Quantum Incompatibility Witnesses

Claudio Carmeli
DIME, Università di Genova, Via Magliotto 2, I-17100 Savona, Italy

Teiko Heinosaari
QTF Centre of Excellence, Turku Centre for Quantum Physics, Department of Physics and Astronomy, University of Turku, FI-20014 Turku, Finland

Alessandro Toigo
Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy and I.N.F.N., Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy

We demonstrate that quantum incompatibility can always be detected by means of a state discrimination task with partial intermediate information. This is done by showing that only incompatible measurements allow for an efficient use of premeasurement information in order to improve the probability of guessing the correct state. The gap between the guessing probabilities with pre- and postmeasurement information is a witness of the incompatibility of a given collection of measurements. We prove that all linear incompatibility witnesses can be implemented as some state discrimination protocol according to this scheme.

I. INTRODUCTION

Quantum incompatibility is one of the key features that separate the quantum from the classical world [1]. It gives rise to several among the most intriguing quantum phenomena, including measurement uncertainty relations [2], contextuality [3] and non-locality [4]. So far, however, the direct experimental verification of quantum incompatibility has been a demanding task, as the known detection methods, based on Bell experiments [5–7] and steering protocols [8–11], rely on entanglement.

In this paper, we show that quantum incompatibility can be detected by means of a state discrimination task with partial intermediate information. More precisely, we consider a scenario where Alice sends Bob a quantum system that she has prepared into a state chosen from one of \( n \) disjoint state ensembles, but she reveals him the chosen ensemble only at a later time. Bob can then decide to perform his measurement either before or after Alice’s announcement and, importantly, the achievable success probabilities can be compared. We show that Bob can benefit from prior compared to posterior measurement information and improve his probability of guessing the correct state only if his measurements are incompatible.

Looking at it from another perspective, the difference between Bob’s guessing probabilities with pre- and post-measurement information is a witness of the incompatibility of the collection of measurements he uses in the discrimination task. Since the complement set of incompatible collections of measurements is the closed and convex set of all the compatible collections of measurements, this observation sets the previous detection scheme for incompatibility within the broader framework of witnesses.

In general, a witness is any experimentally assessable linear function whose value is greater than or equal to zero whenever the measured object does not have the investigated property, but gives a negative value at least for some object with that property. The paradigmatic example of witnesses is that of entanglement witnesses, which have become one of the main methods to detect entanglement [12, 13]. Other examples include the detection of non-Gaussianity of states [14], dimensionality of correlations [15], or for the unital channels the detection of not being a random unitary channel [16]. The fact that witnesses can be applied to detect incompatibility has been first noted in [17].

In this paper, we prove that any incompatibility witness essentially arises as a state discrimination task with intermediate information of the type described above. By standard separation results for convex sets, this implies that all incompatible sets of measurements can be detected by performing some state discrimination where premeasurement information is strictly better than postmeasurement information. This yields a novel operational interpretation of quantum incompatibility, and provides a method to detect it in a physically feasible experiment. In particular, this proves that entanglement is not needed to reveal incompatibility.

II. GENERAL FRAMEWORK OF WITNESSES

We briefly recall the general setting of witnesses as this clarifies our main results on incompatibility witnesses and makes the reasoning behind them easy to follow.

Let \( \mathcal{V} \) be a real linear space and \( C \subset \mathcal{V} \) a compact convex subset which mathematically describes the objects...
that to each witness one are only interested in the set of detected elements. The meaning of the first condition is then clear: the subset of all elements of \( C \) divided into two disjoint subsets \( C_0 \) and \( \overline{C}_0 \), with \( C_0 \) being closed and convex. We can think the subsets \( C_0 \) and \( \overline{C}_0 \) as properties - either an element \( x \in C \) is in \( C_0 \) or in \( \overline{C}_0 \). A witness of the property \( \overline{C}_0 \), or \( \overline{C}_0 \)-witness, is a map \( \xi : C \to \mathbb{R} \) such that

\[
(W1) \quad \xi(x) \geq 0 \quad \text{for all} \quad x \in C_0 \quad \text{and} \quad \xi(x) < 0 \quad \text{at least for some} \quad x \in \overline{C}_0;
\]

\[
(W2) \quad \xi(tx + (1-t)y) = t\xi(x) + (1-t)\xi(y) \quad \text{for all} \quad x,y \in C \quad \text{and} \quad t \in [0,1].
\]

The second condition implies that to each witness one can associate a hyperplane that separates \( \mathcal{V} \) into two halfspaces. The meaning of the first condition is then clear: \( \overline{C}_0 \) is entirely contained in one of the two halves.

We say that an element \( x \in \overline{C}_0 \) is detected by \( \xi \) if \( \xi(x) < 0 \), and we denote by \( \mathcal{D}(\xi) \) the subset of all elements of \( \overline{C}_0 \) that are detected by \( \xi \). Another \( \overline{C}_0 \)-witness \( \xi' \) is called finer than \( \xi \) if \( \mathcal{D}(\xi') \subseteq \mathcal{D}(\xi) \), and in this case we write \( \xi \leq \xi' \). If \( \mathcal{D}(\xi') = \mathcal{D}(\xi) \), we say that \( \xi \) and \( \xi' \) are detection equivalent and denote this by \( \xi \approx \xi' \). As we typically aim to detect as many elements as possible, we favor witnesses that cannot be made any finer. A necessary condition for \( \xi \) being optimal in that sense is that \( \xi \) is tight, meaning that \( \xi(x) = 0 \) for some \( x \in C_0 \).

Any \( \overline{C}_0 \)-witness \( \xi \) can be written in the form

\[
\xi(x) = \delta - v^*(x) \quad \forall x \in C, \tag{1}
\]

where \( v^* : \mathcal{V} \to \mathbb{R} \) is a linear map and \( \delta \in \mathbb{R} \) is a constant. An essential point for our later developments is that the representation (1) of a witness \( \xi \) is not unique but there is some freedom in the choice of \( v^* \) and \( \delta \). In addition, if we are only interested in the set of detected elements \( \mathcal{D}(\xi) \), we have a further degree of freedom, coming from the possibility to switch from \( \xi \) to an equivalent \( \overline{C}_0 \)-witnesses \( \xi' = \alpha \xi \) for some constant \( \alpha > 0 \).

III. DETECTING QUANTUM INCOMPATIBILITY

A measurement with a finite outcome set \( X \) is mathematically described as a positive operator valued measure (POVM), i.e., a map \( \mathcal{A} \) from \( X \) to the set \( \mathcal{L}_s(\mathcal{H}) \) of selfadjoint linear operators on a Hilbert space \( \mathcal{H} \) such that the operators \( \mathcal{A}(x) \) are positive (meaning that \( \langle \psi | \mathcal{A}(x) \psi \rangle \geq 0 \) for all \( \psi \in \mathcal{H} \)) and they satisfy the normalization condition \( \sum_x \mathcal{A}(x) = 1 \).

For clarity, we limit our discussion to pairs of measurements. The treatment of finite collections of measurements is similar. Two measurements \( \mathcal{A} \) and \( \mathcal{B} \), having outcome sets \( X \) and \( Y \), respectively, are compatible if there exists a measurement \( M \), called their joint measurement, with outcome set \( X \times Y \), such that

\[
\sum_y M(x,y) = \mathcal{A}(x), \quad \sum_x M(x,y) = \mathcal{B}(y).
\]

Otherwise, the measurements \( \mathcal{A} \) and \( \mathcal{B} \) are incompatible.

By \( \mathcal{O}_{X,Y} \) we denote the compact set of all pairs of measurements \( (\mathcal{A},\mathcal{B}) \) with outcome sets \( X,Y \), respectively. This set is divided into compatible pairs \( \mathcal{O}^\text{com}_{X,Y} \) and incompatible pairs \( \mathcal{O}^\text{inc}_{X,Y} \). We define convex combinations in \( \mathcal{O}_{X,Y} \) componentwise, and it follows that the subset \( \mathcal{O}^\text{com}_{X,Y} \) of compatible pairs is closed and convex. Hence we can consider \( \mathcal{O}^\text{inc}_{X,Y} \)-witnesses; we call them incompatibility witnesses (IW's).

A physically significant example of an IW for pairs of dichotomic measurements can be derived from the Bell-CHSH inequality [18]. For four dichotomic measurements \( \mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D} \) and a bipartite state \( \varrho \), the CHSH expression is

\[
\mathcal{B} = \langle AC \rangle_\varrho + \langle AD \rangle_\varrho - \langle BC \rangle_\varrho - \langle BD \rangle_\varrho,
\]

where, for instance,

\[
\langle AC \rangle_\varrho = \sum_{x,y} x y \text{tr} [\varrho A(x) \otimes C(y)].
\]

The Bell-CHSH inequality reads |\( \mathcal{B} \)| \( \leq 2 \). If now \( C \), \( D \) and \( \varrho \) are fixed, we can form two IWs by setting

\[
\zeta^\text{CHSH}_{AB}(\mathcal{A}, \mathcal{B}) = 2 \pm \mathcal{B}, \tag{2}
\]

where \( \mathcal{B} \) is understood as a function of \( \mathcal{A} \) and \( \mathcal{B} \). A violation of the Bell-CHSH inequality hence means that either \( \zeta^\text{CHSH}_{AB}(\mathcal{A}, \mathcal{B}) < 0 \) or \( \zeta^\text{CHSH}_{AB}(\mathcal{A}, \mathcal{B}) > 0 \).

It is known that dichotomic measurements \( \mathcal{A} \) and \( \mathcal{B} \) (acting on the same local system) are incompatible if and only if there exists a bipartite state \( \varrho \) and measurements \( \mathcal{C}, \mathcal{D} \) (acting on the other local system), such that the corresponding measurement scenario leads to the violation of the Bell-CHSH inequality [5]. In the current setting, this means that any incompatible pair of dichotomic measurements is detected by some witness of the type \( \zeta^\text{CHSH}_{AB} \).

Remarkably, it has been recently shown that, for measurements with more outcomes, there exist pairs of incompatible measurements that do not lead to a violation of any Bell inequality [19, 20].
IV. DISCRIMINATION SCENARIO AS AN INCOMPATIBILITY WITNESS

In the standard state discrimination scenario \[21,23\], Alice picks a label \( z \) from a given set \( Z \) with probability \( p(z) \). She encodes the label into a quantum state \( \rho_z \) and delivers the state to Bob. Bob knows the set \( \{ \rho_z \}_{z \in Z} \) of states used in the encoding. He is trying to recover the label by making a measurement on the quantum system that he has received. It is convenient to recover the label from the apriori probability distribution \( p \) and the state encoding into a single map \( \mathcal{E} \), given as \( \mathcal{E}(z) = p(z)\rho_z \). We call this map a *state ensemble*; its defining properties are that \( \mathcal{E}(z) \geq 0 \) for all \( z \), and \( \sum_z \text{tr} [\mathcal{E}(z)] = 1 \). The guessing probability depends on the measurement \( M \) that Bob uses, and it is given as
\[
P_{\text{guess}}(\mathcal{E}; M) = \sum_z \text{tr}[\mathcal{E}(z)M(z)] .
\]

Further, we denote
\[
P_{\text{guess}}(\mathcal{E}) = \max_M P_{\text{guess}}(\mathcal{E}; M) ,
\]
where the optimization is done over all measurements with outcome set \( Z \).

We are then considering two modifications of the standard state discrimination scenario, where partial classical information concerning the correct label is given either before or after the measurement is performed \[24,27\]. The form of the partial information is given as a partitioning of \( Z = X \cup Y \) of \( Z \) into two disjoint subsets. By conditioning the state ensemble \( \mathcal{E} \) to the occurrence of a label in \( X \) or \( Y \), we obtain new state ensembles \( \mathcal{E}_X \) and \( \mathcal{E}_Y \), which we call *subensembles* of \( \mathcal{E} \); they are given as
\[
\mathcal{E}_X(x) = \frac{1}{p(X)}\mathcal{E}(x) , \quad \mathcal{E}_Y(y) = \frac{1}{p(Y)}\mathcal{E}(y) ,
\]
and their label sets are \( X \) and \( Y \), respectively. Here we have denoted \( p(X) = \sum_x p(x) \) and \( p(Y) = \sum_y p(y) \).

We write \( \hat{\mathcal{E}} = (\mathcal{E}, \{X, Y\}) \) for the partitioned state ensemble, i.e., the state ensemble \( \mathcal{E} \) with the partitioning of \( Z \) into disjoint subsets \( X \) and \( Y \).

If Alice announces the correct subensemble before Bob chooses his measurement, we call the task *discrimination with premeasurement information*. In this case, Bob can choose a measurement \( A \) with the outcome set \( X \) to discriminate \( \mathcal{E}_X \) and a measurement \( B \) with the outcome set \( Y \) to discriminate \( \mathcal{E}_Y \). At each round of the experiment he measures either \( A \) or \( B \), depending on Alice’s announcement. Hence, Bob’s total guessing probability is given as
\[
P_{\text{guess}}(\hat{\mathcal{E}}; A, B) = p(X)P_{\text{guess}}(\mathcal{E}_X; A) + p(Y)P_{\text{guess}}(\mathcal{E}_Y; B) \tag{4}
\]
and its maximal value is
\[
P_{\text{guess}}(\hat{\mathcal{E}}) = \max_{(A, B) \in \mathcal{O}_{X,Y}} P_{\text{guess}}(\hat{\mathcal{E}}; A, B)
= p(X)P_{\text{guess}}(\mathcal{E}_X) + p(Y)P_{\text{guess}}(\mathcal{E}_Y) .
\tag{5}
\]

In the other variant of the discrimination scenario, Alice announces the correct subensemble only after Bob has performed his measurement. In this case, Bob has to use a fixed measurement at each round but he can post-process the obtained measurement outcome according to the additional information. We call this task *discrimination with postmeasurement information*. It has been shown in \[26\] that now the maximal guessing probability, denoted as \( P_{\text{guess}}(\hat{\mathcal{E}}) \), is given by
\[
P_{\text{guess}}(\hat{\mathcal{E}}) = \max_{(A, B) \in \mathcal{O}_{X,Y}} P_{\text{guess}}(\hat{\mathcal{E}}; A, B) .
\tag{6}
\]

A comparison of (5) and (6) reveals that the maximum guessing probability \( P_{\text{guess}}(\hat{\mathcal{E}}) \) and \( P_{\text{guess}}(\hat{\mathcal{E}}) \) result in optimizing the same mathematical quantity, with the important difference that in the latter the optimization is restricted to compatible pairs of measurements. From this, we already conclude that if \( P_{\text{guess}}(\hat{\mathcal{E}}; A, B) > P_{\text{guess}}(\hat{\mathcal{E}}) \) for some partitioned state ensemble \( \hat{\mathcal{E}} \), then \( A \) and \( B \) are incompatible. In the following, we develop this observation into a necessary and sufficient condition for incompatibility by using the framework of witnesses.

We first notice that, for a partitioned state ensemble \( \hat{\mathcal{E}} = (\mathcal{E}, \{X, Y\}) \) with \( P_{\text{guess}}(\hat{\mathcal{E}}) > P_{\text{guess}}(\hat{\mathcal{E}}) \), the function
\[
\xi(\hat{\mathcal{E}}; A, B) = P_{\text{guess}}(\hat{\mathcal{E}}) - P_{\text{guess}}(\hat{\mathcal{E}}; A, B) \tag{7}
\]
is a tight IW for pairs of measurements in \( \mathcal{O}_{X,Y} \); we call it the *incompatibility witness associated with \( \hat{\mathcal{E}} \). In some cases, the exact evaluation of \( P_{\text{guess}}(\hat{\mathcal{E}}) \) may be a difficult task, but still by finding a number \( \delta \) such that \( P_{\text{guess}}(\hat{\mathcal{E}}) \leq \delta < P_{\text{guess}}(\hat{\mathcal{E}}) \) one obtains an IW by setting
\[
\xi(\hat{\mathcal{E}}; A, B) = \delta - P_{\text{guess}}(\hat{\mathcal{E}}; A, B) .
\tag{8}
\]

Clearly, we then have \( \xi(\hat{\mathcal{E}}; A, B) < \xi(\hat{\mathcal{E}}) \).

An important feature of the witnesses arising from partitioned state ensembles is that their physical implementation is straightforward. Namely, the quantities \( P_{\text{guess}}(\mathcal{E}_X; A) \) and \( P_{\text{guess}}(\mathcal{E}_Y; B) \) are obtained by performing standard state discrimination experiments, and \( P_{\text{guess}}(\hat{\mathcal{E}}; A, B) \) is then given via (4). The constant term \( p_{\text{guess}}(\hat{\mathcal{E}}) \) must be calculated analytically or numerically, or at least upper bounded tightly enough. It has been shown in \[26\] that the calculation of \( P_{\text{guess}}(\hat{\mathcal{E}}) \) reduces to the evaluation of the standard guessing probability \( P_{\text{guess}}(\hat{\mathcal{E}}') \) of an auxiliary state ensemble \( \mathcal{E}' \), and the techniques for calculating the standard guessing probability (see e.g. \[28\]) are thereby applicable.

V. CHARACTERIZATION OF INCOMPATIBILITY WITNESSES

In this section we present our main results.
Theorem 1. For any incompatibility witness $\xi$, there exists a partitioned state ensemble $\hat{\xi}$ such that the associated incompatibility witness $\hat{\xi}_{E}$ is finer than $\xi$. Further, if $\xi$ is tight, there exists a partitioned state ensemble $\hat{\xi}$ such that $\xi$ is detection equivalent to $\hat{\xi}$.

In the case of IWs, the natural choice for the ambient vector space $V$ containing $O_{X,Y}$ is the Cartesian product $F(X) \times F(Y)$, where $F(X)$ is the vector space of all operator valued functions $F:X \to L_{\alpha}(H)$. All linear maps on $F(X) \times F(Y)$ are expressible in terms of scalar products with elements $(F,G) \in F(X) \times F(Y)$, so that the basic representation $\{x\}$ of witnesses takes the form

$$\xi(A, B) = \delta - \sum_{x} \text{tr}[F(x)A(x)] - \sum_{y} \text{tr}[G(y)B(y)]$$

for all $(A, B) \in O_{X,Y}$. The proof of Thm. 1 is based on the freedom in the choice of $(F, G)$ and $\delta$.

Proof of Thm. 1. Starting from an IW $\xi$ of the general form $\{x\}$, we similarly define a map $\xi'$ by choosing $F(x) = \alpha(F(x) - \mu I)$, $G'(y) = \alpha(G(y) - \mu I)$, and $\delta' = \alpha(\delta - 2 \mu d)$, where $d$ is the dimension of the Hilbert space and $\alpha, \mu \in \mathbb{R}$ are constants that we will determine next. A direct calculation shows that $\xi' = \alpha \xi$ on $O_{X,Y}$. Firstly, we fix the value of $\mu$ by setting

$$- \mu = \sum_{x \in X} \|F(x)\| + \sum_{y \in Y} \|G(y)\|,$$

and with this choice all the operators $E(x) = |\alpha|(F(x) - \mu I)$ and $E(y) = |\alpha|(G(y) - \mu I)$ are positive. Secondly, we fix the value of $\alpha$ by setting

$$\frac{1}{\alpha} = \sum_{x \in X} \text{tr}[F(x) - \mu I] + \sum_{y \in Y} \text{tr}[G(y) - \mu I].$$

The right hand side of this expression is strictly positive, as otherwise $F(x) = G(y) = \mu I$ for all $x, y$ and so the original IW $\{x\}$ would be constant on $O_{X,Y}$, which is impossible. Thereby, $\alpha > 0$, hence the map $\xi' = \alpha \xi$ is an IW and $\xi' \approx \xi$. Moreover, in this way we have obtained a partitioned state ensemble $\hat{\xi} = (\xi, \{X, Y\})$, for which the witness $\xi'$ has the form $\{x\}$: $\xi'(A, B) = \delta' - P_{\text{guess}}^{\text{prior}}(\hat{\xi}; A, B)$. Since $\xi'$ is an IW and hence satisfies (W1), we must have $P_{\text{guess}}^{\text{post}}(\hat{\xi}) \leq \delta' < P_{\text{guess}}^{\text{prior}}(\hat{\xi})$. Thereby, $\xi' \leq \hat{\xi}$. If in addition $\xi$ is tight, then $\delta' = P_{\text{guess}}^{\text{post}}(\hat{\xi})$, and thus $\xi = \hat{\xi}$.

A specific implication of Theorem 1 is that, for any IW of the Bell-CHSH type as defined in (2), we can find a detection equivalent or finer IW associated with a partitioned state ensemble. The implementation of the latter one does not require entanglement and is therefore much more practical. A concrete qubit example is given in the Supplemental Material.

An important consequence of Theorem 1 is the following result, which provides a novel operational meaning for quantum incompatibility.

Theorem 2. Two measurements $A$ and $B$ are incompatible if and only if there exists a partitioned state ensemble $\hat{\xi}$ such that $P_{\text{guess}}^{\text{prior}}(\xi; A, B) > P_{\text{guess}}^{\text{post}}(\hat{\xi}).$

Proof. We have already observed earlier that if for some partitioned state ensemble $\hat{\xi}$ we have $P_{\text{guess}}^{\text{prior}}(\xi; A, B) > P_{\text{guess}}^{\text{post}}(\hat{\xi})$, then $A$ and $B$ are incompatible. Conversely, let us assume that $(A, B) \notin O_{X,Y}$. Then, by the usual separation results for compact convex sets [29, Cor. 11.4.2], there exist $(F, G) \in F(X) \times F(Y)$ and $\delta \in \mathbb{R}$ such that, defining $\xi$ as in (9), we have $\xi(A', B') \geq 0$ for all $(A', B') \in O_{X,Y}^\text{com}$ and $\xi(A, B) < 0$. By Thm. 1 there exists a partitioned state ensemble $\hat{\xi}$ such that $\xi < \hat{\xi}$. It follows that $\xi(A, B) < 0$, i.e., $P_{\text{guess}}^{\text{prior}}(\xi; A, B) \geq P_{\text{guess}}^{\text{post}}(\hat{\xi}).$

VI. BOUNDING THE COMPATIBILITY REGION BY MEANS OF TWO MUTUALLY UNBIASED BASES

As we have seen, constructing an IW involves the solution of two convex optimization problems: the evaluation of the maximal guessing probabilities with pre- and postmeasurement information defined in [3] and [6]. In particular, if $\hat{\xi}$ is a partitioned state ensemble for which the two probabilities differ, whenever the maximum in the right hand side of (7) admits an analytical computation, one can insert the resulting value of $P_{\text{guess}}^{\text{post}}(\hat{\xi})$ into (7) and thus write the tight IW associated with $\hat{\xi}$ in an explicit form.

Interestingly, solving the optimization problem (9) yields even more. Indeed, evaluating a constrained maximum typically requires to find some feasible points where the maximum is attained; if the optimization problem is convex, these points are necessarily located on the relative boundary of the feasible domain. In our specific case, it means that, as a byproduct of solving (6), we get points lying on the relative boundary $\partial O_{X,Y}^\text{com}$ of the convex set $O_{X,Y}$. Then, by taking convex combinations of these points, we can even have an insight into the set $O_{X,Y}^\text{com}$ itself. We thus see that the solution of (6) has a twofold purpose: on the one hand, through the IW constructed in (7), it provides a simple method to detect the incompatibility of many measurement pairs; on the other hand, by using the resulting optimal points, some information on the set of compatible pairs can be inferred.

An interesting special case in which the optimization problems (1)-(6) admit an analytical solution is when the partitioned state ensemble $\hat{\xi}$ is made up of two mutually unbiased bases (MUB) of the system Hilbert space $H$, or, more generally, smearings of two MUB. Suppose $\{\varphi_{k}\}_{k \in \{1, \ldots, d\}}$ and $\{\psi_{k}\}_{k \in \{1, \ldots, d\}}$ is a fixed pair of MUB; then, we can use it to construct a partitioned state ensemble as follows. First, we choose $Z = \{1, \ldots, d\} \times \{\varphi, \psi\}$ as the overall label set of the ensemble, and then we partition $Z$ into the subsets $X = \{(1, \varphi, \ldots, (d, \varphi)\}$ and $Y = \{(1, \psi, \ldots, (d, \psi)\}$; here, the letters $\varphi$ and $\psi$ are just
symbols, which are needed to distinguish labels in different subsets. As our subensembles, we consider smearings of the two MUB by means of uniform noise. We assume all basis vectors within each subensemble are equally probable; however, we allow for different noise parameters \( \mu_\varphi, \mu_\psi \in [0,1] \) according to the subensemble at hand. The resulting overall state ensemble then reads

\[
\mathcal{E}_\mu(j, \ell) = \frac{1}{2d} \left[ \mu_\ell \ell_j (\ell_j + (1 - \mu_\rho) \frac{1}{d} \mathbb{I} \right],
\]

where \( \mu = (\mu_\varphi, \mu_\psi) \); we further set \( \hat{\mathcal{E}}_\mu = (\mathcal{E}_\mu, \{X, Y\}) \). In order to obtain some more inequivalent IWs, we can let either \( \mu_\varphi \) or \( \mu_\psi \) take even slightly negative values.

The detailed solution to the optimization problems (5)-(7) for the present case actually yields \( \theta \) (13) becomes a function of \( \gamma / \varphi \), which depend on \( \mu \). The next two theorems then follow by our earlier observations.

**Theorem 3.** Let \( \mu = (\mu_\varphi, \mu_\psi) \in [1/(1 - d), 1] \times [1/(1 - d), 1] \) with \( \mu \neq (0,0) \). Then \( P_{\text{post}}\left(\mathcal{E}_\mu\right) < P_{\text{prior}}\left(\hat{\mathcal{E}}_\mu\right) \) if and only if \( \mu_\varphi \mu_\psi \neq 0 \) and either \( d = 2 \) or \( \max\{\mu_\varphi, \mu_\psi\} > 0 \). In this case, the tight incompatibility witness associated with the partitioned state ensemble \( \hat{\mathcal{E}}_\mu \) is

\[
\epsilon_{\varphi, \psi}(A, B) = \frac{1}{4} \left[ \mu_\varphi + \mu_\psi + \sqrt{\mu_\varphi^2 + \mu_\psi^2 - 2\left(1 - \frac{2}{d}\right)\mu_\varphi \mu_\psi} \right]
\]

\[
- \frac{1}{2d} \sum_{j=1}^d \left[ \mu_\varphi \langle \varphi_j | A(j, \varphi) \varphi_j \rangle + \mu_\psi \langle \varphi_j | B(j, \psi) \psi_j \rangle \right].
\]

Finally, the ensembles \( \hat{\mathcal{E}}_\mu \) and \( \hat{\mathcal{E}}_\mu \) determine detection equivalent incompatibility witnesses if and only if \( \nu = \alpha \mu \) for some \( \alpha > 0 \).

Interestingly, for \( d = 2 \), all IWs given in (14) are of the Bell-CHSH type (2) (see the Supplemental Material).

By the equivalence statement in the previous theorem, no generality is lost if we express the vector \( \mu \) in terms of a single real parameter \( \theta \). Consequently, also the vector \( \gamma = (\gamma_\varphi, \gamma_\psi) \) parametrizing the optimal measurements (13) becomes a function of \( \theta \). Thus, solving the optimization problem (6) for the present case actually yields a curve in the relative boundary \( \partial \mathcal{O}_{X,Y} \).

**Theorem 4.** The pair of measurements \( (A, B) \) of (13) lie on the relative boundary \( \partial \mathcal{O}_{X,Y} \) if

\[
\gamma = \left( \frac{d - 2 - d \cos(\theta + \theta_0)}{2(d - 1)}, \frac{d - 2 - d \cos(\theta - \theta_0)}{2(d - 1)} \right)
\]

for \( \theta \in [-\theta_0, \theta_0] \) and \( \theta_0 = \pi - \arctan\sqrt{d - 1} \).

FIG. 2. The set of \( \gamma = (\gamma_\varphi, \gamma_\psi) \) for which Eq. (15) defines

\[
\text{two measurements (green square), and the one for which these \text{measurements are compatible (blue region) for different values of the dimension } d. \text{ The red line is the curve (15). The case } d = 2 \text{ is special, and was already treated in [30].}
\]

The operators in (13) are positive if and only if \( \gamma \in [1/(1 - d), 1] \times [1/(1 - d), 1] \). Thus, all pairs of measurements of the form (13) constitute a square-shaped section of the set \( \mathcal{O}_{X,Y} \). Remarkably, the lower-left vertex \( (1/(1 - d), 1/(1 - d)) \) of this square corresponds to a compatible pair of measurements if and only if \( d \geq 3 \); on the contrary, when \( d = 2 \) the relative boundary is symmetric around \( (0,0) \) [30]. Combining these considerations and Thm. 1 we can give a partial inspection of the two sets \( \mathcal{O}_{X,Y} \) and \( \mathcal{O}_{X,Y} \), as shown in Fig. 2.

**VII. DISCUSSION**

The framework of witnesses is an effective tool in the detection of properties described by sets with compact and convex complements. We have shown that for incompatibility of measurements, witnesses are not only a mathematical tool, but can be implemented in simple discrimination experiments. An important feature of this implementation scheme is that it does not require entanglement.

A consequence of our characterization result is a novel operational interpretation of incompatibility: a collection of measurements is incompatible if and only if there is a state discrimination task where premeasurement information is strictly better than postmeasurement information.

Entanglement witnesses have been used not only to detect entanglement but also to quantify entanglement [31]. Further, one can drop the condition (W2) and consider nonlinear witnesses [32]. These and other modifications or generalizations will be an interesting matter of investigation in the case of incompatibility witnesses.
ACKNOWLEDGMENT

The authors wish to thank Anna Jenčová for the valuable discussion on incompatibility witnesses. TH acknowledges financial support from the Academy of Finland via the Centre of Excellence program (Project no. 312058) as well as Project no. 287750.

[1] T. Heinosaari, T. Miyadera, and M. Ziman. An invitation to quantum incompatibility. *J. Phys. A: Math. Theor.*, 49:123001, 2016.

[2] P. Busch, P. Lahti, and R.F. Werner. Measurement uncertainty relations. *J. Math. Phys.*, 55:042111, 2014.

[3] Y.-C. Liang, R.W. Spekkens, and H.M. Wiseman. Specker’s parable of the overprotective seer: A road to contextuality, nonlocality and complementarity. *Phys. Rep.*, 506:1–39, 2011.

[4] A. Fine. Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.*, 48:291–295, 1982.

[5] M.M. Wolf, D. Pérez-Garcia, and C. Fernandez. Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory. *Phys. Rev. Lett.*, 103:230402, 2009.

[6] M. Banik, M.R. Gazi, S. Ghosh, and G. Kar. Degree of complementarity determines the nonlocality in quantum mechanics. *Phys. Rev. A*, 87:052125, 2013.

[7] G. Kar, S. Ghosh, S.K. Choudhary, and M. Banik. Role of measurement incompatibility and uncertainty in determining nonlocality. *Mathematics*, 4:52, 2016.

[8] R. Uola, T. Moroder, and O. Gühne. Joint measurability of generalized measurements implies classicality. *Phys. Rev. Lett.*, 113:160403, 2014.

[9] M.T. Quintino, T. Vértesi, and N. Brunner. Joint measurability, Einstein-Podolsky-Rosen steering, and Bell nonlocality. *Phys. Rev. Lett.*, 113:160402, 2014.

[10] R. Uola, C. Budroni, O. Gühne, and J.-P. Pellenpää. One-to-one mapping between steering and joint measurability problems. *Phys. Rev. Lett.*, 115:230402, 2015.

[11] S.-L. Chen, C. Budroni, Y.-C. Liang, and Y.-N. Chen. Natural framework for device-independent quantification of quantum steerability, measurement incompatibility, and self-testing. *Phys. Rev. Lett.*, 116:240401, 2016.

[12] O. Gühne and G. Tóth. Entanglement detection. *Phys. Rep.*, 474:1–75, 2009.

[13] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, 2009.

[14] C. Hughes, M.G. Genoni, T. Tufarelli, M.G.A. Paris, and M.S. Kim. Quantum non-Gaussianity witnesses in phase space. *Phys. Rev. A*, 90:013810, 2014.

[15] N. Brunner, S. Pironio, A. Acín, N. Gisin, A. Méthot, and V. Scarani. Testing the dimension of Hilbert spaces. *Phys. Rev. Lett.*, 100:210503, 2008.

[16] C.B. Mendl and M.M. Wolf. Unital quantum channels—convex structure and revivals of Birkhoff’s theorem. *Comm. Math. Phys.*, 289:1057–1086, 2009.

[17] A. Jenčová. Incompatible measurements in a class of general probabilistic theories. *Phys. Rev. A*, 98:012133, 2018.

[18] J.F. Clauser, M.A. Horne, A. Shimony, and R.A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, 1969.

[19] F. Hirsch, M.T. Quintino, and N. Brunner. Quantum measurement incompatibility does not imply Bell nonlocality. *Phys. Rev. A*, 97:012129, 2018.

[20] E. Bene and T. Vértesi. Measurement incompatibility does not give rise to Bell violation in general. *New Journal of Physics*, 20:013021, 2018.

[21] A.S. Holevo. Statistical decision theory for quantum systems. *J. Multivariate Anal.*, 3:337–394, 1973.

[22] H.P. Yuen, R.S. Kennedy, and M. Lax. Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Trans. Inform. Theory*, IT-21:125–134, 1975.

[23] C.W. Helstrom. *Quantum Detection and Estimation Theory*, Academic Press, New York, 1976.

[24] M.A. Ballester, S. Wehner, and A. Winter. State discrimination with post-measurement information. *IEEE Trans. Inf. Theory*, 54:4183–4198, 2008.

[25] D. Gopal and S. Wehner. Using postmeasurement information in state discrimination. *Phys. Rev. A*, 82:022326, 2010.

[26] C. Carmeli, T. Heinosaari, and A. Toigo. State discrimination with postmeasurement information and incompatibility of quantum measurements. *Phys. Rev. A*, 98:012126, 2018.

[27] N. Marumo S. Akibue, G. Kato. Perfect discrimination of non-orthogonal quantum states with posterior classical partial information. [arXiv:1811.08089 [quant-ph]].

[28] J. Bae. Structure of minimum-error quantum state discrimination. *New J. Phys.*, 15:073037, 2013.

[29] R.T. Rockafellar. *Convex analysis*. Princeton University Press, Princeton, NJ, 1976.

[30] P. Busch. Unsharp reality and joint measurements for spin observables. *Phys. Rev. D*, 33:2253–2261, 1986.

[31] J. Eisert, F.G.S. Brandão, and K.M.R. Audenaert. Quotative entanglement witnesses. *New Journal of Physics*, 9:46, 2007.

[32] O. Gühne and N. Lütkenhaus. Nonlinear entanglement witnesses. *Phys. Rev. Lett.*, 96:170502, 2006.
Supplemental Materials: Quantum Incompatibility Witnesses

This supplement is planned as follows. In Sec. I we provide the general framework for witnesses, recalling the basic related notions from convex analysis and deriving some useful results; in particular, by making a natural assumption upon the convex set to be detected, we fully characterize the equivalence classes of witnesses under the detection equivalence relation. In Sec. II these results are applied to incompatibility witnesses; to do so, we describe the convex structure of the set of all pairs of measurements and its subset of all the pairs that are compatible. In Sec. III we solve the optimization problem for the guessing probability with postmeasurement information for a state ensemble constructed by means of two mutually unbiased bases; we thus provide the detailed proofs of Thms. 3 and 4 of the main paper. Finally, in Sec. IV we show that in dimension \( d = 2 \) all the incompatibility witnesses found in Thm. 3 of the main text have a very simple implementation: indeed, they correspond to a Bell-CHSH measurement scheme.

Notation: In this supplement, the numberings of equations, theorems etc. are preceeded by the letter ‘S’ (e.g.: Eq. (S1), Thm. S1 etc.). When we refer to results in the main paper, we simply drop the ‘S’ (Eq. (1), Thm. 1 etc.).

I. WITNESSES FOR GENERAL CONVEX SETS

A. Preliminaries from convex analysis

In the following, we will need some standard terminology and notations from convex analysis. We refer to [1] for further details.

Suppose \( V \) is a finite dimensional, real and normed linear space. If \( V_0 \subseteq V \) is a linear subspace, we denote by \( V_0^\ast \) the dual linear space of \( V_0 \), and \( \langle v^\ast, v \rangle \) the canonical pairing between an element \( v \in V_0 \) and a dual vector \( v^\ast \in V_0^\ast \).

We recall that

- an affine [respectively, convex] combination of elements \( v_1, \ldots, v_n \in V \) is any linear combination \( \lambda_1 v_1 + \ldots + \lambda_n v_n \) such that \( \lambda_1 + \ldots + \lambda_n = 1 \) [resp., such that \( \lambda_1 + \ldots + \lambda_n = 1 \) and \( \lambda_i \in [0,1] \) for all \( i = 1, \ldots, n \)];

- an affine [resp., convex] set is any subset \( S \subseteq V \) such that all affine [resp., convex] combinations of elements of \( S \) are still contained in \( S \);

- an affine [resp., c-affine] map on an affine [resp., convex] set \( S \) is a function \( \xi : S \to \mathbb{R} \) such that

\[
\xi(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 \xi(v_1) + \ldots + \lambda_n \xi(v_n)
\]

for all \( v_1, \ldots, v_n \in S \) and any affine [resp., convex] combination \( \lambda_1 v_1 + \ldots + \lambda_n v_n \).

If \( M \) is an affine set, there is a unique vector subspace \( V(M) \) such that \( M = v_0 + V(M) \) for some (hence for any) \( v_0 \in M \). Moreover, if \( \xi : M \to \mathbb{R} \) is an affine map and \( v_0 \in M \) is fixed, there exist unique \( v_\xi^\ast \in V(M)^\ast \) and \( \delta_\xi \in \mathbb{R} \) such that

\[
\xi(v + v_\xi) = \delta_\xi - \langle v_\xi^\ast, v \rangle \quad \forall v \in V(M).
\]

In particular, for an affine map \( \Xi : V \to \mathbb{R} \), there exist unique \( v^\ast \in V^\ast \) and \( \delta \in \mathbb{R} \) such that

\[
\Xi(v) = \delta - \langle v^\ast, v \rangle = \Xi[v^\ast, \delta](v) \quad \forall v \in V.
\]  
(S1)

By choosing \( \delta = \delta_\xi - \langle v_\xi^\ast, v_0 \rangle \) and picking any \( v^\ast \in V^\ast \) whose restriction to \( V(M) \) coincides with \( v_\xi^\ast \), we see that the affine map \( \xi : M \to \mathbb{R} \) extends to the affine map \( \Xi[v^\ast, \delta] : V \to \mathbb{R} \). Clearly, this extension is not unique unless \( V(M) = V \).

Now, suppose \( C \subseteq V \) is a convex set. Then, the affine hull of \( C \) is the smallest affine set \( M(C) \) containing \( C \); equivalently, it is the set of all affine combinations of elements of \( C \). We abbreviate \( V(C) = V(M(C)) \); further, we introduce the following annihilator subspace of \( V(C) \) in \( V^\ast \):

\[
V(C)^\perp = \{ v^\ast \in V^\ast \mid \langle v^\ast, v \rangle = 0 \quad \forall v \in V(C) \}.
\]

Any c-affine map \( \xi : C \to \mathbb{R} \) uniquely extends to an affine map \( \hat{\xi} : M(C) \to \mathbb{R} \); as we have seen, such a map \( \hat{\xi} \) can be further extended to an affine map \( \Xi : V \to \mathbb{R} \), the latter extension being in general not unique. Actually, the first assertion is a particular case of the following more general result [2].

Proposition S1. Suppose \( C_1 \) and \( C_2 \) are convex sets, and let \( \phi : C_1 \to C_2 \) be a map such that \( \phi(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 \phi(v_1) + \ldots + \lambda_n \phi(v_n) \) for all affine combinations of elements \( v_1, \ldots, v_n \in C_1 \). Then, there exists a unique map \( \hat{\phi} : M(C_1) \to M(C_2) \) such that the restriction \( \hat{\phi}|_{C_1} = \phi \) and \( \hat{\phi}(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 \hat{\phi}(v_1) + \ldots + \lambda_n \hat{\phi}(v_n) \) for all affine combinations of elements \( v_1, \ldots, v_n \in M(C_1) \).
Proof. Any element element of the affine hull $\mathcal{M}(C_1)$ is an affine combination of elements of $C_1$. Then, if $x = \lambda_1 v_1 + \ldots + \lambda_n v_n \in \mathcal{M}(C_1)$ is any affine combination of $v_1, \ldots, v_n \in C_1$, we can define

$$\tilde{\phi}(x) = \lambda_1 \phi(v_1) + \ldots + \lambda_n \phi(v_n).$$

First of all, we claim that this definition of $\tilde{\phi}(x)$ is independent of the chosen representation of $x$. Indeed, suppose $x = \mu_1 w_1 + \ldots + \mu_m w_m$ for some other $w_1, \ldots, w_m \in C_1$ and $\mu_1, \ldots, \mu_m \in \mathbb{R}$ with $\mu_1 + \ldots + \mu_m = 1$. We set

$$\sigma = \sum_{i | \lambda_i > 0} \lambda_i - \sum_{j | \mu_j < 0} \mu_j = \sum_{j | \mu_j > 0} \mu_j - \sum_{i | \lambda_i < 0} \lambda_i,$$

and observe that $\sigma > 0$ since at least one among $\lambda_1, \ldots, \lambda_n$ is necessarily strictly positive. Then, as

$$\sum_{i | \lambda_i > 0} \frac{\lambda_i}{\sigma} v_i + \sum_{j | \mu_j < 0} \frac{-\mu_j}{\sigma} w_j = \sum_{j | \mu_j > 0} \frac{\mu_j}{\sigma} w_j + \sum_{i | \lambda_i < 0} \frac{-\lambda_i}{\sigma} v_i,$$

and the two sides of the latter equation are convex combinations of elements of $C_1$, it follows that

$$\sum_{i=1}^n \lambda_i \phi(v_i) - \sum_{j=1}^m \mu_j \phi(w_j) = \sigma \left[ \phi \left( \sum_{i | \lambda_i > 0} \frac{\lambda_i}{\sigma} v_i + \sum_{j | \mu_j < 0} \frac{-\mu_j}{\sigma} w_j \right) - \phi \left( \sum_{j | \mu_j > 0} \frac{\mu_j}{\sigma} w_j + \sum_{i | \lambda_i < 0} \frac{-\lambda_i}{\sigma} v_i \right) \right] = 0,$$

thus showing that $\tilde{\phi}$ is well defined.

Secondly, we prove that $\tilde{\phi}$ is an affine map, that is

$$\tilde{\phi} \left( \sum_{i=1}^n \lambda_i u_i \right) = \sum_{i=1}^n \lambda_i \tilde{\phi}(u_i) \quad \text{for all } u_1, \ldots, u_n \in \mathcal{M}(C_1) \text{ and } \lambda_1, \ldots, \lambda_n \in \mathbb{R} \text{ with } \sum_{i=1}^n \lambda_i = 1. \quad (\ast)$$

To do it, write each $u_i$ as an affine combination $u_i = \mu_{i,1} v_{i,1} + \ldots + \mu_{i,m_i} v_{i,m_i}$ of elements $v_{i,1}, \ldots, v_{i,m_i} \in C_1$. Then,

$$\sum_{i=1}^n \lambda_i u_i = \sum_{i=1}^n \sum_{j=1}^{m_i} \lambda_i \mu_{i,j} v_{i,j},$$

where the right hand side is an affine combination of the elements $v_{i,j} \in C_1$. Hence,

$$\tilde{\phi} \left( \sum_{i=1}^n \lambda_i u_i \right) = \sum_{i=1}^n \sum_{j=1}^{m_i} \lambda_i \mu_{i,j} \phi(v_{i,j}) \quad \text{and} \quad \tilde{\phi}(u_i) = \sum_{j=1}^{m_i} \mu_{i,j} \phi(v_{i,j})$$

by definition of $\tilde{\phi}$. This proves $(\ast)$. Finally, the equality $\tilde{\phi}|_{C_1} = \phi$ also follows by the very definition of $\tilde{\phi}$.

The relative interior $\text{ri}(C)$ of the convex set $C$ is the set of all its interior points with respect to the relative topology of $\mathcal{M}(C)$. The relative boundary of $C$ is the set-theoretic difference $\partial C = C \setminus \text{ri}(C)$. An element $z \in C$ is an extreme point of $C$ if the equality $z = \lambda x + (1 - \lambda) y$ with $x, y \in C$ and $\lambda \in (0,1)$ implies $x = y = z$.

If $C$ is a compact convex set, its support function is the map

$$\delta_C : V^* \to \mathbb{R}, \quad \delta_C(v^*) = \max \{ \langle v^*, x \rangle \mid x \in C \}.$$

Clearly, if $C_0 \subseteq C$ is a compact convex subset, then $\delta_{C_0}(v^*) \leq \delta_C(v^*)$ for all $v^* \in V^*$. The support function satisfies $\delta_C(\alpha v^*) = \alpha \delta_C(v^*)$ for all $\alpha > 0$; moreover, if $v^* - u^* \in \mathcal{V}(C)^\perp$, then $\delta_C(v^*) - \delta_C(u^*) = \langle v^* - u^*, v_0 \rangle$ for some (hence for all) $v_0 \in \mathcal{M}(C)$.

B. Detecting a convex property

We are interested in a set of objects $C$, in which we consider the subset $C_0 \subseteq C$ of all the objects sharing some given property. We assume $\sigma \neq C \neq C_0$, and we denote by $\overline{C_0} = C \setminus C_0$ the subset of all the objects which do not possess the property at hand; we want to find simple sufficient conditions guaranteeing that an object $x \in C$ actually belongs to $\overline{C_0}$.

In the following, we always suppose that both $C$ and $C_0$ are convex and compact subsets of a finite dimensional, real and normed linear space $V$. Then, the simplest conditions involve some specific c-affine map $\xi : C \to \mathbb{R}$, related to $C$ and $C_0$, and the value that $\xi$ takes at $x \in C$. 
Definition S1. A $\overline{C}_0$-witness on the convex set $C$ is a c-affine map $\xi: C \to \mathbb{R}$ such that

(i) $\xi(y) \geq 0$ for all $y \in C_0$;

(ii) $\xi(x) < 0$ for some $x \in \overline{C}_0$.

A $\overline{C}_0$-witness is tight if it satisfies the extra condition

(iii) $\xi(z) = 0$ for some $z \in C_0$.

If $\xi$ is a $\overline{C}_0$-witness, the inequality $\xi(x) < 0$ entails that $x \in C$ does not possess the property $C_0$. The set $D(\xi) = \{x \in C \mid \xi(x) < 0\}$ is thus the subset of all the objects of $\overline{C}_0$ which are detected by $\xi$. Note that there always exists a tight $\overline{C}_0$-witness $\xi'$ detecting at least as many objects of $\overline{C}_0$ as $\xi$. Namely, it is enough to set $\xi'(x) = \xi(x) - \min \{\xi(y) \mid y \in C_0\}$. In general, whenever two $\overline{C}_0$-witnesses $\xi$ and $\xi'$ are such that $D(\xi) \subseteq D(\xi')$, we say that $\xi'$ is finer than $\xi$, and in this case we write $\xi \preceq \xi'$. Moreover, if $D(\xi) = D(\xi')$, we say that $\xi$ and $\xi'$ are detection equivalent and write $\xi \equiv \xi'$.

C. Structure of $\overline{C}_0$-witnesses

Proposition S2. A c-affine map $\xi: C \to \mathbb{R}$ is a $\overline{C}_0$-witness if and only if there exist $v^* \in \mathcal{V}^*$ and $\delta \in \mathbb{R}$ such that

(i) $\delta_{C_0}(v^*) \leq \delta < \delta_{C}(v^*)$;

(ii) $\xi(x) = \delta - \langle v^*, x \rangle$ for all $x \in C$.

In $[\mathbb{R}]$, the equality is attained if and only if $\xi$ is tight.

Proof. Any c-affine map $\xi$ extends to an affine map $\Xi = \Xi[v^*, \delta]$ on $\mathcal{V}$, and then $\xi(x) = \delta - \langle v^*, x \rangle \quad \forall x \in C$ by (S1).

In this case, $\xi$ is a $\overline{C}_0$-witness if and only if, for all $y \in C_0$ and some $x \in \overline{C}_0$,

$$
\Xi[v^*, \delta](y) \geq 0 \Leftrightarrow \Xi[v^*, \delta](x) \Leftrightarrow \langle v^*, y \rangle \leq \delta < \langle v^*, x \rangle.
$$

This is equivalent to $\delta_{C_0}(v^*) \leq \delta < \delta_{C}(v^*)$, where the equality is attained if and only if $\langle v^*, z \rangle = \delta$ for some $z \in C_0$, that is, $\xi(z) = 0$. □

If $\mathcal{V}$ is an Euclidean space and $v^*$ is given by the scalar product $\langle v^*, v \rangle = \hat{e} \cdot v \quad \forall v \in \mathcal{V}$ for some unit vector $\hat{e} \in \mathcal{V}$ not belonging to $\mathcal{V}(C_0)^1$, the nonnegative gap $\delta_{C}(v^*) - \delta_{C_0}(v^*)$ is the distance between the affine hyperplane of $\mathcal{V}$ which is orthogonal to $\hat{e}$, touches $C$ at its relative boundary and has $C$ on its side opposite to $\hat{e}$, and the analogous hyperplane touching the relative boundary of $C_0$.

D. Tight witnesses

The following proposition establishes the natural connection between tight $\overline{C}_0$-witnesses and the extremality property for points of the set $C_0$. Although it is an easy consequence of more general and standard results (see e.g. [1] Thm. 32.1 and Cor. 32.3.1), we provide a simple proof adapted to the present specific case.

Proposition S3. Suppose $\xi$ is a tight $\overline{C}_0$-witness on $C$. Then, the following facts hold.

(a) $\xi(z_0) = 0$ for some extreme point $z_0$ of $C_0$.

(b) If $\mathcal{V}(C_0) = \mathcal{V}(C)$ and $z \in C_0$, the equality $\xi(z) = 0$ implies that $z \in \partial C_0$.

The proof relies on the following simple lemma, that is sometimes useful by itself.

Lemma S1. For $v^* \in \mathcal{V}^*$, the following facts hold.

(a) $\delta_{C_0}(v^*) = \langle v^*, z_0 \rangle$ for some extreme point $z_0$ of $C_0$.

(b) Suppose $v^* \in \mathcal{V}^* \setminus \mathcal{V}(C_0)^1$. Then, if $z \in C_0$, the equality $\langle v^*, z \rangle = \delta_{C_0}(v^*)$ implies that $z \in \partial C_0$.
Proof (a) The set $Z = \{z \in C_0 \mid \langle v^*, z \rangle = \delta_{C_0}(v^*)\}$ is a nonempty closed and convex subset of $C_0$. By Krein-Milman theorem [H Cor. 18.5.1], $Z$ has some extreme point $z_0$. We claim that $z_0$ is extreme also for $C_0$. Indeed, suppose that $z_0 = y_1 + (1 - \lambda)y_2$ with $y_1, y_2 \in C_0$ and $\lambda \in (0, 1)$. The conditions $\delta_{C_0}(v^*) = \langle v^*, z_0 \rangle = \lambda \langle v^*, y_1 \rangle + (1 - \lambda)\langle v^*, y_2 \rangle$ and $\langle v^*, y_i \rangle \leq \delta_{C_0}(v^*)$ for $i = 1, 2$ then imply that $y_1, y_2 \in Z$, hence $y_1 = y_2 = z_0$.

(b) Suppose by contradiction that $z \in \ri(C_0)$. Then, for any $v \in \mathcal{V}(C_0)$ with $\langle v^*, v \rangle \neq 0$, there exists $\varepsilon \in \mathbb{R}$ such that $z + \varepsilon v \in C_0$ and $\varepsilon \langle v^*, v \rangle > 0$. It follows that $\langle v^*, z + \varepsilon v \rangle > \delta_{C_0}(v^*)$, which is impossible. 

Proof of Prop. S4 By Prop. S2 for some $v^* \in \mathcal{V}^*$ we have $\xi(x) = \delta_{C_0}(v^*) - \langle v^*, x \rangle$ for all $x \in C$, and $\delta_{C_0}(v^*) < \delta_C(v^*)$; in particular, $v^*$ is not constant on $C$, hence $v^* \in \mathcal{V}^* \setminus \mathcal{V}(C)^\perp$. Since $\xi(z) = 0$ is then equivalent to $\langle v^*, z \rangle = \delta_{C_0}(v^*)$, the two claims follow by the analogous statements of Lemma S1.

E. Detection equivalent witnesses

If the subset $C_0$ is sufficiently large in $C$, we have the following characterization of detection equivalence.

Proposition S4. Suppose $C_0 \cap \ri(C) \neq \emptyset$. Then, for two $C_0$-witnesses $\xi_1$ and $\xi_2$ on $C$, the following facts are equivalent.

(a) $\xi_1 \equiv \xi_2$.

(b) $\xi_1 = \alpha \xi_2$ for some $\alpha > 0$.

(c) If $v_i^* \in \mathcal{V}^*$ and $\delta_i \in \mathbb{R}$ are such that $\xi_i(x) = \delta_i - \langle v_i^*, x \rangle$ for all $x \in C$ and $i = 1, 2$, then there exists $\alpha > 0$ such that

(i) $v_i^* - \alpha v_2^* \in \mathcal{V}(C)^\perp$;

(ii) $\delta_1 - \alpha \delta_2 = \langle v_1^* - \alpha v_2^*, v_0 \rangle$ for some (hence for all) $v_0 \in \mathcal{M}(C)$.

Note that, if the two $C_0$-witnesses in the above proposition are tight, then in statement (c) we have $\delta_i = \delta_{C_0}(v_i^*)$ by Prop. S2 hence item (c) is directly implied by (c) and the properties of the support function $\delta_{C_0}$ recalled at the end of Sec. I A.

Prop. S4 exhibits the degree of freedom one has in choosing detection equivalent $C_0$-witnesses: namely, if we have a $C_0$-witness $\xi(x) = \delta - \langle v^*, x \rangle$ for all $x \in C$, we can turn it into an equivalent one by: (1) replacing the dual vector $v^* \in \mathcal{V}^*$ with $v'^* = \alpha(v^* + u^*)$, where $u^* \in \mathcal{V}(C)^\perp$ and $\alpha > 0$; (2) redefining the constant $\delta$ according to Prop. S4(c). The smaller is the linear space $\mathcal{V}(C)$ inside $\mathcal{V}$, the larger is the freedom in the choice of the dual vector $v^*$. This freedom is the crux of the proof of Thm. in the main paper, and therefore it is at the heart of the interpretation of incomparability witnesses as postmeasurement discrimination problems.

Proof of Prop. S4 (a) $\Rightarrow$ (b): Assuming statement (a), we preliminarily show that the two nonempty sets $Z_i = \{z \in C \mid \xi_i(z) = 0\}$ ($i = 1, 2$) actually coincide. Indeed, suppose by contradiction that $Z_1 \neq Z_2$. We can assume with no restriction that there is some $z \in C$ such that $\xi_1(z) = 0$ and $\xi_2(z) \neq 0$, hence $\xi_2(z) > 0$ since $D(\xi_1) = D(\xi_2) = D$. Picking any $x \in D$, we have $\xi_1(\lambda x + (1 - \lambda)z) = \lambda \xi_1(x) < 0$, or, equivalently, $\xi_2(\lambda x + (1 - \lambda)z) < 0$ for all $\lambda \in (0, 1)$, that contradicts continuity of the mapping $\lambda \mapsto \xi_2(\lambda x + (1 - \lambda)z)$ at $\lambda = 0$.

We next claim that there is $z_0 \in \ri(C)$ such that $\xi_i(z_0) = 0$ for all $i = 1, 2$. To this aim, let $y \in C_0 \cap \ri(C)$. Then, for all $i = 1, 2$, either $\xi_i(y) = 0$ and we are done, or $\xi_i(y) > 0$. In the latter case, again by a continuity argument, for any $x \in D$ there is some $\lambda \in (0, 1)$ such that $\xi_1(\lambda x + (1 - \lambda)z) = \lambda \xi_1(x) + (1 - \lambda)\xi_1(y) = 0$. Setting $z_0 = \lambda x + (1 - \lambda)y$, we thus see that $z_0 \in Z_1 = Z_2$, and $z_0 \in \ri(C)$ by H. Thm. 6.1.

For $i = 1, 2$, let $\xi_i$ be the extension of $\xi_i$ to an affine map on $\mathcal{M}(C)$. Then, the mapping $\mathcal{V}(C) \ni v \mapsto \hat{\xi}_i(z_0 + v) \in \mathbb{R}$ is linear and nonzero, hence there exists a linear basis $\{v_1, \ldots, v_m\}$ of $\mathcal{V}(C)$ such that $\hat{\xi}_i(z_0 + v_k) = 0$ if $k = 1, \ldots, m - 1$, and $\hat{\xi}_i(z_0 + v_m) < 0$. By possibly replacing all the $v_k$'s with $\mu v_k$ for some $\mu \in (0, 1)$, we can assume that $z_k := z_0 + v_k \in \ri(C)$ for all $k = 1, \ldots, m$, and so $\xi_i(z_k) = 0$ if $k \leq m - 1$, and $\xi_i(z_m) < 0$. Hence, also $\xi_2(z_k) = 0$ if $k \leq m - 1$, and $\xi_2(z_m) < 0$. It follows that

$$
\begin{bmatrix}
\hat{\xi}_1 - \xi_1(z_m) \\
\xi_2(z_m)
\end{bmatrix}
\begin{bmatrix}
z_k \\
\end{bmatrix}
= 0 \quad \forall k = 0, \ldots, m,
$$

which implies $\xi_1 - (\xi_1(z_m)/\xi_2(z_m)) \xi_2 = 0$ because any element of $\mathcal{M}(C)$ is an affine combination of $z_0, \ldots, z_m$. This yields statement (b).

(b) $\Rightarrow$ (a): The implication is clear.
Suppose $v_i^* \in \mathcal{V}^*$ and $\delta_i \in \mathbb{R}$ are as in statement (c), and extend the c-convex map $\xi_i$ to an affine map $\tilde{\xi}_i$ on $\mathcal{M}(C)$ by means of the relation $\tilde{\xi}_i(x) = \delta_i \cdot (v_i^*, x) \ \forall x \in \mathcal{M}(C)$. Then, since affine extensions are unique, statement (b) is equivalent to $\tilde{\xi}_1 = \alpha \tilde{\xi}_2$. Picking any $v_0 \in \mathcal{M}(C)$, this is in turn equivalent to

$$
0 = (\tilde{\xi}_1 - \alpha \tilde{\xi}_2)(v_0 + v)
$$

$$
= \left[ (\delta_1 - \alpha \delta_2) \cdot (v_1^* - \alpha v_2^*, v_0) \right] - (v_1^* - \alpha v_2^*, v) \ \forall v \in \mathcal{V}(C),
$$

that is the same as statement (c).



### II. INCOMPATIBILITY WITNESSES

We fix a finite dimensional complex Hilbert space $\mathcal{H}$, with $\dim \mathcal{H} = d$. We denote by $\mathcal{L}_s(\mathcal{H})$ the real linear space of all selfadjoint operators on $\mathcal{H}$, endowed with the uniform operator norm $\| \cdot \|$. We write $\mathbb{I}$ for the identity operator. If $Z$ is a set, we let $|Z|$ be its cardinality. A measurement with outcomes in a finite set $Z$ is any map $M : Z \to \mathcal{L}_s(\mathcal{H})$ such that $M(z) \geq 0$ for all $z \in Z$ and $\sum_{z \in Z} M(z) = \mathbb{I}$. The uniform measurement with outcomes in $Z$ is given by $U_z(z) = \mathbb{I} / |Z|$ for all $z \in Z$.

All measurements with outcomes in $Z$ constitute a closed and bounded convex subset $\mathcal{O}(Z)$ in the real linear space of all operator valued functions $H : Z \to \mathcal{L}_s(\mathcal{H})$. We denote by $\mathcal{F}(Z)$ the latter linear space of functions, and we regard it as a normed space with the sup-norm $\|H\|_{\infty} = \max\{\|H(z)\| : z \in Z\}$; the dimension of $\mathcal{F}(Z)$ is $d^2 |Z|$. For any $A \in \mathcal{L}_s(\mathcal{H})$, we define the affine set $\mathcal{F}_A(Z) = \{H \in \mathcal{F}(Z) \mid \sum_{z \in Z} H(z) = A\}$; the inclusion $\mathcal{O}(Z) \subset \mathcal{F}_1(Z)$ is clear.

If $X$ and $Y$ are finite sets, two measurements $A \in \mathcal{O}(X)$ and $B \in \mathcal{O}(Y)$ are compatible if there exists a third measurement $M \in \mathcal{O}(X \times Y)$ such that

$$
M_X(x) := \sum_{y \in Y} M(x, y) = A(x) \quad \text{and} \quad M_Y(y) := \sum_{x \in X} M(x, y) = B(y)
$$

(S2)

for all $x$ and $y$. In this case, we say that $M$ is a joint measurement of $A$ and $B$. We denote by $\mathcal{O}_{X,Y} = \mathcal{O}(X) \times \mathcal{O}(Y)$ the set of all pairs of measurements on $X$ and $Y$, and by $\mathcal{O}_{X,Y}^{\mathbb{C}}$ the subset of all pairs made up of compatible measurements.

If $(A, B) \in \mathcal{O}_{X,Y}^{\mathbb{C}} = \mathcal{O}_{X,Y} \setminus \mathcal{O}_{X,Y}^\com$, the two measurements $A$ and $B$ are incompatible.

As it is well known, unless $\mathcal{H} = \mathbb{C}$ or $\min(|X|, |Y|) = 1$, the inclusion $\mathcal{O}_{X,Y}^{\mathbb{C}} \subset \mathcal{O}_{X,Y}$ is strict. So, we will always assume this hypothesis in the following.

The sets $\mathcal{O}_{X,Y}$ and $\mathcal{O}_{X,Y}^{\mathbb{C}}$ are convex and compact in the direct product linear space $\mathcal{V} = \mathcal{F}(X) \times \mathcal{F}(Y)$; here, as the norm of $\mathcal{V}$ we choose the $\ell_\infty$-norm $\|(F, G)\|_{\infty} = \max\{|F|_{\infty}, |G|_{\infty}\}$. Indeed, only the compactness of $\mathcal{O}_{X,Y}^{\mathbb{C}}$ needs to be checked; it follows by the compactness of $\mathcal{O}(X \times Y)$ and the continuity of the mapping $\mathcal{O}(X \times Y) \ni M \mapsto (M_X, M_Y) \in \mathcal{O}(X) \times \mathcal{O}(Y)$.

The next proposition gives some further insight into the convex structure of the sets $\mathcal{O}_{X,Y}$ and $\mathcal{O}_{X,Y}^{\mathbb{C}}$.

**Proposition S5.** The following properties hold.

(a) $\mathcal{M}(\mathcal{O}_{X,Y}^{\mathbb{C}}) = \mathcal{M}(\mathcal{O}_{X,Y}) = \mathcal{F}_1(X) \times \mathcal{F}_1(Y)$.

(b) $\mathcal{V}(\mathcal{O}_{X,Y}^{\mathbb{C}}) = \mathcal{V}(\mathcal{O}_{X,Y}) = \mathcal{F}_0(X) \times \mathcal{F}_0(Y)$.

(c) $\text{ri}(\mathcal{O}_{X,Y}^{\mathbb{C}}) \subset \text{ri}(\mathcal{O}_{X,Y})$ and $(U_X, U_Y) \in \text{ri}(\mathcal{O}_{X,Y}^{\mathbb{C}})$.

**Proof.** Clearly,

$$
\mathcal{M}(\mathcal{O}_{X,Y}^{\mathbb{C}}) \subseteq \mathcal{M}(\mathcal{O}_{X,Y}) \subseteq \mathcal{F}_1(X) \times \mathcal{F}_1(Y) = (U_X, U_Y) + \mathcal{F}_0(X) \times \mathcal{F}_0(Y).
$$

In order to prove that the previous inclusions actually are equalities (item (a)), it is enough to show that

$$
(U_X, U_Y) + \varepsilon B \subseteq \mathcal{O}_{X,Y}^{\mathbb{C}} \quad \text{for} \quad \varepsilon = \frac{1}{2} \min \left\{ \frac{1}{|X|}, \frac{1}{|Y|} \right\} \quad \text{and} \quad B = \{(F, G) \in \mathcal{F}_0(X) \times \mathcal{F}_0(Y) \mid \|(F, G)\|_{\infty} < 1\}.
$$

in this expression, $B$ is the open unit ball in $\mathcal{F}_0(X) \times \mathcal{F}_0(Y)$. Indeed, if (**) holds, then

$$
(U_X, U_Y) + \mathcal{F}_0(X) \times \mathcal{F}_0(Y) = \mathcal{M}((U_X, U_Y) + \varepsilon B) \subseteq \mathcal{M}(\mathcal{O}_{X,Y}^{\mathbb{C}}).
$$
Now, \[\{\ast\}\] immediately follows, since for \((F,G) \in E\), the formula
\[
M(x,y) = U_{x \times y}(x,y) + \varepsilon \left[ \frac{1}{|Y|} F(x) + \frac{1}{|X|} G(y) \right]
\]
defines an element \(M \in \mathcal{O}(X \times Y)\) such that \((M_X, M_Y) = (U_X, U_Y) + \varepsilon(F,G)\). Having proven that in \[\{\ast\}\] all inclusions actually are equalities, item \[\{\ast\}\] is obvious, while item \[\{\ast\}\] follows from \[\{\ast\}\] and [1] Cor. 6.5.2.

The dual space \(V^*\) of \(V = \mathcal{F}(X) \times \mathcal{F}(Y)\) can be identified with \(V\) itself by means of the pairing
\[
\langle (F_1, G_1), (F_2, G_2) \rangle = \sum_{x \in X} \text{tr} [F_1(x)F_2(x)] + \sum_{y \in Y} \text{tr} [G_1(y)G_2(y)].
\]
(S3)

With this identification, Prop. S6\{(b)\} and a simple dimension counting lead to the equalities
\[
\mathcal{V}(\mathcal{O}_{X,Y}^\text{com}) = \mathcal{V}(\mathcal{O}_{X,Y}) = \{(F,G) \in \mathcal{F}(X) \times \mathcal{F}(Y) \mid F(x_1) = F(x_2)
\]
and \(G(y_1) = G(y_2)\) \(\forall x_1, x_2 \in X, y_1, y_2 \in Y\). (S4)

We recall from the main paper that any \(\mathcal{O}_{X,Y}^\text{com}\)-witness on the convex set \(\mathcal{O}_{X,Y}\) is an incompatibility witness (IW). Thus, by Prop. S2 and S3, any IW is of the form
\[
\xi(A,B) = \delta - \langle (F,G), (A,B) \rangle \quad \forall (A,B) \in \mathcal{O}_{X,Y}
\]
for some \((F,G) \in \mathcal{F}(X) \times \mathcal{F}(Y)\) and \(\delta \in [\delta_{\mathcal{O}_{X,Y}^\text{com}}(F,G), \delta_{\mathcal{O}_{X,Y}}(F,G)]\). In particular, \(\xi\) is tight if in the above formula \(\delta = \delta_{\mathcal{O}_{X,Y}^\text{com}}(F,G)\).

Combining Props. S3\{(b)\} and S5\{(c)\} immediately proves the following connection between tight IWs and the relative boundary of the set \(\mathcal{O}_{X,Y}^\text{com}\).

**Proposition S6.** Suppose \(\xi\) is a tight IW on the set \(\mathcal{O}_{X,Y}\) and \((A,B) \in \mathcal{O}_{X,Y}^\text{com}\). Then, the equality \(\xi(A,B) = 0\) implies that \((A,B) \in \partial\mathcal{O}_{X,Y}^\text{com}\). Moreover, there always exists some extreme point \((A_0, B_0)\) of \(\mathcal{O}_{X,Y}^\text{com}\) such that \(\xi(A_0, B_0) = 0\).

As a consequence of Prop. S5\{(c)\}, also the hypothesis of Prop. S4 is satisfied by the sets \(\mathcal{O}_{X,Y}\) and \(\mathcal{O}_{X,Y}^\text{com}\). We then obtain the following characterization of detection equivalence for two IWs.

**Proposition S7.** If \(\xi_1, \xi_2\) are two IWs on the set \(\mathcal{O}_{X,Y}\) and \(\xi_i(A,B) = \delta_i - \langle (F_i, G_i), (A,B) \rangle\) for all \((A,B) \in \mathcal{O}_{X,Y}\) and \(i = 1, 2\), then \(\xi_1 \approx \xi_2\) if and only if
\[
F_2(x) = \alpha F_1(x) + A, \quad G_2(y) = \alpha G_2(y) + B, \quad \delta_2 = \alpha \delta_1 + \text{tr}[A + B]
\]
for some \(\alpha > 0\) and \(A, B \in \mathcal{L}(\mathcal{H})\).

**Proof.** By Props. S4 and S5\{(c)\}, the equivalence \(\xi_1 \approx \xi_2\) holds if and only if, for some \(\alpha > 0\), both the following conditions are satisfied:
- \((F_1, G_1) - \alpha(F_2, G_2) \in \mathcal{V}(\mathcal{O}_{X,Y}^\text{com})\) \(\Leftrightarrow\) there exist \(A, B \in \mathcal{L}(\mathcal{H})\) such that \(F_1(x) - \alpha F_2(x) = A\) and \(G_1(y) - \alpha G_2(y) = B\) for all \(x \in X, y \in Y\) by S4;
- \([\delta_1 - \alpha \delta_2 = \langle (F_1, G_1) - \alpha(F_2, G_2), (U_X, U_Y) \rangle\) \(\Leftrightarrow\) \(\delta_1 - \alpha \delta_2 = \text{tr}[A] + \text{tr}[B]\) for \(A\) and \(B\) given by the previous item.

This concludes the proof.

We have already seen in the main paper that, up to detection equivalence, any tight IWs can be associated with a state discrimination problem with postmeasurement information. This problem consists in discriminating the classical labels of some partitioned state ensemble, that is, a couple \(\mathcal{E} = (\mathcal{E}, \{X, Y\})\), in which

(i) \(X\) and \(Y\) are disjoint finite sets, and

(ii) \(\mathcal{E}\) is a state ensemble with label set \(X \cup Y\), i.e., an element of \(\mathcal{F}(X \cup Y)\) such that \(\mathcal{E}(z) \geq 0\) for all \(z \in X \cup Y\) and \(\sum_{z \in X \cup Y} \text{tr}[\mathcal{E}(z)] = 1\).
Notice that the pairing (S3) for the restrictions $F_1 = \mathcal{E}|_X$ and $G_1 = \mathcal{E}|_Y$ rewrites
\begin{equation}
(\langle \mathcal{E}|_X, \mathcal{E}|_Y \rangle, (A, B)) = P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; A, B),
\end{equation}
where $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; A, B)$ is the guessing probability with premeasurement information defined in (4) of the main paper. It follows that
\begin{equation}
\delta_{\mathcal{O}_{X,Y}}(\mathcal{E}|_X, \mathcal{E}|_Y) = P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}), \quad \delta_{\mathcal{O}_{X,Y}^\text{com}}(\mathcal{E}|_X, \mathcal{E}|_Y) = P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}),
\end{equation}
where $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}})$ and $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ are the optimal guessing probabilities with premeasurement and postmeasurement information given by (S3) and (S4) of the main text. In particular, whenever the strict inequality $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) < P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}})$ holds, we can define the tight IW associated with the partitioned state ensemble $\hat{\mathcal{E}}$ as in (7) of the paper:
\begin{equation}
\xi_{\hat{\mathcal{E}}}(A, B) = P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) - P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; A, B) \quad \forall (A, B) \in \mathcal{O}_{X,Y}.
\end{equation}

We remark that the evaluation of $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}})$ consists in solving two separate standard state discrimination problems (one for the subensemble $\mathcal{E}_X = p(X)^{-1} \mathcal{E}|_X$ and the other for $\mathcal{E}_Y = p(Y)^{-1} \mathcal{E}|_Y$). On the other hand, the optimization problem in the definition of $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ can be turned into a single standard state discrimination problem by means of Thm. 2. Therefore, in order to evaluate both probabilities in (S6), one can resort to techniques from standard quantum state discrimination, as those described e.g. in Sec. IV B1, [4,7].

If the partitioned state ensemble $\hat{\mathcal{E}}$ is not trivial, any compatible pair of measurements solving the optimization problem in the definition of $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ necessarily lies on the relative boundary of the set $\mathcal{O}_{X,Y}^\text{com}$. More precisely, we have the following fact.

**Proposition S8.** Suppose the partitioned state ensemble $\hat{\mathcal{E}} = (\mathcal{E}, \{X, Y\})$ is such that either one of the restrictions $\mathcal{E}|_X$ or $\mathcal{E}|_Y$ is not a constant function. Then, for two compatible measurements $A$ and $B$, the equality $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; A, B) = P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$ entails that $(A, B) \in \partial \mathcal{O}_{X,Y}^\text{com}$. Moreover, there always exists some extreme point $(A_0, B_0)$ of $\mathcal{O}_{X,Y}^\text{com}$ such that $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}; A_0, B_0) = P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}})$.

**Proof.** By combining Lemma (S1) and (S5), (S6), the second claim is always true, while the first one holds whenever $(\mathcal{E}|_X, \mathcal{E}|_Y)$ is not an element of $\mathcal{V}(\mathcal{O}_{X,Y}^\text{com})^\times$. By (S4), this is equivalent to either $\mathcal{E}|_X$ or $\mathcal{E}|_Y$ being not constant.

Note that the last proposition does not require $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}) < P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}})$.

**III. INCOMPATIBILITY WITNESSES WITH TWO MUB**

In this section, we suppose $\{\varphi_h | h = 1, \ldots, d\}$ and $\{\psi_k | k = 1, \ldots, d\}$ are two fixed mutually unbiased bases (MUB) of the $d$-dimensional Hilbert space $H$. We will show how these bases can be used to construct a family of IWs for pairs of measurements with outcomes $X = \{(h, \varphi) | h = 1, \ldots, d\}$ and $Y = \{(k, \psi) | k = 1, \ldots, d\}$. Moreover, as a byproduct of this construction, we will also characterize the amount of uniform noise that is needed in order to make the two given MUB compatible. Note that, although in the separate sets $X$ and $Y$ the extra symbols $\varphi$ and $\psi$ are redundant, nonetheless they are needed to define the disjoint union $Z = X \cup Y$, and then consider the partition $\{X, Y\}$ of $Z$.

**A. Construction of the IWs**

For all $\mu = (\mu_\varphi, \mu_\psi) \in [1/(1-d),1] \times [1/(1-d),1]$, we can define the state ensemble $\mathcal{E}_\mu$ with label set $Z$, given as
\begin{equation}
\mathcal{E}_\mu(j, \ell) = \frac{1}{2d} \left[ \mu_\varphi |\ell_j\rangle |\ell_j\rangle + (1 - \mu_\varphi) \frac{1}{d} \right], \quad \forall j \in \{1, \ldots, d\}, \ell \in \{\varphi, \psi\},
\end{equation}
and the corresponding partitioned state ensemble $\hat{\mathcal{E}}_\mu = (\mathcal{E}, \{X, Y\})$. For this state ensemble, all labels $z \in Z$ occur with the same probability $p(z) = \text{tr} [\mathcal{E}_\mu(z)] = 1/(2d)$, and $p(X) = p(Y) = 1/2$. We always assume $\mu \neq (0,0)$ to avoid the trivial case.

In order to use the state ensemble $\mathcal{E}_\mu$ for constructing a tight IW as in (7) of the main paper, first of all we need to evaluate the pre- and postmeasurement guessing probabilities $P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}_\mu)$ and $P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}_\mu)$. To this aim, we recall the following two useful results.
**Proposition S9** (Prop. 2 of [3]). Suppose $\mathcal{E}$ is a state ensemble with label set $X$. For all $x \in X$, denote by $\lambda(x)$ the largest eigenvalue of $\mathcal{E}(x)$, and by $\Pi(x)$ the orthogonal projection onto the $\lambda(x)$-eigenspace of $\mathcal{E}(x)$. Define
\[ \lambda_E = \max_{x \in X} \lambda(x), \quad X_E = \{ x \in X : \lambda(x) = \lambda_E \}. \quad (S9) \]
Then, if there exists $\nu \in \mathbb{R}$ such that
\[ \sum_{x \in X_E} \Pi(x) = \nu \mathbb{1}, \quad (S10) \]
we have the following consequences:
(a) $\nu = \frac{1}{d} \sum_{x \in X_E} \text{rank} \Pi(x)$;
(b) $P_{\text{guess}}(\mathcal{E}) = d \lambda_E$;
(c) a measurement $M_0$ attaining the maximum guessing probability $P_{\text{guess}}(\mathcal{E})$ is
\[ M_0(x) = \begin{cases} \nu^{-1} \Pi(x) & \text{if } x \in X_E, \\ 0 & \text{if } x \notin X_E. \end{cases} \quad (S11) \]

**Theorem S1** (Thm. 2 of [3]). For any partitioned state ensemble $\hat{\mathcal{E}} = (\mathcal{E}, \{X, Y\})$, we have
\[ P_{\text{post}}^{\text{guess}}(\hat{\mathcal{E}}) = (|X| p(Y) + |Y| p(X)) P_{\text{guess}}(\mathcal{F}), \]
where $\mathcal{F}$ is the state ensemble with the Cartesian product label set $X \times Y$, given as
\[ \mathcal{F}(x, y) = \frac{\mathcal{E}(x) + \mathcal{E}(y)}{|X| p(Y) + |Y| p(X)} \quad \forall (x, y) \in X \times Y. \quad (S12) \]
Moreover, for a measurement $M : X \times Y \to \mathcal{L}(\mathcal{H})$, we have the equivalence
\[ P_{\text{guess}}(\mathcal{F}; M) = P_{\text{guess}}(\mathcal{F}) \iff P_{\text{post}}^{\text{guess}}(\hat{\mathcal{E}}; M_X, M_Y) = P_{\text{post}}^{\text{guess}}(\hat{\mathcal{E}}), \]
where $M_X$ and $M_Y$ are the two marginal measurements of $M$ defined in (S2).

We can immediately use Prop. S9 to evaluate the optimal guessing probability for the subensemble $\mathcal{E}_{\mu,X}(x) = (1/p(X)) \mathcal{E}_\mu(x) \quad \forall x \in X$. Indeed, we explicitly have
\[ \mathcal{E}_{\mu,X}(h, \varphi) = \frac{1}{d} \left[ \mu_{\varphi} |\varphi_h \rangle \langle \varphi_h| + (1 - \mu_{\varphi}) \frac{1}{d} \mathbb{1} \right], \]
whose largest eigenvalue is
\[ \lambda(h, \varphi) = \begin{cases} \frac{1}{d^2} \left[ 1 + (d - 1) \mu_{\varphi} \right] & \text{if } \mu_{\varphi} \geq 0, \\ \frac{1}{d^2} (1 - \mu_{\varphi}) & \text{if } \mu_{\varphi} < 0, \end{cases} \]
\[ = \frac{1}{d^2} \left[ 1 + \frac{1}{2} ((d - 2) \mu_{\varphi} + d |\mu_{\varphi}|) \right] \quad \forall h = 1, \ldots, d. \]
Thus, in (S9), the largest eigenvalue $\lambda_{\mu,X}$ is given by the latter expression, and it is attained on the whole label set $X_{\mu,X} = X$. Eq. (S10) is then easily verified, and so Prop. S9 yields
\[ P_{\text{guess}}(\mathcal{E}_{\mu,X}) = \frac{1}{d} \left[ 1 + \frac{1}{2} ((d - 2) \mu_{\varphi} + d |\mu_{\varphi}|) \right]. \]
A similar expression holds also for $P_{\text{guess}}(\mathcal{E}_{\mu,Y})$. We finally obtain
\[ P_{\text{post}}^{\text{guess}}(\hat{\mathcal{E}}) = p(X) P_{\text{guess}}(\mathcal{E}_{\mu,X}) + p(Y) P_{\text{guess}}(\mathcal{E}_{\mu,Y}) \]
\[ = \frac{1}{4} \left[ 2 + |\mu_{\varphi}| + |\mu_{\psi}| + \left( 1 - \frac{2}{d} \right) (\mu_{\varphi} + \mu_{\psi} - 2) \right]. \quad (S13) \]
Now we tackle the more difficult problem of evaluating \( P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}_{\mu}) \). In order to apply Thm. S1 we first write the auxiliary state ensemble \( \text{S12} \), which is

\[
\mathcal{F}((h, \varphi), (k, \psi)) = \frac{1}{2d^2} \left[ \mu_\varphi |\varphi_k\rangle \langle \varphi_k| + \mu_\psi |\psi_k\rangle \langle \psi_k| + \frac{1}{d} (2 - \mu_\varphi - \mu_\psi) \mathbb{1} \right]
\]

(S14)

for the partitioned state ensemble \( \hat{\mathcal{E}}_{\mu} \). Then, we check if we can apply Prop. S9 in order to calculate \( P_{\text{guess}}(\mathcal{F}) \). To this aim, we need to find the spectral decomposition of \( \mathcal{F}((h, \varphi), (k, \psi)) \) for all \( h, k \). The next lemma is useful for this task.

**Lemma S2.** Let \( \varphi, \psi \in \mathcal{H} \) be two unit vectors such that \( |\langle \varphi | \psi \rangle|^2 = 1/d \). Denote \( Q = |\varphi\rangle \langle \varphi| \) and \( P = |\psi\rangle \langle \psi| \), and let

\[
S = qQ + pP \quad \text{with} \quad q, p \in \mathbb{R} \quad \text{and} \quad (q, p) \neq (0, 0).
\]

Then, the eigenvalues of the selfadjoint operator \( S \) are

\[
\lambda_+ = \frac{1}{2} \left[ (q + p) + \sqrt{q^2 + p^2 - 2qp} \right] \quad \text{with multiplicity} \geq 1,
\]

\[
\lambda_- = \frac{1}{2} \left[ (q + p) - \sqrt{q^2 + p^2 - 2qp} \right] \quad \text{with multiplicity} \geq 1,
\]

\[
\lambda_0 = 0 \quad \text{with multiplicity} \geq d - 2,
\]

where

\[
\Delta = 1 - \frac{2}{d} \geq 0.
\]

They satisfy the following inequalities:
- if \( q \geq 0 \) and \( p \geq 0 \), then \( \lambda_+ > \lambda_- \geq 0 \), with equality if and only if \( qp = 0 \);
- if either \( q > 0 \) and \( p < 0 \), or \( q < 0 \) and \( p > 0 \), then \( \lambda_+ > \lambda_- \);
- if \( q \leq 0 \) and \( p \leq 0 \), then \( 0 \geq \lambda_+ > \lambda_- \), with equality if and only if \( qp = 0 \).

Moreover, the three selfadjoint operators

\[
\Pi_+ = \frac{1}{\lambda_+ - \lambda_-} \left[ qQ + pP - \frac{d\lambda_-}{d - 1} \left[ Q + P - (QP + PQ) \right] \right],
\]

\[
\Pi_- = \frac{1}{\lambda_- - \lambda_+} \left[ qQ + pP - \frac{d\lambda_+}{d - 1} \left[ Q + P - (QP + PQ) \right] \right],
\]

\[
\Pi_0 = 1 - \frac{d}{d - 1} \left[ Q + P - (QP + PQ) \right]
\]

constitute an orthogonal resolution of the identity, with \( \text{S} \Pi_k = \Pi_k S = \lambda_k S \) for all \( k \in \{+, -, 0\} \), and \( \text{rank} \Pi_+ = \text{rank} \Pi_- = 1 \), \( \text{rank} \Pi_0 = d - 2 \).

**Proof.** After noticing that the vectors \( \{ \varphi, \psi \} \) are linearly independent, we define the following two subspaces of \( \mathcal{H} \) with respective dimensions \( 2 \) and \( d - 2 \)

\[
\mathcal{H}_1 = \text{span} \{ \varphi, \psi \} , \quad \mathcal{H}_0 = \mathcal{H}_1^\perp.
\]

Then, \( \mathcal{S} \mathcal{H}_1 \subseteq \mathcal{H}_1 \) and \( \mathcal{S} \mathcal{H}_0 = \{0\} \). In particular, the eigenvalue \( 0 \) has multiplicity greater than or equal to \( d - 2 \) in \( S \).

The matrix form of \( S_1 \) with respect to the (nonorthogonal) basis \( \{ \varphi, \psi \} \) of \( \mathcal{H}_1 \) is

\[
S_1 = \begin{pmatrix}
q & q \\
p \langle \psi | \varphi \rangle & p
\end{pmatrix}.
\]

The roots of the characteristic polynomial of \( S_1 \) are the eigenvalues \( \lambda_+ \) and \( \lambda_- \) in \( \text{S16}, \text{S17} \). Since the quadratic form \( q^2 + p^2 - 2qp \) is positive definite, the square root in \( \text{S16} \) and \( \text{S17} \) is nonzero, hence \( \lambda_+ > \lambda_- \). The remaining inequalities involving \( 0 \) are straightforward calculations. The multiplicities of \( \lambda_+ \) and \( \lambda_- \) in \( S \) can not be less than the respective multiplicities in \( S_1 \), which are \( 1 \). This completes the proof of the statements about the eigenvalues.

Now, we claim that

\[
\Pi_1 = \frac{d}{d - 1} \left[ Q + P - (QP + PQ) \right]
\]

is the orthogonal projection onto \( \mathcal{H}_1 \). Indeed,
- $\Pi_1^* = \Pi_1$ (immediate);
- $\Pi_1|_{\mathcal{H}_0} = 0$ (immediate);
- $\Pi_1Q = Q$ (because $Q^2 = Q$ and $QPQ = (1/d)Q$) $\Rightarrow$ $\Pi_1\varphi = \varphi$;
- $\Pi_1P = P$ (because $P^2 = P$ and $PQP = (1/d)P$) $\Rightarrow$ $\Pi_1\psi = \psi$.

By applying the spectral theorem to $S_1$, there exist two rank-1 orthogonal projections $\Pi_+$ and $\Pi_-$ defined on $\mathcal{H}$ and satisfying the relations

$$\Pi_+ + \Pi_- = \Pi_1, \quad S = \lambda_+\Pi_+ + \lambda_-\Pi_-.$$

Inserting the expressions \([S15]\) for $S$ and \([S19]\) for $\Pi_1$ into these relations, and solving with respect to $\Pi_+$, $\Pi_-$, we get \([S19]\) and \([S20]\). To conclude the proof of the last claim of the lemma, note that the operator \([S21]\) is the orthogonal projection $\Pi_0 = 1 - \Pi_1$ onto $\mathcal{H}_0$.

The expression in the square root of \([S16]\), \([S17]\) is the quadratic form associated with the $2 \times 2$ matrix

$$G = \begin{pmatrix} 1 & -\Delta \\ -\Delta & 1 \end{pmatrix}.$$ 

Since $G$ is positive definite, this quadratic form actually is the squared norm

$$\| (q,p) \|^2_G = (q,p)G(q,p)^t = q^2 + p^2 - 2\Delta qp$$

defined by the Euclidean scalar product $\langle (q_1,p_1), (q_2,p_2) \rangle_G = (q_1,p_1)G(q_2,p_2)^t$ in $\mathbb{R}^2$. Remarkably, in the special case $d = 2$, we have $\Delta = 0$, hence $\langle \cdot, \cdot \rangle_G$ is the canonical Euclidean scalar product of $\mathbb{R}^2$. Also note that in this case, independently of the values of $q$ and $p$, the largest eigenvalue of $S$ is $\lambda_+$ with multiplicity 1, its associated eigenprojection is $\Pi_+$, and the orthogonal projection $\Pi_0$ is zero.

Assuming $d \geq 3$, only for $q > 0$ or $p > 0$ the largest eigenvalue of $S$ is still given by $\lambda_+$; otherwise, for $q \leq 0$ and $p \leq 0$, the largest eigenvalue of $S$ is 0. In the former case, the eigenvalue $\lambda_+$ has still multiplicity 1 and associated eigenprojection $\Pi_+$; in the latter case, the eigenvalue 0 has either multiplicity $d - 2$ and associated eigenprojection $\Pi_0$ (subcase $qp = 0$), or multiplicity $d - 1$ and associated eigenprojection $\Pi_0 + \Pi_+$ (subcase $qp \neq 0$).

As a consequence of the last two paragraphs, the spectral decomposition of the operators \([S14]\) is different according to the cases:

(i) $d = 2$ or $\mu_\varphi > 0$ or $\mu_\psi > 0$;
(ii) $d \geq 3$ and $\mu_\varphi \leq 0$ and $\mu_\psi \leq 0$, which in turn includes the subcases:

(a) $\mu_\varphi\mu_\psi \neq 0$;
(b) $\mu_\varphi\mu_\psi = 0$.

More precisely, all the operators $\{\mathcal{F}( (h, \varphi), (k, \psi) ) \mid (h, \varphi) \in X, (k, \psi) \in Y \}$ always have the same largest eigenvalue, which is

$$\lambda_\mathcal{F} = \begin{cases} \frac{1}{2d^2} \left[ 1 - \frac{2}{d^2} \right] (\mu_\varphi + \mu_\psi) + \frac{4}{d^2} & \text{in case (i)} \\ \frac{1}{2d^2} \left[ 2 - (\mu_\varphi + \mu_\psi) \right] & \text{in case (ii)} \end{cases}$$

by Lemma \([S22]\), in particular, the state ensemble $\mathcal{F}$ attains its largest eigenvalue on the whole label set $X \times Y$, that is, $(X \times Y)_\mathcal{F} = X \times Y$. Moreover, combining \([S19]\) and \([S21]\) according to the case at hand, the orthogonal projection onto the $\lambda_\mathcal{F}$-eigenspace of \([S14]\) is

$$\Pi((h, \varphi), (k, \psi)) = a1 + b(\mu_\varphi Q(h) + \mu_\psi P(k)) - \frac{dc}{d - 1} [Q(h)P(k) - (Q(h)P(k) + P(k)Q(h))],$$

where

$$Q(h) = |\varphi_h\rangle \langle \varphi_h|, \quad P(k) = |\psi_k\rangle \langle \psi_k|$$

and
Moreover, since the tight IW associated with the partitioned state ensemble  

Finally, the ensembles  

Proof. By (S13) and (S22),

$$
P_{\text{post}}(\hat{\mathcal{E}}_\mu) - P_{\text{post}}(\hat{\mathcal{E}}_\nu) = \begin{cases} 
\frac{1}{4} (|\mu_\varphi| + |\mu_\psi| - |\mu|_G) & \text{in case (i)} \\
0 & \text{in case (ii)} 
\end{cases}$$

Since

$$
\sum_{h=1}^{d} Q(h) = \sum_{k=1}^{d} P(k) = 1,
$$

we have

$$
\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \Pi(x,y) = \sum_{h,k=1}^{d} \Pi((h, \varphi), (k, \psi)) = \nu I
$$

with

$$
\nu = d^2a + db(\mu_\varphi + \mu_\psi) - 2dc = \begin{cases} 
d & \text{in case (i)} \\
d(d-2) & \text{in case (ii)} \\
d(d-1) & \text{in case (iii)}
\end{cases}.
$$

In all cases, Eq. (S10) holds, hence we can apply Prop. S9 to determine  

Finally, Thm. S1 yields  

Moreover,  

and

$$(\text{in the last expressions, } \delta_{\mu_\varphi,0} \text{ and } \delta_{\mu_\psi,0} \text{ are the usual Kronecker delta's). We can now determine the values of } \mu \text{ for which } P_{\text{post}}(\hat{\mathcal{E}}_\mu) < P_{\text{prior}}(\hat{\mathcal{E}}_\mu), \text{ and for these values explicitly evaluate the tight IW (S7) associated with } \hat{\mathcal{E}}_\mu. \text{ This yields the first main result of the present section.}$$

**Theorem S2** (Thm. 3 of the main paper). Let  

$$\mu = (\mu_\varphi, \mu_\psi) \in [1/(1-d), 1] \times [1/(1-d), 1] \text{ with } \mu \neq (0,0).$$

Then, we have the strict inequality  

$$P_{\text{post}}(\hat{\mathcal{E}}_\mu) < P_{\text{prior}}(\hat{\mathcal{E}}_\mu) \text{ if and only if } \mu_\varphi \mu_\psi \neq 0 \text{ and either } d = 2 \text{ or } \max\{\mu_\varphi, \mu_\psi\} > 0. \text{ In this case, the tight IW associated with the partitioned state ensemble } \hat{\mathcal{E}}_\mu \text{ by means of (S7) is}$$

$$
\xi_{\hat{\mathcal{E}}_\mu}(A, B) = \frac{1}{4} (\mu_\varphi + \mu_\psi + |\mu|_G) - \frac{1}{2d} \left\{ \mu_\varphi \sum_{h=1}^{d} \text{tr}[A(h, \varphi)Q(h)] + \mu_\psi \sum_{k=1}^{d} \text{tr}[B(k, \psi)P(k)] \right\}.
$$

Finally, the ensembles  

$$\hat{\mathcal{E}}_\mu \text{ and } \hat{\mathcal{E}}_\nu \text{ yield detection equivalent IWs if and only if } \nu = \alpha \mu \text{ for some } \alpha > 0.$$
The triangular inequality implies that the expression for case (i) strictly positive unless \( \mu_\varphi = 0 \) or \( \mu_\psi = 0 \). This proves the first claim.

Eq. (S25) then follows by combining (S22) with

\[
P_{\text{guess}}^\text{prior}(\mathcal{E}_\mu; A, B) = p(X)P_{\text{guess}}(\mathcal{E}_{\mu,X}; A) + p(Y)P_{\text{guess}}(\mathcal{E}_{\mu,Y}; B) = \frac{1}{2d} \left\{ \mu_\varphi \sum_h \text{tr}[A(h, \varphi)Q(h)] + \mu_\psi \sum_k \text{tr}[B(k, \psi)P(k)] + 2 - (\mu_\varphi + \mu_\psi) \right\}.
\]

Finally, we prove the equivalence statement. If \( \nu = \alpha \mu \), then \( \xi_{\mathcal{E}_\mu} = \alpha \xi_{\mathcal{E}_\mu} \) by (S25), hence the two witnesses are detection equivalent. Conversely, if \( \xi_{\mathcal{E}_\mu} = \xi_{\mathcal{E}_\mu} \), then by Prop. S7, there exist \( \alpha > 0 \) and \( A, B \in \mathcal{L}_s(\mathcal{H}) \) such that

\[
\mathcal{E}_\nu(h, \varphi) = \alpha \mathcal{E}_\mu(h, \varphi) + A \quad \forall h \quad \text{and} \quad \mathcal{E}_\nu(k, \psi) = \alpha \mathcal{E}_\mu(k, \psi) + B \quad \forall k.
\]

In particular, \( \mu_\ell = 0 \iff \nu_\ell = 0 \), as in this case both the maps \( \mathcal{E}_\mu(\cdot, \ell) \) and \( \mathcal{E}_\nu(\cdot, \ell) \) are constant. Since

\[
\mathcal{E}_\nu(j, \ell) = \frac{\nu_\ell}{\mu_\ell} \mathcal{E}_\mu(j, \ell) + \frac{1}{2d^2} \left( 1 - \frac{\nu_\ell}{\mu_\ell} \right) \quad \forall j, \ell \quad \left( \text{with } 0 \leq \nu_\ell \leq 0 \right),
\]

it must be \( \alpha = \nu_\varphi/\mu_\varphi = \nu_\psi/\mu_\psi \) and \( A = B = (1 - \alpha)/(2d^2) \). In particular, \( \nu = \alpha \mu \).

B. Incompatibility of noisy MUB

As a byproduct, the solution to the optimization problem for \( P_{\text{guess}}^\text{post}(\mathcal{E}_\mu) \) provided in the last subsection allows to characterize the compatibility of the following pair of measurements \( (A_\gamma, B_\gamma) \in \mathcal{O}_{X,Y} \)

\[
A_\gamma(h, \varphi) = \gamma_\varphi Q(h) + (1 - \gamma_\varphi) \frac{1}{d} \mathbb{I}, \quad B_\gamma(k, \psi) = \gamma_\psi P(k) + (1 - \gamma_\psi) \frac{1}{d} \mathbb{I}, \quad (S26)
\]

where \( \gamma = (\gamma_\varphi, \gamma_\psi) \in [1/1/(d-1), 1] \times [1/(d-1), 1] \). The latter set constitutes all the values of the parameters \( \gamma \) such that the operator valued maps \( A_\gamma, B_\gamma \) constitute two POVMs. Note that, for \( \gamma_\varphi, \gamma_\psi \geq 0 \), the measurements (S26) can be interpreted as uniformly noisy versions of the sharp measurements \( Q \) and \( P \).

Since \( P_{\text{guess}}^\text{prior}(\mathcal{E}_{\alpha \mu}; A, B) = \alpha P_{\text{guess}}^\text{prior}(\mathcal{E}_\mu; A, B) \), there is no restriction in parametrizing the family of state ensembles (S8) by means of a single real parameter \( \theta \). More precisely, we can choose the parametrization \( \mu = \mu(\theta) \) given by

\[
\mu_\varphi(\theta) = \alpha \left( \sqrt{d} \cos \theta + \sqrt{\frac{d}{d-1}} \sin \theta \right), \quad \mu_\psi(\theta) = \alpha \left( \sqrt{d} \cos \theta - \sqrt{\frac{d}{d-1}} \sin \theta \right)
\]

in terms of the angle

\[
\theta \in [-\theta_0, \theta_0] \quad \text{with} \quad \theta_0 = \pi - \arctan \sqrt{d-1}.
\]

Here, \( \alpha > 0 \) is fixed in such a way that \( |\mu_\varphi(\theta)| < 1/(d-1) \) and \( |\mu_\psi(\theta)| < 1/(d-1) \) for all \( \theta \in [-\theta_0, \theta_0] \); for example, \( \alpha = 1/(d\sqrt{d-1}) \). With this parametrization, we have \( \|\mu(\theta)\|_\mathcal{L} = 2 \alpha \) for all \( \theta \)'s. Note that \( \theta_0 \in (\pi/2, 3\pi/4) \), hence \([-\theta_0, \theta_0] \subseteq [-3\pi/4, 3\pi/4] \). For \( \theta \in [-\theta_0, \theta_0] \), the inequality \( \mu_\psi(\theta) < 0 \) then holds if and only if \( \theta \in (\pi - \theta_0, 0) \), while \( \mu_\varphi(\theta) = 0 \) if and only if \( \theta \in \{ -\theta_0, \pi - \theta_0 \} \). By the simmetry \( \mu_\varphi(\theta) = \mu_\psi(\pi - \theta) \), this implies that \( \max(\mu_\varphi(\theta), \mu_\psi(\theta)) \geq 0 \); moreover, \( \Theta = \{-\theta_0, \theta_0, \pi - \theta_0, \theta_0\} \) are all the values of \( \theta \) for which \( \mu_\varphi(\theta) \mu_\psi(\theta) = 0 \).

If \( \theta \in [-\theta_0, \theta_0] \), inserting \( \mu_\varphi(\theta) \) and \( \mu_\psi(\theta) \) into (S24), we obtain that the two measurements (S26) are compatible for \( \gamma = \gamma(\theta) = (\gamma_\varphi(\theta), \gamma_\psi(\theta)) \), where

\[
\gamma_\varphi(\theta) = \frac{d - 2 - d \cos(\theta + \theta_0)}{2(d-1)}, \quad \gamma_\psi(\theta) = \frac{d - 2 - d \cos(\theta - \theta_0)}{2(d-1)}; \quad (S27)
\]

Moreover, the equality \( P_{\text{guess}}^\text{post}(\mathcal{E}_{\mu(\theta)}) = P_{\text{guess}}^\text{prior}(\mathcal{E}_{\mu(\theta)}; A_{\gamma(\theta)}), B_{\gamma(\theta)} \) holds with this choice of \( \gamma \). By a continuity argument applied to both sides of the latter equation (where the probability in the left hand side is evaluated by means of (S22)), the equality extends to all \( \theta \in [-\theta_0, \theta_0] \). Therefore, by Prop. S8 we have \( (A_{\gamma(\theta)}, B_{\gamma(\theta)}) \in \partial \mathcal{O}_{X,Y}^\text{com} \) for \( \theta \)'s in the last interval. Note that for \( \lambda > 1 \) the two measurements \( A_{\lambda \gamma(\theta)} \) and \( B_{\lambda \gamma(\theta)} \) are necessarily incompatible, as otherwise we would get the contradiction \( (A_{\gamma(\theta)}, B_{\gamma(\theta)}) = (1 - 1/\lambda)(U_X, U_Y) + (1/\lambda)(A_{\lambda \gamma(\theta)}, B_{\lambda \gamma(\theta)}) \in \partial(\mathcal{O}_{X,Y}^\text{com}) \) by Prop. S8 and [1] Thm. 6.1].
In the \((\gamma, \gamma_0)-plane\), the curve \(\Gamma = \{ \gamma(\theta) \mid \theta \in [-\theta_0, \theta_0] \}\) given by (S27) is the part of the ellipse

\[
d(\gamma_\varphi^2 + \gamma_\psi^2) + 2(d-2)\gamma_\varphi\gamma_\psi - 2(d-2)(\gamma_\varphi + \gamma_\psi) = 4 - d
\]

(S28)

lying above the line \(\gamma_\varphi + \gamma_\psi = (d-3)/(d-1)\), as depicted in Fig. 3. In particular, all pairs of measurements corresponding to points beyond \(\Gamma\) are incompatible.

Now, in dimension \(d = 2\), the ellipse (S28) is actually the unit circle \(S^1 = \{(\gamma_\varphi, \gamma_\psi) \in [-1, 1] \times [-1, 1] \mid \gamma_\varphi^2 + \gamma_\psi^2 = 1\}\), and the curve \(\Gamma\) is the part of \(S^1\) lying in the 1st, 2nd and 4th quadrants. Moreover, in this case we have \(A_\varphi(1, \varphi) = A_\varphi(2, \varphi)\) for all \(\varphi \in [-1, 1] \times [-1, 1]\), and similarly for \(B_\varphi\). Thus, the set of \(\gamma\)'s such that \((A_\varphi, B_\varphi) \in O_{X, Y}^\cong\) is symmetric around the origin \(\gamma_0 = 0\); since also \(\gamma_0\) belongs to this set, by convexity we conclude that the measurements \(A_\varphi\) and \(B_\varphi\) are compatible if and only if \(\gamma_\varphi^2 + \gamma_\psi^2 \leq 1\).

On the other hand, if \(d \geq 3\), then by (S24) the two measurements \(A_\varphi\) and \(B_\varphi\) are compatible for \(\gamma_\varphi = (1/(1-d), 1/(1-d))\). By convexity, it then follows that all \(\gamma\)'s lying between \(\gamma_\varphi\) and the curve \(\Gamma\) correspond to compatible pairs of measurements.

The previous discussion is summarized in the second main result of this section.

**Theorem S3** (Thm. 4 of the main paper). (a) Suppose \(d = 2\). For \(\gamma \in [-1, 1] \times [-1, 1]\), the two measurements \(A_\varphi\) and \(B_\varphi\) of (S26) are compatible if and only if \(\gamma_\varphi^2 + \gamma_\psi^2 \leq 1\).

(b) Suppose \(d \geq 3\). For \(\gamma \in [1/(1-d), 1] \times [1/(1-d), 1]\), the two measurements \(A_\varphi\) and \(B_\varphi\) of (S26) are compatible if and only if

\[
\gamma_\varphi + \gamma_\psi \leq \frac{d-3}{d-1} \quad \text{or} \quad d(\gamma_\varphi^2 + \gamma_\psi^2) + 2(d-2)\gamma_\varphi\gamma_\psi - 2(d-2)(\gamma_\varphi + \gamma_\psi) \leq 4 - d,
\]

that is, \(\gamma\) belongs to the convex hull of the points

\[
\left\{ \left( \frac{d-2 - d\cos(\theta + \theta_0)}{2(d-1)}, \frac{d-2 - d\cos(\theta - \theta_0)}{2(d-1)} \right) \mid \theta \in [-\theta_0, \theta_0] \right\} \cup \left\{ \left( \frac{1}{1-d}, \frac{1}{1-d} \right) \right\},
\]

where \(\theta_0 = \pi - \arctan \sqrt{d-1}\).

Statement (a) of the previous theorem is well known [9]. On the other hand, statement (b) was proved in the particular case in which the two measurements \(A_\varphi\) and \(B_\varphi\) are conjugate by the Fourier transform of the cyclic group \(Z_d\), and restricting only to \(\gamma \in [0, 1] \times [0, 1]\) [8]. Compared with the group theoretical approach of [8], the present derivation of Thm. S3 has the advantage of not requiring any symmetry condition on the two MUB \(\{\varphi_1, \ldots, \varphi_d\}\) and \(\{\psi_1, \ldots, \psi_d\}\); thus, for dimensions \(d \geq 4\) it actually applies to many inequivalent pairs of MUB, and not only to the Fourier conjugate pairs considered in [8] (see [10] for a list of inequivalent pairs in dimensions \(4 \leq d \leq 16\)). In the general (not symmetric) case, a proof of statement (b) for \(\gamma\) constrained on the diagonal \(\gamma_\varphi = \gamma_\psi\) is contained in [11].
We finally remark that the essential differences that lead to separate results for the cases \( d = 2 \) and \( d \geq 3 \) are: (1) the additional symmetry \((A, B) \in \mathcal{O}_{X,Y}^\text{com} \iff (A, B) \in \mathcal{O}_{X,Y}^\text{com}\) which is specific of the \( d = 2 \) case; (2) the fact that, for \( \gamma_1 = (1/(1-d), 1/(1-d)) \), the two measurements \((A_{\gamma_1}, B_{\gamma_1})\) are compatible if and only if \( d \geq 3 \).

\[\text{IV. A BELL-CHSH INCOMPATIBILITY WITNESS}\]

The Bell-CHSH inequality can be regarded as a couple of IWs for pairs of dicotomic measurements with either outcomes \( X = \{ +\hat{a}, -\hat{a} \} \) or \( Y = \{ +\hat{b}, -\hat{b} \} \), where for the moment \( \pm \hat{a}, \pm \hat{b} \) are just symbols. Indeed, the Bell-CHSH inequality states that, if \((A, B) \in \mathcal{O}_{X,Y}\) constitutes a pair of compatible measurements, then both the bounds

\[
\xi^{\text{CHSH}}_c(A, B) = 2 + \text{tr}[g(A \otimes (C + D) + B \otimes (C - D))] \geq 0
\]

\[
\xi^{\text{CHSH}}(A, B) = 2 - \text{tr}[g(A \otimes (C + D) + B \otimes (C - D))] \geq 0
\]

must hold, where \( g \) is any state on the bipartite Hilbert space \( \mathcal{H} \otimes \mathcal{K} \), the maps \( C: X \rightarrow \mathcal{L}(\mathcal{K}) \) and \( D: Y \rightarrow \mathcal{L}(\mathcal{K}) \) are ancillary measurements, and

\[
A = A(+\hat{a}) - A(-\hat{a}) , \quad B = B(+\hat{b}) - B(-\hat{b}) , \quad C = C(+\hat{a}) - C(-\hat{a}) , \quad D = D(+\hat{b}) - D(-\hat{b})
\]

(see [12, Thm. 20.4]). Thus, in order for \( \xi^{\text{CHSH}}_c(k) \) \((k = \pm \) or \(- \)) to be an IW, it is enough to find a pair of incompatible measurements \((A, B)\) such that \( \xi^{\text{CHSH}}_c(A, B) < 0 \). Since \( \xi^{\text{CHSH}}_c(A, B) = \xi^{\text{CHSH}}(A', B') \), where \( A' = A(\pm \hat{a}) \) and \( B' = B(\mp \hat{b}) \), the map \( \xi^{\text{CHSH}}_c \) is an IW if and only if \( \xi^{\text{CHSH}}_c \) is such.

Now, we fix \( \mathcal{H} = \mathbb{C}^2 \), and in this case we provide a particular choice of the ancillary Hilbert space \( \mathcal{K} \), the bipartite state \( \varrho \), and the measurements \( C \) and \( D \), for which \( \xi^{\text{CHSH}}_c \) and \( \xi^{\text{CHSH}} \) turn out to be detection equivalent to the tight IW \( \langle S25 \rangle \) (Eq. \( \langle 14 \rangle \) of the main paper) constructed by means of two MUB in dimension \( d = 2 \). The choice is

\[
\mathcal{K} = \mathbb{C}^2 , \quad \varrho = \frac{1}{4} \left( \mathbb{1} \otimes \mathbb{1} - \sigma_1 \otimes \sigma_1 - \sigma_2 \otimes \sigma_2 - \sigma_3 \otimes \sigma_3 \right) , \quad C(\pm \hat{a}) = \frac{1}{2} \left( \mathbb{1} \pm \hat{a} \cdot \hat{\sigma} \right) , \quad D(\pm \hat{b}) = \frac{1}{2} \left( \mathbb{1} \pm \hat{\sigma} \cdot \hat{b} \right),
\]

where \( \hat{\sigma} = (\sigma_1, \sigma_2, \sigma_3) \) are the usual Pauli matrices, and \( \hat{a}, \hat{b} \) are the unit vectors \( \hat{a} = (\cos \theta, \sin \theta, 0) \) and \( \hat{b} = (\cos \theta, -\sin \theta, 0) \). (Note that \( \varrho \) is actually a pure state, as \( \varrho = \varrho^* = \varrho^2 \).)

With a straightforward calculation,

\[
\xi^{\text{CHSH}}_c(A, B) = 2 - k \left( \text{tr}[A \sigma_1] \cos \theta + \text{tr}[B \sigma_2] \sin \theta \right) = \frac{8}{\sqrt{\mu^2_\varphi + \mu^2_\psi}} \left( 1 + \frac{1}{4} \left( 2 - \sqrt{\mu^2_\varphi + \mu^2_\psi} - P_{\text{guess}}(\hat{c}^{(k)}; A, B) \right) \right),
\]

where we have set

\[
\cos \theta = \frac{\mu_\varphi}{\sqrt{\mu^2_\varphi + \mu^2_\psi}} , \quad \cos \theta = \frac{\mu_\psi}{\sqrt{\mu^2_\varphi + \mu^2_\psi}}
\]

and, for fixed \( k \in \{ +, -, 0 \} \), we have introduced the partitioned state ensemble \( \hat{c}^{(k)} \) \((\hat{c}^{(k)}, \{X, Y\})\), with

\[
\xi^{(k)}_c(\pm \hat{a}) = \frac{1}{8} \left( \mathbb{1} \pm k \mu_\varphi \sigma_1 \right) , \quad \xi^{(k)}_c(\pm \hat{b}) = \frac{1}{8} \left( \mathbb{1} \pm k \mu_\psi \sigma_2 \right) .
\]

Up to the relabeling

\[
+\hat{a} \rightarrow (1, \varphi) , \quad -\hat{a} \rightarrow (2, \varphi) , \quad +\hat{b} \rightarrow (1, \psi) , \quad -\hat{b} \rightarrow (2, \psi),
\]

we have that \( \xi^{(k)}_c \) coincides with the state ensemble \( \langle S8 \rangle \) constructed by means of the two MUB

\[
\varphi^\ast = \frac{e^{\mp i}}{\sqrt{2}} (|0\rangle + k|1\rangle) , \quad \varphi^\ast = \frac{e^{\pm i}}{\sqrt{2}} (|0\rangle - k|1\rangle)
\]

\[
\psi^\ast = \frac{e^{\mp i}}{\sqrt{2}} (|0\rangle + ik|1\rangle) , \quad \psi^\ast = \frac{e^{\pm i}}{\sqrt{2}} (|0\rangle - ik|1\rangle)
\]
where \(|\{0\}, |1\rangle\rangle is the canonical (computational) basis of \(C^2\). Therefore, by (S22),

\[
P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}^{(k)}_\mu) = \frac{1}{4} \left(2 + \sqrt{\mu_\varphi^2 + \mu_\psi^2}\right).
\]

Eq. (S29) then rewrites

\[
\xi_{k}^{\text{CHSH}}(A, B) = \frac{8}{\sqrt{\mu_\varphi^2 + \mu_\psi^2}} \left[ P_{\text{guess}}^{\text{post}}(\hat{\mathcal{E}}^{(k)}_\mu) - P_{\text{guess}}^{\text{prior}}(\hat{\mathcal{E}}^{(k)}_\mu, A, B) \right].
\]

This shows that \(\xi_{k}^{\text{CHSH}}\) is actually an IW, which is detection equivalent to the tight IWs associated with the partitioned state ensemble \(\hat{\mathcal{E}}^{(k)}_\mu\) and given in (S25).

[1] R.T. Rockafellar, *Convex analysis*. Princeton University Press, Princeton, NJ (1997)
[2] L. Vesely, unpublished notes
[3] C. Carmeli, T. Heinosaari, and A. Toigo, *Phys. Rev. A* 98 012126 (2018)
[4] A.S. Holevo, *J. Multivariate Anal.* 3 337–394 (1973)
[5] H.P. Yuen, R.S. Kennedy, and M. Lax, *IEEE Trans. Inform. Theory* IT-21 125–134 (1975)
[6] C.W. Helstrom, *Quantum Detection and Estimation Theory*. Academic Press, New York (1976)
[7] J. Bae, *New J. Phys.* 15 073037 (2013)
[8] C. Carmeli, T. Heinosaari, and A. Toigo, *Phys. Rev. A* 85 012109 (2012)
[9] P. Busch, *Phys. Rev. D* 33 2253–2261 (1986)
[10] W. Tadej and K. Życzkowski, *Open Syst. Inf. Dyn.* 13 133–177 (2006)
[11] R. Uola, T. Moroder, and O. Güne, *Phys. Rev. Lett.* 113 160403 (2014)
[12] P. Busch, P. Lahti, J.-P. Pellonpää, and K. Ylinen, *Quantum Measurement*. Springer International Publishing, Switzerland (2016)