} \quad \text{and} \quad y = \left(\frac{d + 2 - \sqrt{d^2 + 4d - 4}}{2d}\right)^2, \tag{C9}$$

which corresponds to the feasible point presented in equation (28) of the main text. It is easy to check that this choice of x and y saturates equation (C5) for a particular value of c_{ab}, which we refer to as the *critical overlap*

$$c_{crit}^d = \frac{d - 2 + \sqrt{d^2 + 4d - 4}}{2d} \geqslant \frac{1}{\sqrt{d}}. \tag{C10}$$

Note that this coincides with the MUB overlap only in dimension $d = 2$.

There is an easy way to refine this bound in a measurement-dependent way: instead of requiring that equation (C5) holds for all values $c_{ab} \in [0, 1]$, we only require that it holds for the values that appear for the specific pair of rank-one measurements we consider. Imposing fewer constraints means that we are optimising over a larger region, so we might hope to reach a higher value of the objective function.

If we only care about a finite number of overlaps c_{ab}, the lower boundary of the relevant region is piecewise linear (see figure C1). If one of the overlaps equals the critical one, the bound cannot be improved, so in the following we assume that none of the overlaps equals the critical one. It turns out that to determine the optimal assignment of x and y we only need to know the value of the largest overlap that is smaller than the critical one, which we denote by c_-^d, and whether there are any overlaps larger than the critical one. If there are overlaps larger than the critical one, let us denote the smallest of these by c_+^{d} and then the optimal point is reached at the intersection of the two lines defined by c_-^{d} and c_+^{d} in equation (C5), which gives

$$x = c_-^{\mathrm{d}} + c_+^{\mathrm{d}} - 1 \quad \text{and} \quad y = (1 - c_-^{\mathrm{d}})(1 - c_+^{\mathrm{d}}), \tag{C11}$$

so that the measurement-dependent refinement of equation (28) reads

$$\eta_{A,B}^{\mathrm{d}} \geqslant \eta_{A,B}^{\mathrm{d,low}} = \frac{(c_-^{\mathrm{d}} + c_+^{\mathrm{d}} - 1)d + 2}{2 + 2(c_-^{\mathrm{d}} + c_+^{\mathrm{d}} - 1)d + (1 - c_-^{\mathrm{d}})(1 - c_+^{\mathrm{d}})d^2}. \tag{C12}$$

What is particularly interesting about this bound is that whenever c_-^{d} tends to 0 and c_+^{d} tends to 1, the bound tends to 1, i.e. these conditions are strong enough to ensure that the measurements are almost compatible. This is clearly the case for for identical measurements, that is, for $A = B$, for which the bound equals 1.

If none of the overlaps is greater than the critical one, the optimal assignment is given by $x = c_-^{\mathrm{d}}$ and $y = 0$ and the resulting value corresponds to setting $c_+^{\mathrm{d}} = 1$ in the right-hand side of equation (C12).

As an example we can compute the lower bound for the embedding of qubit MUBs into higher dimensions introduced in section 4.3. In this example, $c_-^{\mathrm{d}} = 1/\sqrt{2}$ and $c_+^{\mathrm{d}} = 1$ so that we get

$$\eta_{\mathrm{qMUB}}^{\mathrm{d}}(d) \geqslant \frac{1}{2}\left(1 + \frac{\sqrt{2}}{d + \sqrt{2}}\right), \tag{C13}$$

which turns out to be the correct value, see equation (C23) for a matching upper bound.

C.1.2. Upper bound for embeddings in higher dimensions. Here we investigate how the upper bound on η^{d} is affected by the following procedure, which we refer to as *embedding*. Consider a pair $(\hat{A}, \hat{B})$ of rank-one projective measurements in dimension d_i and create a new pair (A, B) in dimension $d_f \geqslant d_i$ as follows:

$$A_a = \begin{cases} \begin{pmatrix} \hat{A}_a & 0 \\ 0 & 0 \end{pmatrix} & \text{if } 1 \leqslant a \leqslant d_i \\ \begin{pmatrix} 0 & 0 \\ 0 & M_{a-d_i} \end{pmatrix} & \text{if } d_i + 1 \leqslant a \leqslant d_f \end{cases} \quad \text{and} \quad B_b = \begin{cases} \begin{pmatrix} \hat{B}_b & 0 \\ 0 & 0 \end{pmatrix} & \text{if } 1 \leqslant b \leqslant d_i \\ \begin{pmatrix} 0 & 0 \\ 0 & N_{b-d_i} \end{pmatrix} & \text{if } d_i + 1 \leqslant b \leqslant d_f \end{cases}, \tag{C14}$$

where (M, N) is a pair of rank-one projective measurements acting on a $(d_f - d_i)$-dimensional space.

We derive an upper bound on $\eta_{A,B}^{\mathrm{d}}$ which depends only on the quantity λ (defined in equation (18) of the main text) computed for the measurement pair $(\hat{A}, \hat{B})$ and the dimensions d_i and d_f. As long as $\lambda < 2$ the bound decreases as d_f increases and in the limit $d_f \to \infty$ it converges to $\frac{1}{2}$. This can be explained by observing that as d_f increases the noise gets spread out over the entire space and its weight on the subspace relevant for the measurements $(\hat{A}, \hat{B})$ decreases. Note that the bound shows no dependence on the second pair of measurements (M, N).

Let us introduce the following ansatz for the dual in equation (23):

$$\begin{cases} X_a = \begin{pmatrix} \alpha\mathbb{1} - \beta\hat{A}_a & 0 \\ 0 & 0 \end{pmatrix} & \text{if } 1 \leqslant a \leqslant d_i \\ X_a = \begin{pmatrix} \gamma\mathbb{1} & 0 \\ 0 & 0 \end{pmatrix} & \text{if } d_i + 1 \leqslant a \leqslant d_f \end{cases} \quad \text{and} \quad \begin{cases} Y_b = \begin{pmatrix} \alpha\mathbb{1} - \beta\hat{B}_b & 0 \\ 0 & 0 \end{pmatrix} & \text{if } 1 \leqslant b \leqslant d_i \\ Y_b = \begin{pmatrix} \gamma\mathbb{1} & 0 \\ 0 & 0 \end{pmatrix} & \text{if } d_i + 1 \leqslant b \leqslant d_f \end{cases}. \tag{C15}$$

The scalar constraint of the dual in equation (23) reads

$$\begin{aligned} &1 + \sum_{a=1}^{d_i}(\alpha\mathrm{tr}\hat{A}_a - \beta\mathrm{tr}\hat{A}_a^2) + \sum_{b=1}^{d_i}(\alpha\mathrm{tr}\hat{B}_b - \beta\mathrm{tr}\hat{B}_b^2) \geqslant \\ &\sum_{a=1}^{d_i}\frac{\alpha d_i - \beta\mathrm{tr}\hat{A}_a}{d_f}\mathrm{tr}\hat{A}_a + \sum_{a=d_i+1}^{d_f}\frac{\gamma d_i}{d_f}\mathrm{tr}M_a + \sum_{b=1}^{d_i}\frac{\alpha d_i - \beta\mathrm{tr}\hat{B}_b}{d_f}\mathrm{tr}\hat{B}_b + \sum_{b=d_i+1}^{d_f}\frac{\gamma d_i}{d_f}\mathrm{tr}N_b, \end{aligned} \tag{C16}$$

which can be further simplified using the rank-one projective assumption to

$$1 + 2\alpha d_i\left(1 - \frac{d_i}{d_f}\right) \geqslant 2\beta d_i\left(1 - \frac{1}{d_f}\right) + 2\gamma d_i\left(1 - \frac{d_i}{d_f}\right). \tag{C17}$$

It is easy to see that the constraints

$$\gamma \geqslant 0, \quad \alpha + \gamma \geqslant \beta, \quad \text{and} \quad 2\alpha \geqslant \beta\lambda \tag{C18}$$

ensure that $X_a + Y_b \geqslant 0$. More specifically, the first one is required to ensure positivity when both indices are between $d_i + 1$ and d_f, the second one when one of the indices is between 1 and d_i and the other between $d_i + 1$

Table C1. Values obtained for η^{d} using the embedding procedure described in appendix C.1.2. Specifically, the values correspond to the embedding of a complete set of MUBs in dimension d_i into dimension d_f. For example, the value 4/10 in the last column comes from the embedding of 5 MUBs from dimension $d_i = 4$ to dimension $d_f = 6$. Although we present numerical values all values are analytical. Note also that the upper bound obtained via the construction explained in appendix C.1.2 only gives an upper bound on η^{d}. In all cases shown in this table, this bound is tight as there exists a parent POVM reaching exactly the same value. Such a parent is not given in this paper. As they provide upper bounds on the lowest value achievable by η^{d}, they can be compared to table 4 in [33].

d_i \ d_f	2	3	4	5	6
2	0.5774	0.5273	0.4975	0.4778	0.4605
3		0.4818	0.4514	0.4314	0.4114
4			0.4309	0.4128	0.4
5				0.6863	0.3620

and d_f, and the last one when both indices are between 1 and d_i. Requiring that the last two inequalities given in equation (C18) are saturated implies

$$\alpha = \frac{\lambda}{2}\beta, \quad \text{and} \quad \gamma = \left(1 - \frac{\lambda}{2}\right)\beta. \tag{C19}$$

Plugging these back into equation (C16) and requiring that the inequality is saturated allows us to deduce that

$$\frac{1}{\beta d_i} = 2\left[1 - \frac{1}{d_f} - (\lambda - 1)\left(1 - \frac{d_i}{d_f}\right)\right]. \tag{C20}$$

To see that this corresponds to a non-negative value of β note that $\lambda \leqslant 2$ implies that

$$2\left[1 - \frac{1}{d_f} - (\lambda - 1)\left(1 - \frac{d_i}{d_f}\right)\right] \geqslant 2\left[1 - \frac{1}{d_f} - \left(1 - \frac{d_i}{d_f}\right)\right] = \frac{2(d_i - 1)}{d_f} \geqslant 0. \tag{C21}$$

We immediately see that $\gamma \geqslant 0$, which means that the assignment given above is a feasible point for the dual given in equation (23). The resulting upper bound reads

$$\eta^{\mathrm{d}}_{A,B} \leqslant \frac{\lambda - \frac{2}{d_f} - 2(\lambda - 1)\left(1 - \frac{d_i}{d_f}\right)}{2\left[1 - \frac{1}{d_f} - (\lambda - 1)\left(1 - \frac{d_i}{d_f}\right)\right]} = \frac{1}{2}\left[1 + \frac{(\lambda - 1)d_i - 1}{(2 - \lambda)d_f + (\lambda - 1)d_i - 1}\right]. \tag{C22}$$

It is immediate that whenever $d_i = d_f$ we recover exactly the upper bound given in equation (30).

As an example we can compute the upper bound for the embedding of qubit MUBs into higher dimensions introduced in section 4.3. For this example, $d_i = 2$, $d_f = d$, $f = 2$, and $\lambda = 1 + 1/\sqrt{2}$ so that we get

$$\eta^{\mathrm{d}}_{\mathrm{qMUB}}(d) \leqslant \frac{1}{2}\left(1 + \frac{\sqrt{2}}{d + \sqrt{2}}\right), \tag{C23}$$

which turns out to be the correct value, see equation (C13) for a matching lower bound.

Note that this procedure can also be applied to sets of more than two measurements. Although we do not go into the details in this case, table C1 contains the values obtained by embedding a complete set of MUBs in higher dimensions by adding rank-one projective measurements onto the computational basis of the remaining $(d_f - d_i)$-dimensional space, e.g. $M_a = |a\rangle\langle a|$ in equation (C14).

C.2. Incompatibility random robustness

C.2.1. Lower bound. For a pair of rank-one measurements, it is possible to refine the ansatz defined in equation (34) by tuning the relative weight of the anticommutator, but this does not lead to any general bound on η^{r} as this measure is not monotonic under post-processings.

C.2.2. Upper bound with addition of zero outcomes. Here we show how to tighten the upper bound introduced in section 3.2.4 in the presence of zero POVM elements, which we then use in section 4.1. We consider a pair (A, B) of measurements that contain zero POVM elements. Without loss of generality we can assume that the first POVM elements are non-zero. Then, for simplicity, we assume that $n_A = n_B = n_f$ and that the number of non-zero elements of A and B is the same and we denote it by n_i. The other cases, namely, $n_A \neq n_B$ or the number of non-zero elements of A and B being different, can be treated in a similar manner. Therefore we are left with two POVMs with the same number n_f of outcomes such that

$$\begin{cases} A_a \neq 0 & \text{if } 1 \leqslant a \leqslant n_i \\ A_a = 0 & \text{if } n_i + 1 \leqslant a \leqslant n_f \end{cases} \quad \text{and} \quad \begin{cases} B_b \neq 0 & \text{if } 1 \leqslant b \leqslant n_i \\ B_b = 0 & \text{if } n_i + 1 \leqslant b \leqslant n_f \end{cases}. \tag{C24}$$

Then we introduce the following ansatz for the dual in equation (32):

$$\begin{cases} X_a = \alpha \mathbb{1} - \beta A_a & \text{if } 1 \leqslant a \leqslant n_i \\ X_a = \gamma \mathbb{1} & \text{if } n_i + 1 \leqslant a \leqslant n_f \end{cases} \quad \text{and} \quad \begin{cases} Y_b = \alpha \mathbb{1} - \beta B_b & \text{if } 1 \leqslant b \leqslant n_i \\ Y_b = \gamma \mathbb{1} & \text{if } n_i + 1 \leqslant b \leqslant n_f \end{cases}. \tag{C25}$$

Note that the only difference from equation (37) is that the coefficient of the identity in the dual variable depends on whether the outcome corresponds to a zero or non-zero POVM elements. The scalar constraint of the dual in equation (32) reads

$$\begin{aligned} &1 + \sum_{a=1}^{n_i} (\alpha \mathrm{tr} A_a - \beta \mathrm{tr} A_a^2) + \sum_{b=1}^{n_i} (\alpha \mathrm{tr} B_b - \beta \mathrm{tr} B_b^2) \\ &\geqslant \sum_{a=1}^{n_i} \frac{\alpha d - \beta \mathrm{tr} A_a}{n_f} + \sum_{a=n_i+1}^{n_f} \frac{\gamma d}{n_f} + \sum_{b=1}^{n_i} \frac{\alpha d - \beta \mathrm{tr} B_b}{n_f} + \sum_{b=n_i+1}^{n_f} \frac{\gamma d}{n_f}, \end{aligned} \tag{C26}$$

which can be further simplified by introducing f defined in equation (18) of the main text:

$$1 + 2\alpha d \left(1 - \frac{n_i}{n_f}\right) \geqslant \beta d \left(f - \frac{2}{n_f}\right) + 2\gamma d \left(1 - \frac{n_i}{n_f}\right). \tag{C27}$$

Assume that $\beta > 0$ and let λ be the quantity defined in equation (18) of the main text computed for the measurement pair (A, B). It is easy to see that the constraints

$$\gamma \geqslant 0, \quad \alpha + \gamma \geqslant \beta, \quad \text{and} \quad 2\alpha \geqslant \beta\lambda, \tag{C28}$$

ensure that $X_a + Y_b \geqslant 0$. The first one is required to ensure positivity when both indices are between $n_i + 1$ and n_f, the second one when one is between 1 and n_i and the other between $n_i + 1$ and n_f, and the last one when both are between 1 and n_i.

Requiring that the last two inequalities given in equation (C28) are saturated implies

$$\alpha = \frac{\lambda}{2}\beta, \quad \text{and} \quad \gamma = \left(1 - \frac{\lambda}{2}\right)\beta. \tag{C29}$$

Plugging these back into equation (C27) and requiring that the inequality is saturated allows us to deduce that

$$\frac{1}{\beta d} = f - \frac{2}{n_f} - 2(\lambda - 1)\left(1 - \frac{n_i}{n_f}\right). \tag{C30}$$

It is easy to check that $f > 2(\lambda - 1)$ (which is only possible if $\lambda < 2$) guarantees that this assignment leads to strictly positive β. Then, this constitutes a feasible point for the dual given in equation (32) and we obtain

$$\eta^{\mathrm{r}}_{A,B} \leqslant \frac{\lambda - \frac{2}{n_f} - 2(\lambda - 1)\left(1 - \frac{n_i}{n_f}\right)}{f - \frac{2}{n_f} - 2(\lambda - 1)\left(1 - \frac{n_i}{n_f}\right)}. \tag{C31}$$

It is easy to check that if $f = 2$, the right-hand side tends to $\frac{1}{2}$ as $n_f \to \infty$.

C.3. Incompatibility probabilistic robustness

C.3.1. Lower bound. For this measure, a natural idea would be to mix the terms $\mathrm{tr}(B_b)A_a + \mathrm{tr}(A_a)B_b$ used for η^{d} with the terms $\sqrt{n_A/n_B}A_a + \sqrt{n_B/n_A}B_b$ used for η^{r}. Unfortunately, our efforts in this direction did not lead to any universal lower bound. Nevertheless, this procedure can be used for any fixed pair of measurements to obtain improved lower bounds.

C.3.2. Upper bound. In the main text, we mention in section 4.3.2 that the value of η^{d} given by the qubit MUBs construction is also reachable by η^{p} when the dimension is even. Here we show this fact by adapting the procedure explained in section C.1.2 to the measure η^{p}.

Recall that we consider pairs of rank-one projective measurements (A, B) whose d_i first outcomes live in the first d_i dimensions of the total d_f-dimensional space, and whose $d_f - d_i$ remaining outcomes live in the remaining space. For this structure, an ansatz for the dual for η^{d} given in equation (23) has been presented in equation (C14). However, this ansatz does not satisfy the additional constraints present in the dual for η^{p} given in equation (40), namely, $\mathrm{tr}X_a \leqslant \xi$ and $\mathrm{tr}Y_b \leqslant \upsilon$.

Assume now that $d_f = md_i$, where m is a positive integer, and that the structure of the pair of rank-one projective measurements (A, B) is the following

$$A_a = \begin{cases} \begin{pmatrix} \hat{A}_a & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} & \text{if } 1 \leqslant a \leqslant d_i \\ \quad\vdots & \quad\vdots \\ \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{A}_{a-(m-1)d_i} \end{pmatrix} & \text{if } (m-1)d_i + 1 \leqslant a \leqslant md_i \end{cases} \tag{C32}$$

and similarly for B_b with respect to $\hat{B}_b$, where there are m blocks in the matrices we write and where $\hat{A}$ and $\hat{B}$ are rank-one projective measurements acting on a d_i-dimensional space. We can apply the procedure from section C.1.2 to each d_i-dimensional subspace of the total d_f-dimensional space to get a pair of dual variables $(X^{(l)}, Y^{(l)})$ for each $l \in \{1, 2,\dots,m\}$. Then, if we define

$$X_a = \frac{1}{m}\sum_{l=1}^{m} X^{(l)}_{a-(l-1)d_i} \quad \text{and} \quad Y_b = \frac{1}{m}\sum_{l=1}^{m} Y^{(l)}_{b-(l-1)d_i}, \tag{C33}$$

it clearly satisfies all constraints of the dual for η^{p} given in equation (40), including the trace constraint by symmetry. This implies that for the specific block structure of equation (C32), the upper bound obtained in equation (C22) for η^{d} remains valid for η^{p}.

As an example, consider the measurement pair defined in equation (84). For this instance, we have $d_i = 2$ and $d_f = 4$. The above procedure gives the same bound as for η^{d}, which is given in equation (C23) by setting $d = 4$.

C.4. Incompatibility jointly measurable robustness

For this measure, we combine the results of section C.1.1 with the relation between η^{d} and η^{jm} obtained in equation (B7). Specifically, in the primal in equation (45), the parent POVM G_{ab} will be exactly the one we used for η^{d} in equation (27) of the main text, that is, equation (C3) with x and y given in equation (C9), while the parent POVM H_{ab} will be of the form given in (C1) with $\tilde{\alpha}_{ab} = -x\mathrm{tr}B_b$, $\tilde{\beta}_{ab} = -x\mathrm{tr}A_a$, and $\gamma_{ab} = y\mathrm{tr}A_a\mathrm{tr}B_b$, so that

$$H_{ab} = \frac{1}{2(1-dx) + d^2y}(\{A_a, B_b\} - x(A_a\mathrm{tr}B_b + B_b\mathrm{tr}A_a) + y\mathrm{tr}A_a\mathrm{tr}B_b\mathbb{1}). \tag{C34}$$

Note that such a choice gives rise to a valid parent POVM for the noise considered in equation (B3), namely, $(\{[\mathrm{tr}(A_a)\mathbb{1} - \epsilon A_a]/(d-\epsilon)\}_a, \{[\mathrm{tr}(B_b)\mathbb{1} - \epsilon B_b]/(d-\epsilon)\}_b)$, where

$$\epsilon = \frac{dx-2}{dy-x}. \tag{C35}$$

Then we aim at maximising ϵ under the constraint that the operators H_{ab} of equation (C34) are positive. Since the only difference between equations (C34) and (C3) is the sign of the middle term, the allowed region corresponds to the reflection about $x = 0$ of the allowed region given in equation (C7). An analysis very similar to the one detailed in section C.1.1 can be done in order to show that the optimal point is reached for

$$x = \frac{2 + \sqrt{d^2 + 4d - 4}}{d} \quad \text{and} \quad y = \left(\frac{d + 2 + \sqrt{d^2 + 4d - 4}}{2d}\right)^2, \tag{C36}$$

which corresponds to the feasible point presented in equation (50) of the main text. Note that, similarly to the case of η^{d}, these values of x and y correspond to a critical overlap:

$$c^{\mathrm{jm}}_{\mathrm{crit}} = \frac{-d + 2 + \sqrt{d^2 + 4d - 4}}{2d} \leqslant \frac{1}{\sqrt{d}}. \tag{C37}$$

Note that this coincides with the MUB overlap only in dimension $d = 2$.

To obtain a measurement-dependent refinement of the universal bound given in equation (50), we follow the approach described in section C.1.1, i.e. we maximise ϵ over a larger region determined by the values of c_{ab} present in the specific measurement pair we consider. In an analogous manner we introduce c_-^{jm} and c_+^{jm}, where the former is taken to be 0 if no overlap is smaller that the critical one. Finally we obtain the following measurement-dependent bound:

$$\eta^{\mathrm{jm}}_{A,B} \geqslant \eta^{\mathrm{d,low}}_{A,B} + \frac{1-\eta^{\mathrm{d,low}}_{A,B}}{d} \cdot \frac{(1+c^{\mathrm{jm}}_{-}+c^{\mathrm{jm}}_{+})d-2}{(1+c^{\mathrm{jm}}_{-}+c^{\mathrm{jm}}_{+})(d-1)+c^{\mathrm{jm}}_{-}c^{\mathrm{jm}}_{+}d}, \tag{C38}$$

where $\eta^{\mathrm{d,low}}_{A,B}$ was defined in equation (C12). Note that the optimisations of the two parent POVMs appearing in the primal given in equation (45) were performed separately. A better bound could in principle be obtained by optimising over both POVMs at the same time, but we leave this task open for future work.

C.5. Incompatibility generalised robustness

For a pair (A, B) of rank-one measurements, an ansatz of the form (C1) that is easy to analyse is defined by $\tilde{\alpha}_{ab} = (x + yc_{ab}^2)\mathrm{tr}B_b$, $\tilde{\beta}_{ab} = (x + yc_{ab}^2)\mathrm{tr}A_a$, and $\gamma_{ab} = 0$, where

$$c_{ab} = \sqrt{\frac{\mathrm{tr}(A_a B_b)}{\mathrm{tr}A_a \mathrm{tr}B_b}} \tag{C39}$$

if $\mathrm{tr}A_a\mathrm{tr}B_b > 0$ and $c_{ab} = 0$ otherwise. Then

$$G_{ab} = \frac{1}{2(1+dx+y)}(\{A_a, B_b\} + (x+yc_{ab}^2)(A_a\mathrm{tr}B_b + B_b\mathrm{tr}A_a)). \tag{C40}$$

If $\mathrm{tr}A_a\mathrm{tr}B_b = 0$ we immediately see that $G_{ab} = 0$, so we only need to check positivity in the case $\mathrm{tr}A_a\mathrm{tr}B_b > 0$. Under the assumption that $x, y \geqslant 0$ we deduce from equation (C2) that in order to have $G_{ab} \geqslant 0$, we should have

$$x + yc_{ab}^2 + c_{ab}^2 \pm (x + yc_{ab}^2 + 1)c_{ab} \geqslant 0. \tag{C41}$$

As shown in section 3.5.3 of the main text the corresponding visibility reads

$$\eta = \frac{2 + \left(1+\frac{1}{d}\right)(dx+y)}{2(1+dx+y)}. \tag{C42}$$

The goal is to maximise this η in the positivity region of all G_{ab}. Then a similar analysis to that of η^{d} leads to the maximum $\eta = (1 + 1/\sqrt{d})/2$ achieved by the point $x = 1/(2\sqrt{d})$ and $y = \sqrt{d}/2$ presented in equation (58).

As before to obtain a measurement-dependent refinement we define the critical overlap $c^{\mathrm{g}}_{\mathrm{crit}} = 1/\sqrt{d}$ (note that this coincides with the MUB overlap in every dimension). If one of the overlaps equals $c^{\mathrm{g}}_{\mathrm{crit}}$ no improvement can be obtained, so from now we assume all the overlaps to be different from $c^{\mathrm{g}}_{\mathrm{crit}}$. Let c^{g}_{-} be the biggest overlap smaller than $c^{\mathrm{g}}_{\mathrm{crit}}$ and c^{g}_{+} the smallest bigger than $c^{\mathrm{g}}_{\mathrm{crit}}$. The optimal point corresponds to

$$x = \frac{c^{\mathrm{g}}_{-}c^{\mathrm{g}}_{+}}{c^{\mathrm{g}}_{+}+c^{\mathrm{g}}_{-}} \quad \text{and} \quad y = \frac{1}{c^{\mathrm{g}}_{+}+c^{\mathrm{g}}_{-}} \tag{C43}$$

and gives the following measurement-dependent refinement:

$$\eta^{\mathrm{g}}_{A,B} \geqslant \frac{2(c^{\mathrm{g}}_{-}+c^{\mathrm{g}}_{+})d + (1+c^{\mathrm{g}}_{-}c^{\mathrm{g}}_{+}d)(d+1)}{2d(1+c^{\mathrm{g}}_{-}+c^{\mathrm{g}}_{+}+c^{\mathrm{g}}_{-}c^{\mathrm{g}}_{+}d)}. \tag{C44}$$

Contrary to the measurement-dependent bounds on η^{d} and η^{jm}, namely, equations (C12) and (C38), whenever c^{g}_{-} tends to 0 and c^{g}_{+} tends to 1, this bound tends to $(3d+1)/(4d) \neq 1$. This is due to the fact that the ansatz given in equation (C40) does not contain the identity term, as including such a term makes the optimisation procedure difficult. Therefore in some cases a better measurement-dependent lower bound on η^{g} is obtained by plugging equation (C12) into (B9), which gives

$$\eta^{\mathrm{g}}_{A,B} \geqslant \frac{1 + c^{\mathrm{d}}_{-} + c^{\mathrm{d}}_{+} + c^{\mathrm{d}}_{-}c^{\mathrm{d}}_{+}d}{2 + 2(c^{\mathrm{d}}_{-}+c^{\mathrm{d}}_{+}-1)d + (1-c^{\mathrm{d}}_{-})(1-c^{\mathrm{d}}_{+})d^2}. \tag{C45}$$

Appendix D. Details of the path used in figure 5

In figure 5 of the main text, we plot the value of the studied incompatibility measures on a continuous path. Recall that we fix the first measurement to correspond to the computational basis, so the path is determined by the second measurement and it leads from B^{dev} through B^{qMUB} to B^{MUB}. In this section we provide an explicit description of this path.

The trajectory from B^{dev} to B^{qMUB} corresponds to the interval $\theta \in [\pi/4,\ \pi/2]$ for

$$B_b(\theta) = U(\theta)|b\rangle\langle b|U(\theta)^\dagger, \quad \text{where} \quad U(\theta) = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{\sin\theta}{\sqrt{2}} & \frac{\cos\theta}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{\sin\theta}{\sqrt{2}} & -\frac{\cos\theta}{\sqrt{2}} \\ 0 & -\cos\theta & \sin\theta \end{pmatrix}. \tag{D1}$$

It is easy to check that $\theta = \pi/4$ corresponds to B^{dev} defined in equation (79), while $\theta = \pi/2$ corresponds to B^{qMUB} defined in equation (78).

For the second part of the path, let us first explicitly state our choice of the basis B unbiased to A in dimension $d = 3$:

$$B_b^{\mathrm{MUB}} = U|b\rangle\langle b|U^\dagger, \quad \text{where} \quad U = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 1 & 1 \\ 1 & \mathrm{e}^{\frac{4i\pi}{3}} & \mathrm{e}^{\frac{2i\pi}{3}} \\ 1 & \mathrm{e}^{\frac{2i\pi}{3}} & \mathrm{e}^{\frac{4i\pi}{3}} \end{pmatrix}. \tag{D2}$$

We now choose a particular unitary V that maps B^{qMUB} to B^{MUB}:

$$V = \begin{pmatrix} \frac{\sqrt{2}}{\sqrt{3}} & \frac{\sqrt{3}+3i}{6\sqrt{2}} & \frac{\sqrt{3}-3i}{6\sqrt{2}} \\ 0 & \frac{\sqrt{3}-i}{2\sqrt{2}} & \frac{\sqrt{3}+i}{2\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{-\sqrt{3}-3i}{6} & \frac{-\sqrt{3}+3i}{6} \end{pmatrix}. \tag{D3}$$

To generate a continuous path we compute the principal matrix logarithm of V, i.e. we find a Hermitian matrix H that satisfies $V = \mathrm{e}^{iH}$ and whose spectrum is contained in $(-\pi,\ \pi]$. The path is given by $\mathrm{e}^{itH}B^{\mathrm{qMUB}}\mathrm{e}^{-itH}$ for $t \in [0,\ 1]$, which clearly gives B^{qMUB} for $t = 0$ and B^{MUB} for $t = 1$.

Appendix E. Larger sets of measurements

In this appendix, we generalise some notions and techniques introduced in the main text to larger sets of measurements. The notation of pairs used through the main text, namely, A_a and B_b, was useful for clarity. However, for more measurements we opt for another notation taken from nonlocality: $A_{a|x}$, where $x = 1\ \ldots\ k$ labels the measurement performed and $a = 1\ \ldots\ n_x$ is its outcome. In the following, we will refer to the set of measurements $\{\{A_{a|x}\}_a\}_x$ simply as $\{A_{a|x}\}$, dropping the indices, and we will use $\sum_{a,x}$ as a shorthand for $\sum_{x=1}^{k}\sum_{a=1}^{n_x}$. Similarly to definition 1 in the main text, we say that a set of POVMs $\{A_{a|x}\}$ is compatible if there exists a parent POVM $G_{\vec{j}}$, where $\vec{j} = j_1 j_2\ \ldots\ j_k$ and $j_x \in \{1,\ldots,n_x\}$, such that $\sum_{\vec{j}}\delta_{j_x,a}G_{\vec{j}} = A_{a|x}$, that is, we obtain the original POVM elements as marginals of the parent POVM.

Similarly to section 2.2, we can define noise models through the maps $\mathbf{N}\colon \mathbf{POVM}_d^{n_1,\ldots,n_k} \to \mathbb{P}(\mathbf{POVM}_d^{n_1,\ldots,n_k})$ such that $\mathbf{N}\colon \{A_{a|x}\} \mapsto \mathbf{N}_{\{A_{a|x}\}} \subseteq \mathbf{POVM}_d^{n_1,\ldots,n_k}$. Given a noise model such that each noise set contains at least one jointly measurable set of measurements, we can define the corresponding incompatibility robustness measure, similarly to definition 3,

$$\eta^*_{\{A_{a|x}\}} = \sup_{\substack{\eta\in[0,1] \\ \{N_{a|x}\}\in\mathbf{N}_{\{A_{a|x}\}}}} \{\eta \mid \eta\cdot\{A_{a|x}\} + (1-\eta)\cdot\{N_{a|x}\} \in \mathbf{JM}\}.$$

For these measures, the properties discussed in sections 2.2 and 2.3 can also be naturally generalised to larger sets of measurements, together with the corresponding properties of the noise models $\mathbf{N}$. Then it is straightforward to see that the general measures satisfy the same properties as the ones discussed for pairs in section 3. These general versions can also be formulated as SDPs, and in the remainder of this appendix we present these SDP formulations and provide lower and upper bounds on the measures.

E.1. SDP

Here we write the formulations of all the measures introduced in the main text as SDPs.

$$
\eta^{\mathrm{d}}_{\{A_{a|x}\}} = \begin{cases} \max\limits_{\eta,\{G_{\vec{j}}\}_{\vec{j}}} & \eta \\ \text{s.t.} & G_{\vec{j}} \geqslant 0, \quad \eta \leqslant 1 \\ & \sum_{\vec{j}} \delta_{j_x,a} G_{\vec{j}} = \eta A_{a|x} + (1-\eta)\mathrm{tr}A_{a|x}\frac{\mathbb{1}}{d} \end{cases} = \begin{cases} \min\limits_{\{X_{a|x}\}_{a,x}} & 1 + \sum_{a,x}\mathrm{tr}(X_{a|x}A_{a|x}) \\ \text{s.t.} & X_{a|x} = X^{\dagger}_{a|x}, \quad \sum_{a,x}\delta_{j_x,a}X_{a|x} \geqslant 0 \\ & 1 + \sum_{a,x}\mathrm{tr}(X_{a|x}A_{a|x}) \geqslant \sum_{a,x}\frac{\mathrm{tr}A_{a|x}}{d}\mathrm{tr}X_{a|x} \end{cases}
$$

$$
\eta^{\mathrm{r}}_{\{A_{a|x}\}} = \begin{cases} \max\limits_{\eta,\{G_{\vec{j}}\}_{\vec{j}}} & \eta \\ \text{s.t.} & G_{\vec{j}} \geqslant 0, \quad \eta \leqslant 1 \\ & \sum_{\vec{j}} \delta_{j_x,a} G_{\vec{j}} = \eta A_{a|x} + (1-\eta)\frac{\mathbb{1}}{n_x} \end{cases} = \begin{cases} \min\limits_{\{X_{a|x}\}_{a,x}} & 1 + \sum_{a,x}\mathrm{tr}(X_{a|x}A_{a|x}) \\ \text{s.t.} & X_{a|x} = X^{\dagger}_{a|x}, \quad \sum_{a,x}\delta_{j_x,a}X_{a|x} \geqslant 0 \\ & 1 + \sum_{a,x}\mathrm{tr}(X_{a|x}A_{a|x}) \geqslant \sum_{a,x}\frac{1}{n_x}\mathrm{tr}X_{a|x} \end{cases}
$$

$$
\eta^{\mathrm{p}}_{\{A_{a|x}\}} = \begin{cases} \max\limits_{\substack{\eta,\{G_{\vec{j}}\}_{\vec{j}} \\ \{\tilde{p}_{a|x}\}_{a,x}}} & \eta \\ \text{s.t.} & G_{\vec{j}} \geqslant 0, \quad \tilde{p}_{a|x} \geqslant 0, \quad \sum_a \tilde{p}_{a|x} = 1-\eta \\ & \sum_{\vec{j}} \delta_{j_x,a} G_{\vec{j}} = \eta A_{a|x} + \tilde{p}_{a|x}\mathbb{1} \end{cases} = \begin{cases} \min\limits_{\substack{\{X_{a|x}\}_{a,x} \\ \{\xi_x\}_x}} & 1 + \sum_{a,x}\mathrm{tr}(X_{a|x}A_{a|x}) \\ \text{s.t.} & X_{a|x} = X^{\dagger}_{a|x}, \quad \sum_{a,x}\delta_{j_x,a}X_{a|x} \geqslant 0 \\ & 1 + \sum_{a,x}\mathrm{tr}(X_{a|x}A_{a|x}) \geqslant \sum_x \xi_x \\ & \xi_x \geqslant \mathrm{tr}X_{a|x} \end{cases}
$$

$$
\eta^{\mathrm{jm}}_{\{A_{a|x}\}} = \begin{cases} \max\limits_{\substack{\eta, \{G_{\vec{j}}\}_{\vec{j}} \\ \{\tilde{H}_{\vec{j}}\}_{\vec{j}}}} & \eta \\ \text{s.t.} & G_{\vec{j}} \geqslant 0, \quad \sum_{\vec{j}} G_{\vec{j}} = \mathbb{1}, \quad \tilde{H}_{\vec{j}} \geqslant 0 \\ & \sum_{\vec{j}} \delta_{j_x,a}(G_{\vec{j}} - \tilde{H}_{\vec{j}}) = \eta A_{a|x} \end{cases} = \begin{cases} \min\limits_{N,\{X_{a|x}\}_{a,x}} & \mathrm{tr}N \\ \text{s.t.} & N = N^{\dagger}, \quad X_{a|x} = X^{\dagger}_{a|x} \\ & N \geqslant \sum_{a,x}\delta_{j_x,a}X_{a|x} \geqslant 0 \\ & \sum_{a,x}\mathrm{tr}(X_{a|x}A_{a|x}) \geqslant 1 \end{cases}
$$

$$
\eta^{\mathrm{g}}_{\{A_{a|x}\}} = \begin{cases} \max\limits_{\eta,\{G_{\vec{j}}\}_{\vec{j}}} & \eta \\ \text{s.t.} & G_{\vec{j}} \geqslant 0, \quad \sum_{\vec{j}} G_{\vec{j}} = \mathbb{1} \\ & \sum_{\vec{j}} \delta_{j_x,a} G_{\vec{j}} \geqslant \eta A_{a|x} \end{cases} = \begin{cases} \min\limits_{N,\{X_{a|x}\}_{a,x}} & \mathrm{tr}N \\ \text{s.t.} & N = N^{\dagger}, \quad X_{a|x} = X^{\dagger}_{a|x} \\ & N \geqslant \sum_{a,x}\delta_{j_x,a}X_{a|x}, \quad X_{a|x} \geqslant 0. \\ & \sum_{a,x}\mathrm{tr}(X_{a|x}A_{a|x}) \geqslant 1 \end{cases}
$$

E.2. Lower bounds

Here we derive lower bounds on some of the above measures in this general setting.

For η^{d}, the following bound is presented in [16], equation (11)

$$
\eta^{\mathrm{d}}_{\{A_{a|x}\}} \geqslant \frac{1}{k}\left(1 + \frac{k-1}{d+1}\right), \tag{E1}
$$

and from (68) this same bound holds for η^{jm} and η^{g} as well. and from (68) this same bound holds for η^{jm} and η^{g} as well. Here we outline a few ways to improve on this bound.

One option is to apply the universal lower bounds for pairs, derived in the main text, successively on subsets of pairs of measurements. Starting from k measurements, we group them into $k/2$ or $(k+1)/2$ pairs, depending on the parity of k, and we compute the parent POVMs for these pairs defined in equations (28), (50), and (59), corresponding to the universal lower bound. Therefore we end up with $k/2$ or $(k+1)/2$ measurements, which are the parent POVMs. We repeat this process until we end up with only one pair of measurements. Since we use universal lower bounds, the specific pairings do not matter, and we obtain a bound that depends only on k and d. When $k = 2^n$ for instance, we get that the lower bounds on η^{d} and η^{g} are the nth power of the corresponding lower bound for pairs, namely, equations (28) and (59), respectively. Note that whenever k is odd, an asymmetry is introduced by the choice of which measurement is not paired with another one, but we can overcome this problem by symmetrisation.

Let us illustrate this procedure on a triplet of measurements denoted by (A, B, C). For any pair (A, B) of POVMs, we denote by $G(A, B)$ their parent POVM used to derive universal lower bounds in section 3, for instance, equation (28) for η^{d}. Then the following POVM is a parent POVM for noisy versions of A, B, and C, with respect to the noise of η^{d} in this case:

$$\frac{1}{3}[G(G(A, B), C) + G(G(C, A), B) + G(G(B, C), A)]. \tag{E2}$$

For η^{d} and any number of measurements $k \geqslant 3$ and any dimension $d \geqslant 2$, this procedure never improves on equation (E1), except for triplets of qubit measurements for which it gives the bound $(1 + 1/\sqrt{2})/3$. Note that we outperform this bound by completely solving this case of three measurements in dimension two in section E.4.

For η^{jm}, the above procedure is made more complex by the fact that two parent POVMs are necessary. An alternative bound can be obtained by plugging equation (E1) into (B7). Note that this requires the equivalent of ϵ in equation (B3) for more measurements, namely, a dimension-dependent number such that the set $\{(\mathrm{tr}(A_{a|x})\mathbb{1} - \epsilon A_{a|x})/(d - \epsilon)\}$ is jointly measurable. Both procedures are possible and involve suitable combinations of the parent POVMs introduced in this work. However, due to their complexity, we do not present the resulting bounds.

For η^{g}, we should compare the above procedure and the bound obtained by plugging equation (E1) into (B9). For instance, for $k = 3$ and $d = 4$, the former gives $5/8$ and the latter $3/5$.

E.3. Upper bounds

The various upper bounds presented throughout the main text naturally generalise to more measurements. We introduce the generalised quantities corresponding to equation (18)

$$f = \sum_{a,x} \frac{\mathrm{tr}A_{a|x}^2}{d} \quad \text{and} \quad \lambda = \max_{\vec{j}} \left\{ \max \mathrm{Sp}\left(\sum_{a,x} \delta_{j_x,a} A_{a|x} \right) \right\}, \tag{E3}$$

and also those corresponding to equation (19)

$$g^{\mathrm{d}} = \sum_{a,x} \left(\frac{\mathrm{tr}A_{a|x}}{d} \right)^2, \quad g^{\mathrm{r}} = \sum_{x} \frac{1}{n_x}, \quad g^{\mathrm{p}} = \sum_{x} \min_{a} \frac{\mathrm{tr}A_{a|x}}{d}, \quad \text{and} \quad g^{\mathrm{jm}} = \min_{\vec{j}} \left\{ \min \mathrm{Sp}\left(\sum_{a,x} \delta_{j_x,a} A_{a|x} \right) \right\}. \tag{E4}$$

Using these definitions, the feasible points for the duals in section E.1 are

$$\begin{aligned}
X_{a|x} &= \frac{\frac{\lambda}{k}\mathbb{1} - A_{a|x}}{(f - g^{\mathrm{d}})d}, \quad X_{a|x} = \frac{\frac{\lambda}{k}\mathbb{1} - A_{a|x}}{(f - g^{\mathrm{r}})d}, \quad \text{and} \quad X_{a|x} = \frac{\frac{\lambda}{k}\mathbb{1} - A_{a|x}}{(f - g^{\mathrm{p}})d}, \quad \text{for } \eta^{\mathrm{d}}, \eta^{\mathrm{r}}, \text{ and } \eta^{\mathrm{p}}, \text{ respectively,} \\
X_{a|x} &= \frac{A_{a|x} - \frac{g^{\mathrm{jm}}}{k}\mathbb{1}}{(f - g^{\mathrm{jm}})d} \quad \text{and} \quad N = \frac{\lambda - g^{\mathrm{jm}}}{f - g^{\mathrm{jm}}} \cdot \frac{\mathbb{1}}{d} \quad \text{for } \eta^{\mathrm{jm}}, \\
X_{a|x} &= \frac{A_{a|x}}{fd} \quad \text{and} \quad N = \frac{\lambda}{f} \cdot \frac{\mathbb{1}}{d} \quad \text{for } \eta^{\mathrm{g}}.
\end{aligned} \tag{E5}$$

Note that we have implicitly assumed that $f \neq g^*$ for all the measures. From the discussion below it turns out that the equality holds only when all measurement elements are proportional to the identity operator, in which case the set is trivially compatible. These feasible points give rise to the following bounds:

$$\begin{aligned}
&\eta^{\mathrm{d}}_{\{A_{a|x}\}} \leqslant \frac{\lambda - g^{\mathrm{d}}}{f - g^{\mathrm{d}}} = \eta^{\mathrm{d,up}}_{\{A_{a|x}\}}, \quad \eta^{\mathrm{r}}_{\{A_{a|x}\}} \leqslant \frac{\lambda - g^{\mathrm{r}}}{f - g^{\mathrm{r}}} = \eta^{\mathrm{r,up}}_{\{A_{a|x}\}}, \quad \eta^{\mathrm{p}}_{\{A_{a|x}\}} \leqslant \frac{\lambda - g^{\mathrm{p}}}{f - g^{\mathrm{p}}} = \eta^{\mathrm{p,up}}_{\{A_{a|x}\}}, \\
&\eta^{\mathrm{jm}}_{\{A_{a|x}\}} \leqslant \frac{\lambda - g^{\mathrm{jm}}}{f - g^{\mathrm{jm}}} = \eta^{\mathrm{jm,up}}_{\{A_{a|x}\}}, \quad \text{and} \quad \eta^{\mathrm{g}}_{\{A_{a|x}\}} \leqslant \frac{\lambda}{f} = \eta^{\mathrm{g,up}}_{\{A_{a|x}\}}.
\end{aligned} \tag{E6}$$

Note that, from the inequalities in equation (E9) below and under the assumption $f > \lambda$, we have

$$\max\{\eta^{\mathrm{d,up}}_{\{A_{a|x}\}}, \eta^{\mathrm{r,up}}_{\{A_{a|x}\}}\} \leqslant \eta^{\mathrm{p,up}}_{\{A_{a|x}\}} \leqslant \eta^{\mathrm{jm,up}}_{\{A_{a|x}\}} \leqslant \eta^{\mathrm{g,up}}_{\{A_{a|x}\}}. \tag{E7}$$

E.3.1. Discussion of the feasible points. Here we first show that for all sets of measurements, the inequalities

$$f \geqslant g^{\mathrm{d}} \quad \text{and} \quad f \geqslant g^{\mathrm{r}} \tag{E8}$$

hold, with equality if and only if all POVM elements involved are proportional to the identity. Then we also derive the hierarchy used to derive equation (E7), namely,

$$\min\{g^{\mathrm{d}}, g^{\mathrm{r}}\} \geqslant g^{\mathrm{p}} \geqslant g^{\mathrm{jm}} \geqslant 0. \tag{E9}$$

These two inequalities imply that unless all POVM elements are proportional to the identity we have $f > g^*$ and the bounds given in equation (E6) hold (which are generalisations to larger sets of measurements of the upper bounds given in equations (30), (38), (43), (52), and (66) of the main text).

In order to prove the inequalities in equation (E8), we use the Cauchy–Schwarz inequality:

$$(\mathrm{tr}A_{a|x})^2 = [\mathrm{tr}(\mathbb{1}\cdot A_{a|x})]^2 \leqslant \mathrm{tr}(\mathbb{1}^2)\mathrm{tr}(A_{a|x}^2) = d\,\mathrm{tr}(A_{a|x}^2). \tag{E10}$$

For g^{d}, this implies that

$$f = \sum_{a,x}\frac{\mathrm{tr}(A_{a|x}^2)}{d} \geqslant \sum_{a,x}\left(\frac{\mathrm{tr}A_{a|x}}{d}\right)^2 = g^{\mathrm{d}}. \tag{E11}$$

For g^{r}, we also use the concavity of the square-root, which implies that

$$\sqrt{\sum_{a=1}^{n_x}\mathrm{tr}(A_{a|x}^2)} = \sqrt{n_x\sum_{a=1}^{n_x}\frac{1}{n_x}\mathrm{tr}(A_{a|x}^2)} \geqslant \sqrt{n_x}\sum_{a=1}^{n_x}\frac{1}{n_x}\sqrt{\mathrm{tr}(A_{a|x}^2)} \geqslant \frac{1}{\sqrt{n_x}}\sum_{a=1}^{n_x}\sqrt{\frac{(\mathrm{tr}A_{a|x})^2}{d}} = \sqrt{\frac{d}{n_x}}, \tag{E12}$$

where we have used equation (E10) to get the second inequality. This gives

$$f = \sum_{a,x}\frac{\mathrm{tr}(A_{a|x}^2)}{d} = \frac{1}{d}\sum_{x=1}^{k}\left[\sum_{a=1}^{n_x}\mathrm{tr}(A_{a|x}^2)\right] \geqslant \sum_x\frac{1}{n_x} = g^{\mathrm{r}}. \tag{E13}$$

Note that to have equality in the inequality in equation (E10), the eigenvalues of A_a should all be equal, that is, $A_a \propto \mathbb{1}$. This shows that in order to have equality in the inequalities of equation (E8), all measurement operators need to be proportional to the identity.

Regarding equation (E9), the inequality $g^{\mathrm{d}} \geqslant g^{\mathrm{p}}$ comes from

$$g^{\mathrm{d}} = \sum_{x=1}^{k}\sum_{a=1}^{n_x}\left(\frac{\mathrm{tr}A_{a|x}}{d}\right)^2 \geqslant \sum_{x=1}^{k}\left(\min_a\frac{\mathrm{tr}A_{a|x}}{d}\right)\left(\sum_{a=1}^{n_x}\frac{\mathrm{tr}A_{a|x}}{d}\right) = \sum_x\min_a\frac{\mathrm{tr}A_{a|x}}{d} = g^{\mathrm{p}}. \tag{E14}$$

The inequality $g^{\mathrm{r}} \geqslant g^{\mathrm{p}}$ comes from

$$g^{\mathrm{r}} = \sum_x\frac{1}{n_x} = \sum_{x=1}^{k}\frac{1}{n_x}\left(\sum_{a=1}^{n_x}\frac{\mathrm{tr}A_{a|x}}{d}\right) \geqslant \sum_{x=1}^{k}\min_a\frac{\mathrm{tr}A_{a|x}}{d} = g^{\mathrm{p}}. \tag{E15}$$

The inequality $g^{\mathrm{p}} \geqslant g^{\mathrm{jm}}$ comes from $\mathrm{tr}(M) \geqslant d\min\mathrm{Sp}(M)$ for every $d\times d$ Hermitian matrix M, so that

$$\begin{aligned} g^{\mathrm{p}} &= \sum_x\min_a\frac{\mathrm{tr}A_{a|x}}{d} \\ &= \min_{\vec{j}}\left\{\sum_{a,x}\delta_{j_x,a}\frac{\mathrm{tr}A_{a|x}}{d}\right\} = \min_{\vec{j}}\left\{\frac{1}{d}\mathrm{tr}\sum_{a,x}\delta_{j_x,a}A_{a|x}\right\} \geqslant \min_{\vec{j}}\left\{\min\mathrm{Sp}\left(\sum_{a,x}\delta_{j_x,a}A_{a|x}\right)\right\} = g^{\mathrm{jm}}. \end{aligned} \tag{E16}$$

Lastly, the inequality $g^{\mathrm{jm}} \geqslant 0$ comes from the positivity of the POVM elements involved in its definition, which concludes the proof of equation (E9).

E.3.2. Alternative upper bounds. Here we provide alternative feasible points for the duals in Section (E.1) that give rise to upper bounds that are in some cases tighter than the ones discussed above. Let us consider sets of POVMs $\{A_{a|x}\}$ such that no POVM element is zero. We can define new quantities very similar to the ones of equations (E3) and (E4), namely,

$$\begin{aligned} f_{\mathrm{tr}} &= \sum_{a,x}\frac{\mathrm{tr}A_{a|x}^2}{d\,\mathrm{tr}A_{a|x}}, \quad \lambda_{\mathrm{tr}} = \max_{\vec{j}}\left\{\max\mathrm{Sp}\left(\sum_{a,x}\delta_{j_x,a}\frac{A_{a|x}}{\mathrm{tr}A_{a|x}}\right)\right\}, \\ g_{\mathrm{tr}}^{\mathrm{d}} &= g_{\mathrm{tr}}^{\mathrm{r}} = g_{\mathrm{tr}}^{\mathrm{p}} = \frac{k}{d}, \quad \text{and } g_{\mathrm{tr}}^{\mathrm{jm}} = \min_{\vec{j}}\left\{\min\mathrm{Sp}\left(\sum_{a,x}\delta_{j_x,a}\frac{A_{a|x}}{\mathrm{tr}A_{a|x}}\right)\right\}. \end{aligned} \tag{E17}$$

Using these we can derive bounds similar to those in equation (E6):

$$\text{for } \eta^{\mathrm{d}}, \eta^{\mathrm{r}}, \text{ and } \eta^{\mathrm{p}}, \quad X_{a|x} = \frac{\frac{\lambda_{\mathrm{tr}}}{k}\mathbb{1} - \frac{A_{a|x}}{\mathrm{tr}A_{a|x}}}{(f_{\mathrm{tr}} - g_{\mathrm{tr}}^{\mathrm{p}})d} \quad \text{so that} \quad \max\{\eta^{\mathrm{d}}_{\{A_{a|x}\}}, \eta^{\mathrm{r}}_{\{A_{a|x}\}}\} \leqslant \eta^{\mathrm{p}}_{\{A_{a|x}\}} \leqslant \frac{\lambda_{\mathrm{tr}} - g_{\mathrm{tr}}^{\mathrm{p}}}{f_{\mathrm{tr}} - g_{\mathrm{tr}}^{\mathrm{p}}}, \tag{E18}$$

$$\text{for } \eta^{\mathrm{jm}}, \quad X_{a|x} = \frac{\frac{A_{a|x}}{\mathrm{tr}A_{a|x}} - \frac{g_{\mathrm{tr}}^{\mathrm{jm}}}{k}\mathbb{1}}{(f_{\mathrm{tr}} - g_{\mathrm{tr}}^{\mathrm{jm}})d} \quad \text{and} \quad N = \frac{\lambda_{\mathrm{tr}} - g_{\mathrm{tr}}^{\mathrm{jm}}}{f_{\mathrm{tr}} - g_{\mathrm{tr}}^{\mathrm{jm}}}\cdot\frac{\mathbb{1}}{d} \quad \text{so that} \quad \eta^{\mathrm{jm}}_{\{A_{a|x}\}} \leqslant \frac{\lambda_{\mathrm{tr}} - g_{\mathrm{tr}}^{\mathrm{jm}}}{f_{\mathrm{tr}} - g_{\mathrm{tr}}^{\mathrm{jm}}}, \tag{E19}$$

$$\text{for } \eta^{\mathrm{g}}, \quad X_{a|x} = \frac{A_{a|x}}{f_{\mathrm{tr}}\, d \operatorname{tr} A_{a|x}} \quad \text{and} \quad N = \frac{\lambda_{\mathrm{tr}}}{f_{\mathrm{tr}}} \cdot \frac{\mathbb{1}}{d} \quad \text{so that} \quad \eta^{\mathrm{g}}_{\{A_{a|x}\}} \leqslant \frac{\lambda_{\mathrm{tr}}}{f_{\mathrm{tr}}}. \tag{E20}$$

Similarly to as in section E.3.1, the inequalities $f \geqslant g_{\mathrm{tr}}^{\mathrm{d}} = g_{\mathrm{tr}}^{\mathrm{r}} = g_{\mathrm{tr}}^{\mathrm{p}} \geqslant g_{\mathrm{tr}}^{\mathrm{jm}} \geqslant 0$ hold and give natural relations between the bounds.

For the qubit measurements mentioned in section 4.2, namely, any rank-one POVM pair such that $A_a = |a\rangle\langle a|$ and the Bloch vectors of B lie on the xy-plane of the Bloch sphere, the parameters in equation (E17) are $f_{\mathrm{tr}} = 2$ as such a pair is rank-one, $\lambda_{\mathrm{tr}} = 1 + 1/\sqrt{2}$, and $g_{\mathrm{tr}}^{\mathrm{jm}} = 1 - 1/\sqrt{2}$ due to the orthogonality of the Bloch vectors of the POVM elements of A and B. Therefore the upper bounds in equations (E18)–(E20) coincide with the MUB values given in equation (69).

For rank-one projective pairs of measurements the bounds in equations (E18)–(E20) coincide with their counterparts in equation (E6), but in general they are incomparable, that is, for different measurement pairs one or the other might give the lower value. For the pair $(A^{\Lambda}, B^{\Lambda})$ used in Counterexample 1, the bound on η^{d} in equation (E18) gives $3(\sqrt{13} + 1)/10 \approx 1.3817$, whereas the one in equation (E6) gives $(9\sqrt{2} - 1)/14 \approx 0.8377$. On the other hand, for the pair (A^{β}, B) used in Counterexample 3, the bound on η^{d} in equation (E18) gives $1/\sqrt{2} \approx 0.7071$, whereas the one in equation (E6) gives $(4\sqrt{2} + 1)/7 \approx 0.9510$. This incomparability suggests that there may exist a more general way to construct such upper bounds, e.g. involving polynomials in $A_{a|x}$ in the definition of $X_{a|x}$. We leave this question open for further work.

E.3.3. Tightness of the upper bound on η^{g} for MUBs. We investigate the tightness of the upper bound on η^{g} in equation (E6) for various MUB constructions. The relation (B9) between η^{d} and η^{g} is obviously also valid for more than two measurements. Therefore, the cases in which the bounds on η^{d} in [36] are tight, that is, $\eta^{\mathrm{d}} = (\lambda - k/d)/(k - k/d)$, give rise to tight upper bounds on η^{g} as well. This is because in this case equation (B9) reads $\lambda/k \leqslant \eta^{\mathrm{g}}$, which saturates the upper bound for η^{g} in equation (E6). In particular, for the standard construction of MUBs in prime power dimensions [20] the bound on η^{g} in equation (E6) is tight when $k = d$ and $k = d + 1$.

The methods developed in [36] can also be applied to show the tightness of the upper bound on η^{g} in equation (E6) in some additional cases. Specifically, applying the ansatz [36], equation (11) to the incompatibility generalised robustness primal leads to optimal constructions in some cases. In particular, when the dimension is $d = 2^r$, all subsets of size $k \in \{2, 3,\ldots,d + 1\}$ of the standard construction of complete sets of MUBs saturate the upper bound on η^{g} in equation (E6).

To show this, we use the notation of [36], appendix D. In this work the authors show that for the standard MUB construction the marginals along j_x of the operator $G_{\vec{j}}$ defined in [36], equation (11) are diagonal in the basis $\{|\varphi_a^x\rangle\}_a$. Thus, the corresponding value of η in the incompatibility generalised robustness primal is

$$\eta = \min_{a,x} \langle\varphi_a^x|\left(\sum_{\vec{j}} \delta_{j_x,a} G_{\vec{j}}\right)|\varphi_a^x\rangle. \tag{E21}$$

Moreover, by definition [36], equation (11) we have

$$\sum_{a,x}\langle\varphi_a^x|\left(\sum_{\vec{j}} \delta_{j_x,a} G_{\vec{j}}\right)|\varphi_a^x\rangle = \sum_{\vec{j}} \operatorname{tr}\left(G_{\vec{j}} \sum_{a,x} \delta_{j_x,a}|\varphi_a^x\rangle\langle\varphi_a^x|\right) = \sum_{\vec{j}} \operatorname{tr}(G_{\vec{j}} S_{\vec{j}}) = \sum_{\vec{j}} \operatorname{tr}(\lambda G_{\vec{j}}) = \lambda d. \tag{E22}$$

Therefore, if all $\langle\varphi_a^x|\sum_{\vec{j}} \delta_{j_x,a} G_{\vec{j}}|\varphi_a^x\rangle$ are equal, regardless of a and x, we can replace the minimum in equation (E21) by the total sum divided by the number of terms:

$$\eta = \frac{1}{kd}\sum_{a,x}\langle\varphi_a^x|\left(\sum_{\vec{j}} \delta_{j_x,a} G_{\vec{j}}\right)|\varphi_a^x\rangle = \frac{\lambda}{k}, \tag{E23}$$

and λ/k is a lower bound for $\eta^{\mathrm{g}}_{\{|\varphi_a^x\rangle\langle\varphi_a^x|\}}$. As it coincides with the upper bound for η^{g} in equation (E6) (recall that $f = k$ for rank-one measurements), the tightness of this upper bound follows.

When $d = 2^r$, one can see from [36], appendix D 3 that $\langle\varphi_a^x|\sum_{\vec{j}} \delta_{j_x,a} G_{\vec{j}}|\varphi_a^x\rangle$ is indeed independent of a and x. This shows that if $d = 2^r$, then for the standard construction of MUBs, we have $\eta^{\mathrm{g}} = \lambda/k$ for all sets of $k \in \{2, 3,\ldots,d + 1\}$ projective measurements onto k MUBs.

Another interesting example is given by triplets of MUBs in dimension $d = 4$. From [54], we know that all possible triplets can be parametrised by three (real) parameters. For η^{d}, depending on the choice of these parameters, we get a different robustness, whereas for η^{g}, they all give the same value, namely, 2/3.

Note however that the bound on η^{g} in equation (E6) is not always tight for MUBs. For $k = 4$ MUBs in dimension $d = 5$, we get $\eta^{\mathrm{g}} \approx 0.5692 < 0.5693 \approx \lambda/3$.

E.3.4. Upper bound on η^{jm} for MUBs. Below we show that for the standard construction of MUBs in odd prime power dimensions [55], the bounds on η^{jm} and η^{g} in equation (E6) coincide. In order to show this, we need to prove that in this case $g^{\mathrm{jm}} = 0$ for all $k \in \{2, 3,\ldots,d + 1\}$. When $k < d$, this is clear. For $k = d$ and $k = d + 1$, this minimum eigenvalue is reached, for instance, when we pick the first POVM element of each measurement. We first use the notations of [36], appendix D 2 to prove the general case and then give a simplified proof in the case of prime dimensions.

Here we give the proof when $k = d = p^r$, with p prime and $x \in \mathbb{F}_d$, the Galois field with d elements, which singles out a particular choice of d MUBs that does not include the computational basis. The other cases, namely, $k = d + 1$ and $k = d$ with one of the bases being the computational basis, can be treated similarly. Recall that g^{jm} concerns the spectra of the operators $\sum_{a,x}\delta_{j_x,a}A_{a|x}$ for every $\vec{j}$, see equation (E4). If we choose $\vec{j} = \vec{0}$, we get

$$\sum_{x\in\mathbb{F}_d}|\varphi_0^x\rangle\langle\varphi_0^x| = \frac{1}{d}\sum_{l,l'\in\mathbb{F}_d}\sum_{x\in\mathbb{F}_d}\mathrm{e}^{\frac{2\mathrm{i}\pi}{p}\mathrm{Tr}[x(l^2-l'^2)]}|l\rangle\langle l'| = \mathbb{1} + \sum_{l\in\mathbb{F}_d^*}|l\rangle\langle -l|, \tag{E24}$$

where the trace over the Galois field $\mathbb{F}_d$ is defined by $\mathrm{Tr}\,a = a + a^2 + \cdots + a^{p-1}$ so that it belongs to $\{0, 1,\ldots,p - 1\}$. Note that the convention used here to label the POVM elements is different than the one in the main text, as it starts from 0 instead of 1. For the operator in equation (E24), the vector $(|l\rangle - |-l\rangle)$ is an eigenvector with eigenvalue 0 for $l \in \mathbb{F}_d\backslash\{0\}$, which concludes the proof.

As an easier illustration, we consider the case when d is an odd prime. In this case, a complete set of MUBs is given by the computational basis $\{|l\rangle\}_{l=0}^{d-1}$ and

$$|\varphi_a^x\rangle = \frac{1}{\sqrt{d}}\sum_{l=0}^{d-1}\mathrm{e}^{\frac{2\mathrm{i}\pi}{d}(xl^2+al)}|l\rangle, \tag{E25}$$

where x labels the bases and a the vectors. equation (E24) then takes the form

$$\sum_{x=0}^{d-1}|\varphi_0^x\rangle\langle\varphi_0^x| = \frac{1}{d}\sum_{l,l'=0}^{d-1}\sum_{x=0}^{d-1}\mathrm{e}^{\frac{2\mathrm{i}\pi}{d}[x(l^2-l'^2)]}|l\rangle\langle l'| = \mathbb{1} + \sum_{l=1}^{d-1}|l\rangle\langle -l|, \tag{E26}$$

for which $|l\rangle - |-l\rangle$ is an eigenvector with eigenvalue 0 for $l \in \{1, 2,\ldots,d - 1\}$.

E.4. Most incompatible triplets of qubit measurements

Below we analyse the incompatibility robustness of a triplet of qubit MUBs, and show that they are among the most incompatible triplets in dimension 2 under η^{d}, η^{p}, η^{jm}, and η^{g}. For a triplet of projective measurements onto three qubit MUBs (A^{MUB}, B^{MUB}, C^{MUB}), the quantities defined in equations (E3) and (E4) are $f = 3$, $\lambda = (3 + \sqrt{3})/2$, $g^{\mathrm{d}} = g^{\mathrm{p}} = 3/2$, and $g^{\mathrm{jm}} = (3 - \sqrt{3})/2$, so that the bounds of equation (E6) read

$$\eta^{\mathrm{d}}_{3\mathrm{MUB}} \leqslant \eta^{\mathrm{p}}_{3\mathrm{MUB}} \leqslant \frac{1}{\sqrt{3}},\quad \eta^{\mathrm{jm}}_{3\mathrm{MUB}} \leqslant \sqrt{3} - 1,\quad \text{and}\quad \eta^{\mathrm{g}}_{3\mathrm{MUB}} \leqslant \frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right), \tag{E27}$$

where we write $\eta^{*}_{3\mathrm{MUB}}$ to denote the incompatibility robustness of (A^{MUB}, B^{MUB}, C^{MUB}).

Now we derive universal lower bounds for the above measures for triplets of qubit measurements, and show that a triplet of MUBs saturates these. We start with η^{d}, which is post-processing monotonic, and therefore it is enough to derive bounds on it for rank-one triplets (A, B, C), for which we introduce

$$\begin{aligned} G_{abc} &= \frac{1}{2(9 - \sqrt{3})}\Bigg\{[A_aB_bC_c + A_aC_cB_b + B_bC_cA_a + B_bA_aC_c + C_cA_aB_b + C_cB_bA_a] + \frac{3\sqrt{3} - 4}{2} \\ &\quad \times [\mathrm{tr}(B_b)\mathrm{tr}(C_c)A_a + \mathrm{tr}(A_a)\mathrm{tr}(C_c)B_b + \mathrm{tr}(A_a)\mathrm{tr}(B_b)C_c] + \frac{9 - 5\sqrt{3}}{2}\mathrm{tr}(A_a)\mathrm{tr}(B_b)\mathrm{tr}(C_c)\mathbb{1}\Bigg\}. \end{aligned} \tag{E28}$$

We show that this is a valid feasible point for the primal for η^{d} in section E.1 together with $\eta = 1/\sqrt{3}$. The correctness of the marginals is immediate. The positivity follows from a tedious but straightforward computation in which we express the eigenvalues of G_{abc} as functions of the overlaps between A_a and B_b, B_b and C_c, and C_c and A_a (which is possible, because we are dealing with 2×2 matrices). This shows that

$$\eta^{\mathrm{d}}_{A,B,C} \geqslant \frac{1}{\sqrt{3}}, \tag{E29}$$

which also holds for non-rank-one triplets by post-processing monotonicity of this measure.

Regarding the other measures, the above inequality immediately holds for η^{p} due to the obvious generalisation of equation (68) to triplets of measurements.

For η^{jm}, the method described in equation (B2) can be used for triplets as well to get $\eta^{\mathrm{d}}_{A,B,C} + (1 - \eta^{\mathrm{d}}_{A,B,C})\epsilon/2 \leqslant \eta^{\mathrm{g}}_{A,B,C}$, where $\epsilon = \sqrt{3} - 1$ because

$$\frac{\mathrm{tr}(A_a)\mathbb{1}-(\sqrt{3}-1)A_a}{3-\sqrt{3}}=\frac{1}{\sqrt{3}}[\mathrm{tr}(A_a)\mathbb{1}-A_a]+\left(1-\frac{1}{\sqrt{3}}\right)\mathrm{tr}[\mathrm{tr}(A_a)\mathbb{1}-A_a]\frac{\mathbb{1}}{2}, \tag{E30}$$

and similarly for B_b and C_c. The validity of ϵ is then guaranteed by applying the bound obtained just above on η^{d} to the measurements $(\{\mathrm{tr}(A_a)\mathbb{1}-A_a\}_a, \{\mathrm{tr}(B_b)\mathbb{1}-B_b\}_b, \{\mathrm{tr}(C_c)\mathbb{1}-C_c\}_c)$.

For η^{g}, the method described in equation (B9) can be used for triplets as well to get $\eta^{\mathrm{d}}_{A,B,C}+(1-\eta^{\mathrm{d}}_{A,B,C})/2\leqslant\eta^{\mathrm{g}}_{A,B,C}$.

Therefore, we have proven that

$$\frac{1}{\sqrt{3}}\leqslant\eta^{\mathrm{d}}_{A,B,C}\leqslant\eta^{\mathrm{p}}_{A,B,C},\quad\sqrt{3}-1\leqslant\eta^{\mathrm{jm}}_{A,B,C},\quad\text{and}\quad\frac{1}{2}\left(1+\frac{1}{\sqrt{3}}\right)\leqslant\eta^{\mathrm{g}}_{A,B,C}. \tag{E31}$$

As a triplet of projective measurements onto three qubit MUBs reaches these lower bounds from equation (E27), they are among the most incompatible triplets of qubit measurements with respect to η^{d}, η^{p}, η^{jm}, and η^{g}.

ORCID iDs

Sébastien Designolle https://orcid.org/0000-0003-0303-3556
Jędrzej Kaniewski https://orcid.org/0000-0003-1133-3786

References

[1] Ludwig G 1954 *Die Grundlagen der Quantenmechanik* (Berlin: Springer)
[2] Busch P, Lahti P J, Pellonpää J-P and Ylinen K 2016 *Quantum Measurement* (Berlin: Springer) (https://doi.org/10.1007/978-3-319-43389-9)
[3] Heinosaari T and Wolf M M 2010 Non-disturbing quantum measurements *J. Math. Phys.* **51** 092201
[4] Reeb D, Reitzner D and Wolf M M 2013 Coexistence does not imply joint measurability *J. Phys. A: Math. Theor.* **46** 462002
[5] Wolf M M, Perez-Garcia D and Fernandez C 2009 Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory *Phys. Rev. Lett.* **103** 230402
[6] Quintino M T, Vértesi T and Brunner N 2014 Joint measurability, Einstein–Podolsky–Rosen steering, and Bell nonlocality *Phys. Rev. Lett.* **113** 160402
[7] Uola R, Budroni C, Gühne O and Pellonpää J-P 2015 One-to-one mapping between steering and joint measurability problems *Phys. Rev. Lett.* **115** 230402
[8] Tavakoli A and Uola R 2019 Measurement incompatibility and steering are necessary and sufficient for operational contextuality arXiv:1905.03614
[9] Leonardo Guerini M T Q and Aolita L 2019 Distributed sampling, quantum communication witnesses, and measurement incompatibility *Phys. Rev.* A **100** 042308
[10] Coecke B, Fritz T and Spekkens R W 2016 A mathematical theory of resources *Inf. Comput.* **250** 59–86
[11] Fritz T 2017 Resource convertibility and ordered commutative monoids *Math. Struct. Comput. Sci.* **27** 850–938
[12] Heinosaari T, Kiukas J and Reitzner D 2015 Noise robustness of the incompatibility of quantum measurements *Phys. Rev.* A **92** 022115
[13] Skrzypczyk P, Šupić I and Cavalcanti D 2019 All sets of incompatible measurements give an advantage in quantum state discrimination *Phys. Rev. Lett.* **122** 130403
[14] Chitambar E and Gour G 2019 Quantum resource theories *Rev. Mod. Phys.* **91** 025001
[15] Oszmaniec M and Biswas T 2019 Operational relevance of resource theories of quantum measurements *Quantum* **3** 133
[16] Heinosaari T, Miyadera T and Ziman M 2016 An invitation to quantum incompatibility *J. Phys. A: Math. Theor.* **49** 123001
[17] D'Ariano G M 2004 Extremal covariant quantum operations and positive operator valued measures *J. Math. Phys.* **45** 3620–35
[18] Heinosaari T, Reitzner D and Stano P 2008 Notes on joint measurability of quantum observables *Found. Phys.* **38** 1133–47
[19] Guerini L and Cunha M Terra 2018 Uniqueness of the joint measurement and the structure of the set of compatible quantum measurements *J. Math. Phys.* **59** 042106
[20] Durt T, Englert B-G, Bengtsson I and Życzkowski K 2010 On mutually unbiased bases *Int. J. Quantum Inf.* **08** 535–640
[21] Haapasalo E 2015 Robustness of incompatibility for quantum devices *J. Phys. A: Math. Theor.* **48** 255303
[22] Boyd S and Vandenberghe L 2004 *Convex Optimization* (Cambridge: Cambridge University Press)
[23] Skrzypczyk P and Linden N 2019 Robustness of measurement, discrimination games, and accessible information *Phys. Rev. Lett.* **122** 140403
[24] Saha D, Oszmaniec M, Czekaj M, Horodecki and Horodecki R 2018 Operational foundations of complementarity and uncertainty relations arXiv:1809.03475
[25] Löfberg J 2004 YALMIP : a toolbox for modeling and optimization in MATLAB *Proc. CACSD Conf. (Taipei, Taiwan)*
[26] Toh K, Todd M and Tutuncu R 1999 SDPT3–a MATLAB software package for semidefinite programming *Optim. Methods Softw.* **11** 545–81
[27] *MOSEK ApS, The MOSEK optimization toolbox for MATLAB manual*
[28] Carmeli C, Cassinelli G and Toigo A 2019 Constructing extremal compatible quantum observables by means of two mutually unbiased bases arXiv:1904.09451
[29] Heinosaari T, Schultz J, Toigo A and Ziman M 2014 Maximally incompatible quantum observables *Phys. Lett.* A **378** 1695–9
[30] Heinosaari T, Kiukas J, Reitzner D and Schultz J 2015 Incompatibility breaking quantum channels *J. Phys. A: Math. Theor.* **48** 435301
[31] Carmeli C, Heinosaari T and Toigo A 2018 State discrimination with postmeasurement information and incompatibility of quantum measurements *Phys. Rev.* A **98** 012126
[32] Carmeli C, Heinosaari T and Toigo A 2012 Informationally complete joint measurements on finite quantum systems *Phys. Rev.* A **85** 012109

[33] Bavaresco J, Quintino M T, Guerini L, Maciel T O, Cavalcanti D and Cunha M T 2017 Most incompatible measurements for robust steering tests *Phys. Rev.* A **96** 022110

[34] Bluhm A and Nechita I 2018 Joint measurability of quantum effects and the matrix diamond *J. Math. Phys.* **59** 112202

[35] Bluhm A and Nechita I 2018 Compatibility of quantum measurements and inclusion constants for the matrix jewel arXiv:1809.04514

[36] Designolle S, Skrzypczyk P, Fröwis F and Brunner N 2019 Quantifying measurement incompatibility of mutually unbiased bases *Phys. Rev. Lett.* **122** 050402

[37] Oszmaniec M, Guerini L, Wittek P and Acín A 2017 Simulating positive-operator-valued measures with projective measurements *Phys. Rev. Lett.* **119** 190501

[38] Busch P 1986 Unsharp reality and joint measurements for spin observables *Phys. Rev.* D **33** 2253–61

[39] Uola R, Luoma K, Moroder T and Heinosaari T 2016 Adaptive strategy for joint measurements *Phys. Rev.* A **94** 022109

[40] Cavalcanti D and Skrzypczyk P 2016 Quantitative relations between measurement incompatibility, quantum steering, and nonlocality *Phys. Rev.* A **93** 052112

[41] Banik M, Gazi M R, Ghosh S and Kar G 2013 Degree of complementarity determines the nonlocality in quantum mechanics *Phys. Rev.* A **87** 052125

[42] Jenčová A and Plávala M 2017 Conditions on the existence of maximally incompatible two-outcome measurements in general probabilistic theory *Phys. Rev.* A **96** 022113

[43] Addis C, Heinosaari T, Kiukas J, Laine E-M and Maniscalco S 2016 Dynamics of incompatibility of quantum measurements in open systems *Phys. Rev.* A **93** 022114

[44] Heinosaari T 2016 Simultaneous measurement of two quantum observables: compatibility, broadcasting, and in-between *Phys. Rev.* A **93** 042118

[45] Jenčová A 2018 Incompatible measurements in a class of general probabilistic theories *Phys. Rev.* A **98** 012133

[46] Busch P, Heinosaari T, Schultz J and Stevens N 2013 Comparing the degrees of incompatibility inherent in probabilistic physical theories *Europhys. Lett.* **103** 10002

[47] Plávala M 2016 All measurements in a probabilistic theory are compatible if and only if the state space is a simplex *Phys. Rev.* A **94** 042108

[48] Kiukas J, Budroni C, Uola R and Pellonpää J-P 2017 Continuous-variable steering and incompatibility via state-channel duality *Phys. Rev.* A **96** 042331

[49] Carmeli C, Heinosaari T and Toigo A 2019 Quantum incompatibility witnesses *Phys. Rev. Lett.* **122** 130402

[50] Uola R, Kraft T, Shang J, Yu X-D and Gühne O 2019 Quantifying quantum resources with conic programming *Phys. Rev. Lett.* **122** 130404

[51] Bronzan J B 1988 Parametrization of SU(3) *Phys. Rev.* D **38** 1994–9

[52] Acín A, Durt T, Gisin N and Latorre J I 2002 Quantum nonlocality in two three-level systems *Phys. Rev.* A **65** 052325

[53] Yu S, Liu N-l, Li L and Oh C H 2010 Joint measurement of two unsharp observables of a qubit *Phys. Rev.* A **81** 062116

[54] Brierley S, Weigert S and Bengtsson I 2010 All mutually unbiased bases in dimensions two to five *Quantum Inf. Comput.* **10** 803–20

[55] Klappenecker A and Rötteler M 2004 *Constructions of Mutually Unbiased Bases* (Berlin: Springer)