http://eprints.gla.ac.uk/140294/

Deposited on: 08 May 2017

Optimal sequential measurements for bipartite state discrimination

Sarah Croke,[1] Stephen M. Barnett,[1] and Graeme Weir[1, *]

[1]*School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom*

(Dated: May 5, 2017)

State discrimination is a useful test problem with which to clarify the power and limitations of different classes of measurement. We consider the problem of discriminating between given states of a bipartite quantum system via sequential measurement of the subsystems, with classical feed-forward of measurement results. Our aim is to understand when sequential measurements, which are relatively easy to implement experimentally, perform as well, or almost as well, as optimal joint measurements, which are in general more technologically challenging. We construct conditions that the optimal sequential measurement must satisfy, analogous to the well-known Helstrom conditions for minimum error discrimination in the unrestricted case. We give several examples and compare the optimal probability of correctly identifying the state via global versus sequential measurement strategies.

I. INTRODUCTION

The problem of quantum state discrimination is most naturally thought of as a task in quantum communications, although it also has applications elsewhere in quantum information theory and quantum metrology [1–5]. The communications scenario is as follows: a sender, traditionally called Alice, chooses a quantum state ρ_i drawn from a given set $\{\rho_j\}$ with associated *a priori* probabilities $\{p_j\}$, and sends a system prepared in this state to a receiver, Bob. Bob knows the allowed set of states and their associated probabilities, and his task is to determine which state was sent, thereby recovering the message sent by Alice. The task was first considered in the pioneering work of Helstrom, Holevo, and others in the late 1960s and 1970s [6–12]. Various strategies exist, each optimising a different figure of merit (see, e.g., [10, 13–20]), and for arguably the simplest such figure, minimising the probability of error in identifying the state, necessary and sufficient conditions for a quantum measurement strategy to be optimal are known [8, 9].

More recently, state discrimination has proved a useful test problem with which to clarify the power and limitations of different classes of measurement. For information encoded across multiple quantum systems, the ability to measure jointly is strictly more powerful (but in general technologically more challenging) than the ability to measure each subsystem independently, even if many rounds of classical communication between systems are allowed. Intuitively, one might expect the difference in performance to be more pronounced when information is encoded in entangled states. That this is not necessarily the case was first revealed through two state discrimination problems. The first, so-called "nonlocality without entanglement," gave a set of multipartite orthogonal *product* states between which perfect discrimination is not possible using only local measurements and classical communication [21]. The second, complementary

and no less surprising, showed that any two orthogonal pure states, regardless of entanglement or multipartite structure, may be perfectly discriminated using only sequential measurement, i.e., local measurement on each system, with classical feed-forward [22]. This was later extended to show that any two generally non-orthogonal pure states may be discriminated optimally by sequential measurement of the subsystems, according to the commonly used minimum error [23] and unambiguous discrimination strategies [24–26].

Beyond the two-state examples, the situation becomes much less clear: for the next simplest example of discriminating three possible qubit states given two copies, it was postulated by Peres and Wootters in 1991 that local measurement was strictly weaker than joint measurement on both copies [27], and only 20 years later was it finally proved that such a gap exists for this problem, for the minimum error strategy [28].

In this paper we consider sequential measurements on a bipartite system; i.e. subsystem A and B are measured in turn, and the choice of measurement performed on subsystem B is allowed to depend in general on the result of measurement of A. This is often a physically relevant class of measurement; for example, if A and B are in different laboratories it is easy to imagine that feed-forward of measurement results from laboratory A to laboratory B would be practical, but many rounds of classical communication could become unfeasible. Alternatively, if A and B interact only weakly or not at all (e.g., photons), joint measurements are difficult to perform, while classical feed-forward from one detector to another apparatus is relatively easily achieved with current technology (see, e.g., Ref. [29] for such an experiment in the state discrimination context). It is natural then to ask how well information can be retrieved with this restriction on the measurement strategy that may be employed. Furthermore, implementations of joint measurement strategies for extracting information may provide applications for small quantum processors [30], and it is useful to understand when the additional experimental challenge of joint measurement may provide a significant advantage over local measurement strategies. For simplicity, we restrict

to bipartite instead of the more general multipartite state discrimination.

We begin with the case where the bipartite state is simply a two-copy state. We construct necessary conditions that a given sequential measurement must satisfy to be optimal in the sense of minimising the error in determining the state, analogous to the well-known Helstrom conditions [8, 9]. We further find a condition which is both necessary and sufficient, but which requires optimisation over an arbitrary measurement on one subsystem. We illustrate the two-copy case through the example of the trine states considered in [27, 28], and give the probabilities of correctly identifying the state for sequential and global strategies, as well as discussing features of the optimal measurements in each case.

We extend the discussion to arbitrary bipartite states, and as an example give the optimal sequential strategies for discriminating three Bell states, and for discriminating the so-called domino states introduced by Bennett *et al.* in [21]. Finally, we discuss an interpretation for our necessary and sufficient condition in terms of a related discrimination problem.

II.   REVIEW: HELSTROM CONDITIONS

We first recall the minimum error problem, where there are no restrictions on the allowed measurement: a quantum system is prepared in one of a known set of states $\{\rho_i\}$ with associated probabilities $\{p_i\}$. Any physically allowed measurement may be represented by a POVM (positive operator-valued measure) [31], also referred to as a POM (probability operator measure) [11], that is, a set of Hermitian operators $\{\pi_i\}$ satisfying

$$\pi_i \geq 0,$$
$$\sum_i \pi_i = \mathbb{1}.$$

For a measurement described by operators $\{\pi_i\}$, if outcome i is taken to indicate state ρ_i, the probability of correctly identifying the state is given by

$$P_{\mathrm{corr}} = \sum_i p_i \mathrm{Tr}(\rho_i \pi_i). \tag{1}$$

The operators $\{\pi_i\}$ describing the optimal measurement satisfy the following conditions [5, 8, 9, 32]:

$$\sum_i p_i \rho_i \pi_i - p_j \rho_j \geq 0, \quad \forall j \tag{2}$$

$$\pi_i (p_i \rho_i - p_j \rho_j) \pi_j = 0, \quad \forall i, j. \tag{3}$$

It is worth noting that the conditions are not independent, as the second follows from the first. Condition (3) may be thought of as analogous to the condition in an optimisation problem that the first derivative vanish at a stationary point, while condition (2) is analogous to the second derivative condition: it is the sign of the second derivative which determines whether the corresponding point is a local maximum or local minimum. Condition (3) is therefore necessary but not sufficient for $\{\pi_i\}$ to be an optimal measurement, however (2) is both necessary and sufficient. We give here a sketch of the proof, following the treatment of [32], which is extended to the sequential case in the rest of the paper.

If $\{\pi_i\}$ is optimal then for all other physically allowed measurements $\{\pi_i'\}$ we require

$$P_{\mathrm{corr}}(\{\pi_i\}) \geq P_{\mathrm{corr}}(\{\pi_i'\}).$$

From this we obtain

$$\sum_i p_i \mathrm{Tr}(\rho_i \pi_i) - \sum_j p_j \mathrm{Tr}(\rho_j \pi_j') \geq 0$$

$$\sum_j \mathrm{Tr}\left[\left(\sum_i p_i \rho_i \pi_i - p_j \rho_j\right)\pi_j'\right] \geq 0. \tag{4}$$

Note that for positive operators A, B it is always true that $\mathrm{Tr}(AB) \geq 0$, which may be seen by evaluating the trace in the eigenbasis of A:

$$\mathrm{Tr}(AB) = \sum_i \langle a_i | AB | a_i \rangle = \sum_i a_i \langle a_i | B | a_i \rangle \geq 0, \tag{5}$$

where $a_i \geq 0$ are the eigenvalues of A, $\{|a_i\rangle\}$ are the eigenkets of A, and the inequality follows from the positivity of B. As π_j' is a positive operator it is therefore clear that condition (2) is sufficient in order for the inequality (4) to be satisfied. That this condition is also necessary may be shown by introducing the Hermitian operators

$$G_j = \sum_i p_i \frac{1}{2}\{\rho_i, \pi_i\} - p_j \rho_j.$$

Now if $\exists |\lambda\rangle$ such that $\langle \lambda | G_j | \lambda \rangle < 0$, the variation

$$\pi_i' = (\mathbb{1} - \epsilon|\lambda\rangle\langle\lambda|)\pi_i(\mathbb{1} - \epsilon|\lambda\rangle\langle\lambda|) + \epsilon(2+\epsilon)|\lambda\rangle\langle\lambda|\delta_{ij} \tag{6}$$

results in a measurement with higher probability of success than $\{\pi_i\}$, which therefore cannot be an optimal measurement. Finally it is possible to show [32] that

$$\sum_j G_j \pi_j = 0$$

and thus

$$\sum_i p_i \frac{1}{2}\{\rho_i, \pi_i\} = \sum_j p_j \rho_j \pi_j.$$

Thus the requirement $G_j \geq 0$ reduces to condition (2), which is therefore both necessary and sufficient.

It is useful to denote $\Gamma = \sum_i p_i \rho_i \pi_i$. We finish by noting that for an optimal measurement $\{\pi_j'\}$, we require

$$P_{\mathrm{corr}} = \mathrm{Tr}(\Gamma) = \sum_j p_j \mathrm{Tr}(\rho_j \pi_j'),$$

and therefore

$$\sum_j \text{Tr}[(\Gamma - p_j\rho_j)\pi'_j] = 0.$$

As discussed above, for positive operators A, B, $\text{Tr}(AB) \geq 0$, and it is clear from Eq. (5) that equality holds if and only if $AB = 0$. Thus we require that each term in the sum be identically zero, which further requires

$$(\Gamma - p_j\rho_j)\pi'_j = 0, \quad \forall j \qquad (7)$$

for any optimal measurement $\{\pi'_j\}$. This is an alternative necessary (but not sufficient) condition and is sometimes useful for finding optimal measurements. It also implies, on summing over j, that Γ is unique, $\Gamma = \sum_j p_j\rho_j\pi'_j$ for any optimal $\{\pi'_j\}$ (see also [33]).

III. TWO-COPY STATE DISCRIMINATION WITH SEQUENTIAL MEASUREMENT

A. Necessary conditions

Now let us consider the two-copy case, with sequential measurement. Suppose therefore we are provided with two copies of a state drawn from a known set $\{\rho_i\}$ with associated probabilities $\{p_i\}$. The allowed measurement procedures are as follows: make a measurement described by some POVM $\{M_j^A\}$ on system A; given outcome j, make a measurement on system B, as shown in the tree in Fig. 1. As the choice of measurement on system B
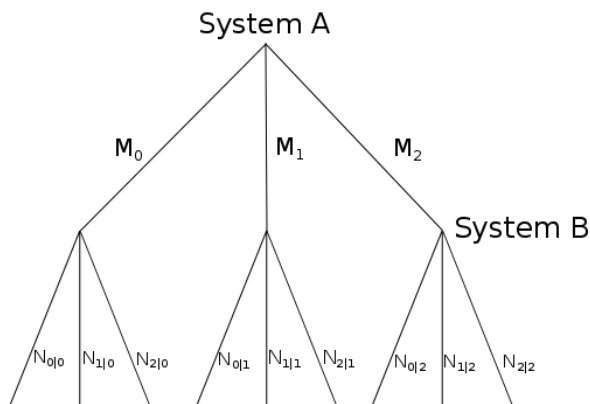

FIG. 1. Probability tree showing sequential measurement notation. The measurement described by POVM $\{M_j\}$ is performed on system A. Given outcome j, the measurement described by POVM $\{N_{i|j}\}$ is performed on system B.

can in general depend on the outcome of measurement on A, we denote the associated POVM $\{N_{i|j}^B\}$, where for all i and j, $N_{i|j}^B \geq 0$, and for each j,

$$\sum_i N_{i|j}^B = \mathbb{1}^B.$$

The measurement on the joint AB system is thus of the form $\{\pi_i = \sum_j M_j^A \otimes N_{i|j}^B\}$, with the probability of correctly identifying the state given by

$$\begin{aligned}
\text{P}_{\text{corr}} &= \sum_{ij} p_i \text{Tr}_{AB}\left(\rho_i^A \otimes \rho_i^B M_j^A \otimes N_{i|j}^B\right) \\
&= \sum_{ij} p_i \text{Tr}_A\left(\rho_i^A M_j^A\right) \text{Tr}_B\left(\rho_i^B N_{i|j}^B\right). \qquad (8)
\end{aligned}$$

In the following we drop the superscripts A, B, whenever it is not confusing to do so. We begin by pointing out that each of $\{M_j\}$, $\{N_{i|j}\}$ may be interpreted as an optimal measurement for an appropriately defined discrimination problem, as follows. We first note that, given measurement result j on system A, we can update the probabilities as follows, using Bayes' rule:

$$\Pr(i|M_j) = \frac{\Pr(i, M_j)}{\Pr(M_j)} = \frac{p_i\text{Tr}_A(\rho_i M_j)}{\sum_k p_k\text{Tr}_A(\rho_k M_j)} = p_{i|j}. \quad (9)$$

Thus given result j on system A, the possible states $\{\rho_i\}$ of system B occur with probabilities $p_{i|j}$. Clearly $\{N_{i|j}\}$ should thus be optimal for discriminating the states ρ_i with the updated priors $p_{i|j}$, and thus a necessary condition is

$$\sum_i p_{i|j}\rho_i N_{i|j} - p_{k|j}\rho_k \geq 0, \quad \forall k,$$

or equivalently, using Eq. (9),

$$\sum_i p_i\text{Tr}_A(\rho_i M_j)\rho_i N_{i|j} - p_k\text{Tr}_A(\rho_k M_j)\rho_k \geq 0, \quad \forall k,$$
$$(10)$$

which must hold for each j. This set of conditions is necessary, but not sufficient (we have not done any optimisation over M_j). Finally, summing over j gives

$$\text{Tr}_A\left(\sum_{i,j} p_i(\rho_i \otimes \rho_i)(M_j \otimes N_{i|j}) - p_k\rho_k \otimes \rho_k\right) \geq 0, \quad \forall k,$$
$$(11)$$

which is rather similar to the Helstrom condition (2), but with a partial trace over system A.

Conversely, we can rewrite Eq. (8) as follows:

$$\begin{aligned}
\text{P}_{\text{corr}} &= \sum_j \text{Tr}_A\left[\sum_i p_i\text{Tr}_B\left(\rho_i^B N_{i|j}^B\right)\rho_i^A M_j^A\right] \\
&= \sum_j c_j\text{Tr}_A\left(\sigma_j^A M_j^A\right),
\end{aligned}$$

where we have defined

$$\sigma_j^A = \frac{\sum_i p_i\text{Tr}_B\left(\rho_i^B N_{i|j}^B\right)\rho_i^A}{\sum_k p_k\text{Tr}_B\left(\rho_k^B N_{k|j}^B\right)}, \qquad (12)$$

$$c_j = \sum_i p_i\text{Tr}_B\left(\rho_i^B N_{i|j}^B\right). \qquad (13)$$

We can interpret the trace one operators $\{\sigma_j\}$ as density operators, and if we further define probabilities $q_j = c_j / (\sum_i c_i)$, it follows that $\{M_j^A\}$ must be optimal for discriminating the states $\{\sigma_j^A\}$ with probabilities $\{q_j\}$. The Helstrom condition (2) then gives

$$\sum_j q_j \sigma_j^A M_j^A - q_k \sigma_k^A \geq 0,$$

which may be rewritten as

$$\sum_j \left[\sum_i p_i \mathrm{Tr}_B \left(\rho_i^B N_{i|j}^B \right) \rho_i^A \right] M_j^A$$
$$- \sum_i p_i \mathrm{Tr}_B \left(\rho_i^B N_{i|k}^B \right) \rho_i^A \geq 0. \quad (14)$$

Finally, we obtain

$$\mathrm{Tr}_B \left(\sum_{ij} p_i (\rho_i^A \otimes \rho_i^B)(M_j^A \otimes N_{i|j}^B) \right.$$
$$\left. - \sum_i p_i (\rho_i^A \otimes \rho_i^B)(\mathbb{1}^A \otimes N_{i|k}^B) \right) \geq 0. \quad (15)$$

Again, this is necessary, but not sufficient (this time we have not done any optimisation over $N_{i|j}$). One might hope that the conditions (10) and (15) when taken together are also sufficient, and could then imagine that it may be possible to construct an iterative procedure for numerical solution of the optimization problem. However, this turns out not to be the case; we will return to this point later. Each of conditions (10) and (15), however, have a clear interpretation. Note that it might have been expected that $\{N_{i|j}^B\}$ should be optimal for the updated priors given measurement of A; that M_j^A plays a complementary role for a different discrimination problem is less obvious *a priori*. We return to give an interpretation of this discrimination problem later.

B. A necessary and sufficient condition

We now turn to the problem of simultaneously optimising both the measurement on A and that on system B. We find that the condition

$$\sum_{i,j} p_i \mathrm{Tr}_B(\rho_i N_{i|j}) \rho_i M_j - \sum_k p_k \mathrm{Tr}_B(\rho_k \widetilde{N}_k) \rho_k \geq 0, \quad (16)$$

where $\{\widetilde{N}_k\}$ is any physically allowed measurement on system B, is both necessary and sufficient for optimality of $\{\pi_i = \sum_j M_j \otimes N_{i|j}\}$. Unfortunately, this still contains an arbitrary measurement on system B, and thus is not as readily applicable as the original Helstrom conditions to verify optimality of a candidate measurement. Nevertheless we will give examples in which it can be used to prove optimality analytically. We also note that the

inclusion of an arbitrary measurement on one subsystem means that analysis beyond the bipartite case becomes complicated and our method is not readily extended to multipartite discrimination.

We begin by proving the sufficiency of condition (16). If $\{\pi_i = \sum_j M_j \otimes N_{i|j}\}$ is optimal among sequential measurements, we require

$$\mathrm{Tr}_{AB} \left(\sum_{i,j} p_i (\rho_i \otimes \rho_i)(M_j \otimes N_{i|j}) \right) \geq$$
$$\mathrm{Tr}_{AB} \left(\sum_{k,l} p_k (\rho_k \otimes \rho_k)(M_l' \otimes N_{k|l}') \right),$$

for all $\{\pi_k' = \sum_l M_l' \otimes N_{k|l}'\}$. Inserting the identity $\sum_l M_l' \otimes \mathbb{1}$ and rearranging gives

$$\sum_l \mathrm{Tr}_{AB} \left[\left(\sum_{i,j} p_i (\rho_i \otimes \rho_i)(M_j \otimes N_{i|j}) \right. \right.$$
$$\left. \left. - \sum_k p_k (\rho_k \otimes \rho_k)(\mathbb{1} \otimes N_{k|l}') \right) M_l' \right] \geq 0,$$

$$\sum_l \mathrm{Tr}_A \left[\left(\sum_{i,j} p_i \mathrm{Tr}_B(\rho_i N_{i|j}) \rho_i M_j \right. \right.$$
$$\left. \left. - \sum_k p_k \mathrm{Tr}_B(\rho_k N_{k|l}') \rho_k \right) M_l' \right] \geq 0.$$

Condition (16) is therefore sufficient, if $\{\widetilde{N}_k\}$ is any allowed measurement on B.

That condition (16) is also necessary may be seen as follows: as in the unrestricted case, we introduce the manifestly Hermitian operator

$$\Gamma_{\mathrm{sym}}^A = \sum_{i,j} p_i \mathrm{Tr}_B \left(\rho_i N_{i|j} \right) \frac{1}{2} \{\rho_i, M_j\}.$$

Suppose now that there exists some $|\lambda\rangle$ and some $\{\widetilde{N}_k\}$ such that

$$\langle \lambda | \Gamma_{\mathrm{sym}}^A - \sum_k p_k \mathrm{Tr}_B(\rho_k \widetilde{N}_k) \rho_k | \lambda \rangle < 0.$$

We can construct a variation of $\{\pi_i = \sum_j M_j \otimes N_{i|j}\}$ as follows:

$$M_j' = (\mathbb{1} - \epsilon|\lambda\rangle\langle\lambda|)M_j(\mathbb{1} - \epsilon|\lambda\rangle\langle\lambda|), \quad 0 \leq j < n$$
$$N_{i|j}' = N_{i|j}, \quad 0 \leq j < n$$
$$M_n = \epsilon(2 + \epsilon)|\lambda\rangle\langle\lambda|,$$
$$N_{i|n} = \widetilde{N}_i,$$

where $0 < \epsilon \ll 1$. Note that if $\{M_j^A\}$ has n outcomes, the primed measurement on system A has $n + 1$ outcomes.

Now note that

$$P_{\text{corr}}\left(\{M'_j \otimes N'_{i|j}\}\right) = P_{\text{corr}}\left(\{M_j \otimes N_{i|j}\}\right)$$

$$-\epsilon \text{Tr}_{AB}\left(\sum_{i,j} p_i \rho_i \otimes \rho_i \left(|\lambda\rangle\langle\lambda|M_j + M_j|\lambda\rangle\langle\lambda|\right) \otimes N_{i|j}\right)$$

$$+2\epsilon \text{Tr}_{AB}\left(\sum_i p_i \rho_i \otimes \rho_i(|\lambda\rangle\langle\lambda| \otimes \widetilde{N}_i)\right) + O(\epsilon^2)$$

$$= P_{\text{corr}}\left(\{M_j \otimes N_{i|j}\}\right)$$

$$-2\epsilon\langle\lambda|\Gamma^A_{\text{sym}} - \sum_i p_i \text{Tr}_B(\rho_i\widetilde{N}_i)\rho_i|\lambda\rangle + O(\epsilon^2)$$

$$> P_{\text{corr}}\left(\{M_j \otimes N_{i|j}\}\right).$$

Finally, we note that, by virtue of the fact that $\{M_j\}$ is an optimal measurement for discriminating the states σ_j, it follows that $\Gamma^A_{\text{sym}} = \Gamma^A$, where Γ^A is defined as

$$\Gamma^A = \sum_{i,j} p_i \text{Tr}_B\left(\rho_i N_{i|j}\right)\rho_i M_j.$$

Thus we require

$$\Gamma^A - \sum_k p_k \text{Tr}_B\left(\rho_k \widetilde{N}_k\right)\rho_k \geq 0,$$

which completes our proof.

IV. EXAMPLE: THE DOUBLE TRINE ENSEMBLE

As an example we consider the so-called double trine ensemble: two copies of the trine states, for which $\rho_j = |\psi_j\rangle\langle\psi_j|$, and

$$|\psi_j\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{2\pi ji/3}|1\rangle\right).$$

These each occur with prior probabilities $p_j = \frac{1}{3}$ and have the symmetry property

$$|\psi_j\rangle = U^j|\psi_0\rangle$$

where U is a rotation of $\frac{2\pi}{3}$ around the z axis in the Bloch sphere.

A. Optimal sequential measurement

For the two-copy case, Chitambar and Hsieh [28] showed that the optimal sequential measurement rules out one state of the three in the first step, and corresponds to the Helstrom measurement to distinguish between the remaining two states in the second step. We first briefly present this optimal measurement and then use it to demonstrate our conditions.

The optimal sequential measurement thus makes the measurement $\{M_j = \frac{2}{3}|\psi_j^\perp\rangle\langle\psi_j^\perp|\}$ on the first copy, where the states $\{|\psi_j^\perp\rangle\}$ form the so-called antitrine ensemble:

$$|\psi_j^\perp\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - e^{2\pi ji/3}|1\rangle\right).$$

Following this measurement, the updated priors become $p_{i|j} = \frac{1}{2}(1 - \delta_{ij})$, and $\{N_{i|j}\}$ is then the optimal measurement to distinguish the two remaining equiprobable pure states $\{|\psi_i\rangle, |\psi_k\rangle, i \neq j \neq k\}$. This is a case of the well-known Helstrom measurement and is a projective measurement in a basis located symmetrically around the signal states (see, e.g., [1]). Thus for $i = j$, $N_{i|j} = 0$, and for $i \neq j$ we denote $N_{i|j} = |\phi_{i|j}\rangle\langle\phi_{i|j}|$, where

$$|\phi_{1|0}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + i|1\rangle\right),$$

$$|\phi_{2|0}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - i|1\rangle\right),$$

$$|\phi_{0|1}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{i\pi/6}|1\rangle\right) = U|\phi_{2|0}\rangle,$$

$$|\phi_{2|1}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - e^{i\pi/6}|1\rangle\right) = U|\phi_{1|0}\rangle,$$

$$|\phi_{0|2}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{-i\pi/6}|1\rangle\right) = U^2|\phi_{1|0}\rangle,$$

$$|\phi_{1|2}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - e^{-i\pi/6}|1\rangle\right) = U^2|\phi_{2|0}\rangle.$$

These states, along with the trine and antitrine states, are shown in the Bloch sphere picture in Fig. 2.
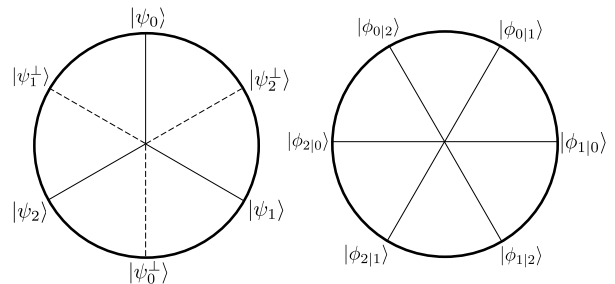
FIG. 2. Trine and antitrine states shown in the equator of the Bloch sphere (left). Bases defined by the optimal Helstrom measurements in step two of the optimal sequential measurement procedure for discriminating the two-copy trine ensemble (right).

B. Necessary and sufficient conditions

We now use this strategy to illustrate the conditions presented in the previous section. From the symmetry we find that $\text{Tr}(\rho_i N_{i|j}) = p_H(1 - \delta_{ij})$, for all i, j, where

p_H is the probability of success of the Helstrom measurement distinguishing between two equiprobable states with overlap $|\langle\psi_i|\psi_k\rangle| = |\langle\psi_0|\psi_1\rangle| = 1/2$, i.e., from [11]:

$$p_H = \frac{1}{2}\left(1 + \sqrt{1 - |\langle\psi_0|\psi_1\rangle|^2}\right) = \frac{1}{2}\left(1 + \frac{\sqrt{3}}{2}\right).$$

By construction, this measurement strategy satisfies condition (10). To evaluate (15) and the necessary and sufficient condition (16), we first calculate Γ^A:

$$\begin{aligned}\Gamma^A &= \sum_{i,j} p_i \mathrm{Tr}\left(\rho_i N_{i|j}\right)\rho_i M_j\\
&= \sum_{i,j} \frac{1}{3}p_H(1 - \delta_{ij})\left(|\psi_i\rangle\langle\psi_i|\right)\left(\frac{2}{3}|\psi_j^\perp\rangle\langle\psi_j^\perp|\right)\\
&= \frac{1}{3}p_H\left(\sum_i |\psi_i\rangle\langle\psi_i|\right)\left(\sum_j \frac{2}{3}|\psi_j^\perp\rangle\langle\psi_j^\perp|\right)\\
&= \frac{1}{2}p_H\mathbb{1} = \frac{1}{4}\left(1 + \frac{\sqrt{3}}{2}\right)\mathbb{1},\end{aligned}$$

where in the last line we have used $\sum_j \frac{2}{3}|\psi_j^\perp\rangle\langle\psi_j^\perp| = \sum_j \frac{2}{3}|\psi_j\rangle\langle\psi_j| = \mathbb{1}$. We first show that the strategy satisfies condition (15). We obtain

$$\begin{aligned}\sum_i p_i \mathrm{Tr}\left(\rho_i N_{i|j}\right)\rho_i^A &= \frac{1}{3}\sum_i p_H(1 - \delta_{ij})\rho_i\\
&= \frac{1}{2}p_H\left(\mathbb{1} - \frac{2}{3}\rho_j\right),\end{aligned}$$

from which it is clear that condition (15) is satisfied for each j. Finally, to prove that this is indeed the optimal strategy, we must show that it satisfies the necessary and sufficient condition (16). As we have shown that Γ^A is proportional to the identity, this amounts to showing that for any allowed measurement $\{\widetilde{N}_k\}$ on system B, the largest eigenvalue of the operator

$$\sum_k p_k \mathrm{Tr}\left(\rho_k \widetilde{N}_k\right)\rho_k$$

is bounded by $\frac{1}{2}p_H = \frac{1}{4}\left(1 + \frac{\sqrt{3}}{2}\right)$. The proof that this holds is straight-forward but needs a few steps, and the details are given in Appendix A.

The probability of correctly identifying the state using the optimal sequential measurement is given by

$$\mathrm{P}_{\mathrm{corr}}^{\mathrm{seq}} = \mathrm{Tr}(\Gamma^A) = p_H = \frac{1}{2}\left(1 + \frac{\sqrt{3}}{2}\right) \simeq 0.933.$$

C. Comparison of global and sequential schemes

For comparison we recall the globally optimal measurement strategy, also discussed in [28]. Recall that the double-trine ensemble satisfies $|\psi_i\rangle|\psi_i\rangle = (U \otimes U)^i|\psi_0\rangle|\psi_0\rangle$, where U is a rotation of $\frac{2\pi}{3}$ around the z axis in the Bloch sphere. For sets with such symmetry the optimal measurement was shown by Ban et al to be given by the so-called square-root measurement [34] (also known as the "pretty-good measurement" [35]). In this case, the optimal measurement corresponds to a projective measurement, with operators $\{\Pi_j = |\Phi_j\rangle\langle\Phi_j|\}$, where

$$\begin{aligned}|\Phi_j\rangle = \frac{1}{\sqrt{3}}&\left(|0\rangle|0\rangle + e^{2\pi ji/3}\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)\right.\\
&\left. + e^{4\pi ji/3}|1\rangle|1\rangle\right).\end{aligned} \tag{17}$$

The probability of correctly identifying the state is

$$\mathrm{P}_{\mathrm{corr}}^{\mathrm{glob}} = \frac{1}{2} + \frac{\sqrt{2}}{3} \simeq 0.971.$$

Note that the probability of identifying the state correctly achieved by the optimal sequential measurement is greater than 96% of that achieved by the optimal global measurement. In systems where joint measurement is technologically challenging it is thus perhaps difficult to argue that the additional experimental effort is merited by the improvement in performance in this case.

We comment finally on the optimal sequential measurement as an approximation to the optimal global measurement. For the optimal sequential measurement, given above, we obtain

$$\begin{aligned}\pi_0 &= \frac{2}{3}\left(|\psi_1^\perp\rangle\langle\psi_1^\perp| \otimes |\phi_{0|1}\rangle\langle\phi_{0|1}|\right.\\
&\left. + |\psi_2^\perp\rangle\langle\psi_2^\perp| \otimes |\phi_{0|2}\rangle\langle\phi_{0|2}|\right)\\
\pi_1 &= (U \otimes U)\pi_0(U \otimes U)^\dagger\\
\pi_2 &= (U \otimes U)^2\pi_0((U \otimes U)^\dagger)^2.\end{aligned} \tag{18}$$

Considering π_0, after a little algebra we find

$$\begin{aligned}|\psi_1^\perp\rangle \otimes |\phi_{0|1}\rangle = \frac{1}{2}e^{-\pi i/12}&\left[\sqrt{1 + 2\cos^2\frac{\pi}{12}}|\alpha_0\rangle\right.\\
&\left. + i\sqrt{1 + 2\sin^2\frac{\pi}{12}}|\beta_0\rangle\right],\\
|\psi_2^\perp\rangle \otimes |\phi_{0|2}\rangle = \frac{1}{2}e^{\pi i/12}&\left[\sqrt{1 + 2\cos^2\frac{\pi}{12}}|\alpha_0\rangle\right.\\
&\left. - i\sqrt{1 + 2\sin^2\frac{\pi}{12}}|\beta_0\rangle\right],\end{aligned}$$

where

$$\begin{aligned}|\alpha_0\rangle = \left(1 + 2\cos^2\frac{\pi}{12}\right)^{-1/2}&\left(\cos\frac{\pi}{12}|00\rangle\right.\\
&\left. + \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) + \cos\frac{\pi}{12}|11\rangle\right),\\
|\beta_0\rangle = \left(1 + 2\sin^2\frac{\pi}{12}\right)^{-1/2}&\left(\sin\frac{\pi}{12}|00\rangle\right.\\
&\left. + \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) - \sin\frac{\pi}{12}|11\rangle\right).\end{aligned}$$

Thus we can write

$$\pi_0 = \frac{1}{3}\left(1 + 2\cos^2\frac{\pi}{12}\right)|\alpha_0\rangle\langle\alpha_0|$$
$$+\frac{1}{3}\left(1 + 2\sin^2\frac{\pi}{12}\right)|\beta_0\rangle\langle\beta_0|,$$
$$= \frac{1}{3}\left(2 + \frac{\sqrt{3}}{2}\right)|\alpha_0\rangle\langle\alpha_0|$$
$$+\frac{1}{3}\left(2 - \frac{\sqrt{3}}{2}\right)|\beta_0\rangle\langle\beta_0|.$$

Note that $\langle\alpha_0|\beta_0\rangle = 0$, and hence this is the eigende-composition of the operator. We further note that $|\beta_0\rangle$ is orthogonal to the signal state $|\psi_0\rangle|\psi_0\rangle$ and thus does not contribute to the probability of identifying the state. The remaining eigenvector $|\alpha_0\rangle$ is an approximation to $|\Phi_0\rangle$, the state onto which the optimal global measurement projects, which is, in fact, an amazingly good one: it turns out $|\langle\alpha_0|\Phi_0\rangle|^2 = 0.9997$. Due to the weighting factor, the overlap between $|\Phi_0\rangle$ and π_0 is given by $\langle\Phi_0|\pi_0|\Phi_0\rangle = \frac{1}{3}\left(2 + \frac{\sqrt{3}}{2}\right)|\langle\alpha_0|\Phi_0\rangle|^2 = 0.9551$.

The state $|\Phi_0\rangle$ is thus very close to a superposition of $|\psi_1^\perp\rangle \otimes |\phi_{0|1}\rangle$ and $|\psi_2^\perp\rangle \otimes |\phi_{0|2}\rangle$, with appropriate normalisation:

$$|\Phi_0\rangle \simeq |\alpha_0\rangle = \left(1 + 2\cos^2\frac{\pi}{12}\right)^{-1/2}\left(e^{\pi i/12}|\psi_1^\perp\rangle \otimes |\phi_{0|1}\rangle\right.$$
$$\left. + e^{-\pi i/12}|\psi_2^\perp\rangle \otimes |\phi_{0|2}\rangle\right).$$

The optimal sequential measurement, on the other hand, is formed from a *mixture* of projectors onto these same states. It gives additional information — one state is ruled out with certainty — at the expense of a slightly lower probability of success.

D. A non-optimal sequential measurement

The example of the trine states is further illuminating, as there exists another measurement strategy which satisfies both necessary conditions (10) and (15) but which is not an optimal strategy, thus demonstrating that these two conditions, when taken together, are not sufficient to define the optimal measurement. This strategy is to perform the optimal minimum error measurement at each step, with Bayesian update of the probabilities in between measurements. Note that such a strategy is known to be optimal (and in fact performs as well as the best joint measurement) for a different set of states — the case of just two pure states [36, 37]. For the trine states, the measurement is as follows: $\{M_j\}$ is the optimal one-copy minimum error measurement, which consists of weighted projectors onto the trine states themselves [11, 34], $M_j = \frac{2}{3}|\psi_j\rangle\langle\psi_j|$. Note that for the trine states $|\langle\psi_i|\psi_j\rangle|^2 = \frac{1}{4}(1 + 3\delta_{ij})$, and thus the updated

priors upon obtaining outcome j are, using Eq. (9),

$$p_{i|j} = \frac{\frac{2}{3}|\langle\psi_i|\psi_j\rangle|^2}{\frac{2}{3}\sum_k|\langle\psi_k|\psi_j\rangle|^2} = \frac{1}{6} + \frac{1}{2}\delta_{ij}.$$

For each j, the states with these probabilities have so-called mirror symmetry — the set is invariant under reflection about $|\psi_j\rangle$. For such a set, the minimum error problem was considered by Andersson *et al.* [38]. Using their results we find for $j = 0$ the optimal measurement is of the form:

$$N_{0|0} = (1 - a^2)|\psi_0\rangle\langle\psi_0|,$$
$$N_{1|0} = \frac{1}{2}\left(a|\psi_0\rangle - i|\psi_0^\perp\rangle\right)\left(a\langle\psi_0| + i\langle\psi_0^\perp|\right),$$
$$N_{2|0} = \frac{1}{2}\left(a|\psi_0\rangle + i|\psi_0^\perp\rangle\right)\left(a\langle\psi_0| - i\langle\psi_0^\perp|\right),$$

where a depends on the geometry of the set and the prior probabilities [38]. For our case we find $a = \frac{1}{5\sqrt{3}}$. The optimal measurements for $j = 1, 2$ are obtained by symmetry $\{N_{i|j} = U^j N_{i|0}(U^j)^\dagger\}$. Note that condition (10) is satisfied by construction. Turning to condition (15), we find

$$\text{Tr}\left(\rho_0 N_{0|0}\right) = \frac{74}{75},$$
$$\text{Tr}\left(\rho_1 N_{1|0}\right) = \text{Tr}\left(\rho_2 N_{2|0}\right) = \frac{32}{75},$$

with analogous results for $j = 1, 2$. Concisely, $\text{Tr}\left(\rho_i N_{i|j}\right) = \frac{32}{75} + \frac{42}{75}\delta_{ij}$. Finally, we can calculate Γ^A:

$$\Gamma^A = \sum_{i,j} p_i \text{Tr}\left(\rho_i N_{i|j}\right)\rho_i M_j$$
$$= \sum_{i,j}\frac{1}{3}\left(\frac{32}{75} + \frac{42}{75}\delta_{ij}\right)(|\psi_i\rangle\langle\psi_i|)\left(\frac{2}{3}|\psi_j\rangle\langle\psi_j|\right)$$
$$= \frac{30}{75}\mathbb{1} = \frac{2}{5}\mathbb{1}.$$

For $c_j\sigma_j$ we obtain

$$\sum_i p_i \text{Tr}\left(\rho_i N_{i|j}\right)\rho_i^A = \frac{1}{3}\sum_i\left(\frac{32}{75} + \frac{42}{75}\delta_{ij}\right)\rho_i$$
$$= \frac{16}{75}\mathbb{1} + \frac{14}{75}\rho_j$$
$$= \frac{2}{5}|\psi_j\rangle\langle\psi_j| + \frac{16}{75}|\psi_j^\perp\rangle\langle\psi_j^\perp|$$

from which it is clear that condition (15) is satisfied for each j.

An analogous situation arises in state discrimination maximising the mutual information between sender and receiver. A necessary but not sufficient condition is known, and for the example of the trine states, is satisfied by both the trine measurement, which is not optimal [34], and the antitrine measurement, which is optimal [14]. We finally note that the probability of correctly identifying the state using this scheme, $\text{Tr}(\Gamma^A) = \frac{4}{5}$, is considerably worse than that given by the optimal sequential measurement from above.

V. GENERAL BI-PARTITE CASE

A. Necessary and sufficient conditions

Above, for simplicity, we confined our discussion of optimal sequential measurement strategies to the case of two-copy state discrimination. The conditions obtained, however, are easily extended to the general bi-partite case. Suppose, therefore, we are provided with a bi-partite state drawn from a known set $\{\rho_i^{AB}\}$, with known *a priori* probabilities $\{p_i\}$. If our measurement strategy is restricted to sequential measurements on each subsystem, with feed-forward, what is the best measurement to make? The allowed measurements on the joint AB system are again described by POVMs of the form $\{\pi_i = \sum_j M_j^A \otimes N_{i|j}^B\}$, and the probability of correctly identifying the state is expressed:

$$\mathrm{P_{corr}} = \sum_{ij} p_i \mathrm{Tr}_{AB} \left(\rho_i^{AB} M_j^A \otimes N_{i|j}^B \right).$$

Following the same reasoning as in Sec. III, the necessary conditions Eqs. (10) and (15), become

$$\sum_i p_i \mathrm{Tr}_A \left(\rho_i^{AB} M_j \right) N_{i|j} - p_k \mathrm{Tr}_A \left(\rho_k^{AB} M_j \right) \geq 0,$$

$$\sum_{i,j} p_i \mathrm{Tr}_B \left(\rho_i^{AB} N_{i|j} \right) M_j - \sum_i p_i \mathrm{Tr}_B \left(\rho_i^{AB} N_{i|k} \right) \geq 0,$$

with the following interpretation. Given a measurement M_j^A on system A, $\{N_{i|j}^B\}$ must be optimal for discriminating the updated states $\sigma_{i|j}^B$, occurring with probabilities $p_{i|j}$:

$$\sigma_{i|j}^B = \frac{\mathrm{Tr}_A \left(\rho_i^{AB} M_j^A \right)}{\mathrm{Tr}_{AB} \left(\rho_i^{AB} M_j^A \right)},$$

$$p_{i|j} = \frac{p_i \mathrm{Tr}_{AB} \left(\rho_i^{AB} M_j^A \right)}{\sum_k p_k \mathrm{Tr}_{AB} \left(\rho_k^{AB} M_j^A \right)}.$$

Similarly, given measurements $\{\{N_{i|j}^B\}\}$ on system B, $\{M_j^A\}$ must be optimal for discriminating the states σ_j^A, occurring with probabilities q_j:

$$\sigma_j^B = \frac{\sum_i p_i \mathrm{Tr}_B \left(\rho_i^{AB} N_{i|j}^B \right)}{\sum_k p_k \mathrm{Tr}_{AB} \left(\rho_k^{AB} N_{k|j}^B \right)},$$

$$q_j = \frac{\sum_i p_i \mathrm{Tr}_{AB} \left(\rho_i^{AB} N_{i|j}^B \right)}{\sum_l \sum_k p_k \mathrm{Tr}_{AB} \left(\rho_k^{AB} N_{k|l}^B \right)}.$$

Finally, following the same argument as in Sec. III, the necessary and sufficient condition for optimality of $\{\pi_j = \sum_j M_j \otimes N_{i|j}\}$ for discriminating the general bipartite states $\{\rho_i^{AB}\}$ becomes

$$\sum_{i,j} p_i \mathrm{Tr}_B \left(\rho_i^{AB} N_{i|j} \right) M_j - \sum_k p_k \mathrm{Tr}_B \left(\rho_k^{AB} \widetilde{N}_k \right) \geq 0, \tag{19}$$

where $\{\widetilde{N}_k\}$ is any physically allowed measurement on system B.

B. Example: Three Bell states

As an example of the general case, we consider the simple case of discriminating between three Bell states $\rho_i^{AB} = |\Psi_i\rangle\langle\Psi_i|$:

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle|0\rangle + |1\rangle|1\rangle \right),$$

$$|\Psi_1\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle|1\rangle + |1\rangle|0\rangle \right),$$

$$|\Psi_2\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle|0\rangle - |1\rangle|1\rangle \right),$$

occurring with equal probabilities $p_i = \frac{1}{3}$. Although perfect discrimination between any two Bell states is possible by only local measurements and feed-forward (for example, to distinguish between $|\Psi_0\rangle$ and $|\Psi_1\rangle$ one need only measure both systems in the $\{|0\rangle, |1\rangle\}$ basis and look at the correlations between outcomes) it is known that for more than two states this is no longer possible [39, 40]. To distinguish between all three states, one strategy is to simply perform the measurement that perfectly distinguishes any two states and never identify the third. We show that this strategy is optimal in terms of minimising the probability of error.

Consider, therefore, the measurement

$$M_0 = |0\rangle\langle 0|, \ M_1 = |1\rangle\langle 1|,$$
$$N_{0|0} = |0\rangle\langle 0|, \ N_{1|0} = |1\rangle\langle 1|, \ N_{2|0} = 0,$$
$$N_{0|1} = |1\rangle\langle 1|, \ N_{1|1} = |0\rangle\langle 0|, \ N_{2|1} = 0,$$

that is, both Alice and Bob measure in the $\{|0\rangle, |1\rangle\}$ basis. Bob takes outcome 0 to indicate state $|\Psi_0\rangle$, and outcome 1 to indicate state $|\Psi_1\rangle$. State $|\Psi_2\rangle$ is never identified.

It is useful to rewrite Eq. (19) as follows:

$$\Gamma^A - \widetilde{c}\widetilde{\sigma} \geq 0,$$

where

$$\Gamma^A = \sum_{i,j} p_i \mathrm{Tr}_B \left(\rho_i^{AB} N_{i|j} \right) M_j,$$

$$\widetilde{c} = \sum_k p_k \mathrm{Tr}_{AB} \left(\rho_k^{AB} \widetilde{N}_k \right),$$

$$\widetilde{\sigma} = \frac{1}{\widetilde{c}} \sum_k p_k \mathrm{Tr}_B \left(\rho_k^{AB} \widetilde{N}_k \right).$$

Note that $\widetilde{\sigma}$ is a density operator. Furthermore, it is straight-forward to show that

$$\Gamma^A = \sum_{i,j} p_i \mathrm{Tr}_B \left(\rho_i^{AB} N_{i|j} \right) M_j = \frac{1}{3} \mathbb{1}^A$$

while

$$\widetilde{c} = \frac{1}{3} \sum_k \mathrm{Tr}_B \left(\mathrm{Tr}_A(\rho_k^{AB}) \widetilde{N}_k \right)$$

$$= \frac{1}{3} \sum_k \mathrm{Tr}_B \left[\left(\frac{1}{2} \mathbb{1}^B \right) \widetilde{N}_k \right]$$

$$= \frac{1}{3},$$

where the second line follows as the reduced density operator for system B in all cases is proportional to the identity $\mathbb{1}^B$, and the last line follows from the POVM condition $\sum \widetilde{N}_k = \mathbb{1}^B$. Thus the condition (19) becomes

$$\mathbb{1} - \widetilde{\sigma} \geq 0,$$

which is true for any arbitrary density operator $\widetilde{\sigma}$. Thus $\{M_j \otimes N_{i|j}\}$ is an optimal measurement among sequential strategies for discrimination of the three Bell states. A similar approach could be taken to the problem of discriminating all four Bell states, but this complicates the analysis without changing the basic conclusions.

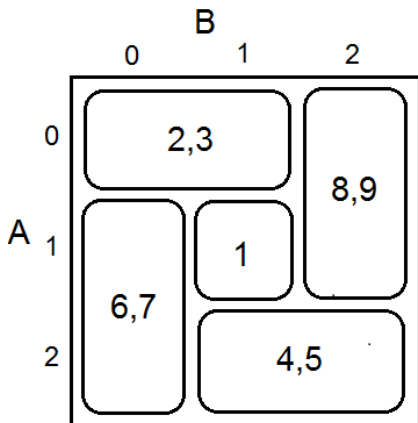### C.   Example: Domino states

FIG. 3.   Graphical representation of the domino states as domino tiles.

The domino states are an orthonormal basis of a two-qutrit system, composed entirely of product states, which nevertheless are not perfectly distinguishable by local measurement and classical communication (LOCC) [21]. These were the first set of states for which it was shown explicitly that information encoded in product states is not all available through local measurements. The set of states is given by

$$|\psi_1\rangle = |1\rangle|1\rangle,$$
$$|\psi_2\rangle = |0\rangle|0 + 1\rangle,$$
$$|\psi_3\rangle = |0\rangle|0 - 1\rangle,$$

$$|\psi_4\rangle = |2\rangle|1 + 2\rangle,$$
$$|\psi_5\rangle = |2\rangle|1 - 2\rangle,$$
$$|\psi_6\rangle = |1 + 2\rangle|0\rangle,$$
$$|\psi_7\rangle = |1 - 2\rangle|0\rangle,$$
$$|\psi_8\rangle = |0 + 1\rangle|2\rangle,$$
$$|\psi_9\rangle = |0 - 1\rangle|2\rangle,$$

where $|i \pm j\rangle = \frac{1}{\sqrt{2}} (|i\rangle \pm |j\rangle)$. A useful graphical representation of the states is given in Fig. 3. As the states are orthonormal, they are perfectly distinguishable by a joint measurement. While it is known that any local measurement strategy, possibly consisting of many rounds of measurement and classical communication, cannot achieve perfect discrimination, in practice known lower bounds are really very small — the information deficit of any local protocol is at least 5.31×10^{-6} (compared to an achievable value of $\log_2 9 = 3.17$ bits) [21], while the probability of error is at least 1.9×10^{-8} [41]. Although this is of theoretical interest, if these bounds are achievable it is fair to say that for all practical purposes the states are distinguishable by a local strategy, albeit one with many rounds of communication.

1.   A candidate sequential measurement

We recently gave the achievable probability of correct discrimination using sequential measurements for these states [42]. Here we give an alternative proof, which uses the conditions introduced earlier. Motivated by the double-copy trine example, in which the best strategy rules out one state in the first step and discriminates optimally between the remaining two in the second step, we consider strategies that partition the allowed states into subsets in the first step and discriminate between the states within a subset in the second step. Note that among the domino states there are a total of eight subsets that are perfectly distinguishable on the second system alone:

$$\mathcal{S}_1 = \{|\psi_1\rangle, |\psi_6\rangle, |\psi_8\rangle\},$$
$$\mathcal{S}_2 = \{|\psi_1\rangle, |\psi_6\rangle, |\psi_9\rangle\},$$
$$\mathcal{S}_3 = \{|\psi_1\rangle, |\psi_7\rangle, |\psi_8\rangle\},$$
$$\mathcal{S}_4 = \{|\psi_1\rangle, |\psi_7\rangle, |\psi_9\rangle\},$$
$$\mathcal{S}_5 = \{|\psi_2\rangle, |\psi_3\rangle, |\psi_8\rangle\},$$
$$\mathcal{S}_6 = \{|\psi_2\rangle, |\psi_3\rangle, |\psi_9\rangle\},$$
$$\mathcal{S}_7 = \{|\psi_4\rangle, |\psi_5\rangle, |\psi_6\rangle\},$$
$$\mathcal{S}_8 = \{|\psi_4\rangle, |\psi_5\rangle, |\psi_7\rangle\}.$$

We therefore begin with the conjecture that the best measurement for discriminating the domino states optimally assigns the state to one of these subsets at the first step, and discriminates perfectly between the remaining three states in the second step. Thus the only error is introduced in the first step. Denoting now $|\psi_i\rangle = |\phi_i\rangle|\chi_i\rangle$, the

probability of identifying the correct state is

$$P_{corr} = \sum_{ij} p_i \langle \phi_i | M_j | \phi_i \rangle \langle \chi_i | N_{i|j} | \chi_i \rangle$$

$$= \frac{1}{9} \sum_j \sum_{i \in \mathcal{I}_j} \langle \phi_i | M_j | \phi_i \rangle$$

$$= \frac{8}{3} \sum_j \frac{1}{8} \mathrm{Tr} \left(\frac{1}{3} \sum_{i \in \mathcal{I}_j} | \phi_i \rangle \langle \phi_i | M_j \right),$$

where in the second line we have used the fact that $\langle \chi_i | N_{i|j} | \chi_i \rangle = 1$ if $|\chi_i\rangle$ is in the subset $\mathcal{S}_j$, and is zero otherwise, and we have defined the index set $\mathcal{I}_j$ such that $i \in \mathcal{I}_j$ if $|\psi_i\rangle \in \mathcal{S}_j$. Thus $\{M_j\}$ is the optimal measurement for discriminating the states

$$\rho_j = \frac{1}{3} \sum_{i \in \mathcal{I}_j} |\phi_i\rangle\langle\phi_i|, \tag{20}$$

formed by taking an equal mixture of states in the subsets $\{\mathcal{S}_j\}$ and occurring with equal probabilities $p_j = \frac{1}{8}$.

2. Optimal subset discrimination

The optimal measurement $\{M_j\}$ and probability of success may be derived analytically and are given in Appendix B. Here we discuss some of the symmetries of the states which allow us to simplify the problem, before proving that this measurement scheme results in an optimal sequential strategy.

The states $\left\{ \rho_j = \mathrm{Tr}_B \left(\sum_{i \in \mathcal{I}_j} |\psi_i\rangle\langle\psi_i| \right) \right\}$ have a lot of symmetry, in particular, the set is invariant under the unitaries

$$U_0 = -|0\rangle\langle 0| + |1\rangle\langle 1| + |2\rangle\langle 2|,$$
$$U_1 = |0\rangle\langle 2| + |1\rangle\langle 1| + |2\rangle\langle 0|, \tag{21}$$

which simply permute the states within the set. To make this point clear, it is useful to list the states explicitly:

$$\rho_1 = \frac{1}{3} \mathrm{Tr}_B \left(|\psi_1\rangle\langle\psi_1| + |\psi_6\rangle\langle\psi_6| + |\psi_8\rangle\langle\psi_8| \right)$$
$$= \frac{1}{3} \left(|1\rangle\langle 1| + |1+2\rangle\langle 1+2| + |0+1\rangle\langle 0+1| \right),$$

$$\rho_2 = \frac{1}{3} \left(|1\rangle\langle 1| + |1+2\rangle\langle 1+2| + |0-1\rangle\langle 0-1| \right),$$

$$\rho_3 = \frac{1}{3} \left(|1\rangle\langle 1| + |1-2\rangle\langle 1-2| + |0+1\rangle\langle 0+1| \right),$$

$$\rho_4 = \frac{1}{3} \left(|1\rangle\langle 1| + |1-2\rangle\langle 1-2| + |0-1\rangle\langle 0-1| \right),$$

$$\rho_5 = \frac{2}{3} |0\rangle\langle 0| + \frac{1}{3} |0+1\rangle\langle 0+1|,$$

$$\rho_6 = \frac{2}{3} |0\rangle\langle 0| + \frac{1}{3} |0-1\rangle\langle 0-1|,$$

$$\rho_7 = \frac{2}{3} |2\rangle\langle 2| + \frac{1}{3} |1+2\rangle\langle 1+2|,$$

$$\rho_8 = \frac{2}{3} |2\rangle\langle 2| + \frac{1}{3} |1-2\rangle\langle 1-2|. \tag{22}$$

Clearly, applying either U_0 or U_1 to any of the states in the set simply results in another state from the same set. $U_k \rho_j U_k^\dagger = \rho_{\sigma_k(j)}$ for some permutation $\sigma_k(j)$. We may therefore expect the optimal measurement $\{M_j\}$ to have the same symmetries — indeed, for any measurement $\{M_j\}$, we can construct another measurement $\{M'_{\sigma_k(j)} = U_k M_j U_k^\dagger\}$ which has the same probability of success. As the figure of merit is linear, a probabilistic mixture of such strategies $\{M''_j = \frac{1}{2}\left(M'_j + M_j\right)\}$ achieves the same probability of success, and further, has the symmetry property $U_k M''_j U_k^\dagger = M''_{\sigma_k(j)}$. Thus we can restrict attention to measurements which have the same symmetry properties as the states.

Turning now to the operator

$$\Gamma^A = \sum_{ij} p_i \langle \chi_i | N_{i|j} | \chi_i \rangle \left(|\phi_i\rangle\langle\phi_i| M_j \right)$$

$$= \frac{1}{9} \sum_j \sum_{i \in \mathcal{I}_j} |\phi_i\rangle\langle\phi_i| M_j$$

$$= \frac{8}{3} \sum_j \frac{1}{8} \rho_j M_j, \tag{23}$$

we further note that

$$U_k \Gamma^A U_k^\dagger = \frac{8}{3} \sum_j \frac{1}{8} U_k \rho_j M_j U_k^\dagger$$

$$= \frac{8}{3} \sum_j \frac{1}{8} \left(U_k \rho_j U_k^\dagger \right) \left(U_k M_j U_k^\dagger \right)$$

$$= \frac{8}{3} \sum_j \frac{1}{8} \rho_{\sigma_k(j)} M_{\sigma_k(j)} = \Gamma^A.$$

Thus Γ^A is invariant under both U_0, and U_1, and it follows that it must have the following form:

$$\Gamma^A = \frac{1}{3} \sum_j \rho_j M_j$$
$$= p \left(|0\rangle\langle 0| + |2\rangle\langle 2| \right) + q |1\rangle\langle 1|. \tag{24}$$

In Appendix B we show that $p = \frac{1}{9}\frac{1}{12}(17 + \sqrt{7}\sqrt{31}) \simeq 0.294$, $q = \frac{1}{9}\frac{1}{16}(21 + \sqrt{7}\sqrt{31}) \simeq 0.248$, giving a probability of correctly identifying the state of

$$P_{corr} = 2p + q \simeq 0.836.$$

3. An optimal sequential strategy

We can now use condition (19) to prove that this is the best possible sequential measurement. Again, we outline here some of the symmetry arguments that allow us to simplify the problem, and give the remaining details in Appendix B. Equation (19) then becomes

$$\Gamma^A - \frac{1}{9} \sum_k \langle \chi_k | \widetilde{N}_k | \chi_k \rangle |\phi_k\rangle\langle\phi_k| \geq 0.$$

It is useful to define the non-normalized state

$$\widetilde{\sigma} = \sum_k \langle \chi_k | \widetilde{N}_k | \chi_k \rangle |\phi_k\rangle\langle\phi_k|$$

$$= \langle 1|\widetilde{N}_1|1\rangle |1\rangle\langle 1|$$
$$+ \left(\langle 0+1|\widetilde{N}_2|0+1\rangle + \langle 0-1|\widetilde{N}_3|0-1\rangle \right) |0\rangle\langle 0|$$
$$+ \left(\langle 1+2|\widetilde{N}_4|1+2\rangle + \langle 1-2|\widetilde{N}_5|1-2\rangle \right) |2\rangle\langle 2|$$
$$+ \langle 0|\widetilde{N}_6|0\rangle |1+2\rangle\langle 1+2| + \langle 0|\widetilde{N}_7|0\rangle |1-2\rangle\langle 1-2|$$
$$+ \langle 2|\widetilde{N}_8|2\rangle |0+1\rangle\langle 0+1| + \langle 2|\widetilde{N}_9|2\rangle |0-1\rangle\langle 0-1|,$$

$$(25)$$

and we wish to show that $\Gamma^A - \frac{1}{9}\widetilde{\sigma} \geq 0$, for all $\widetilde{\sigma}$ resulting from an allowed measurement $\{\widetilde{N}_j\}$.

We can again use symmetry properties of the states to considerably restrict the set of measurements we need to consider. Consider this time the unitaries

$$V_0 = -|0\rangle\langle 0| + |1\rangle\langle 1| + |2\rangle\langle 2|,$$
$$V_1 = |0\rangle\langle 0| + |1\rangle\langle 1| - |2\rangle\langle 2|$$

Note that the operator $\widetilde{\sigma}$ is unchanged if we make the substitution

$$\widetilde{N}'_j = V_0 \widetilde{N}_j V_0^\dagger, \quad j \neq 2, 3$$
$$\widetilde{N}'_2 = V_0 \widetilde{N}_3 V_0^\dagger,$$
$$\widetilde{N}'_3 = V_0 \widetilde{N}_2 V_0^\dagger,$$

or indeed, under a strategy which is formed from a mixture of such strategies. We can make similar arguments for V_1, with the result that we only need consider strategies such that the following combinations of operators are invariant under both V_0 and V_1, and therefore must be diagonal in the $\{|0\rangle, |1\rangle, |2\rangle\}$ basis: $\{\widetilde{N}_1, \widetilde{N}_2 + \widetilde{N}_3, \widetilde{N}_4 + \widetilde{N}_5, \widetilde{N}_6, \widetilde{N}_7, \widetilde{N}_8, \widetilde{N}_9\}$. Further, among such strategies it is clear that in order to maximise the coefficients $\langle \chi_k | \widetilde{N}_k | \chi_k \rangle$ we only need consider those such that

$$\widetilde{N}_1 = c_1 |1\rangle\langle 1|,$$
$$\widetilde{N}_2 = c_2 \left(\cos\alpha |0\rangle + \sin\alpha |1\rangle \right) \left(\cos\alpha \langle 0| + \sin\alpha \langle 1| \right),$$
$$\widetilde{N}_3 = c_2 \left(\cos\alpha |0\rangle - \sin\alpha |1\rangle \right) \left(\cos\alpha \langle 0| - \sin\alpha \langle 1| \right),$$
$$\widetilde{N}_4 = c_4 \left(\cos\beta |2\rangle + \sin\beta |1\rangle \right) \left(\cos\beta \langle 2| + \sin\beta \langle 1| \right),$$
$$\widetilde{N}_5 = c_4 \left(\cos\beta |2\rangle - \sin\beta |1\rangle \right) \left(\cos\beta \langle 2| - \sin\beta \langle 1| \right),$$
$$\widetilde{N}_6 = c_6 |0\rangle\langle 0|,$$
$$\widetilde{N}_7 = c_7 |0\rangle\langle 0|,$$
$$\widetilde{N}_8 = c_8 |2\rangle\langle 2|,$$
$$\widetilde{N}_9 = c_9 |2\rangle\langle 2|, \quad (26)$$

where $0 \leq \alpha, \beta \leq \pi/2$ and for constants $c_1, c_2, c_4, c_6, c_7, c_8, c_9 \geq 0$ satisfying

$$c_1 + 2c_2 \sin^2\alpha + 2c_4 \sin^2\beta = 1,$$
$$2c_2 \cos^2\alpha + c_6 + c_7 = 1,$$
$$2c_4 \cos^2\beta + c_8 + c_9 = 1.$$

First note that for $\alpha = \beta = \frac{\pi}{4}$ all such measurements are convex combinations of the eight fiducial measurements

$$\{\widetilde{N}_1 = |1\rangle\langle 1|, \widetilde{N}_i = |0\rangle\langle 0|, \widetilde{N}_j = |2\rangle\langle 2|\},$$
$$\{\widetilde{N}_2 = |0+1\rangle\langle 0+1|, \widetilde{N}_3 = |0-1\rangle\langle 0-1|, \widetilde{N}_j = |2\rangle\langle 2|\},$$
$$\{\widetilde{N}_4 = |1+2\rangle\langle 1+2|, \widetilde{N}_5 = |1-2\rangle\langle 1-2|, \widetilde{N}_i = |0\rangle\langle 0|\},$$

where $i = 6, 7$, $j = 8, 9$. Each such measurement leads to $\widetilde{\sigma} = 3\rho_j$ for some j. Since for ρ_j defined in Eq. (20), $\Gamma^A - \frac{1}{3}\rho_j \geq 0$ by construction, any convex combination of such measurements satisfies $\Gamma^A - \frac{1}{9}\widetilde{\sigma} \geq 0$, as required.

The remaining cases, in which $\alpha, \beta \neq \frac{\pi}{4}$ must be considered separately. It is straight-forward to show that $\Gamma^A - \frac{1}{9}\widetilde{\sigma} \geq 0$ for these cases also. Details of all possible measurements and the corresponding calculations are given in Appendix B.

Thus we find that the optimal sequential strategy for discriminating between the domino states with equal priors identifies the state correctly with a probability of 83.6%. As the optimal joint measurement discriminates the states perfectly, there is a significant gap in performance, which could in principle be demonstrated experimentally.

VI. A RELATED DISCRIMINATION PROBLEM

We conclude by discussing an interpretation of the states σ_j, defined in Eq. (12), and thus a second discrimination problem for which $\{M_j \otimes N_{i|j}\}$ provides an optimal measurement. For simplicity we refer to the two-copy case throughout, but the discussion applies equally to the general bipartite case. Recall that, in a measurement on system B described by POVM $\{N_{i|j}\}$, the probability of identifying a given state ρ_i correctly from the set $\{\rho_k\}$ with prior probabilities $\{p_k\}$ is given by the joint probability

$$\Pr(\rho_i, i) = \Pr(\rho_i)\Pr(i|\rho_i) = p_i \text{Tr}(\rho_i N_{i|j}),$$

while the overall probability of correctly identifying the prepared state is obtained by summing over each possible i: $\Pr(\text{corr}|\{N_{i|j}\}) = \sum_i p_i \text{Tr}(\rho_i N_{i|j})$, where we have chosen the notation to make explicit the dependence on the choice of measurement $\{N_{i|j}\}$. We can thus interpret the ratio of these as the conditional probability that state ρ_i was prepared given that measurement $\{N_{i|j}\}$ was performed and the state was identified correctly as a result. Explicitly,

$$\Pr(\rho_i|\text{corr}, \{N_{i|j}\}) = \frac{p_i \text{Tr}\left(\rho_i N_{i|j}\right)}{\sum_k p_k \text{Tr}\left(\rho_k N_{k|j}\right)}.$$

We thus arrive at an interpretation for the state σ_j^A – it is the state we should assign to system A, given knowledge that a measurement $\{N_{i|j}\}$ performed on system B identified the state correctly (but importantly, *without*

knowledge of the particular outcome of measurement on B):

$$\sigma_j^A = \sum_i \mathrm{P}(\rho_i|\mathrm{corr}, \{N_{i|j}\})\rho_i^A.$$

Similarly, consider the probabilities q_j: if a measurement is chosen from the set $\{\{N_{i|j}\}\}$ with equal probabilities $1/n$, then q_j represents the conditional probability that the measurement used was $\{N_{i|j}\}$, given the knowledge that the state was correctly identified:

$$q_j = \frac{c_j}{\sum_k c_k} = \frac{\frac{1}{n}\mathrm{Pr}(\mathrm{corr}|\{N_{i|j}\})}{\frac{1}{n}\sum_k \mathrm{Pr}(\mathrm{corr}|\{N_{i|k}\})}$$
$$= \mathrm{Pr}(\{N_{i|j}\}|\mathrm{corr}).$$

$\{M_j\}$ is thus the optimal measurement for discriminating the residual states on system A, given that a measurement chosen from the set $\{\{N_{i|j}\}\}$ was performed on system B, and the state of system B was correctly identified as a result. $\{M_j\}$ is the measurement that allows us to optimally guess which measurement was performed on B.

Suppose further that another measurement $\{\widetilde{N}_i\}$ is added to the set of possible measurements $\{N_{i|j}\}$, on B. If we now define

$$\widetilde{\sigma}^A = \sum_i \mathrm{Pr}(\rho_i|\mathrm{corr}, \{\widetilde{N}_i\})\rho_i^A,$$

$$q_j' = \frac{\frac{1}{n+1}\mathrm{Pr}(\mathrm{corr}|\{N_{i|j}\})}{\frac{1}{n+1}\left(\sum_k \mathrm{Pr}(\mathrm{corr}|\{N_{i|k}\}) + \mathrm{Pr}(\mathrm{corr}|\{\widetilde{N}_i\})\right)}$$
$$= \mathrm{Pr}(\{N_{i|j}\}|\mathrm{corr}),$$

$$\widetilde{q} = \frac{\frac{1}{n+1}\mathrm{Pr}(\mathrm{corr}|\{\widetilde{N}_i\})}{\frac{1}{n+1}\left(\sum_k \mathrm{Pr}(\mathrm{corr}|\{N_{i|k}\}) + \mathrm{Pr}(\mathrm{corr}|\{\widetilde{N}_i\})\right)}$$
$$= \mathrm{Pr}(\{\widetilde{N}_i\}|\mathrm{corr}),$$

we can rewrite condition (16) as follows:

$$\sum_{i,j} q_j'\sigma_j^A M_j^A - \widetilde{q}\widetilde{\sigma}^A \geq 0,$$

where σ_j^A are defined as before. This tells us that the measurement which discriminates optimally between the states $\{\sigma_j^A\}$, but which never identifies $\{\widetilde{\sigma}\}$ remains optimal for the new set. That is, the optimal measurement $\{M_j\}$ is unchanged by the addition of a new measurement to the set used to measure system B.

We thus arrive at a new discrimination problem, seemingly unrelated to our original problem of interest, but for which the measurement $\{M_j^A \otimes N_{i|j}\}$ also provides an optimal strategy and for which the conditions (15) and (16) have a natural interpretation. Consider the following game involving three parties, Alice, Bob, and Claire. Claire prepares two copies of a state ρ_i, drawn from a given set $\{\rho_j\}$ with priors $\{p_j\}$, and sends one copy to Alice and one to Bob. Claire also sends an index j to

Bob (this is classical information and may be sent over a classical channel), chosen with equal probabilities from the set $0 \leq j \leq n-1$, where n is an integer chosen by Alice and Bob, and is the payout they will receive if they win the game. Alice and Bob each perform a measurement on their system, following which Bob must make a guess as to the state ρ_i, while Alice must guess the index j. Alice and Bob can preagree on a strategy but cannot communicate with one another once the game begins. They win only if both guesses are correct. Their strategy is as follows: Alice always makes the measurement $\{M_j\}$, while Bob's chosen measurement depends on the index j received from Claire — given j, Bob makes measurement $\{N_{i|j}\}$. The probability that Alice and Bob win the game is simply

$$\mathrm{Pr}(\mathrm{win}) = \sum_{i,j} \mathrm{Pr}(i)\mathrm{Pr}(j)\mathrm{Tr}(\rho_i N_{i|j})\mathrm{Tr}(\rho_i M_j)$$
$$= \frac{1}{n}\sum_{i,j} p_i \mathrm{Tr}(\rho_i N_{i|j})\mathrm{Tr}(\rho_i M_j),$$

and the expected payout is thus

$$\langle \mathrm{Payout} \rangle = n\mathrm{Pr}(\mathrm{win}) = \sum_{i,j} p_i \mathrm{Tr}(\rho_i N_{i|j})\mathrm{Tr}(\rho_i M_j).$$

This is the same figure of merit as that given in Eq. (8). Thus the optimal measurement strategy maximising the expected payout for Alice and Bob is the same as the optimal sequential measurement discriminating the bipartite states $\{\rho_i \otimes \rho_i\}$.

VII. DISCUSSION

We have discussed the problem of extracting classical information from a set of bipartite states, when the measurement strategy is restricted to sequential measurements of each subsystem, with feed-forward of classical information in between measurements. As this is a physically well-motivated class, it is useful to understand how well it performs compared to the ability to perform arbitrary joint measurements, which in many physical systems is still technologically challenging. We have constructed an analogue of the Helstrom conditions for sequential measurement strategies. Like the Helstrom conditions, it is not obvious how to use this condition to construct an optimal measurement, but we show how for certain examples it is possible to use the condition to prove optimality of a candidate measurement procedure.

Our necessary and sufficient condition for optimality of a given sequential measurement still contains an arbitrary measurement on one subsystem. We have been unable to find a condition which is both necessary and sufficient and requires only the set of states and a candidate measurement. It would certainly be useful to find one, but in the absence of such, given a candidate optimal measurement our condition reduces the complexity

of checking optimality from optimising over both systems to optimising over just one. It would also be interesting in the future to extend this analysis to other figures of merit, such as those which interpolate between minimum-error and unambiguous discrimination [19], or which maximise the success rate of discrimination while allowing for inconclusive results [43, 44].

For the two-copy trine case, the probability of success of the optimal sequential measurement is 96 % of the value achieved by the optimal global measurement [28]. The optimal sequential measurement sometimes rules out one of the states with certainty, thus providing information not given by the optimal global strategy, at the expense of a slightly higher probability of failure. Nonetheless, the difference in performance is arguably too small to motivate experimental implementation of the joint measurement. In fact, although we do not give the details here, we have found that for two copies of any set of symmetric qubit states, the optimal sequential measurement performs almost as well as the optimal joint measurement.

We have given the optimal sequential measurement for discriminating the domino states. This set of orthogonal product states has played an important role in quantifying the difference between local measurements and separable measurements [21], but the more practical question of how well a sequential measurement performs seems not to have been considered in the literature until recently [42]. We gave here an alternative proof for the optimal sequential measurement that reveals a significant gap between the probability of success of discrimination of the optimal sequential measurement (83.6%) and the optimal global measurement, which discriminates the states perfectly.

We further introduced a complementary discrimination problem, in the form of a three-party game, which requires optimisation of the same figure of merit as our original problem of minimising the error over sequential measurement strategies. This game arises naturally in providing an interpretation for the conditions an optimal measurement must satisfy and provides a different perspective on the sequential measurement problem.

Appendix A: Proof of optimality for the double trine ensemble

To prove optimality of the sequential measurement scheme given in the text, we wish to show that the largest eigenvalue of

$$\widetilde{\sigma} = \frac{1}{3} \sum_k \mathrm{Tr}(\rho_k \widetilde{N}_k) \rho_k$$

is less than or equal to $\frac{1}{2} p_H = \frac{1}{4} \left(1 + \frac{\sqrt{3}}{2} \right)$ for any physically allowed measurement $\{\widetilde{N}_k\}$. We begin by writing the trine states ρ_j in the Bloch sphere representation:

$$\rho_j = \frac{1}{2} \left[I + \cos\left(\frac{2\pi j}{3} \right) \sigma_x + \sin\left(\frac{2\pi j}{3} \right) \sigma_y \right].$$

Writing $s_k = \frac{1}{3} \mathrm{Tr} \left(\rho_k \widetilde{N}_k \right)$ we thus obtain

$$\widetilde{\sigma} = \frac{1}{2} \left\{ (s_0 + s_1 + s_2)I \right.$$
$$\left. + \left[s_0 - \frac{1}{2}(s_1 + s_2) \right] \sigma_x + \frac{\sqrt{3}}{2}(s_1 - s_2)\sigma_y \right\}$$

with eigenvalues

$$\lambda_\pm = \frac{1}{2}(s_0 + s_1 + s_2)$$
$$\pm \frac{1}{2} \left(\sqrt{ \left[s_0 - \frac{1}{2}(s_1 + s_2) \right]^2 + \left[\frac{\sqrt{3}}{2}(s_1 - s_2) \right]^2 } \right)$$
$$= \frac{1}{2}(s_0 + s_1 + s_2)$$
$$\pm \frac{1}{2} | s_0 + e^{\frac{2\pi i}{3}} s_1 + e^{-\frac{2\pi i}{3}} s_2 |.$$

Thus it follows that there exists some θ such that the largest eigenvalue λ_+ may be written

$$\lambda_+ = \frac{1}{2}(s_0 + s_1 + s_2)$$
$$+ \frac{1}{2} e^{i\theta} \left(s_0 + e^{\frac{2\pi i}{3}} s_1 + e^{-\frac{2\pi i}{3}} s_2 \right),$$
$$= \frac{1}{2}(1 + \cos\theta)s_0 + \frac{1}{2} \left[1 + \cos\left(\theta + \frac{2\pi}{3} \right) \right] s_1$$
$$+ \frac{1}{2} \left[1 + \cos\left(\theta - \frac{2\pi}{3} \right) \right] s_2$$
$$= \frac{1}{2} \left[\sum_k q_k \mathrm{Tr} \left(\rho_k \widetilde{N}_k \right) \right],$$

where in the second equality we use the fact that λ_+ is real, and in the last line we have substituted for s_k, and defined $q_k = \frac{1}{3} \left[1 + \cos\left(\theta + \frac{2\pi k}{3} \right) \right]$. Each strategy $\{\widetilde{N}_k\}$ thus defines a θ such that the above equalities hold. For each such θ, we can find an upper bound for λ_+ by considering the optimisation problem of discriminating the states $\{\rho_k\}$ occurring with priors q_k:

$$\lambda_+ \leq \frac{1}{2} \mathrm{P_{corr}} \left(\{q_k \rho_k\} \right).$$

We thus wish to find the optimal strategy $\{\pi_k\}$ for discriminating the trine states with *a priori* probabilities $\frac{1}{3}\left[1 + \cos\left(\theta + \frac{2\pi k}{3}\right)\right]$, ultimately maximising the probability of correctness also over θ. Finally, if this maximum is achievable then we have succeeded in finding the optimal λ_+.

We first choose, without loss of generality, to consider $q_0 \geq q_1 \geq q_2$. This corresponds to $-\frac{\pi}{3} \leq \theta \leq 0$. We use the strategy in [28] of bounding the probability of correctly identifying the state by considering the problem of discriminating $q_0\rho_0$ from $q_1\rho_1 + q_2\rho_2$. As any strategy which discriminates all three states also discriminates these two states, the optimal probability of correctly identifying the state for this problem is greater than or equal to that for discriminating all three states:

$$P_{\mathrm{corr}}(\{q_k\rho_k\}) \leq P_{\mathrm{corr}}(q_0\rho_0, q_1\rho_1 + q_2\rho_2)$$
$$= \frac{1}{2}[1 + \mathrm{Tr}|q_0\rho_0 - (q_1\rho_1 + q_2\rho_2)|]$$

Substituting for q_k, ρ_k, after a little algebra we obtain:

$$P_{\mathrm{corr}}(\{q_k\rho_k\}) \leq \frac{1}{2}\left[1 + \frac{2}{3}\sqrt{\left(1 + \frac{1}{4}\cos\theta\right)^2 + \left(\frac{3}{4}\sin\theta\right)^2}\right].$$

For θ in the range $-\frac{\pi}{3} \leq \theta \leq 0$, this is a monotonically decreasing function of θ, thus the maximum occurs at the boundary of the allowed domain, at $\theta = -\frac{\pi}{3}$, corresponding to

$$P_{\mathrm{corr}}(\{q_k\rho_k\}) \leq \frac{1}{2}\left(1 + \frac{\sqrt{3}}{2}\right).$$

Thus we obtain $\lambda_+ \leq \frac{1}{4}\left(1 + \frac{\sqrt{3}}{2}\right)$, as desired. Furthermore, this bound is achievable by the strategy given by [28] and outlined in the main text.

Appendix B: Domino states

We begin with the optimal measurement $\{M_j\}$ to discriminate the states $\{\rho_j\}$ given by Eq. (22), occurring with equal probabilities $p_j = p = \frac{1}{8}$. Our strategy is to search for $\Gamma = \frac{1}{8}\sum_i \rho_i M_i$ satisfying $\Gamma - \frac{1}{8}\rho_j \geq 0$, and then use Eq. (7) to find $\{M_j\}$. In fact, as Γ^A defined in Eq. (23) is simply proportional to Γ, $\Gamma^A = \frac{8}{3}\Gamma$, and Eq. (7) still holds when multiplied by a constant, we work directly with Γ^A and search for Γ^A, $\{M_j\}$ satisfying

$$\Gamma^A - \frac{1}{3}\rho_j \geq 0,$$
$$\left(\Gamma^A - \frac{1}{3}\rho_j\right)M_j = 0$$

for all j.

Recall, as argued in the main text, that Γ^A is invariant under the unitaries U_0, U_1 from Eq. (21), and thus has the form $\Gamma^A = p\left(|0\rangle\langle 0| + |2\rangle\langle 2|\right) + q|1\rangle\langle 1|$. We further

note that it suffices to check the inequality for ρ_1 and ρ_5, as all other states can be obtained from these by application of some combination of U_0 and U_1, and by symmetry it then holds for all j. For ρ_1, this condition, expressed in matrix form in the $\{|0+2\rangle, |1\rangle, |0-2\rangle\}$ basis, becomes

$$\Gamma^A - \frac{1}{3}\rho_1 = \begin{pmatrix} p - \frac{1}{18} & -\frac{1}{9\sqrt{2}} & 0 \\ -\frac{1}{9\sqrt{2}} & q - \frac{2}{9} & 0 \\ 0 & 0 & p - \frac{1}{18} \end{pmatrix} \geq 0.$$

Thus we require $p \geq \frac{1}{18}$, $q \geq \frac{2}{9}$ and

$$\left(p - \frac{1}{18}\right)\left(q - \frac{2}{9}\right) - \left(\frac{1}{9\sqrt{2}}\right)^2 \geq 0. \qquad \text{(B1)}$$

For ρ_5, the condition is more conveniently expressed in the $\{|0\rangle, |1\rangle, |2\rangle\}$ basis, in which it has matrix form

$$\Gamma^A - \frac{1}{3}\rho_5 = \begin{pmatrix} p - \frac{5}{18} & -\frac{1}{18} & 0 \\ -\frac{1}{18} & q - \frac{1}{18} & 0 \\ 0 & 0 & p \end{pmatrix} \geq 0,$$

leading to the additional conditions $p \geq \frac{5}{18}$, and

$$\left(p - \frac{5}{18}\right)\left(q - \frac{1}{18}\right) - \left(\frac{1}{18}\right)^2 \geq 0. \qquad \text{(B2)}$$

Equality in (B1) and (B2) means that each of the operators $\Gamma^A - \frac{1}{3}\rho_j$ has a zero eigenvalue: M_j is then a weighted projector onto the corresponding eigenvector in each case. Setting each equal to zero, therefore, and solving for p and q gives

$$p = \frac{1}{9}\frac{1}{12}(17 + \sqrt{7}\sqrt{31}) \simeq 0.294,$$
$$q = \frac{1}{9}\frac{1}{16}(21 + \sqrt{7}\sqrt{31}) \simeq 0.248. \qquad \text{(B3)}$$

$\{M_j\}$ are then fixed up to multiplying factors, which may be chosen such that $\sum_j M_j = I$.

We now turn to the proof that the sequential measurement strategy given in the main text is indeed the optimal strategy. We wish to show that $\Gamma^A - \frac{1}{9}\widetilde{\sigma} \geq 0$ for all $\widetilde{\sigma}$, of the form given in Eq. (25), where according to symmetry arguments we only need consider measurements $\{\widetilde{N}_j\}$ of the form (26). Note that *all* measurements of this form may be considered to be probabilistic mixtures of a smaller set of fiducial measurements. An exhaustive list of those that cannot be decomposed as a mixture of other measurements are those in which the following elements are the only nonzero ones:

$$\{\widetilde{N}_1, \widetilde{N}_i, \widetilde{N}_j\}, i = 6, 7, j = 8, 9, \alpha = \beta = \pi/4,$$
$$\{\widetilde{N}_2, \widetilde{N}_3, \widetilde{N}_j\}, j = 8, 9, \alpha = \beta = \pi/4,$$
$$\{\widetilde{N}_4, \widetilde{N}_5, \widetilde{N}_i\}, i = 6, 7, \alpha = \beta = \pi/4,$$
$$\{\widetilde{N}_1, \widetilde{N}_2, \widetilde{N}_3, \widetilde{N}_j\}, j = 8, 9, \alpha < \pi/4,$$
$$\{\widetilde{N}_2, \widetilde{N}_3, \widetilde{N}_i, \widetilde{N}_j\}, i = 6, 7, j = 8, 9, \alpha > \pi/4,$$

$\{\widetilde{N}_1, \widetilde{N}_2, \widetilde{N}_3, \widetilde{N}_4, \widetilde{N}_5\}, \tan^2 \alpha + \tan^2 \beta < 1,$

$\{\widetilde{N}_2, \widetilde{N}_3, \widetilde{N}_4, \widetilde{N}_5, \widetilde{N}_i\}, i = 6, 7, \tan^2 \alpha + \tan^2 \beta > 1,$

$\{\widetilde{N}_2, \widetilde{N}_3, \widetilde{N}_4, \widetilde{N}_5, \widetilde{N}_j\}, j = 8, 9, \tan^2 \alpha + \tan^2 \beta > 1,$

$\{\widetilde{N}_1, \widetilde{N}_4, \widetilde{N}_5, \widetilde{N}_i\}, j = 6, 7, \beta < \pi/4,$

$\{\widetilde{N}_4, \widetilde{N}_5, \widetilde{N}_i, \widetilde{N}_j\}, i = 6, 7, j = 8, 9, \beta > \pi/4,$

The first three correspond to the case $\widetilde{\sigma} = 3\rho_j$ for some j, and the operator inequality holds by construction. The last three are obtained from the others by applying U_1, and so by symmetry we only need check the inequality for the remaining four. We consider the remaining possibilities in turn:

Case 1. $\alpha < \frac{\pi}{4}$:

All measurements in which $\{\widetilde{N}_1, \widetilde{N}_2, \widetilde{N}_3, \widetilde{N}_j\}, j = 8, 9$ are the only non-zero elements may be parametrised in the following way, where $r = \tan^2 \alpha$, $0 \le r < 1$:

$$\widetilde{N}_1 = (1 - r)|1\rangle\langle 1|,$$
$$\widetilde{N}_2 = \frac{1}{2}\left(|0\rangle + \sqrt{r}|1\rangle\right)\left(\langle 0| + \sqrt{r}\langle 1|\right),$$
$$\widetilde{N}_3 = \frac{1}{2}\left(|0\rangle - \sqrt{r}|1\rangle\right)\left(\langle 0| - \sqrt{r}\langle 1|\right),$$
$$\widetilde{N}_j = |2\rangle\langle 2|.$$

For $j = 8$, this leads to

$$\widetilde{\sigma} = (1 - r)|1\rangle\langle 1| + \frac{1}{2}(1 + \sqrt{r})^2|0\rangle\langle 0| + |0 + 1\rangle\langle 0 + 1|$$
$$\le (1 - r)|1\rangle\langle 1| + (1 + r)|0\rangle\langle 0| + |0 + 1\rangle\langle 0 + 1|,$$

where we have used the inequality $(a + b)^2 \le 2(a^2 + b^2)$. Note that the $j = 9$ case is obtained by applying U_0 to $\widetilde{\sigma}$ and thus by symmetry we only need check the $j = 8$ case. For this case, in matrix notation, in the $\{|0\rangle, |1\rangle, |2\rangle\}$ basis, we obtain

$$\Gamma^A - \frac{1}{9}\widetilde{\sigma} \ge \begin{pmatrix} p - \frac{1}{9}(1 + r) - \frac{1}{18} & -\frac{1}{18} & 0 \\ -\frac{1}{18} & q - \frac{1}{9}(1 - r) - \frac{1}{18} & 0 \\ 0 & 0 & p \end{pmatrix}.$$

Note that with the values of p and q calculated in eqn (B3), the diagonal elements are all strictly positive, and we further require $\left(p - \frac{1}{9}(1 + r) - \frac{1}{18}\right)\left(q - \frac{1}{9}(1 - r) - \frac{1}{18}\right) - \left(\frac{1}{18}\right)^2 \ge 0$. Re-writing the first term in brackets as $p - \frac{5}{18} + \frac{1}{9}(1 - r)$ we obtain

$$\left(p - \frac{5}{18}\right)\left(q - \frac{1}{18}\right) - \left(\frac{1}{18}\right)^2$$
$$+ \frac{1}{9}(1 - r)\left[q - \frac{1}{18} - \left(p - \frac{5}{18}\right) - \frac{1}{9}(1 - r)\right] \ge 0.$$

The combination of the first two terms is positive according to Eq. (B2), and the last term is also positive, as is readily seen by noting $0 < r < 1$ and by explicit substitution for p and q (or as a quick check, note that $p < \frac{6}{18}$, while $q > \frac{2}{9}$). Thus the inequality holds.

Case 2. $\alpha > \frac{\pi}{4}$:

All measurements in which $\{\widetilde{N}_2, \widetilde{N}_3, \widetilde{N}_i, \widetilde{N}_j\}, i = 6, 7, j = 8, 9$ are the only non-zero elements may be parametrised in the following way, where $r = \frac{1}{\tan^2 \alpha}$, $0 \le r < 1$:

$$\widetilde{N}_2 = \frac{1}{2}\left(\sqrt{r}|0\rangle + |1\rangle\right)\left(\sqrt{r}\langle 0| + \langle 1|\right),$$
$$\widetilde{N}_3 = \frac{1}{2}\left(\sqrt{r}|0\rangle - |1\rangle\right)\left(\sqrt{r}\langle 0| - \langle 1|\right),$$
$$\widetilde{N}_i = (1 - r)|0\rangle\langle 0|,$$
$$\widetilde{N}_j = |2\rangle\langle 2|.$$

Note that by symmetry, we only need explicitly consider one value of i and j. For $i = 6$, $j = 8$, we obtain

$$\widetilde{\sigma} = \frac{1}{2}(1 + \sqrt{r})^2|0\rangle\langle 0| + (1 - r)|1 + 2\rangle\langle 1 + 2|$$
$$+ |0 + 1\rangle\langle 0 + 1|,$$
$$\le (1 + r)|0\rangle\langle 0| + (1 - r)|1 + 2\rangle\langle 1 + 2|$$
$$+ |0 + 1\rangle\langle 0 + 1|,$$
$$\le (1 + r)|0\rangle\langle 0| + (1 - r)|1\rangle\langle 1| + (1 - r)|2\rangle\langle 2|$$
$$+ |0 + 1\rangle\langle 0 + 1|,$$

where in the second line we have used the inequality $(a + b)^2 \le 2(a^2 + b^2)$, and in the third line we have added the positive operator $(1 - r)|1 - 2\rangle\langle 1 - 2|$. Thus in matrix notation, in the $\{|0\rangle, |1\rangle, |2\rangle\}$ basis, we obtain

$$\Gamma^A - \frac{1}{9}\widetilde{\sigma} \ge$$
$$\begin{pmatrix} p - \frac{1}{9}(1 + r) - \frac{1}{18} & -\frac{1}{18} & 0 \\ -\frac{1}{18} & q - \frac{1}{9}(1 - r) - \frac{1}{18} & 0 \\ 0 & 0 & p - \frac{1}{9}(1 - r) \end{pmatrix}.$$

The diagonal elements are strictly positive, while $\left\{[p - \frac{1}{9}(1 + r) - \frac{1}{18}][q - \frac{1}{9}(1 - r) - \frac{1}{18}]\right\} - \left(\frac{1}{18}\right)^2 \ge 0$ according to the discussion in case 1 above. Thus the operator inequality holds.

Case 3. $\tan^2 \alpha + \tan^2 \beta < 1$:

All measurements in which $\{\widetilde{N}_1, \widetilde{N}_2, \widetilde{N}_3, \widetilde{N}_4, \widetilde{N}_5\}$ are the only non-zero elements may be parametrised in the following way, where $r = \tan^2 \alpha$, $0 \le r < 1$, $s = \tan^2 \beta$, $0 \le s < 1$:

$$\widetilde{N}_1 = (1 - r - s)|1\rangle\langle 1|,$$
$$\widetilde{N}_2 = \frac{1}{2}\left(|0\rangle + \sqrt{r}|1\rangle\right)\left(\langle 0| + \sqrt{r}\langle 1|\right),$$
$$\widetilde{N}_3 = \frac{1}{2}\left(|0\rangle - \sqrt{r}|1\rangle\right)\left(\langle 0| - \sqrt{r}\langle 1|\right),$$
$$\widetilde{N}_4 = \frac{1}{2}\left(|2\rangle + \sqrt{s}|1\rangle\right)\left(\langle 2| + \sqrt{s}\langle 1|\right),$$
$$\widetilde{N}_5 = \frac{1}{2}\left(|2\rangle - \sqrt{s}|1\rangle\right)\left(\langle 2| - \sqrt{s}\langle 1|\right).$$

From Eq. (25) we obtain

$$\widetilde{\sigma} = (1 - r - s)|1\rangle\langle 1| + \frac{1}{2}(1 + \sqrt{r})^2|0\rangle\langle 0|$$

$$+\frac{1}{2}(1+\sqrt{s})^2|2\rangle\langle 2|$$
$$\leq (1+r)|0\rangle\langle 0| + (1-r-s)|1\rangle\langle 1| + (1+s)|2\rangle\langle 2|.$$

where we have used again the inequality $(a+b)^2 \leq 2(a^2 + b^2)$. This is diagonal in the $\{|0\rangle, |1\rangle, |2\rangle\}$ basis, and it is easy to see that $\Gamma^A - \frac{1}{9}\widetilde{\sigma} \geq 0$ holds.

Case 4. $\tan^2\alpha + \tan^2\beta > 1$:

All measurements in which $\{\widetilde{N}_2, \widetilde{N}_3, \widetilde{N}_4, \widetilde{N}_5, \widetilde{N}_i\}$, $i = 6,7$, are the only non-zero elements may be parametrised in the following way, where $r = \frac{1}{\tan^2\alpha}$, $0 \leq r < 1$, $s = \tan^2\beta$, $0 \leq s < 1$:

$$\widetilde{N}_2 = \frac{1-s}{2}\left(\sqrt{r}|0\rangle + |1\rangle\right)\left(\sqrt{r}\langle 0| + \langle 1|\right),$$
$$\widetilde{N}_3 = \frac{1-s}{2}\left(\sqrt{r}|0\rangle - |1\rangle\right)\left(\sqrt{r}\langle 0| - \langle 1|\right),$$
$$\widetilde{N}_4 = \frac{1}{2}\left(|2\rangle + \sqrt{s}|1\rangle\right)\left(\langle 2| + \sqrt{s}\langle 1|\right),$$
$$\widetilde{N}_5 = \frac{1}{2}\left(|2\rangle - \sqrt{s}|1\rangle\right)\left(\langle 2| - \sqrt{s}\langle 1|\right),$$
$$\widetilde{N}_i = [1 - r(1-s)]|0\rangle\langle 0|,$$

From Eq. (25) we obtain

$$\widetilde{\sigma} = \frac{1-s}{2}(1+\sqrt{r})^2|0\rangle\langle 0| + \frac{1}{2}(1+\sqrt{s})^2|2\rangle\langle 2|$$
$$+[1 - r(1-s)]|1+2\rangle\langle 1+2|$$
$$\leq (1-s)(1+r)|0\rangle\langle 0| + (1+s)|2\rangle\langle 2|$$
$$+[1 - r(1-s)]|1+2\rangle\langle 1+2|,$$
$$< (1-s)(1+r)|0\rangle\langle 0|$$
$$2|2\rangle\langle 2| + |1+2\rangle\langle 1+2|,$$

where we have used again the inequality $(a+b)^2 \leq 2(a^2 + b^2)$, and in the last line use $s < 1$. Recall that Γ^A is diagonal in the $\{|0\rangle, |1\rangle, |2\rangle\}$ basis, so it suffices to check

$$\Gamma^A - \frac{1}{9}(1-s)(1+r)|0\rangle\langle 0| \geq 0,$$

which clearly holds as $\frac{1}{9}(1-s)(1+r) < \frac{2}{9} < p$, and

$$\Gamma^A - \frac{2}{9}|2\rangle\langle 2| + \frac{1}{9}|1+2\rangle\langle 1+2| \geq 0,$$

which we recognise as $\Gamma^A - \frac{1}{3}\rho_7 \geq 0$, and therefore holds by construction.

[1] S.M. Barnett and S. Croke, Adv. Opt. Phot. **1**, 238 (2009).
[2] J. Bae, and L.-C. Kwek, J. Phys. A: Math. Theor. **48**, 083001 (2015).
[3] A. Chefles, Contemporary Physics **41**, 401 (2000)
[4] J. Bergou, J. Mod. Opt. **57**, 160 (2010).
[5] S.M. Barnett, "Quantum Information" (Oxford University Press, Oxford, 2009).
[6] C.W. Helstrom, Inf. Control **10**, 254 (1967).
[7] C.W. Helstrom, Inf. Control **13**, 156 (1968).
[8] A.S. Holevo, J. Multivariate Analysis **3**, 337 (1973).
[9] H.P. Yuen, R.S. Kennedy, and M. Lax, IEEE Trans. Inf. Theory **IT-21**, 125 (1975).
[10] E.B. Davies, IEEE Trans. Inf. Theory **IT-24**, 596 (1978).
[11] C.W. Helstrom, "Quantum Detection and Estimation Theory", (Academic Press, New York, 1976).
[12] A.S. Holevo, "Probabilistic and Statistical Aspects of Quantum Theory" (North Holland Publishing Company, Amsterdam, 1982).
[13] L.B. Levitin, in "Quantum Communications and Measurement", V.P. Belavkin, O. Hirota, and R.L. Hudson, eds. (Plenum, 1995), pp. 439-448.
[14] M. Sasaki, S.M. Barnett, R. Jozsa, M. Osaki, and O. Hirota, Phys. Rev. A **59**, 3325 (1999).
[15] I.D. Ivanovic, Phys. Lett. A **123**, 257 (1987).
[16] D. Dieks, Phys. Lett. A **126**, 303 (1988).
[17] A. Peres, Phys. Lett. A **128**, 19 (1988).
[18] G. Jaeger and A. Shimony, Phys. Lett. A **197**, 83 (1995).
[19] A. Chefles and S.M. Barnett, J. Mod. Opt. **45**, 1295 (1998).
[20] S. Croke, E. Andersson, S.M. Barnett, C.R. Gilson, and J. Jeffers, Phys. Rev. Lett. **96**, 070401 (2006).
[21] C.H. Bennett, D.P. DiVincenzo, C.A. Fuchs, T. Mor, E. Rains, P.W. Shor, J.A. Smolin, and W.K. Wootters, Phys. Rev. A **59**, 1070 (1999).
[22] J. Walgate, A. J. Short, L. Hardy and V. Vedral, Phys. Rev. Lett. **85**, 4972 (2000).
[23] S. Virmani, M. F. Sacchi, M. B. Plenio and D. Markham, Phys. Lett. A **288**, 62 (2001).
[24] Y.-X. Chen and D. Yang, Phys. Rev. A **64**, 064303 (2001).
[25] Y.-X. Chen and D. Yang, Phys. Rev. A **65**, 022320 (2002).
[26] Z. Ji, H. Cao and M. Ying, Phys. Rev. A **71**, 032323 (2005).
[27] A. Peres and W.K. Wootters, Phys. Rev. Lett. **66**, 1119 (1991).
[28] E. Chitambar and M.-H. Hsieh, Phys. Rev. A **88**, 020302(R) (2013).
[29] Y. Lu, N. Coish, R. Kaltenbaek, D.R. Hamel, S. Croke, and K.J. Resch, Phys. Rev. A **82**, 042340 (2010).
[30] R. Blume-Kohout, S. Croke, and M. Zwolak, Sci. Rep. **3**, 1800 (2013).
[31] A. Peres, "Quantum Theory: Concepts and Methods", (Kluwer Academic Publishers, Dordrecht, 1995).
[32] S.M. Barnett and S. Croke, J. Phys. A: Math. Theor. **42**, 062001 (2009).
[33] J. Bae, New J. Phys. **15**, 073037 (2013).
[34] M. Ban, K. Kurokawa, R. Momose, and O. Hirota, Int. J. Theor. Phys. **36**, 1269 (1997).
[35] P. Hausladen and W..K. Wootters, J. Mod. Opt. **41**, 2385 (1994).
[36] D. Brody and B. Meister, Phys. Rev. Lett. **76**, 1 (1996).
[37] A. Acin, E. Bagan, M. Baig, Ll. Masanes, and R. Munoz-Tapia, Phys. Rev. A **71**, 032338 (2005).
[38] E. Andersson, S.M. Barnett, C.R. Gilson, and K. Hunter,

Phys. Rev. A **65**, 052308 (2002).

[39] S. Ghosh, G. Kar, A. Roy, A. Sen(De), and U. Sen, Phys. Rev. Lett. **87**, 277902 (2001).

[40] J. Walgate and L. Hardy, Phys. Rev. Lett. **89**, 147901 (2002).

[41] A.M. Childs, D. Leung, L. Mančinska, and M. Ozols,

Comm. Math. Phys. **323**, 1121 (2013).

[42] S. Croke and S.M. Barnett, Phys. Rev. A **95**, 012337 (2017).

[43] U. Herzog, Phys. Rev. A. **91**, 042338 (2015).

[44] J. Fiurášek, M. Ježek, Phys. Rev. A. **67**, 012321 (2003).